mahesh.sha@aeronomie.be 18 March 2020

Letter to the Editor

for manuscript by **Sha et al.**, MS No. **amt-2019-371**; title: "**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**".

Dear Editor,

We would like to thank you for acting as the Editor for our manuscript and providing positive comments during the start of the review process.

We have revised our manuscript carefully taking into account all of the comments made by the two referees.

Both referees suggested a restructuring of the paper and to put emphasis on the non-linearity corrected TCCON data. We found this suggestion useful. In the revised version of the paper, we therefore start our results section with the description of the non-linearity (section 5.2), followed by the discussion on the performance of the Bruker IFS 125HR (both standard and non-linearity corrected data sets) against the AirCore. We then discuss the intercomparison results of the low-resolution test instruments with respect to the non-linearity corrected TCCON data set. In addition, we have moved the discussions of the uncorrected reference data sets to the appendix. The sections related to the solar zenith angle (SZA) and humidity dependence of the bias have also been moved to the appendix following the suggestion of the referee. Keeping in line with the above-mentioned changes, all associated figures have been updated. The restructuring only changed the ordering of the sections, but our lessons learned and the conclusions drawn from the campaign remain the same.

We hope that the revised version of the manuscript now meets the high quality standards for publication in the journal "Atmospheric Measurement Techniques".

With kind regards,
Mahesh Kumar Sha (on behalf of all authors)

Response to comments from Referee 1 (Dr. Ruediger Lang)

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

We thank referee #1, Dr. Ruediger Lang for the review and for providing useful feedback.

Referee:
Accurate ground based observations of carbon dioxide and methane are becoming increasingly important in the context of continuous satellite remote sensing validation. While ground based in-situ and remote sensing networks measuring CO_2 and CH4 concentrations in the atmosphere exist, and are instrumental in measuring the continuous increase in greenhouse gases over the past decades, it has been recognised that the existing network of stations is still lacking with respect to its expected role in future monitoring and verification system-of-systems (MVS) of greenhouse gas emissions. Ground based remote sensing instruments like the Total Carbon Column Observing Network (TCONN) FTS instruments measuring total column dry air mole fractions of carbon dioxide and methane (XCO_2 and XCH4) are key for the validation and calibration of future operational satellite based observing systems, targeting greenhouse gas emissions. However, the current distribution of stations and their representability for the validation of global satellite based measurements, as well as their maintenance and operability for providing a continuous flow of quality-monitored data are providing big challenges ahead for the network to become the much aimed for fiducial greenhouse-gas reference measurements within an operational MVS context. One of the key-challenges is achieving and maintaining the very high accuracies needed – significantly below 0.5 ppm for XCO_2 – in order to become useful for the MVS and the Cal/Val of its satellite components.

The paper by Sha et al. is an important and significant contribution towards the latter aspect, by addressing the question on how to secure high accuracies across the network, e.g. through travelling instrument standards, while identifying, correcting and potentially even reducing instrument and measurement biases. The key finding of the paper is, from my point of view, the potential of at least one of the instruments taking part in the campaign exercise at Sodankyla, Finnland, functioning as a "travelling standard" in a potential future operational network of reference FTS instruments as currently operated under the TCCON umbrella. The paper also addresses important open questions with respect to remaining systematic biases and points at, maybe even more significant, remaining issues in measurement bias and precision. The paper is well written, although I think the paper could benefit from some restructuring of the results sections. I therefore recommend the paper for publishing in ACP, but would like to highlight in the following a couple of observations (apart from some additional minor comments), which the authors, from my point of view, should address.
Authors' response:
Thanks for the positive comments. As the paper focuses on the measurements of atmospheric components using ground-based instruments, we think it is better suited for publication in AMT rather than ACP and would like to keep it here.

Referee:
1) Ari-core comparisons and non-linearity effects
The discussion of the non-linearity effect found in the measurements of the Bruker IFS HR reference instruments (referred to as TCCON "reference", 125HR), as well as the comparison to the AirCore measurements, as taken during the campaign at Sodankyla - which are considered to represent the "true atmospheric" state - are both presented only at the end of the major results

section 5. This is first of all confusing, since the section on non-linearity effects implies that the non-linearity correction "has been applied to TCCON" data as a whole. At this stage, the reader wonders if this therefore now applies to all previous results, but the answer is probably not, since the results and different labelling then implies that the main comparisons results are not (yet) non-linearity corrected.

Second, the performance of the "reference instruments" with respect to what is considered the "true state of the atmosphere" represented by the AirCore measurements is an important result against which also the results of the measurements from the other systems have to be evaluated and interpreted. So both aspects have to be taken together. I would therefore recommend to present the results on the 125HR instrument non-linearity and the performance of the reference instruments against the "truth" (5.8 and 5.9) at the beginning of Section 5 (and ideally then present only the non-linearity corrected results for the TCONN reference – if this is considered a stable result – in the comparison against the other systems). This would help to interpret the results of the other systems when compared to the "reference" 125HR better with respect to the AirCore "truth". In addition, establishing the TCCON 125HR as a "reference" and therefore then talking about a "bias" with respect to the reference for the other instruments (and not in terms of "differences"), requires, in my view, presenting the AirCore results first anyhow.
Authors' response:
Thanks for this suggestion.
Our original idea was to present the comparison results first with the official TCCON data and then show the results of the non-linearity corrected TCCON data, which was part of the lessons learned during the campaign. We agree to the referee's viewpoint and have moved the (former) sections 5.8 and 5.9 to the beginning of Section 5. In addition, we have moved the other intercomparison results with the official TCCON data and the HR125LR data to the appendix. These results are useful as a reference comparison to the official TCCON data. They also provide useful analysis on the resolution dependent effects on the Xgas retrievals of the target gases.

Referee:
2) EM27 systematic biases with respect to the two references The EM27 (COCCON) instruments show a convincing performance with respect to the "reference" TCCON (125HR), both with respect to the systematic biases and precision (for the latter see below). This is alas true especially with respect to the non- corrected TCCON results. It seems peculiar that this bias is even lower than -0.2 ppm for the non-corrected comparisons, but then gets worse for the TCCON non-linearity corrected ones (-0.73 ppm). At the same time, the low-resolution "reference" measurements (125LR) show a high bias to EM27 of roughly the same order of magnitude than the effect of the non-linearity correction. Considering that the low resolution 125HR are more comparable in terms of information content (and probably AKs, although those have unfortunately not been presented –see below), one wonders if the non-linearity correction should not also be applied to the 125LR measurements (or has this been done?), which then may lead to a consistency between EM27 and the 125LR. This would then also physically make a lot of sense considering the information content of both measurements. Also it would be important to rule out any link (e.g. in retrieval processing) between the standard 125HR measurements (not corrected) and the EM27 results, which potentially make them similar to the standard reference

by default (and therefore more biased with respect to the corrected 125HR results and the AirCore "truth"). In this context, the processing algorithm ProFIT vs GFIT performances and their potentially relevant "peculiarities" are not much discussed at all. ProFIT is used both for the TCCON 125LR and the COCCON EM27s but not for the uncorrected TCCON "reference" it seems. Is the bias correction scheme used in ProFit maybe somehow related to TCCON measurements (or associated climatologies?).

I think it would be important to add a discussion of the (potential) relationship between the observed systematic biases between "reference" HR, LR, and non-linearity correction on one side, and the EM27 measurement results on the other side, which could shine some light on the underlying mechanisms.

Authors' response:

We have added a description of the PROFFAST code in the paper and the underlying differences with respect to the GFIT code. 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."

The Xgas values, which are calculated using GFIT are scaled to the WMO standards using calibration factors (airmass dependent and independent). The calibration factors were derived from dedicated campaigns at different TCCON sites. However, the XCO_2 and XCH4 products from the EM27/SUN are bias-corrected based on the scaling factors calculated from the extensive COCCON development. The residual bias between the non-linearity corrected TCCON data and the EM27/SUN might be due to several reasons, such as (1) effect of non-linearity correction of the TCCON data on the retrieved Xgas values, (2) bias of the EM27/SUN due to the imperfect instrument specific scaling factor used which has been determined independently prior to this study from long-term intercomparison measurements performed at the KIT TCCON site. (3) bias of the Sodankylä TCCON station which may come from the imperfect use of the airmass

independent calibration factor derived for the global TCCON. As more comparisons of the COCCON spectrometers with respect to the TCCON stations takes place in the future we need to verify the site-to-site bias of the TCCON with respect to the COCCON spectrometer. As mentioned by the referee as well, the use of the low-resolution spectrometer as a travelling standard will be very useful in this respect.
The HR125LR measurements (low-resolution measurements with the Bruker IFS 125HR) are also affected by the non-linearity of the InGaAs detector. As this is not our standard data set, no non-linearity correction has been done. It is used to check the resolution dependent effects of the measurements as discussed in the paper. For example, the intercomparison results help us to verify the bias in XCH4 during the springtime (as high as 0.01 ppm) which is due to the large difference between the a prior profile to the true atmospheric state and the influence of the different AKs between the low- and high-resolution spectrometers. The XCO_2 bias between the EM27SUN vs the HR125LR is 0.56 ppm. If we apply similar non-linearity correction to the HR125LR data, then the expected XCO_2 bias is about 0.06 ppm, XCH4 is about 0.006 ppm and XCO is about 4.5 ppb.

3) Averaging kernels, low res versus high res

The differences in performance and measurement information content for the low-resolution measurements (125LR and EM27) are of high importance, especially with respect to the fact that some of the forthcoming remote sensing satellite based systems may be operated at lower spectral resolutions, i.e. being more similar to the LR measurements in terms of spectroscopy. Also some of the differences observed between AirCore truth, high resolution and low resolution measurements, as discussed under 2), may be interpreted in this respect. The performance of these lower resolution systems are of importance also with respect to the knowledge of spectroscopy and potential future research needs there. But mainly the differences in performance between LR and HR (also with respect to AirCore "truth") may be better interpreted if the differences in their respective AK would be discussed and addressed - at least to some extend (and potentially also differences in GFIT and ProFIT a priori profiles if any – see also 2)). This could be done e.g., and ideally, in a corresponding figure to Fig. 17 (or adding such results to the latter).
Authors' response:
We completely agree to the referee's point of view. It is very important to identify differences caused by the resolution differences of the spectrometers and their related effects. As a result we have kept the HR125LR comparison results in the appendix. The AKs of the TCCON and the EM27/SUN for all SZA are shown in Figure 6 of Hedelius et al. (2016). The retrieval of the low and high-resolution measurement data sets were done using the TCCON a priori as the common a priori. As discussed in point 2 above, other differences between the PROFFAST and GFIT codes have been added to the paper.

Referee:
4) Precision of TCCON versus COCCON

The striking difference in measurement noise (Fig. 18) between TCCON and COCCON EM27 is not much discussed it seems. Is this feature a question of retrieval algorithm performance (or/and applied constraints therein), spectral resolution (like the more stable LR with respect to HR measurements), or is it solely related to instrument noise?
Authors' response:
The EM27/SUN shows a lower scatter as compared to the TCCONmod due to the low noise resulting from the averaging of the individual measurements. The individual measurement time for each instrument is provided in Table 2 in the paper. Within the period of five minutes, it is possible to average five measurements for the EM27/SUN data set. This is equivalent to 50 scans in total in 5 min with 10 scans for each measurement. Whereas a maximum of only two measurements are possible for the TCCONmod data set. This is equivalent to a maximum of 8 scans in 5 min with 4 scans for each measurement.
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.

Additional minor comments:

Referee:
p. 2, l.36ff: "Carbon dioxide (CO2) and methane (CH4) are the two main components of the carbon cycle of the earth's atmosphere. They absorb and retain the heat in the atmosphere causing the greenhouse effect and global warming" – I would remove "greenhouse effect" here. Otherwise, water vapour would have to be mentioned in this context.
Authors' response:
Done

Referee:
Section 5.3: I think it would be would be very helpful to introduce here the logic of the following sections (titles content) to guide the reader through what is coming, apart from potentially shifting the AirCore results and non-linearity sections first (see above).
Authors' response:
We agree. We now begin the results section with a description of the non-linearity issue with our reference measurement "TCCON". This is followed by a comparison of the standard and non-linearity corrected TCCON results with respect to the AirCore.

Referee:
Figure 1 to 4: It would be helpful to use the same axis limits in the scatter plots.
Authors' response:
Done. Note that these are now Figures 7 – 10 which are made for the TCCONmod case with the same axis for the scatter plots for each instrument except for the XCO_2 plot for the LHR instrument where the scatter is significantly higher as compared to the other instruments.

Referee:
Figure 16: A horizontal dashed line or similar at zero would really help to interpret the results. Especially since the range of the y-axis has to be quite large.

Authors' response:
Done

Referee:
Section 5.4.4. I think here it should be highlighted and maybe discussed in context of the discussion on biases between TCCON, TCCONmod, LR and EM27, why the EM27 is so much closer to one than TCCON (see also point 2) above).
Authors' response:
We have added an explanation in response to point 2 above and we included it in the last paragraph of section 5.5, see lines 631 – 642: "The Xgas biases between the low-resolution test instruments and the TCCONmod data sets as reference may be due to effects such as different responses to a priori profiles, interfering species in the retrieval windows or different averaging kernels. Furthermore it is important to note that, TCCON uses a network-wide constant scaling factor to scale its Xgas values to the WMO standards. The scaling factors specific to each gas for TCCON had been determined from several measurement campaigns where vertically distributed measurements of the gases were performed from airborne platforms using WMO calibrated instruments. The EM27/SUN uses species dependent scaling factors for XCO2 and XCH4 which had been calculated from long-term intercomparison measurements performed at the KIT TCCON site. However, no such instrument specific calibration factors were applied for the other instruments and also for the XCO results from the EM27/SUN measurements. This also contributes to the residual bias which is observed in this intercomparison result. The biases which are purely due to resolution differences are addressed by performing low-resolution measurements with the same TCCON instrument. These data are then used for an intercomparison relative to the TCCON as well as for the intercomparison with other low-resolution test instruments. Further details of the intercomparison results are given in appendix C and D, respectively."

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.

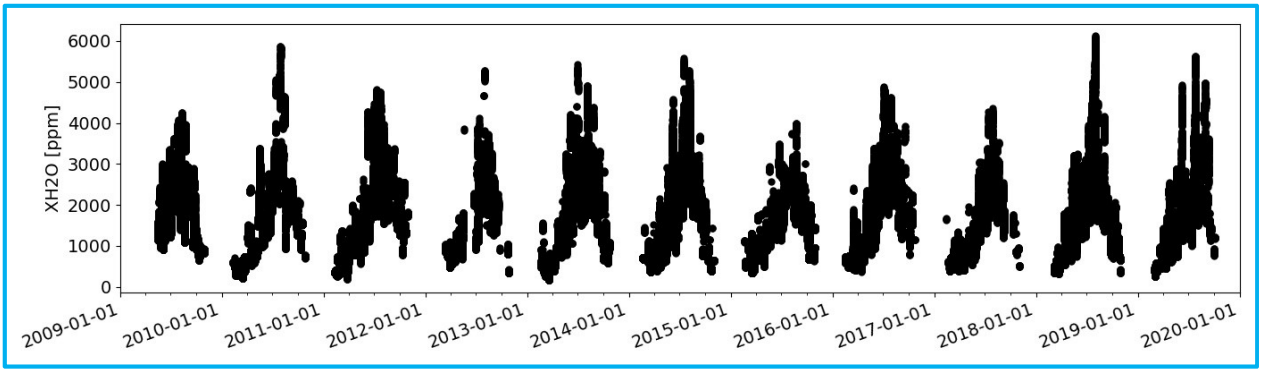

Figure 1: Timeseries of the XH2O measured at the Sodankylä TCCON site.

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.

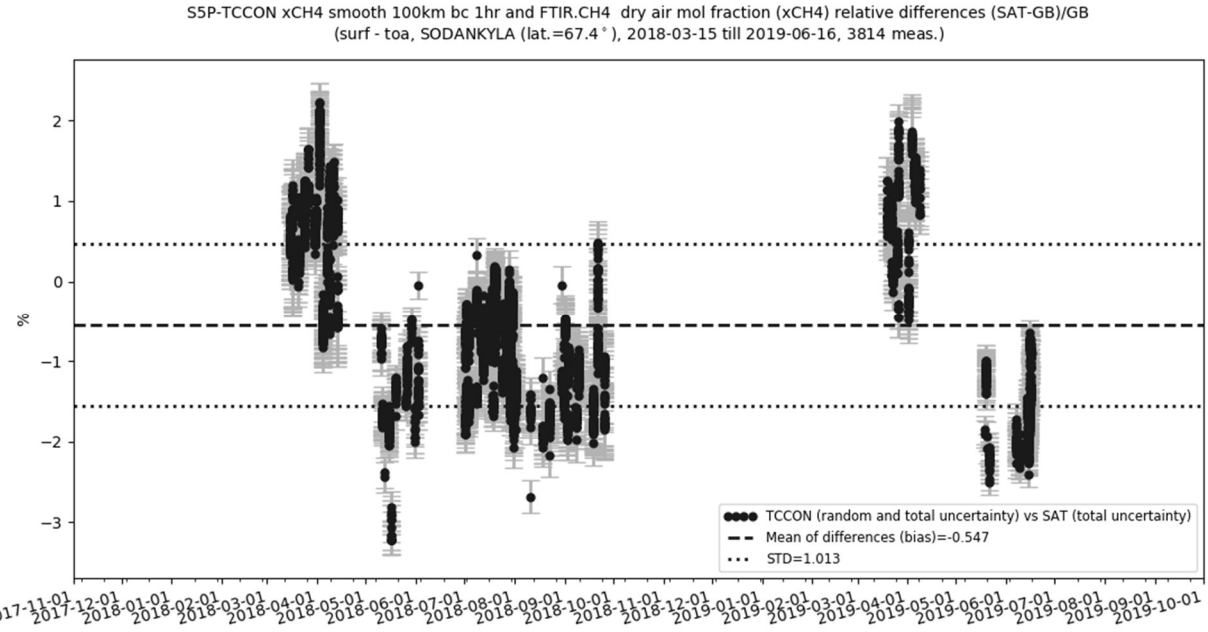

Figure 2: Relative difference of the XCH4 from the (S5P-TCCON)/TCCON data sets for the Sodankylä site. The coincidence criteria for the validation are given in the presentation by Sha et al. 2019.

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

Intercomparison of low and high resolution infrared spectrometers for ground-based solar remote sensing measurements of total column concentrations of CO_2, CH_4 and CO

Mahesh Kumar Sha[1], Martine De Mazière[1], Justus Notholt[2], Thomas Blumenstock[3], Huilin Chen[4], Angelika Dehn[5], David W T Griffith[6], Frank Hase[3], Pauli Heikkinen[7], Christian Hermans[1], Alex Hoffmann[8], Marko Huebner[8], Nicholas Jones[6], Rigel Kivi[7], Bavo Langerock[1], Christof Petri[2], Francis Scolas[1], Qiansi Tu[3], and Damien Weidmann[8]

[1]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium
[2]Institute of Environmental Physics, University of Bremen, Bremen, Germany
[3]Karlsruhe Institute of Technology, IMK-ASF, Karlsruhe, Germany
[4]Centre for Isotope Research, University of Groningen, Groningen, The Netherlands
[5]European Space Agency, ESA/ESRIN
[6]University of Wollongong, Wollongong, Australia
[7]Finnish Meteorological Institute, Sodankylä, Finland
[8]Rutherford Appleton Laboratory, United Kingdom

Correspondence: Mahesh Kumar Sha (mahesh.sha@aeronomie.be)

Abstract. The Total Carbon Column Observing Network (TCCON) has been the baseline ground-based network of instruments that record solar absorption spectra from which accurate and precise column-averaged dry air mole fractions of CO_2 (XCO_2), CH_4 (XCH_4), CO (XCO) and other gases are retrieved. The TCCON data have been widely used for carbon cycle science and validation of satellites measuring greenhouse gas concentrations globally. The number of stations in the network (currently about 25) is limited and ~~the stations~~ has a very uneven geographical coverage: the stations in the Northern hemisphere are distributed mostly in ~~Northern~~ North America, Europe, Japan and ~~Oceania~~ 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 ~~cover~~ improve the representativeness of the measurement data for various atmospheric conditions (humid, dry, polluted, presence of aerosol), various surface conditions ~~(high~~ such as high albedo (> 0.4) and very low albedo and ~~low albedo) and~~ a larger latitudinal distribution. More stations in the southern hemisphere are also needed but a further expansion of the network is limited by its costs and logistical requirements. For this reason, several groups are investigating supplemental portable low-cost instruments. The European Space Agency (ESA) funded campaign Fiducial Reference Measurements for Ground-Based Infrared Greenhouse Gas Observations (FRM4GHG) at the Sodankylä TCCON site in northern Finland aims at characterising the assessment of several low-cost portable instruments for precise solar absorption measurements of XCO_2, XCH_4 and XCO. The test instruments under investigation are three Fourier transform spectrometers (FTS): a Bruker EM27/SUN, a Bruker IRcube and a Bruker Vertex70; as well as a Laser Heterodyne spectro-Radiometer (LHR) developed by the UK Rutherford Appleton Laboratory. All four remote sensing instruments performed measurements simultaneously next to the reference TCCON instrument, a Bruker IFS 125HR, for a full year in 2017. The TCCON FTS was operated in its normal

high-resolution mode (TCCON data set) and in a special low-resolution mode (HR125LR data set), similar to the portable spectrometers. The remote sensing measurements have been complemented by regular AirCore launches performed from the same site. They provide in-situ vertical profiles of the target gas concentrations as auxiliary reference data for the column retrievals which is traceable to the WMO SI standards. 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. The timeseries, the bias relative to the reference instrument and its scatter and the seasonal and the day-to-day variations of the target gases are shown and discussed. The comparisons with the HR125LR data set gave useful analysis of the resolution dependent effects on the target gas retrieval. The solar zenith angle dependence of the retrievals is shown and discussed. The ~~reference measurements performed with the Bruker IFS 125HR (TCCON and HR125LR data sets) were found to be affected by non-linearity. A non-linearity correction of the TCCON data was performed and compared with the test instruments and AirCore. The non-linearity corrected TCCON data show a better match with the test instruments and AirCore data as compared to the reference TCCON data. The~~ intercomparison results show that the LHR data have a large scatter and biases with a strong diurnal variation relative to the TCCON and other FTS instruments. The LHR is a new instrument under development and these biases are being currently investigated and addressed. The campaign helped to characterise and identify the instrumental biases and possibly retrieval biases which are currently under investigation. Further improvements of the instrument are ongoing. The EM27/SUN, the IRcube, the modified Vertex70 and the HR125LR provided stable and precise measurements of the target gases during the campaign with quantified small biases. The bias dependence on the humidity along the measurement line-of-sight has been investigated and no dependence was found. These three portable low-resolution FTS instruments are suitable to be used for campaign deployment or long-term measurements from any site and offer the ability to complement the TCCON and expand the global coverage of ground-based reference measurements of the target gases.

1 Introduction

Carbon dioxide (CO_2) and methane (CH_4) are the two main components of the carbon cycle of the earth's atmosphere. They absorb and retain the heat in the atmosphere causing the ~~greenhouse effect and~~ global warming. CH_4 has a global warming potential of about 28 times greater than CO_2 over a 100 year time period. It however exists in much lower concentrations and has a significantly shorter lifetime compared to CO_2. CH_4 also plays an important role in atmospheric chemistry by reacting with hydroxyl radicals (OH), thereby reducing the oxidation capacity of the atmosphere and producing ozone (Kirschke et al., 2013). The atmospheric concentration of both these gases has been steadily increasing in ~~the~~ recent years caused by anthropogenic activities (Stocker et al., 2013; Dlugokencky and Tans). The third gas focused on is carbon monoxide (CO). It is a poisonous reactive gas considered principally as a man-made pollutant. The volatile organic compounds (VOCs) plays an important role in the production of CO. It plays an important role in atmospheric chemistry by reacting with the atmospheric oxidants, ozone (O_3), hydroperoxy (HO_2) and hydroxyl radicals (OH). The lifetime of CO ranges from weeks to months (Novelli et al.,

1998). An increase of CO would imply that more OH will be lost through chemical reaction with CO and that less OH will be available for reaction with CH_4. Therefore CO has an indirect but important influence in determining the chemical composition and radiative properties of the atmosphere. The emissions of CO are virtually certain to have a positive radiative forcing~~. It is therefore~~ , therefore it is considered as an indirect greenhouse gas (Stocker et al., 2013). Continuous monitoring of precise and accurate measurements of these gases is of utmost importance to determine their sources and sinks, and trends. Currently this is one of the major challenges within climate research which will help in understanding the carbon cycle.

Atmospheric measurements of CO_2, CH_4 and CO have been performed by in-situ surface based networks for many decades. These have been complemented by sparse aircraft measurement campaigns providing important additional measurements. However, both these measurement types have been performed at only a few locations and the atmosphere sampled non-uniformly. In recent years, satellite based remote sensing measurements have been able to provide global coverage of these gases. The nadir looking satellites detecting scattered sunlight in the near-infrared (NIR) spectral region provide the most powerful method for global mapping of these gases. These measurements cover the whole atmospheric column providing the total column concentrations of the trace gases of interest and add important measurements to the in-situ networks. However, satellite measurements require accurate validation. These accurate reference measurements can be performed from surface based, air-borne (e.g., balloon or aircraft) or already validated satellites. To ensure equal dependency on the ~~input spectral data~~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. Moreover, the total column observations are much less sensitive to boundary layer effects compared to the in-situ surface measurements.

The current state-of-the-art validation system for greenhouse gases (GHGs) is the Total Carbon Column Observing Network (TCCON). TCCON is a network of ground-based Fourier transform spectrometers (FTS), of the type Bruker IFS 125HR, that record solar absorption spectra in the NIR spectral range to retrieve accurate and precise column-averaged abundances of atmospheric constituents including CO_2, CH_4 and CO amongst other species (Wunch et al., 2011). There are currently about 25 TCCON stations distributed globally which form the backbone of the validation data set for the GHG measuring satellites (e.g., GOSAT, OCO-2, Sentinel-5 Precursor, ...) and model comparisons (Inoue et al., 2016; Wunch et al., 2016; Borsdorff et al., 2018; Kivimäki et al., 2019; Ostler et al., 2016; Jing et al., 2018; Kong et al., 2019). The distribution of the TCCON stations currently lacks global coverage with a majority of its stations located in ~~Northern~~ North America, Europe and Japan, and currently only five stations in the southern hemisphere. The lack of stations close to important source areas and the limited number of stations in general is unable to resolve global GHG gradients. Furthermore, for the complete validation of the satellite data set, ~~measurements covering a wide range of~~ 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~~) are missing within the current set up of the TCCON network~~, humid, dry).

An extension of the TCCON network is limited by high start-up, maintenance and operational costs, as well as difficulties of campaign based transportability. The maintenance of the instrument requires skill and experience. All these factors resulted

in the development of a number of cheap and easily deployable instruments for remote sensing measurements of greenhouse gases, mainly driven by scientific research institutes in collaboration with industrial partners. Some of these instruments have been in operation for several years. However, there has been little characterisation, intercomparison and harmonisation of these new instruments in comparison to the standard instrument used in TCCON, except for the EM27/SUN were some previous characterisation works are done (Gisi et al., 2012; Frey et al., 2015; Hedelius et al., 2016, 2017; Frey et al., 2019). These comparisons however are mandatory for using these individual data sets independently for science.

For this reason in 2017 ~~an intercomparison campaign, initiated by~~ , 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) ~~to assess~~ with the aim to assess the performance of different spectrometric instruments for remote sensing of atmospheric trace gases ~~as to~~ and to quantify their performances regarding precise measurements of column-averaged dry-air volume mole fractions of CO_2, CH_4 and CO. The instruments were deployed at the meteorological observatory Sodankylä where measurements took place between March and October 2017. The remote sensing measurements were complemented by regular AirCore (Karion et al., 2010) launches from the same site. AirCore measurements provide vertical profiles of the target gas concentrations as auxiliary reference data for the column measurements. The performances of the instruments were compared between themselves and to a reference TCCON instrument. The goal of this campaign was the characterisation of less expensive and more portable FTSs to complement TCCON for the establishment of a wider and denser network.

This paper is organised as follows: Section 2 gives a description of the campaign site, lists the details of the instruments taking part in the campaign and their evolution. Section 3 gives a description of the measurement strategy that was used to ensure comparable observations. Section 4 gives the description of the data and its availability. Section 5 gives the campaign results showing the intercomparison ~~between the test instruments with respect to the reference TCCON and low-resolution measurements performed by the TCCON instrument. It also gives the intercomparison~~ results between the TCCON, non-linearity corrected TCCON (TCCONmod) and AirCore data and ~~the~~ results using the AirCore profile as a priori for the FTS retrievals. It also gives the intercomparison results between the test instruments with respect to the reference TCCONmod. Section 6 concludes the paper by giving a summary of the results.

2 Measurements at Sodankylä and campaign instrumentation

2.1 Description of the campaign site

The Finnish Meteorological Institute (FMI) Sodankylä facility was selected as the campaign site as it fulfilled all selection criteria: (I) availability of TCCON measurements at the site, (II) possibility to launch, retrieve and analyse AirCore, (III) infrastructure to host all participating instruments and (IV) local support by scientists / engineers in case of problems occurring with the instruments during the campaign. The Sodankylä facility is located above the Arctic Circle in northern Finland, 67.3668° N, 26.6310° E, 188 m.a.s.l about 6 km south of Sodankylä. ~~The~~ 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. The site is equipped with a stratospheric balloon launch facility. The AirCore system has been operated by FMI to perform regular balloon launches since early September 2013. AirCore and other balloon payloads can be launched within 200 meters from the TCCON instrument. In addition, the site also has a mobile system to launch payloads from an upwind site in order to retrieve them in the vicinity of the TCCON site. Upon its recovery, the analysis of the AirCore is done on site using a Picarro G2401 analyser. Continuous surface in-situ measurements of CO_2, CH_4 and CO are performed from a 50 meter tower located at 500 meters away from the TCCON instrument. Further details on the site can be found in Kivi and Heikkinen (2016). An air-conditioned laboratory container ($\sim$ 9.1 meter long) was set up for the deployment of the visiting instruments for the campaign. The laboratory was placed about 30 meters south from the building hosting the TCCON instrument.

2.2 Instruments

The TCCON spectrometer, a Bruker IFS 125HR, was the main reference instrument for this campaign. Four ~~low resolution~~ low-resolution portable instruments participated in the campaign: a Bruker EM27/SUN, a Bruker Vertex70, a Bruker IRcube, and a homemade Laser Heterodyne spectro-Radiometer (LHR). Each of the three Bruker ~~low resolution~~ low-resolution instruments is based on a RockSolid^{TM} corner-cube pendulum interferometer. This allows for comparable sampling quality and robustness amongst the instruments. However, the instruments differ in the use of the surrounding imaging optics and their geometric arrangement which defines the interferometric field-of-view (FOV) and thus determines the instrumental line shape (ILS) of the respective instrument. The position of the centre burst, which determines the resolution differs for each instrument. The EM27/SUN records double sided and the IRcube single sided interferograms yielding a maximum resolution of 0.5 cm^{-1}. The Vertex70 records single sided interferograms giving a maximum resolution of 0.16 cm^{-1}. 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. All instruments used solar trackers with an active feedback loop to track the sun with an accuracy better than 0.1 mrad either with the help of active quadrant diodes or by active camera positioning. All low-resolution test instruments have the advantage that they do not need to be disassembled for transport. A detailed description of the instruments is given in the following sub-sections and some of the key features of the instruments, measurement properties and retrieval strategies during the campaign are listed in Tables 1 & 2.

2.2.1 Bruker IFS 125HR

The instrumental and operational setting of the Bruker IFS 125HR in the TCCON mode of operation can be found in detail in Kivi and Heikkinen (2016). The TCCON instrument's operation, maintenance and data analysis was performed by FMI. The measurements were performed at a spectral resolution of 0.02 cm^{-1} in vacuum (<1 hPa) to improve the stability and to reduce water vapour in the system. They were recorded using ~~room-temperature (RT) Indium Gallium Arsenide (InGaAs)~~ RT InGaAs

and RT Silicon (Si) detectors. The recorded signal (interferogram) was stored in DC mode in order to make corrections for the solar intensity variations (Keppel-Aleks et al., 2007). The interferogram upon DC correction was then Fourier transformed to get the corresponding spectrum. Column abundances of CO_2, CH_4, CO, N_2O, H_2O, HDO, O_2 and HF were retrieved from the spectra based on the TCCON GFIT retrieval code GGG2014 software version (Wunch et al., 2015). The instrument was also equipped with a liquid nitrogen (LN2) cooled ~~Indium Antimonide (InSb)~~ InSb detector. This detector enhances the possibilities to expand the wavelength region covered by the instrument (see Table 2) and to retrieve more atmospheric species. In addition to the TCCON and InSb measurements, the instrument was also used to record double-sided DC coupled interferograms at 0.5 cm^{-1} using the InGaAs detector. These measurements are henceforth called HR125LR. These measurements provide low-resolution data sets performed with the same TCCON instrument to be compared to the results of the other tested low-resolution instruments. The sequence of measurements was that first one InGaAs / Si forward-backward scan (standard TCCON measurement) was recorded, then two forward-backward HR125LR scans, after that again one standard TCCON measurement and two forward-backward HR125LR scans followed by one InSb forward-backward scan. This cycle was repeated for the whole measurement day. This paper focuses on the measurements performed with only the InGaAs detector (standard TCCON and HR125LR data sets). The instrument was operated in an automated way with the possibility of manual intervention. The ILS characterisation was performed using a HCl (hydrogen chloride) gas cell following the recommendations of TCCON (Hase et al., 2013) using the LINEFIT software (Hase et al., 1999).

2.2.2 Bruker EM27/SUN

The EM27/SUN spectrometer has been developed by Karlsruhe Institute of Technology (KIT) in cooperation with Bruker starting in 2011 (Gisi et al., 2012). The spectrometer is available as commercial item from Bruker since 2014, an additional channel for CO detection has been assigned in 2016 (Hase et al., 2016). Today already more than 40 units are operated by working groups around the globe (Frey et al., 2019). The EM27/SUN used during the campaign was provided by KIT. The EM27/SUN records double-sided DC coupled interferograms making an average of 10 scans in about 58 sec at a spectral resolution of 0.5 cm^{-1}. A double-sided recording of the interferograms largely reduces the sensitivity to residual phase error. The measurements were performed using a RT InGaAs detector (5500–11000 cm^{-1}) and a DC-coupled wavelength extended RT InGaAs detector (4000–5500 cm^{-1}) (Hase et al., 2016). In this extended configuration, the EM27/SUN covers the same spectral region as TCCON and encompasses the spectral section as observed by TROPOspheric Monitoring Instrument (TROPOMI) (TRO, a, b). Spectra were generated from raw interferograms using the ~~preprocessor~~ pre-processor tool developed by KIT in the frame-work of the COCCON-PROCEEDS project funded by ~~ESA~~the European Space Agency (ESA). Column abundances of CO_2, CH_4, CO, H_2O, and O_2 were retrieved from the resulting spectra using the PROFFAST retrieval code. ~~The~~ 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 H_2O, CH_4, N_2O, HITRAN 2008 line lists are used (in case of H_2O including some minor empirical adjustments), for CO_2 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. The XCO_2 and XCH_4 products are bias-corrected based on the extensive COCCON development. The bias correction is only done for the EM27/SUN and not for any other test data sets. The ~~codes are open source and licence free and~~ PROFFAST and the PREPROCESSOR tools can be downloaded from the KIT webpage http://www.imk-asf.kit.edu/english/3225.php. The characterisation of the ILS was performed using an open path measurement as described in Frey et al. (2015). The solar tracker of the EM27/SUN is attached to the body of the spectrometer. It was operated outside the FRM4GHG laboratory container at ambient conditions for the whole campaign period. This mode of deployment showed the capability of the instrument to be operated even under harsh campaign conditions. The day-to-day instrument operation was performed by KIT with local support from FMI for some measurement days. Once deployed, the instrument operation is automated. The EM27/SUN was supported by a pressure sensor and a GPS sensor for accurate timekeeping and position acquisition.

2.2.3 Bruker Vertex70

The Vertex70 spectrometer was purchased from Bruker to take part in the campaign. It records single-sided DC coupled interferograms making an average of 2 scans in about 17.3 sec at a spectral resolution of 0.2 cm^{-1}. The intensity of the interferogram varies during the scan and the incident angle on the two interferometer mirrors of the pendulum changes during the scan due to the large optical path covered by the pendulum drive, leading to self-apodisation. Both these factors were taken into account while performing the retrieval. Several scans were co-added for one measurement ($\sim$ 2.5 min) with a comparable signal-to-noise ratio (SNR) to the reference TCCON measurements. The Vertex70 has the advantage of accommodating and measuring with two detectors covering a wide spectral range. An extended RT InGaAs detector (3500–15000 cm^{-1}) and a LN2 cooled InSb detector (2500–10000 cm^{-1}) were used. This paper focuses on the measurements performed with only the InGaAs detector. The GFIT retrieval code was used to analyse the measured spectra and retrieve column abundances of CO_2, CH_4, CO, H_2O, and O_2. The characterisation of the ILS was performed using a HCl gas cell similar to TCCON. The Vertex70 was operated from inside the dedicated FRM4GHG air-conditioned laboratory container regulated at about 20 °C with the solar beam being fed to the instrument using a homemade BIRA-IASB solar tracker mounted on top of the container. The distance between the solar tracker and the spectrometer was 3 m. The tracking of the sun was performed using a camera-based active feedback option. The instrument operation was automated using the BARCOS system (Neefs et al., 2007) and a homemade

automated control unit system ~~build~~ built by BIRA-IASB with the possibility of a manual intervention at any time. The solar tracker was equipped with sun intensity and rain detection sensors which facilitated the automatic opening and closing of the solar tracker cover depending on the weather conditions. This facilitated performing atmospheric measurements on every occasion with good weather conditions. The data analysis was performed by University of Bremen and the maintenance was shared between BIRA-IASB and University of Bremen.

2.2.4 Bruker IRcube

The IRcube is a compact portable FTS manufactured by Bruker Optics. It records single-sided DC coupled interferograms using a RT extended InGaAs detector (4500–15000 cm^{-1}) making an average of 33 scans (17 forward and 16 backward) in about 1.7 mins at a spectral resolution of 0.5 cm^{-1}. It has an internal full angle FOV of 72 mrad. The novel design of the IRcube for this field campaign was the use of a fibre-optic ~~(FO)~~ feed from an independent solar tracker (a STR-21G, Eko instruments Ltd of Japan) mounted on top of the FRM4GHG laboratory container to receive the solar beam. A 50 cm focal length F/5 telescope (glass lens) focuses the solar beam onto a 20 m long, 600 μm core fibre with a numerical aperture ~~(NA)~~ of 0.22. This defines the external FOV on the solar disk at 1.2 mrad. The coupling of light from the optical fibre to the IRcube was chosen to optically match the input optics of the IRcube as closely as possible; coupling the power from the ~~FO~~ fibre-optic to the spectrometer so that the signal-to-noise ratio is comparable to TCCON, while avoiding unwanted spectral features that are present in NIR optical fibres. There is a limited range of ~~NAs~~ numerical apertures commercially available, out of which the best compromise for the IRcube with good spectral characteristics was the low OH Thorlabs FG550LEC. A glass lens and aperture in front of the IRcube refocuses the solar beam from the fibre into the entrance aperture (0.5 mm). A small part of the main beam reflected from the CaF_2 entrance window was used to monitor the solar radiation for cloud filtering. The IRcube can be housed anywhere within the length of the fibre-optic cable (here 20 m). This design concept is of significant importance for certain applications where the spectrometer can be placed far away from the solar tracker e.g., inside a weather proof enclosure. During this campaign the IRcube was set up by the University of Wollongong inside the FRM4GHG container and the operation of both tracker and IRcube was automatic. The characterisation of the ILS was performed using an open path measurement similar to the procedures followed by the EM27/SUN. The data analysis was performed by the University of Wollongong using the GFIT retrieval code.

2.2.5 Laser Heterodyne spectro-Radiometer (LHR)

The LHR is a research instrument developed by the Spectroscopy Group of the Space Science and Technology Department of the Rutherford Appleton laboratory (RAL) (Weidmann et al., 2007; Tsai et al., 2012; Hoffmann et al., 2016). The principle of operation is similar to that of a heterodyne radio-receiver; however, the LHR operates in the mid-infrared region of the spectrum. The benefits of such an approach to spectroscopy include: (I) High spectral resolution (up to >500,000 resolving power), (II) ideally shot noise limited radiometric noise, (III) intrinsic narrow FOV and (IV) scalable down to ultra-miniaturised packages through optical integration.

Compared to the laboratory instrument reported in (Hoffmann et al., 2016), the LHR was re-engineered to the requirements of the FRM4GHG campaign with the following modifications: I) The optical path was reworked to bring the instrument package down to 40x40x20 cm^3. II) A secondary laser channel (to be equipped in future) was integrated. III) A thermoelectrically cooled Mercury Cadmium Telluride (HgCdTe) photodiode for photomixing was installed to avoid LN2 usage. IV) A solar disk imager was installed for FOV monitoring and optional solar tracking operations. V) Acquisition, instrument control hardware and software were integrated to allow full unattended operation, except for switch on and off procedures.

The LHR was installed inside the FRM4GHG container and operated under ambient conditions. The incoming solar beam had a 12 mm diameter and was side-sampled from the BIRA-IASB solar tracker. The LHR has no entrance window. Inside the instrument, the incoming beam is split into a transmitted mid-infrared component for heterodyning and a visible component for solar imaging. To that end, a Germanium (Ge) long-wave infrared bandpass filter is used. To carry out the fine spectral analysis, the incoming mid IR field is superimposed with that of an optical local oscillator ~~(LO)~~ by a Zinc Selenide (ZnSe) beamsplitter. The ~~LO~~ local oscillator consists of a continuously tunable semiconductor laser source, in this case a quantum cascade laser, operating in the narrow spectral range between 952–955 cm^{-1} ($v_1 \leftarrow v_3$ CO_2 band) optimised through prior analysis to optimise atmospheric state retrieval information. The spectra were resolved through ~~LO~~ the local oscillator continuous frequency tuning. The superimposed atmospheric ~~beam~~ and local oscillator ~~one~~ beams are mixed onto the high speed photodiode, effectively transposing the middle infrared spectral information into the radio-frequency (RF) domain. The spectral resolution is determined by electronic filters. For the FRM4GHG campaign, the spectral resolution was set to 0.02 cm^{-1}. Each spectrum was recorded over 30 s. The start and stop operation of the LHR was performed manually by the local support staff at the measurement site. A typical atmospheric spectrum showing the CO_2 window as measured by the LHR can be seen in Fig. 6 in Hoffmann et al. (2016). The data analysis was performed by the RAL team using the optimum estimation atmospheric retrieval method, in which the Reference Forward Model was used (Dudhia, 2017).

2.2.6 AirCore

The AirCore is a novel innovative technique to sample high-altitude profiles of atmospheric concentrations of trace gases. A detailed description of the technique can be found in Karion et al. (2010). The AirCore system used for this campaign was originally built by the University of Groningen (RUG) and was further developed together with the Finnish Meteorological Institute (FMI). The total length of the AirCore is 100 m. It consists of two types of stainless steel tubing with outer diameters of 1/4" and 1/8", respectively. The vertical resolution of measurements from the AirCore is 13.4 mbar for ~~an ambient pressure higher than~~ ambient pressures between the surface and 232 mbar, and 3.9 mbar for ~~an ambient pressure~~ ambient pressures lower than 232 mbar. A custom-made data logger by FMI was used to record temperature and ambient pressure of the AirCore tubing. An automatic valve was developed and installed prior to the campaign, which closed the inlet valve of the AirCore system upon landing. The AirCore was packed in a styrofoam box to protect it from damage during landing, with its inlet valve protruding through the styrofoam box. Magnesium perchlorate ($Mg(ClO_4)_2$) was used as dryer in the AirCore. The AirCore package includes tubing, connectors, valves, data logger and a box. The air volume of the AirCore is approximately 1400 ml. The AirCore was launched hanging on a 3000 g meteorological balloon (Totex TX3000). The payload included a Vaisala

RS92-SGPL radiosonde (Dirksen et al., 2014), an Iridium and GPS/GSM positioning device and a light-weight transponder. The balloon burst after reaching the ceiling height (typically about 30–35 km). A large parachute was used to slow down the descent speed of the AirCore while a tracking system located its position. Upon landing, the AirCore was recovered and brought to the laboratory to obtain mole fractions of CO_2, CH_4 and CO with a Picarro G2401-m Cavity Ring-Down Spectrometer (CRDS). The precision/accuracy of the CO_2, CH_4 and CO are 0.05 ppm/0.1 ppm, 0.5 ppb/1 ppb and 8 ppb/3 ppb, respectively. An orifice (Sapphire, Type A, size 0.18 mm) was placed between the pump and the analyser to achieve a constant flow of 40 ml/min. The sample was analysed starting from the stratospheric part (the closed end) to minimise the diffusion. Before each flight, the AirCore was flushed with dry air from a fill cylinder for several hours. This procedure dries the inner surface of the AirCore and fills it with air of known mole fractions. The mole fraction of CO in the fill cylinder was ~~12000 ppb~~12 ppm. The fill air was used as an indicator of air mixing and as a diagnostic tool. Radiosonde (Vaisala RS92-SGPL) ambient pressure, temperature and AirCore temperature were available for each AirCore flight. AirCore vertical profiles were retrieved based on the measured time series of mole fractions and the recorded in-flight information, e.g., coil temperature, ambient pressure and ambient altitude using a custom made retrieval software by RUG.

2.2.7 In-situ

The in-situ measurements used for this work were provided by the FMI. The concentration of CO_2, CH_4 and CO were measured on a 50 meter tower at 3 levels (2 m, 22 m and 48 m) above surface using a Picarro G2401 system. More information about the site can be found on the webpage at http://fmiarc.fmi.fi/index.php.

3 Description of the measurement strategy to ensure comparable observations

3.1 Measurement set-up

The campaign took place between March and October 2017. The site is located at high latitude therefore it was not possible to measure beyond this period due to the high solar zenith angle (SZA). Solar measurements were recorded between sunrise and sunset, depending on the SZA limits set by the local scene and weather conditions (cloud, fog and strong winds). The FMI team monitored the operation of the instruments during the campaign period. Depending on the weather conditions, all spectrometers performed as many measurements as possible to improve the measurement statistics. The ~~continuous operation in the presence of the sun and good weather condition~~ measurements preformed helped to observe diurnal variation of the target gases. The campaign began with an initial blind intercomparison phase where the instruments were operated with the optimised settings best known to their PIs to get a good SNR comparable to the TCCON instrument. The measurements performed by the different remote sensing instruments were submitted to the chosen referee BIRA-IASB.

The intercomparison study of the blind phase showed that the Vertex70 instrument was not optimised and needed a modification. The aperture was reduced on 6 July 2017, such that the beam diameter changed from 40 mm to 20 mm, reducing the intensity of the light reaching the detector. This helped to reduce the scatter in the retrieved column values by ensuring

the operation of the instrument in the linear region of the detector. This configuration was used until almost the end of the measurement period when the aperture stop was further reduced with an iris to 9 mm. However, we did not have any solar measurements with this setting due to unsuitable weather conditions.

The IRcube did not have to undergo any internal modifications however ~~a broken optical fibre had to be~~ an optical fibre which was broken on 23 March 2017 was replaced in April 2017 and the measurements resumed again as of 25 April 2017. The first ~~FO~~ fibre-optic used for the IRcube was an ultra-low OH silica optical fibre from Polymicron Technologies, part FIA8008801100 with a ~~NA~~ numerical aperture of 0.22 and a core diameter of 800 μm. Due to a long delivery time of this ~~FO~~fibre-optic, a replacement ~~FO~~ fibre-optic as discussed in section 2.2.4 was ordered and used ~~after~~ since the end of April 2017.

The EM27/SUN was operated without any modifications during the whole campaign period. The exact dates of all performed modifications are shown in Table 3.

A total of 10 AirCore launches were performed during the campaign and these were used as an in-situ reference data set to better understand the intercomparison of the remote sensing data. Further details are discussed in section ~~5.4~~5.2.1 and section 5.3.

3.2 Instrument characterisation

All teams performed a full functionality test of their respective instruments and accessories before shipping and upon arrival at the campaign site in Sodankylä. The functionality test included quality checks as well as performing ILS measurements of the instruments. These measurements serve as reference to check the effects (if any) of transport on the instrumental properties and to ensure nominal operation in case of new set ups. During the campaign all teams performed ILS measurements when possible to monitor the long-term stability of the participating instruments. The modulation efficiency ~~(ME)~~ of the TCCON instrument at the ~~highest~~ maximum OPD was <1.02 with a phase error ~~(PE)~~ in the range of $\pm$2 mrad throughout the year. The ~~ME~~ modulation efficiency of the EM27/SUN at the ~~highest~~ maximum OPD was about 1.02 with a ~~PE~~ phase error in the range between -3 mrad and 1 mrad throughout the year. The ~~ME~~ modulation efficiency of the Vertex70 before shipping and upon arrival at the Sodankylä site was about 0.935 at 4.5 cm OPD and the ~~PE~~ phase error was changing between -16 and -36 mrad. The ~~ME~~ modulation efficiency improved significantly from 0.935 to about 0.973 and the ~~PE~~ phase error improved to about -13 mrad after the modification of the Vertex70 with the introduction of the additional aperture. The IRcube has a ~~ME~~ modulation efficiency of about 0.95 with the ~~PE~~ phase error in the range between -5 and +1.5 mrad. A summary of the ILS properties of the FTS is given in Table 3. The ILS of the LHR was determined by the radio frequency (RF) filter characteristics used to limit the detector bandwidth and hence the spectral resolution of the instrument and is therefore an inherent property of the instrument. A detailed description of the ILS validation of the LHR with C_2H_4 gas cell measurements can be found in a technical document by Hoffmann et al. (2017). None of the instruments showed any sign of degradation of the instrumental properties during the whole campaign.

4 Data description

The raw measurements (level 0 data) from all participating remote sensing instruments are made publicly available with the DOI https://doi.org/10.18758/71021040 (Sha et al., 2018). The atmospheric concentration of the trace gases (level 2 data) together with the auxiliary data are made publicly available with the DOI https://doi.org/10.18758/71021048 (Sha et al., 2019). All data sets and the ~~documentations~~ documentation are also made publicly available via the project webpage (http://frm4ghg.aeronomie.be) as well as via the ESA Atmospheric Validation Data Centre (EVDC).

5 Campaign results

5.1 Intercomparison data

Sodankylä is located within the Arctic Circle therefore solar measurements with sufficiently low SZA are only possible from the beginning of March to the end of October. During the month of September and October we had mostly overcast sky. Only three days of measurements were possible with the TCCON instrument during ~~the periodbetween the middle of September and the end of October. These measurements however have to be filtered out from the intercomparison study as they were measured at~~ this period. However, these measurements were recorded with SZA > 75°.

Based on the measurement capabilities by the individual instruments, the groups were asked to provide some or preferably all of the following parameters: Measurement day/time; ground pressure; total column amounts of O_2, H_2O, CO_2, CH_4, CO; and column averaged dry air mole fraction of the gas (Xgas) values for XCO_2, XCH_4, XCO. Xgas is defined by the following equation:

$$\mathrm{Xgas} = \frac{\mathrm{gas}_{\mathrm{column,dry}}}{\mathrm{O}_{2,\mathrm{column,dry}}} \times 0.2095 \tag{1}$$

where 0.2095 is the dry air O_2 mole fraction.

For the FTIR instruments also ~~Xair was submitted, which is~~ the column averaged dry air mole fraction of dry air (Xair) was submitted. Xair is dependent on the total column amounts of measured oxygen, surface pressure and water vapour column. It is calculated following equation 3 described in Wunch et al. (2015). Xair is a measure of the instrument's performance ~~. Xair is calculated from the scaled ratio of the surface pressure to the O_2 column amount analogous to equation~~ and is used by TCCON to examine station-to-station biases. Ideally, the Xair values should be 1 ~~described in Deutscher et al. (2010) with the value of 1.0 for a perfect measurement and a perfect pressure measurement~~for measurements of total column amounts of oxygen with accurate spectroscopy, surface pressure and water vapour retrievals. Typical Xair values for TCCON measurements are 0.98 which is because of a 2% bias in the O_2 spectroscopy. A summary of the data sets and the corresponding retrieval methods is provided in Table 2. The spectrometers used an identical set of ground-pressure data collected at the Sodankylä site for the retrieval. The Xgas values which were calculated using GFIT were scaled to the WMO standards using the calibration factors used by TCCON and as discussed in Wunch et al. (2015). The recent values of the correction factors (airmass dependent correction factor (ADCF) and airmass independent correction factor (AICF)) for the respective gases were taken from Table 4

in Wunch et al. (2015). The scaling factors for the Xgas values which were calculated using PROFFAST for the EM27/SUN are discussed in detail in Frey et al. (2015).

~~The optical fibre of the solar tracker of the Bruker IRcube was broken on 23 March~~ All interventions performed on the respective instruments and as discussed in section 3.1 are marked in the time-series plots with vertical lines and the colours corresponding to the respective instrument. The dates are given in Table 3. In the following sections the intercomparison results will be shown, the long-term stability will be discussed and cases where clear deviations of the retrieval results from the participating instruments w.r.t. the reference data set are observed will be explained.

5.2 Detector non-linearity effects

The reference measurements performed with the Bruker IFS 125HR during the campaign in 2017 ~~, which led to an interruption of measurements until it was replaced by the end of April. The blind intercomparison study revealed that the optical setting of the Bruker Vertex70 was not optimised. It was further improved on 06 July~~ are found to be affected by the non-linearity of the InGaAs detector. The non-linearity was identified towards the very end of the campaign in 2017 ~~when a20 mm aperture stop was introduced in the parallel light beam falling on the detector. This configuration was used until almost the end of the measurement period when the aperture stop was further reduced with an iris to 9 mm. However, we did not have any solar measurements with this setting due to unsuitable weather conditions. All three interventions on~~ while checking the interferogram signal measured by the TCCON and comparing it to the ~~respective instruments are marked in the time-series plots with vertical lines and the colours corresponding to the respective instrument .~~ EM27/SUN. The detector non-linearity is dependent on the photon load incident on the detector and influences the Xgas values dependent on the signal strength of the measurements. The non-linearity being a signal dependent function, can be avoided by keeping the signal level within the linear domain of the detector. To test the non-linearity, a metal grid was placed in the parallel light beam at the entrance port to reduce the signal by about 20%. Figure 1 shows two spectra measured with the standard TCCON configuration with no grid (red) and with a grid (black) placed in the parallel light beam. These spectra cover the complete spectral regions measured by the detector and are zoomed in to highlight the signal of the out-of-band spectral regions. The non-linearity effect leads to out-of-band artefacts in the spectrum falsely indicating the presence of energy where the detector is insensitive. The signal between 0 cm^{-1} and the lower cutoff of the detector at 4000 cm^{-1} as well as the signal between the upper cutoff at about 12000 cm^{-1} and the end of the detector bandpass at about 16000 cm^{-1} show non-zero values for the no grid case indicating that the measurements performed were affected by the detector non-linearity. However the measurements performed with the reduced intensity by introducing the grid in the parallel beam do not show such high out-of-band intensities. The lower wavenumber out-of-band region shows only noise values and the higher wavenumber region close to the detector bandpass shows values which are higher than the noise but much lower than the signal of the standard measurements. 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. The presence of the signal, as a result of this double passing, is superimposed on to the non-linearity artifact of the detector in this wavenumber region which makes this spectral region unusable for the determination of non-linearity. The high signal in the out-of-band spectral regions confirms that the TCCON

measurements performed during 2017 are affected by the detector non-linearity. A correction method has been developed based on the method described in Hase (2000, chap. 5), it has been tested and applied to the TCCON data. The results of this are shown in the appendix A. The non-linearity corrected TCCON data are henceforth referred to as TCCONmod in this paper. The AirCore measurements performed during the campaign were used to compare with the TCCON and TCCONmod data sets. These results are discussed further in the next section.

5.2.1 Intercomparison results of the Xgas calculated from AirCore relative to the TCCON and TCCONmod data set

AirCore measurements performed in 2017 at the Sodankylä site are listed in Table 4. The retrieval of the TCCON and TCCONmod data set were performed using the TCCON a priori. The daily a priori files were automatically generated during the GFIT run. In addition, the tool to generate the daily TCCON a priori for any given location is available using a stand-alone program via a DOI link provided by Toon and Wunch (2017). The AirCore measurements are in-situ measurements of the targeted species calibrated to the WMO scale and serve as a better reference for the vertical profile of the measured species. However, the AirCore profiles are limited to a vertical sampling height of about 25–30 km depending on the ceiling height reached by the launching balloon. Given this height limitation, the AirCore profiles cover only a part of the atmosphere relative to the TCCON a priori profile which covers a larger range starting from the site altitude up to 70 km. The lowermost layer of an AirCore profile is contaminated as the sampled air of the lowermost part of the atmosphere gets mixed with the reference push-gas. The push-gas is needed to let the sampled air pass through the analyser. The in-situ measurements performed at 2 m height above ground-level from a nearby ICOS (Integrated Carbon Observation System; https://www.icos-ri.eu/) ecosystem tower (https://en.ilmatieteenlaitos.fi/GHG-measurement-sites#Sodankyla) were used to substitute the concentrations of the lowermost layer of the measured AirCore profile. The AirCore profile above the top most measured layer was further extended by a scaled TCCON a priori profile to cover the missing profile information up to 70 km of altitude. This is equivalent to a filling of $< 5\%$ of the total column above the top height of an AirCore measurement. The modified profile constructed using the ground-based in-situ measurement, AirCore measurement and the scaled TCCON a priori profile for three sample days on 24 April, 15 May and 28 August 2017 are shown in Fig. 2. The three rows show the measured AirCore profiles (blue rectangles), the a priori profiles from the GFIT run (black plus), the tower mast measurements (green rectangle) and the extended AirCore profiles (red circles) for three days. These three days were chosen to show the variability of the a priori profile during the different seasons at the Sodankylä site. The left-column represents the data plotted for XCO_2, the middle-column represents the plots for XCH_4 and the right-column represents the plots for XCO as a function of the altitude for 24 April (top-row), 15 May (middle-row) and 28 August 2017 (bottom-row), respectively.

The Xgas values are calculated directly from the modified AirCore profiles by using the TCCON averaging kernels (AKs). These Xgas values are then used to compare to the Xgas values retrieved from the standard TCCON and the non-linearity corrected TCCON data sets. Any difference in the intercomparison results is a direct reflection of the difference between the measured AirCore profile and the ground-based in-situ data relative to the TCCON a priori for the same altitude coverage. The time corresponding to 90% of the profile (starting at the top of the atmosphere) acquisition time is taken as the AirCore timestamp for the intercomparison of the Xgas values. A 3h time window around the AirCore measurement time was used as

the co-incidence limit. All Xgas values from TCCON data and TCCONmod data in this time window were averaged and taken as the coincident data sets for the intercomparison. The 3hr time window was selected for the remote sensing measurement as it is a good representation of the AirCore measurements. Reducing the time window resulted in the reduction of co-located measurement days and increasing the time window introduced the true variability of the atmospheric state in the remote sensing data.

The mean bias, the standard deviation of the difference and the correlation coefficient of the Xgas values calculated from the AirCore relative to the TCCON and the TCCONmod are shown in Table 5. The XCO_2 mean bias between AirCore and TCCON is 0.47 ppm with a standard deviation of 0.66 ppm and a correlation coefficient of 0.994. The mean bias is reduced significantly to -0.03 ppm for the intercomparison between the AirCore and the TCCONmod. The standard deviation of the difference is very similar, however the correlation coefficient improved slightly for the TCCONmod. This shows that the XCO_2 values from the TCCONmod data set are a better representation of the true atmospheric state.

The XCH_4 mean bias between the AirCore and the TCCONmod increases to -0.007 ppm as compared to the mean bias of -0.004 ppm between AirCore and TCCON. The scatter remains the same with an improvement in the correlation for the TCCONmod. The improvement in the correlation indicates that the TCCONmod data is a better representation of the true atmospheric state. The increase in the mean bias is due to the difference in the TCCON a priori profiles used for the retrieval relative to the true atmospheric profiles. The ~~dates are given in Table 3. In the following sections the intercomparison results will be shown , the long-term stability will be discussed and cases where clear deviations~~ top-panel of Fig. 3 shows the time series of a 30 min averaged TCCONmod XCH_4 data set and XCH_4 calculated from the AirCore measurements. The bottom-panel plot shows the difference in the XCH_4 bias. The large difference between the two data sets in April is due to the difference between the a priori from the true atmospheric state. The bias reduces significantly for all later AirCore measurement days.

The XCO mean bias between AirCore and TCCONmod is slightly reduced to 6.25 ppb as compared to the mean bias of 6.4 ppb between AirCore and TCCON. The scatter is almost the same with very similar correlation coefficients. The CO retrieval from the AirCore has a large uncertainty. As a result, the impact due to the change of the data set from the TCCON to the TCCONmod is within the uncertainty budget of the AirCore measurements.

The direct intercomparison results of the Xgas calculated from AirCore relative to the TCCON and non-linearity corrected TCCON data sets indicate clearly that the non-linearity corrected data set gives Xgas amounts which are closer to the AirCore amounts and hence closer to our best estimate of the true atmospheric conditions. We will therefore use the TCCONmod data set as our reference data set for further intercomparison studies in the main section of our paper. However, in the appendix B we also show the intercomparison results of the low-resolution measurements relative to the standard TCCON product which is not yet non-linearity corrected.

5.3 Intercomparison results using AirCore as a priori profile

The extended AirCore vertical profiles for the targeted gases derived from the AirCore flights have been fed as input a priori profiles for the retrieval of the respective gases from the measurements performed with the remote sensing instruments on the respective days. The retrieval results with the modified AirCore profiles have been given the suffix "AC" at the end of

the instrument name. As the remote sensing instruments covered a larger range of SZAs on 15 May and 28 August than on 24 April, those two days were selected for the intercomparison study. 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.

5.3.1 XCO_2 intercomparison results

The intercomparison results for XCO_2 retrieved using TCCON a priori and modified AirCore a priori for the TCCONmod and EM27/SUN data sets are shown in Fig. 4. The top and the bottom subfigures show the results for measurements performed on 15 May and on 28 August 2017, respectively. The same figures for the Vertex70 and the IRcube are shown in Fig. A4. The difference between the TCCON a priori and the modified AirCore a priori profiles is relatively small on 15 May as compared to the high difference of the profiles on 28 August (see Fig. 2). This implies that the TCCON a priori is closer to the true atmospheric state on 15 May than on 28 August. As a result the difference between the standard retrievals from each instrument using TCCON a priori and the retrievals using the modified AirCore a priori is smaller on 15 May compared to the difference on 28 August. The retrieval results for all instruments for the measurements on 28 August show a bias between the TCCON a priori and the modified AirCore a priori retrievals. The bias shows a strong dependency of the retrieval on the SZA of the measurements. This is due to the TCCON CO_2 AK dependence on the SZA as seen in the Figure 6 of Hedelius et al. (2016) . With these AK the a priori information is very relevant. The AirCore a priori is in principle the closest a priori to the truth. When applying the AirCore a priori and doing the retrieval we see that the airmass dependence is much reduced.

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 measurements as compared to the 15 May measurements. The intercomparison plots also show the scatter of the retrieval results from the individual instruments for two days. The EM27/SUN shows a lower scatter as compared to the TCCONmod due to the low noise resulting from the averaging of the individual measurements. 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.

5.3.2 XCH_4 intercomparison results

The intercomparison results for XCH_4 retrieved using TCCON a priori and modified AirCore a priori for the TCCONmod and EM27/SUN data sets are shown in Fig. 5. The top and the bottom subfigures show the results for measurements performed on 15 May and on 28 August 2017. The same figures for the Vertex70 and the IRcube are shown in Fig. A5. The difference between the TCCON a priori and the modified AirCore a priori profiles of CH_4 is the highest for 24 April followed by 15 May and the smallest for 28 August (see Fig. 2). The vertical distribution of the CH_4 concentration during the winter and spring period is poorly modelled by the TCCON a priori tool. The a priori during the summer is in better agreement with the AirCore measurements as seen for the 28 August profiles. As a result the difference between the standard retrievals from each instrument using TCCON a priori and the retrievals using the modified AirCore as a priori is smaller for 28 August than for 15 May.

The TCCON CH_4 AK dependence as a function of the SZA is shown in the Figure 6 of Hedelius et al. (2016). The AK values are >1 for measurements at a lower SZA which means that the retrieval overestimates the contribution from all layers above 10 km. However, the AK values are <1 for measurements with SZA >65° which means that the retrieval underestimates the contribution from all layers above 10 km. The TCCONmodAC results are higher than the TCCONmod results for the lower SZA values and vice versa. This effect is stronger for the retrieval results of 15 May compared to the results of 28 August where the TCCON a priori is closer to the AirCore a priori. The retrieval results for 15 May measurements for all instruments show a bias between the TCCON a priori and the modified AirCore a priori. The bias shows a strong dependency of the retrieval ~~results from the participating instrumentsw. r.t.the reference data set are observed will be explained.~~ on the SZA. The EM27/SUNAC results show a small bias compared to the EM27/SUN. The difference plot shows that the change of the retrieved XCH_4 values with the modified AirCore a priori has the same sign as compared to the TCCONmod. The same feature is also seen by the Vertex70 and IRcube results. The bias for the 28 August is largely reduced compared to that of the 15 May. The small remaining bias is due to the difference in the a priori and the AK of the instruments. The AK for the low-resolution instrument e.g., the EM27/SUN is shown in the top row of Figure 6 in Hedelius et al. (2016).

5.3.3 XCO intercomparison results

The intercomparison results for XCO retrieved using TCCON a priori and modified AirCore a priori for the TCCONmod and EM27/SUN data sets are shown in Fig. 6. The top and the bottom subfigures show the results for measurements performed on 15 May and on 28 August 2017. The same figures for the Vertex70 are shown in Fig. A6. The TCCON a priori and modified AirCore a priori profiles of CO for three days in 2017 are shown in Fig. 2. The AirCore measured CO profiles are provided for altitudes up to 17 km and in some cases as high as 19 km. 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. The difference in the profiles in the stratosphere is the largest for 24 April followed by 15 May and the difference is the smallest for 28 August. As a result the difference between the standard retrievals using TCCON a priori and the retrievals using the modified AirCore as a priori is slightly higher for 15 May than

for 28 August. The TCCON CO AK dependence as a function of the SZA is shown in the Figure 6 of Hedelius et al. (2016) . The AK contribution to the retrieval results are underestimated (AK values <1) for layers below 5 km and overestimated for layers above 5 km with AK values >1, even increasing up to above 2 for higher layers. The bias dependence on SZA is significant for measurements performed only at high SZA. The TCCONmodAC XCO retrievals show a constant bias relative to the TCCONmod XCO retrievals for most of the SZA and the deviation is seen only for measurements performed at the high SZAs. The EM27/SUN and the Vertex70 results also show a slight dependency of the XCO retrieved using the TCCON a priori and the modified AirCore a priori on the measurements performed at a high SZA and a constant bias for measurements performed at a low SZA.

5.4 Methodology for ~~data~~ the intercomparisons of the remote sensing data

The data acquisition of the level 2 products were different for each instrument (see Table 2 for details). 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., ~~TCCON~~TCCONmod) was matched with the same timestamp ~~of~~ as the other instruments to find the coincident data pairs which were used for the difference and the correlation calculation. The TCCON as well as the low-resolution instruments showed a strong air-mass dependence for measurements with SZA>75°; these data were therefore not included in this study. Filtering these data removed only a very limited fraction of the data set (about 5% for EM27/SUN and LHR, about 10% for IRcube and about 13% for Vertex70). Statistical values were computed from the coincident data set to obtain the bias, scatter and seasonal variation of the individual instruments with respect to a reference data set from the Bruker IFS 125HR. A linear regression line was fitted to the correlation data set for each gas. The slope, the intercept, the correlation coefficient and the standard error are shown on the respective correlation plots.

5.5 Intercomparisons with reference ~~TCCON~~ TCCONmod data

The intercomparison results with the ~~TCCON~~ TCCONmod data as reference and data from other ~~low resolution~~ low-resolution remote sensing instruments are discussed in this section species-by-species. All instruments performed the retrievals following their standard procedure and using TCCON a priori as the common prior. The statistical values for the intercomparison results (mean of the bias, the standard deviation of the difference and the Pearson correlation coefficient) are given in Table ~~7~~ 6 and plotted in the overview summary plot of Fig. 11.

5.5.1 XCO_2 intercomparison results

The timeseries of the coincident XCO_2 values measured during the year 2017 by each test instrument and the reference ~~TCCON instrument~~ TCCONmod are shown in the top-panel of Fig. ~~??~~7. The corresponding differences relative to the ~~TCCON instrument~~ TCCONmod are shown in the second row panel. The correlation plots between the test instruments and the ~~TCCON instrument~~ TCCONmod are shown in the individual panels of the third and the fourth rows of Fig. ~~??~~7. The measured XCO_2

values are high during the early winter and low during the summer season which represents the annual seasonal cycle at the site. All instruments captured the annual summer drawdown.

Amongst the test FTS instruments, the EM27/SUN has the lowest mean bias of ~~-0.18~~-0.73 ppm with a standard deviation of ~~0.45~~0.47 ppm and a very high correlation coefficient of ~~0.995~~0.996. The difference plot (second row panel of Fig. ~~??~~7) as well as the correlation plot (bottom-right panel of Fig. ~~??~~7) show a small seasonal dependency of the bias relative to the ~~TCCON instrument~~TCCONmod.

The correlation plot of the bottom-left panel of Fig. ~~??~~ 7 shows a step change of the XCO_2 values for the IRcube in March as a result of the replacement of the fibre-optic which caused a change of the ILS of the instrument. The IRcube data show high bias and have a small seasonal dependency. This may be because of the poorly defined ILS due to compact short focal length optics or detector non-linearity.

The Vertex70 also has shown a step change relative to the ~~TCCON instrument~~ TCCONmod since its modification in July 2017. The data set after the instrument modification shows a significant reduction in scatter and bias as compared to the earlier data from the campaign. As a result, data from the period between 06 July and 12 September 2017 are compared separately to characterise the behaviour of the Vertex70 relative to the ~~TCCON instrument~~ TCCONmod and the other test instruments. The statistics for the data for the selected period are shown in the lower part of Table ~~7~~ 6 and are plotted in the summary plot of Fig. 11. The data from the Vertex70 show a significant reduction in bias from ~~1.93~~1.46 ppm to ~~0.22~~-0.16 ppm, and the standard deviation from ~~1.72~~1.63 ppm to ~~0.58~~0.57 ppm, while the correlation coefficient remained still high. The Vertex70 and EM27/SUN measurements are comparable to each other for this period. The mean bias and the standard deviation of the other instruments are quite similar for the July-September period as compared to the full year. However due to the limited data set, the correlation coefficient is slightly poorer for the shorter period.

The LHR instrument is in its developmental phase and measured only CO_2 and H_2O. XCO_2 data were found to be affected by two clearly different noise processes: A high frequency random error, mostly determined by detector noise, was found ranging from 2 to 5 ppm (one sigma) depending on the instrument SNR. On top of this random error, large slowly varying diurnal biases were observed to be up to $\sim$10 ppm. All bias included and averaged, the biases against the ~~TCCON instrument~~ TCCONmod for the full year were found to be ~~18.4~~-18.9 $\pm$ 5.3 ppm. These biases were found to be inherent to the re-engineered LHR instrument in contrast to the better controlled laboratory one (Hoffmann et al., 2016). They are under studies; some instrumental ones have been identified to step from laser optical feedback and laser excess noise producing a variable offset in the heterodyne demodulated signal.

5.5.2 XCH_4 intercomparison results

The timeseries of the coincident XCH_4 values measured during the year 2017 by each test instrument and the reference ~~TCCON instrument~~ TCCONmod (top panel plot), the corresponding differences relative to the ~~TCCON instrument~~ TCCONmod (middle row panel) and the correlation plots (bottom row panels) are shown in Fig. ~~??~~8. XCH_4 values are high during the late winter followed by a dip during the spring and further rising during the summer period. The annual cycle can be seen for the

~~TCCON~~TCCONmod, EM27/SUN and IRcube measurements. The Vertex70 data, after the instrument modification in July, are also representative of the ~~TCCON measurements~~TCCONmod data set.

The EM27/SUN has the lowest mean bias of ~~0.003 ppm~~ zero with a standard deviation of ~~0.005~~0.004 ppm and a correlation coefficient of ~~0.962~~0.973. The difference plot shown in the middle row panel of Fig. ~~??~~ 8 shows that both the EM27/SUN and IRcube have a high bias of about 0.01 ppm with respect to the ~~TCCON instrument~~ TCCONmod in the period between early March and the end of May. Also the correlation plots relative to the ~~TCCON data~~ TCCONmod shown in the bottom-middle and the bottom-right panel of Fig. ~~??~~ 8 for the IRcube and the EM27/SUN show the monthly deviation very clearly.

The Vertex70 data show a step change in bias of about 0.03 ppm and a significant reduction in the measurement standard deviation after the instrument modification. The statistical values for all instruments between 06 July and 12 September 2017 are shown in Table ~~7~~ 6 and are plotted in Fig. 11. The Vertex70 data ~~have the largest bias of 0.011 ppmfollowed by the IRcube with a bias of -0.008~~has a bias of 0.01 ppm, the IRcube data has the bias of -0.01 ppm and the EM27/SUN ~~with a zero bias~~ data has a bias of -0.002 ppm relative to the ~~TCCON instrument~~TCCONmod. The positive bias of the Vertex70 still remains after the instrument modification and the annual cycle is also captured. The standard deviations of the measurements from the three test instruments are comparable. ~~The correlation coefficient for this shorter period is slightly improved for the EM27/SUN relative to the full period. This points to the large difference in the XCH_4 values between March and May as mentioned earlier. The reasons for this will be discussed in section C2.~~

5.5.3 XCO intercomparison results

Carbon monoxide is measured by TCCON, Vertex70 and EM27/SUN instruments. The timeseries of the coincident XCO values measured during the year 2017 by these instruments (top panel), the corresponding differences (middle row panel) and the correlation plots (bottom row panels) relative to the ~~TCCON instrument~~ TCCONmod are shown in Fig. ~~??~~9. XCO values during the start of the measurement period in late winter are high, followed by a dip during summer and rising values during the late summer period. The annual cycle is seen by all instruments.

The EM27/SUN has the ~~lowest~~ mean bias of ~~4.54~~4.38 ppb with a standard deviation of ~~1.37~~1.36 ppb and a high correlation coefficient of 0.993. The difference plot shows that the bias is seasonally dependent with high scatter during the summer period due to measurements with large SZA variation performed on long summer days.

The Vertex70 data show a significant improvement in the scatter after the instrument modification. The statistics showing the mean, standard deviation and correlation coefficient are given in the bottom part of Table ~~7~~6. The Vertex70 result has a high correlation coefficient of 0.991 for the period after the instrument modification. The scatter and outliers are reduced, the comparison shows mean bias of ~~1.47~~1.34 ppb and standard deviation of 1.04 ppb relative to the ~~TCCON instrument~~TCCONmod data set.

5.5.4 Xair intercomparison results

Xair values were submitted by all FTIR instruments. The timeseries of the coincident Xair values for the year 2017 for the instruments are shown in the top panel plot of Fig. ~~??~~ 10 and the corresponding differences relative to the ~~TCCON instrument~~

TCCONmod are shown in the middle panel plot. Ideally Xair being the scaled ratio of the surface pressure divided by the retrieved total column of oxygen values should be 1. Any difference relative to the ideal case is an indicator for the instrument and retrieval code performance and the spectroscopy.

The Xair values of the Vertex70 show two distinct groups due to the instrument modification in July. After the instrument modification the scatter in the Xair values is significantly reduced. The EM27/SUN shows a slightly lower scatter compared to the IRcube for the full year. However, the scatter of the EM27/SUN, Vertex70 and IRcube are similar for the shorter time period. There is a small offset relative to the ~~TCCON instrument. However~~TCCONmod. However, the small offset in bias is less important than a stable Xair over the long timeseries. The correlation plots between the test instruments and the ~~TCCON instrument~~ TCCONmod are shown in the bottom panel plots of Fig. ~~??~~10. The bottom-left and bottom-middle panels show that the spread of the Xair values in the y-axis (representing Vertex70 and IRcube, respectively) are higher than those in the x-axis (representing ~~TCCON~~TCCONmod). The bottom-right panel shows that the spread in the x-axis (representing ~~TCCON) is slightly higher than~~ TCCONmod) is similar to the spread in the y-axis (representing EM27/SUN) except for a few outliers. The EM27/SUN shows the smallest airmass dependence whereas the Vertex70 and the IRcube show decreasing Xair values with an increasing SZA similar to the ~~TCCON instrument~~TCCONmod. This may reflect the difference between the GFIT and PROFFAST results and the use of the different spectroscopic line list as standardly used by the TCCON and COCCON communities.

The Xgas biases between the ~~low resolution~~ low-resolution test instruments and the ~~TCCON instrument~~ TCCONmod data sets as reference may be due to effects such as different responses to a priori profiles, interfering species in the retrieval windows or different averaging kernels. Furthermore it is important to note that, TCCON uses a network-wide constant scaling factor to scale its Xgas values to the WMO standards. The scaling factors specific to each gas for TCCON had been determined from several measurement campaigns where vertically distributed measurements of the gases were performed from airborne platforms using WMO calibrated instruments. The EM27/SUN uses species dependent scaling factors for XCO_2 and XCH_4 which had been calculated from long-term intercomparison measurements performed at the ~~KIT~~ Karlsruhe TCCON site. However, no such instrument specific calibration factors were applied for the other instruments and also for the XCO results from the EM27/SUN measurements. This also contributes to the residual bias which is observed in this intercomparison result. The biases which are purely due to resolution differences are addressed by performing ~~low resolution~~ low-resolution measurements with the same TCCON instrument. These data are then used for an intercomparison relative to the ~~standard~~ TCCON as well as for the intercomparison with other ~~low resolution~~ low-resolution test instruments. Further details of the intercomparison results are given in ~~sections~~appendix C and D, respectively.

5.6 ~~Comparisons between TCCON~~ Humidity dependencies of bias

The presence of water vapour lines in the retrieval windows can lead to errors in the determination of the Xgas values unless they are fitted well in the forward model. It is therefore necessary to check the influence of the water vapour lines for retrievals performed with the low-resolution instruments. Sodankylä is not the most humid TCCON site. The maximum XH_2O measured

by the TCCON is < 6000 ppm during the summer period. In comparison, the TCCON site at Darwin, which is a relatively humid site, show maximum measured XH_2O of < 10000 ppm during the summer period. The year 2017 was relatively dry where the range of XH_2O measured at the Sodankylä site was between 500 and ~~low resolution measurements performed with Bruker IFS~~ 4500 ppm. A detailed discussion of the bias dependence on the humidity present along the measurement line-of-sight is presented in section F. The results show that the Xgas values derived from the low-resolution instruments during the campaign period showed no dependencies on the humidity along the measurement line-of-sight.

6 Summary and outlook

The FRM4GHG campaign was successfully executed by comparing four portable remote sensing instruments against the reference TCCON instrument at the Sodankylä site during the year 2017. The EM27/SUN was set up every day at the ambient temperature and pressure and was operated without configuration changes during the whole campaign. The other low-resolution FTIR and the LHR were operated from inside a dedicated temperature controlled container. The instruments needed optimisation and behaved better with a low bias and a high correlation relative to the TCCON instrument afterwards.

In the course of the campaign not only the Vertex70, IRcube and LHR instruments have been improved but also the TCCON instrument by detecting and correcting non-linearity of the detector response. Detecting this issue by comparison with the EM27/SUN shows the potential of this instrument as a traveling standard for TCCON.

The intercomparison results using AirCore profiles as a priori provided interesting insights to the FTS retrievals, its sensitivity to the resolution and the averaging kernels. The AirCore profiles also showed the differences relative to the TCCON a prioris and the resulting biases in the retrievals of the target species. The Xgas calculated from AirCore and compared to the TCCON and the non-linearity corrected TCCON (TCCONmod) data sets show that the latter data set is a better representation of the true atmospheric state.

The EM27/SUN Xgas biases relative to the TCCONmod data were low for the target species except for the high XCH_4 bias during the March–May period which is due to the difference in the sensitivity of the high and low-resolution instruments and the a prioris not matching well with the actual profile shape. The EM27/SUN results include a instrument specific bias correction for XCO_2 and XCH_4 using scaling factors which has been determined independently prior to this study from long-term intercomparison measurements performed at the Karlsruhe TCCON site. It may be that the scaling factor is not optimal for the current location and is also contributing to the bias. This needs to be verified for comparison measurements performed at other TCCON locations. The EM27/SUN Xgas values show high precision and good correlation relative to the reference data sets.

The IRcube Xgas values show relatively high biases which are related to the possible dependence of the signal level on the extended InGaAs detector known to have non-linearity characteristics. The ILS of the IRcube is also less ideal compared to other larger instruments due to the compact short focal length optics. The impact of the ILS on the biases is being further investigated. However the comparison shows low scatter and a good correlation relative to the TCCONmod data.

The Vertex70 was equipped with an extended InGaAs detector which led to identifiable non-linearity effects. The optical path was modified by introducing an aperture stop to avoid saturation and operate in the linear region of the detector to improve the ILS. The bias of the Xgas values, the standard deviation of the difference and the correlation of the modified Vertex70 instrument relative to the TCCONmod data were significantly lower after this instrument modification and comparable to the EM27/SUN results relative to the TCCONmod data.

The LHR was a new instrument deployed under test during this campaign. It showed large scatter and large biases with a strong diurnal variation relative to the TCCONmod and other FTS instruments. The LHR data for the 2017 campaign are not yet able to provide meaningful geophysical information. However this comparison has proven to be invaluable to characterise and understand the instrumental biases and possibly the retrieval biases. Both aspects are currently under investigation and improvements are being developed.

The intercomparison results showed that the non-linearity corrected TCCON data gave a better match with the low-resolution instruments. The standard deviation of the bias and the correlation coefficient was similar for the target species for the non-linearity corrected TCCON data relative to the standard TCCON data.

The intercomparison results of the low-resolution measurements performed with the Bruker IFS 125HR relative to the standard TCCON and other low-resolution instruments provided useful analysis of the resolution dependent effects on the Xgas retrieved for the target gases. The low-resolution measurements performed with the Bruker IFS 125HR also helped to identify that the high bias in the XCH_4 during the March–May period was caused by the resolution difference and the corresponding different sensitivities to the vertical profile shape as seen in the averaging kernels.

The airmass dependence of the retrievals is an effect of the software and spectroscopy. The EM27/SUN and the HR125LR results retrieved with PROFFAST do not show SZA dependence for species where an airmass correction factor was applied. Both these data sets were retrieved using PROFFAST. All other results show SZA dependence to some degree. The correction for the SZA dependence is a long-standing and ongoing issue for TCCON and relevant to all instruments. In order to minimise the effect of the SZA, measurements with SZA <75° were used for the intercomparison of the different data sets.

The bias dependence on the humidity along the measurement line-of-sight was investigated for each target species. However no dependence was found.

The EM27/SUN, the IRcube, the modified Vertex70 and the HR125LR provided stable and precise measurements of the target gases during the campaign. The portable low-resolution instruments can be used for campaigns or long-term measurements from any site, and complement the TCCON network. The Xgas measurements from these instruments will be of similar quality as the TCCON Xgas data.

Appendix A: Non-linearity of the Sodankylä TCCON InGaAs detector

The TCCON measurements performed during the campaign in 2017 are found to be affected by the non-linearity of the InGaAs detector used for the measurements. A correction method has been developed based on the method described in

Hase (2000, chap. 5) to correct the signal in the interferogram domain using the following equation:

$$I' = I + a \times I^2 + b \times I^3 + c \times I^4 \quad \text{(A1)}$$

where I is the original interferogram and I′ is the non-linearity corrected interferogram. The coefficients a, b and c are determined for each measurement by minimising the cost function. The cost function is defined as the ratio of the signal sum in the out-of-band regions covering 100–3600 cm^{-1} and 14200–15750 cm^{-1} to the signal sum for the in-band spectral region covering 4100–9700 cm^{-1}. The individual correction functions are plotted in Fig. A1 as red curves, their mean is plotted as blue open circles and the black line is the fourth order polynomial fit of the mean with the coefficient values of a = 0.00138, b = 0.00302 and c = 0.0166. The coefficients of the fitted curve representing the averaged correction function are used to correct the non-linearity of all TCCON measurements. The top-left and bottom-left panel plots of Fig. A2 show the original and the non-linearity corrected spectra for the measurement day of 06 September 2017. The colours of the spectrum depend on the interferogram maximum signal at the centre burst. The highest values represented by the dark red colour are for measurements performed with the highest solar intensity during noon time and the measurements performed with the lowest solar intensity are represented by blue as the minimum. The top-right and bottom-right panel plots of Fig. A2 show the original and the non-linearity corrected spectra for the out-of-band region between 100 and 3600 cm^{-1}. The signal reduction in the out-of-band spectral region clearly shows that the non-linearity correction worked for the spectra. In order to quantify the effect of the non-linearity correction on the Xgas values, explicit results are shown in detail for the 06 September 2017. The top panel plot of Fig. A3 shows the retrieved XCO_2 values from the original spectra (red) and the non-linearity corrected spectra (black). The middle and bottom row panels show the difference in ppm and the relative difference in percentage for the individual measurements. The non-linearity correction is dependent on the signal intensity of the recorded interferograms. The interferograms with less signal intensity are affected less as compared to the signal with high intensity. The maximum of the correction is therefore applied for the measurements with the highest signal during the noon time and the minimum correction is needed for measurements with the lowest signal when the sun is near the horizon.

Appendix B: Effect of the non-linearity on the Xgas values of TCCON

In this section the intercomparison results of the low-resolution test instruments are shown in comparison to the standard TCCON data. The mean bias, the standard deviation of the difference and the correlation coefficient between the individual low-resolution data sets and the TCCON for the full year of measurements in 2017 and for the period between 06 July and 12 September 2017 are shown in Table 7 and are plotted in the summary plot of Fig. 11. The comparison of the shorter time period has been made in order to check the statistics relative to the period where the Vertex70 was operated with improved settings.

B1 XCO_2 intercomparison results

The mean bias changed by -0.5 ppm with a standard deviation change of 0.23 ppm for the intercomparison between the TCCON and the TCCONmod data sets for the full measurement period. The intercomparison of the low-resolution measurements with

TCCONmod and TCCON as reference also show similar changes in the mean bias values. The standard deviation of the difference and the correlation coefficient remained similar for the intercomparison with TCCONmod as compared to TCCON. As an example, the mean bias values for the EM27/SUN changed from -0.18 ppm to -0.73 ppm, the standard deviation of the difference changed from 0.45 ppm to 0.47 ppm and the correlation coefficient changed from 0.995 to 0.996 for the comparison relative to TCCON where the latter values are for the comparison relative to TCCONmod.

B2 XCH_4 intercomparison results

The mean bias changed by -0.003 ppm with a standard deviation change of 0.001 ppm for the intercomparison between the TCCON and the TCCONmod data sets for the full measurement period. The intercomparison of the low-resolution measurements with TCCONmod and TCCON as reference also show similar changes in the mean bias values. The standard deviation of the difference and the correlation coefficient remained similar or improved slightly for the intercomparison with TCCONmod as compared to TCCON. As an example, the mean bias values for the EM27/SUN changed from 0.003 ppm to 0.000 ppm, the standard deviation of the difference changed from 0.005 ppm to 0.004 ppm and the correlation coefficient changed from 0.962 to 0.973 for the comparison relative to TCCON where the latter values are for the comparison relative to TCCONmod.

B3 XCO intercomparison results

The mean bias changed by -0.14 ppm with a standard deviation change of 0.08 ppm for the intercomparison between the TCCON and the TCCONmod data sets for the full measurement period. The intercomparison of the low-resolution measurements with TCCONmod and TCCON as reference also show similar changes in the mean bias values. The standard deviation of the difference and the correlation coefficient remained similar for the intercomparison with TCCONmod as compared to TCCON. As an example, the mean bias values for the EM27/SUN changed from 4.54 ppb to 4.38 ppb, the standard deviation of the difference changed from 1.37 ppb to 1.36 ppb and the correlation coefficient remained the same at 0.993 for the comparison relative to TCCON where the latter values are for the comparison relative to TCCONmod.

The above results show the difference of the TCCON data relative to the TCCONmod and the intercomparison with respect to the low-resolution data sets. This difference of the TCCON data has to be taken into account when using the official TCCON data set until the non-linearity corrected results are uploaded. The standard deviation of the difference and the correlation coefficient remained similar or improved slightly for the target species for the TCCONmod case. The TCCONmod data set is a better representation of the true atmospheric signal. Having this method implemented and tested for one year of data during this campaign, it will help in dealing with many years of historic TCCON data measured at the Sodankylä site. The HR125LR data were compromised by the non-linearity in a similar way to the high-resolution TCCON spectra.

Appendix C: Comparisons between TCCON and low-resolution measurements performed with Bruker IFS 125HR

The Bruker IFS 125HR was configured to record regular ~~low resolution~~ low-resolution measurements (spectral resolution of 0.5 cm^{-1}) together with the standard TCCON measurements (spectral resolution of 0.02 cm^{-1}). The ~~low resolution~~ low-resolution measurements are henceforth referred to as HR125LR. The ~~HR125LR measurements~~ low-resolution measurements (HR125LR data set), which are similar double-sided interferograms to the EM27/SUN, were processed by the KIT-group using PROFFAST. The results were post-processed in the same way as the results of the EM27/SUN. The comparison results with TCCON data as reference and HR125LR data recorded with the same instrument are discussed in this section species-by-species.

The timeseries of the coincident XCO_2, XCH_4, XCO and Xair values measured during the year 2017 for the HR125LR and the reference TCCON instrument are shown in the top panel, third row panel, fifth row panel and the seventh row panel plots of Fig. A7. The corresponding difference for each species relative to the TCCON data are shown in the second row panel, fourth row panel, sixth row panel and eight row panel plots, respectively. The mean bias, the standard deviation of the difference and the correlation coefficient between HR125LR and TCCON data sets for the full year of measurements in 2017 and those for the period between 6 July and 12 September 2017 ~~(henceforth referred to as the shorter period)~~ are given in Table 8 and are plotted in the summary plot of Fig. 11. The shorter time period was chosen in order to compare the results with the improved Vertex70 instrument.

C0.1 ~~XCO_2 comparison results~~

C1 XCO_2 comparison results

The seasonal cycle including the summer drawdown of the CO_2 was captured well by the HR125LR. The mean bias for the full year and the shorter period of measurements are -0.69 ppm and -0.4 ppm, respectively. The relatively small difference in the bias for the two time scales indicates that the bias is quite constant over the year. The high bias of -0.69 ppm may be due to the choice of the constant scaling factor used for the calculation of the XCO_2 values for the HR125LR data set. The same calibration factors as used by the EM27/SUN were used for the scaling of gases retrieved from the HR125LR measurements. However these calibration factors are specific to the EM27/SUN and therefore may not be accurate for the HR125LR. This is the reason for the high bias which is understood and not a problem as long as it remains constant and is not varying over the season. The difference plot (second row panel plot of Fig. A7) shows that there is a quite constant bias over the whole period with no seasonal dependencies. The standard deviation of the difference (0.53 ppm) as well as the correlation coefficient (0.993) between the HR125LR and the TCCON data sets are very similar to those when comparing the EM27/SUN and the modified Vertex70 data sets relative to the TCCON data. This implies similar behaviour of the above mentioned ~~low resolution~~ low-resolution instruments.

C1.1 ~~XCH_4 comparison results~~

C2 XCH_4 comparison results

The seasonal cycle of CH_4 is well captured by the HR125LR except for the March–May period. The mean bias for the full year is -0.005 ppm with a standard deviation of -0.004 ppm and a correlation coefficient of 0.975. The difference plot of the XCH_4 values (see fourth row panel plot of Fig. A7) shows a relatively high bias during the March–May period. This feature is also seen in the intercomparison results between other ~~low resolution~~ low-resolution instruments and the ~~TCCON~~ TCCONmod (see middle panel plot of Fig. ~~??~~8). The reason for this is the difference in resolution and the column averaging kernel (AK) between the TCCON and the ~~low resolution~~ low-resolution instruments. During the March–May period the TCCON a priori profiles show large differences to the AirCore profiles (see ~~top three middle panels~~ middle panels plots of Fig. ~~??~~2) where the latter give a better representation of the true atmospheric state. The AK represents the sensitivity of the retrieved total column to the true partial column profile. An AK value of 1 for all altitudes is the ideal case, which implies perfect sensitivity for the whole atmosphere. In such case the retrieved total column represents the true atmospheric state. An AK value of <1 or >1 for a given altitude implies that the retrieval underestimates or overestimates the contribution from that particular layer in the total column calculation budget, respectively. The TCCON AK for CH_4 as a function of the SZA is shown in the ~~middle plot of the bottom-row of Fig. ??.~~ Figure 6 of Hedelius et al. (2016). The AKs for the high SZAs at the lower layers are underestimating and those above the troposphere are overestimating the contribution. Any deviation of the CH_4 a priori profile from the true atmospheric state will be affecting the retrieval results with the higher SZA (mostly during the winter seasons) more as compared to those with the lower SZA. The deviation of the TCCON a priori from the true profile in combination with the overestimation of the retrieval values due to AKs >1 for high SZAs during the spring season is the reason for the high bias for the March–May period. This is further discussed in detail in section 5.3. The mean bias of the shorter time period is -0.007 ppm with a standard deviation of 0.002 ppm and a correlation coefficient of 0.97. The bias difference and the low standard deviation for the shorter period are due to the selected data set being outside the March–May period which does not cover the high values. The standard deviation of the difference between the HR125LR and TCCON data are very similar to those when comparing the EM27/SUN, IRcube and modified Vertex70 data sets relative to the TCCON data. This implies again similar behaviour of the ~~low resolution~~ low-resolution instruments.

C2.1 ~~XCO comparison results~~

C3 XCO comparison results

The seasonal cycle of CO is captured well by the HR125LR. The mean bias for the full year is 0.03 ppb with a standard deviation of 1.02 ppb and a correlation coefficient of 0.996. The values for the shorter time period are similar to those of the full year. However a slight seasonal dependency is seen. The bias is due to the choice of the scaling factor used for the calculation of the XCO values. The TCCON AK for ~~CO is shown in the right panel plot of the bottom-row of Fig. ??. The AK for~~ most of the atmospheric layers overestimates its contribution to the retrieval results. However the concentration of CO

in the atmosphere decreases rapidly with increasing altitude implying that the contribution in the total column is low and not strongly dependent on the SZA of the measurements. The intercomparison results of the ~~low resolution~~ low-resolution data set from all instruments relative to the TCCON data show that their performance was very similar in relation to the standard deviation of the difference and the correlation coefficient relative to the TCCON data.

C3.1 ~~Xair comparison results~~

C4 Xair comparison results

The Xair values for the HR125LR over the whole year are constant with a mean bias of 0.03 relative to the TCCON data and a standard deviation of 0.003. The mean bias for the shorter time period is the same as that for the full year. As the Xair values should be constant over the year, there is no correlation of Xair values expected between the two data sets. This is seen for the shorter time period. The complete time period shows a slightly negative correlation between the two data sets. The constant Xair values over the longer time period show that the performance of the Bruker IFS 125HR operated in the ~~low resolution~~ low-resolution mode was stable.

C5 ~~Intercomparisons with HR125LR data~~

Appendix D: Intercomparisons with HR125LR data

In this section we discuss the intercomparison results between the HR125LR as reference in relation to other ~~low resolution~~ low-resolution remote sensing instruments species-by-species. The timeseries of the coincident XCO_2, XCH_4 and XCO values for the HR125LR and the other test instruments are shown in the top panel, third row panel and fifth row panel plots of Fig. A8. The corresponding differences relative to the HR125LR are shown in the second row panel, fourth row panel and sixth row panel plots, respectively. The mean bias, standard deviation of the difference and the correlation coefficient between the individual instruments and the HR125LR for the full year of measurements in 2017 and that for the shorter time period are given in Table 9 and are plotted in the summary plot of Fig. 11.

The mean biases of the target species for the test instruments (see Table 9) are close to the difference of the biases of the species in Table 7 minus the biases in Table 8 for the full year of 2017 and for the shorter time period. The Vertex70 shows a significant improvement of the bias, scatter and correlation coefficient for the intercomparison results during the shorter time period performed after the instrument modification.

D0.1 ~~XCO_2 intercomparison results~~

D1 XCO_2 intercomparison results

The EM27/SUN and the IRcube show slight improvement, while the modified Vertex70 shows a slight degradation of the standard deviation of the difference and the correlation coefficient for the intercomparison results relative to the HR125LR as

compared to TCCON for the full year and the shorter time period (see Table 9 and Table 7). The scatter of the LHR instrument is very high and it is the dominating component of the intercomparison with other reference instruments.

D1.1 ~~XCH_4 intercomparison results~~

D2 XCH_4 intercomparison results

The high bias observed in the March–May period for the intercomparison of the test instruments with the TCCON instrument is not seen in the intercomparison results of the test instruments with the HR125LR. This indicates that the ~~high resolution~~ high-resolution and the AK of TCCON ~~(Fig. ??)~~ is the cause of the large bias during this period where the TCCON a priori is further away from the true atmospheric state. The standard deviation of the difference and the correlation coefficient improved for the EM27/SUN comparison. The IRcube has no significant bias, also the scatter and the correlation coefficient improved for the HR125LR intercomparison in relation to the TCCON intercomparison. The modified Vertex70 has the same standard deviation of the difference and a similar correlation coefficient for the HR125LR as compared to the TCCON intercomparison results.

D2.1 ~~XCO intercomparison results~~

D3 XCO intercomparison results

The EM27/SUN results for the full period and the Vertex70 results for the short period show similar values for the standard deviation of the difference and the correlation coefficient for the intercomparison results relative to the HR125LR as compared to the TCCON results.

The intercomparison results show the Xgas dependence on the resolution of the instrument, the averaging kernels and on the a priori. The ~~low resolution~~ low-resolution measurements helped to identify that the high bias in the ~~XCH4~~ XCH_4 during the March–May period was caused by the resolution difference and its sensitivity to the different averaging kernels.

D4 ~~SZA dependencies of bias~~

Appendix E: SZA dependencies of bias

The TCCON Xgas values are known to be affected by the SZA during the measurements. In this section we check the SZA dependence of the ~~low resolution~~ low-resolution instruments with respect to the TCCON and HR125LR data sets.

E0.1 ~~XCO_2 intercomparison results~~

E1 XCO_2 intercomparison results

The XCO_2 biases as a function of the measurement SZA for the ~~low resolution~~ low-resolution test instruments relative to the TCCON and HR125LR for all measurements in 2017 are shown in Fig. A9 and Fig. A10, respectively.

The top-left panel plots in the two figures show no SZA dependence of the bias for the LHR. The scatter is very high and the dominant instrumental biases mask any possible SZA dependence.

The plots for the Vertex70 are shown in the top-right panels. The TCCON comparison results show no SZA dependence of the bias after the instrument modification which implies that the two instruments have the same SZA dependence. However the HR125LR comparison results show decreasing bias values for an increasing SZA of the measurements. This implies that the XCO_2 values retrieved from the HR125LR measurements have a smaller SZA dependency as compared to the Vertex70.

The plots for the IRcube and the EM27/SUN are shown in the bottom-left and the bottom-right panels. The IRcube bias shows an increase with SZA relative to the TCCON data set and is quite constant relative to the HR125LR data set. The EM27/SUN bias shows a slight increase with SZA relative to the TCCON and it shows a rather constant bias relative to the HR125LR. This implies that the XCO_2 values retrieved from the HR125LR measurements show a similar SZA dependence as compared to the retrieval results for the EM27/SUN and the IRcube.

The retrievals for the HR125LR and the EM27/SUN were performed by PROFFAST and the retrievals of other instruments were performed by GFIT. The data set of the HR125LR, the IRcube and the EM27/SUN were all measured at the same spectral resolution of 0.5 cm^{-1}, whereas the data set for the Vertex70 were measured at a higher spectral resolution of 0.2 cm^{-1} and that of the TCCON were measured at 0.02 cm^{-1}. From the plots it can be seen that the SZA dependency of the retrievals is related to the spectral resolution and the AK of the instruments. This explains the decrease of the standard deviation of the bias for the EM27/SUN from 0.45 ppm to 0.4 ppm while the correlation coefficient improved slightly where the first values are intercomparison results relative to the TCCON and the latter values are relative to the HR125LR. For the IRcube the standard deviation of the bias decreased from 1.06 ppm to 1.03 ppm while the correlation coefficient improved from 0.971 to 0.978. For the modified Vertex70 we see an increase in the standard deviation of the bias from 0.582 ppm to 0.657 ppm while the correlation coefficient decreased from 0.931 to 0.911.

E1.1 ~~XCH_4 intercomparison results~~

E2 XCH_4 intercomparison results

The XCH_4 biases as a function of the measurement SZA for the ~~low resolution~~ low-resolution test instruments relative to the TCCON and the HR125LR for all measurements in 2017 are shown in the left panels and right panels of Fig. A11, respectively.

The Vertex70 XCH_4 biases are shown in the top panel plots. The TCCON comparison results show a slight decrease in the bias values for an increasing SZA of the measurements whereas the bias remains constant with an increasing SZA for the HR125LR comparison for the measurements performed after the instrument modification. The correlation coefficient and the scatter of the bias for the two comparison results are very similar.

The plots for the IRcube are shown in the middle panels. The IRcube bias shows a stronger dependence on SZA leading to a slightly poorer correlation coefficient of 0.924 for the TCCON comparison relative to a correlation coefficient of 0.949 for the HR125LR comparison. The standard deviation of the difference is slightly better for the HR125LR comparison.

The bottom-left panel plot shows a dependence of the bias as a function of the SZA for the EM27/SUN comparison relative to the TCCON. The bottom-right panel plot however does not show any dependence of the EM27/SUN comparison relative to the HR125LR. The correlation coefficient of the TCCON comparison was 0.962 and improved slightly to 0.978 for the HR125LR comparison. The scatter in the bias is slightly low leading to a significant improvement of the standard deviation of the difference from 0.005 ppm for the TCCON to 0.003 ppm for the HR125LR.

E2.1 ~~XCO intercomparison results~~

E3 XCO intercomparison results

The XCO biases as a function of the measurement SZA for the Vertex70 and the EM27/SUN relative to the TCCON and the HR125LR for all measurements in 2017 are shown in the left panels and right panels of Fig. A12, respectively.

The Vertex70 comparison results relative to the TCCON show a slight decrease in the bias with an increasing SZA of the measurements performed after the instrument modification. However, the bias increases slightly with an increasing SZA of the measurement for the HR125LR comparison. The standard deviation of the bias changed from 1.04 ppb to 1.17 ppb whereas the correlation coefficient changed from 0.991 to 0.988 when comparing the Vertex70 results relative to TCCON and relative to HR125LR. This shows that the Vertex70 comparison relative to the TCCON is slightly better than that of the HR125LR.

The EM27/SUN comparison results show a larger bias dependency on the SZA for the TCCON relative to the HR125LR comparison. The standard deviation of the bias decreased from 1.37 ppb to 1.27 ppb and the correlation coefficient improved from 0.993 to 0.995 when comparing the results relative to the HR125LR as compared to the TCCON.

This shows that the Vertex70 comparison with the TCCON shows better results whereas the EM27/SUN shows better results in comparison with the HR125LR.

E3.1 ~~Xair intercomparison results~~

E4 Xair intercomparison results

The Xair values as a function of the SZA of the measurements for the ~~low resolution~~ low-resolution instruments and the TCCON for all measurements in 2017 are shown in Fig. A13. The plots show the SZA dependence of the retrieved oxygen column for the performed measurements. The EM27/SUN and the HR125LR show no SZA dependence. However, the TCCON, the Vertex70 and the IRcube show a decreasing Xair value for an increasing SZA of the measurements.

The EM27/SUN and the HR125LR results retrieved with PROFFAST do not show ~~a~~ 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.

E5 ~~Humidity dependencies of bias~~

Appendix F: Xgas bias dependence on the humidity

~~The presence of water vapour lines in the retrieval windows can lead to errors in the determination of the Xgas values unless they are fitted well in the forward model. It is therefore necessary to check the influence of the water vapour lines for retrievals performed with the low resolution instruments.~~ In this section the Xgas bias dependence on the humidity present along the measurement line-of-sight is discussed~~for the intercomparison of the low resolution test instruments with respect to the HR125LR~~.

F0.1 ~~XCO_2 intercomparison results~~

F1 XCO_2 intercomparison results

The XCO_2 bias as a function of the humidity along the measurement line-of-sight for the ~~low resolution~~ low-resolution test instruments relative to the HR125LR for all measurements in 2017 are shown in Fig. A14.

The top-right panel plot shows no bias dependency on the humidity for the Vertex70. The scatter in the bias shows a significant reduction after the instrument modification.

The IRcube bias plot is shown in the bottom-left panel. The measurements in March, which were performed with the original optic fibre, show a high scatter in the values. The scatter is reduced after the change of the fibre and no humidity dependency of the bias is seen.

The bottom-right panel shows the bias plot for the EM27/SUN. The measurements in March show a higher scatter in the bias values compared to the rest of the year however no dependency on the humidity is seen.

The top-left panel plot shows no bias dependency on the humidity for the LHR. However, owing to the instrumental biases the scatter is very high.

This concludes that the XCO_2 retrieved from the ~~low resolution~~ low-resolution instruments have no dependencies on the humidity in the line-of-sight of the measurements.

F1.1 ~~XCH_4 intercomparison results~~

F2 XCH_4 intercomparison results

The XCH_4 bias as a function of the humidity along the measurement line-of-sight for the ~~low resolution~~ low-resolution test instruments relative to the HR125LR for all measurements in 2017 are shown in Fig. A15.

The top-left panel plot shows no bias dependency on the humidity for the Vertex70. Also here we see a significant reduction of the scatter in the bias values after the instrument modification.

The top-right panel plot shows no dependency of the bias on the humidity for the IRcube. Also here we see a high scatter before the change of the fibre.

The middle row-left panel plot shows the bias for the EM27/SUN. The scatter in the bias values during the dry period is slightly higher as compared to the measurements from other months but there is no dependency on the humidity seen. The high scatter of the bias during the dry period (March–May) may be due to the difference of the a prioris as compared to the true atmospheric state. This concludes that the XCH_4 retrieved from the ~~low resolution~~ low-resolution instruments has no dependencies on the humidity in the line-of-sight of the measurements.

F2.1 ~~XCO intercomparison results~~

F3 XCO intercomparison results

The XCO bias as a function of the humidity along the measurement line-of-sight for the ~~low resolution~~ low-resolution test instruments relative to the HR125LR for all measurements in 2017 are shown in Fig. A15.

The middle row-right panel plot shows no bias dependency on the humidity for the Vertex70. A reduction in the scatter of the bias values after the instrument modification is also seen here.

The bottom panel shows the bias plot for the EM27/SUN. The scatter in the values is high, however no dependency on the humidity can be seen. This concludes that the XCO retrieved from the ~~low resolution~~ low-resolution instruments has no dependencies on the humidity in the line-of-sight of the measurements.

F3.1 ~~Xair intercomparison results~~

F4 Xair intercomparison results

The Xair values as a function of the humidity along the measurement line-of-sight for the ~~low resolution~~ low-resolution test instruments, the HR125LR and the TCCON for all measurements in 2017 are shown in Fig. A16. The plot shows the humidity dependency of the retrieved oxygen column for the measurements performed. The Vertex70 and the IRcube show a significant reduction in the scatter of the Xair values after the instrument modification. The ~~low resolution~~ low-resolution instruments show no dependencies on the humidity.

F5 ~~Detector non-linearity effects~~

~~The TCCON measurements performed during the campaign in 2017 are found to be affected by the non-linearity of the detector. The non-linearity was identified towards the very end of the campaign in 2017 while checking the interferogram signal measured by the TCCON and comparing it to the EM27/SUN. The detector non-linearity is dependent on the photon load incident on the detector and influences the Xgas values dependent on the signal strength of the measurements. The non-linearity being a signal dependent function, can be avoided by keeping the signal level within the linear domain of the detector. To test the non-linearity, a metal grid was placed in the parallel light beam at the entrance port to reduce the signal by about 20%. Figure 1 shows two spectra measured with the standard TCCON configuration with no grid (red) and with a grid (black) placed in the parallel light beam. These spectra cover the complete spectral regions measured by the detector and are~~

zoomed in to highlight the signal of the out-of-band spectral regions. The non-linearity effect leads to out-of-band artifacts in the spectrum falsely indicating the presence of energy where the detector is insensitive. The signal between 0 cm^{-1} and the lower cutoff of the detector at 4000 cm^{-1} as well as the signal between the upper cutoff at about 12000 cm^{-1} and the end of the detector bandpass at about 16000 cm^{-1} show non-zero values for the no grid case indicating that the measurements performed were affected by the detector non-linearity. However the measurements performed with the reduced intensity by introducing the grid in the parallel beam do not show such high out-of-band intensities. The lower wavenumber out-of-band region shows only noise values and the higher wavenumber region close to the detector bandpass shows values which are higher than the noise but much lower than the signal of the standard measurements. These higher values could come from the spectral double passing of the signal within the interferometer. The high signal in the out-of-band spectral regions confirms that the TCCON measurements performed during 2017 are affected by the detector non-linearity. A correction method has been developed based on the method described in Hase (2000, chap. 5), it has been tested and applied to the TCCON data. The results of this are shown in the annex 1. In this section the intercomparison results of the low resolution test instruments are shown in comparison to the non-linearity corrected data measured by the TCCON, henceforth referred to as TCCONmod.

The non-linearity correction strongly affects the measured signal at both high and very low intensities. The mean bias, the standard deviation of the difference and the correlation coefficient between the individual low resolution data sets and the TCCONmod for the full year of measurements in 2017 and for the period between 06 July and 12 September 2017 are shown in Table 6 and are plotted in the summary plot of Fig. 11. The comparison of the shorter time period has been made in order to check the statistics relative to the period where the Vertex70 was operated with improved settings.

F4.1 XCO_2 intercomparison results

The mean bias changed by -0.5 ppm with a standard deviation change of 0.23 ppm for the intercomparison between the TCCON and the TCCONmod data sets for the full measurement period. The intercomparison of the low resolution measurements with TCCONmod and TCCON as reference also show similar changes in the mean bias values. The standard deviation of the difference and the correlation coefficient remained similar for the intercomparison with TCCONmod as compared to TCCON. As an example, the mean bias values for the EM27/SUN changed from -0.18 ppm to -0.73 ppm, the standard deviation of the difference changed from 0.45 ppm to 0.47 ppm and the correlation coefficient changed from 0.995 to 0.996 for the comparison relative to TCCON where the latter values are for the comparison relative to TCCONmod.

F4.1 XCH_4 intercomparison results

The mean bias changed by -0.003 ppm with a standard deviation change of 0.001 ppm for the intercomparison between the TCCON and the TCCONmod data sets for the full measurement period. The intercomparison of the low resolution measurements with TCCONmod and TCCON as reference also show similar changes in the mean bias values. The standard deviation of the difference and the correlation coefficient remained similar or improved slightly for the intercomparison with TCCONmod as compared to TCCON. As an example, the mean bias values for the EM27/SUN changed from 0.003 ppm to 0.000 ppm, the standard deviation of the difference changed from 0.005 ppm to 0.004 ppm and the correlation coefficient

changed from 0.962 to 0.973 for the comparison relative to TCCON where the latter values are for the comparison relative to TCCONmod.

F4.1 XCO intercomparison results

The mean bias changed by -0.14 ppm with a standard deviation change of 0.08 ppm for the intercomparison between the TCCON and the TCCONmod data sets for the full measurement period. The intercomparison of the low resolution measurements with TCCONmod and TCCON as reference also show similar changes in the mean bias values. The standard deviation of the difference and the correlation coefficient remained similar for the intercomparison with TCCONmod as compared to TCCON. As an example, the mean bias values for the EM27/SUN changed from 4.54 ppb to 4.38 ppb, the standard deviation of the difference changed from 1.37 ppb to 1.36 ppb and the correlation coefficient remained the same at 0.993 for the comparison relative to TCCON where the latter values are for the comparison relative to TCCONmod.

All of the results above for the individual species show that the non-linearity corrected TCCONmod data set has a different bias value which has to be taken into account. The standard deviation of the difference and the correlation coefficient remained similar or improved slightly for the target species. 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. Having this method implemented and tested for one year of data during this campaign, it will help in dealing with many years of historic TCCON data measured at the Sodankyläsite. The HR125LR data were compromised by the non-linearity in a similar way to the high resolution TCCON spectra. This explains part of the residual deviations of the low resolution test instruments compared to the HR125LR data.

F5 Comparisons with AirCore data

In this section the comparison results between the low resolution test instruments and the TCCON are shown using the AirCore data. Table 4 gives a list of all AirCore measurements performed in 2017 at the Sodankyläsite. The standard retrieval of each remote sensing instrument taking part in the intercomparison exercise were performed using the TCCON a priori. The daily a priori files were automatically generated during the GFIT run. In addition, the tool to generate the daily TCCON a priori for any given location is available using a stand-alone program via a DOI link provided by Toon and Wunch (2017). The AirCore measurements are in-situ measurements of the targeted species calibrated to the WMO scale and serve as a better reference for the vertical profile of the measured species. However, the AirCore profiles are limited to a vertical sampling height of about 25–30 km depending on the ceiling height reached by the launching balloon. Given this height limitation, the AirCore profiles cover only a part of the atmosphere relative to the TCCON a priori profile which covers a larger range starting from the site altitude up to 70 km. The lowermost layer of an AirCore profile is contaminated as the sampled air of the lowermost part of the atmosphere gets mixed with the reference push-gas. The push-gas is needed to let the sampled air pass through the analyser. The in-situ measurements performed at 2 m height above ground-level from a nearby ICOS (Integrated Carbon Observation System; https://www.icos-ri.eu/) ecosystem tower (https://en.ilmatieteenlaitos.fi/GHG-measurement-sites#Sodankyla) were used to substitute the concentrations of the lowermost layer of the measured AirCore profile. The AirCore profile above the top most measured

layer was further extended by a scaled TCCON a priori profile to cover the missing profile information up to 70 km of altitude. This is equivalent to a filling of < 5% of the total column above the top height of an AirCore measurement. The modified profile constructed using the ground-based in-situ measurement, AirCore measurement and the scaled TCCON a priori profile for each AirCore measurement day was further used for the intercomparison study.

The first three rows of Fig. **??** show the measured AirCore profiles (small blue dots), the a priori profiles from the GFIT run (small red dots), the tower mast measurements (green rectangle) and the extended AirCore profiles (large blue dots) for three sample days on 24 April, 15 May and 28 August 2017. The above mentioned three days were chosen to show the variability of the a priori profile during the different seasons at the Sodankyläsite. The left-column represents the data plotted for XCO_2, the middle-column represents the plots for XCH_4 and the right-column represents the plots for XCO as a function of the altitude for 24 April (top-row), 15 May (2nd-row) and 28 August 2017 (3rd-row), respectively. The column averaging kernels for CO_2, CH_4 and CO retrievals for TCCON plotted as a function of pressure and altitude are shown in the bottom-row of Fig. **??** from left to right. The different colours of the AK correspond to the varying SZAs of the measurements.

F4.1 Intercomparison results using AirCore as a priori profile

The extended AirCore vertical profiles for the targeted gases derived from the AirCore flights have been fed as input a priori profiles for the retrieval of the respective gases from the measurements performed with the remote sensing instruments on the respective days. The retrieval results with the modified AirCore profiles have been given the suffix "AC" at the end of the instrument name. As the remote sensing instruments covered a larger range of SZAs on 15 May and 28 August than on 24 April, those two days were selected for the intercomparison study.**XCO_2 intercomparison results**The intercomparison results for XCO_2 retrieved using TCCON a priori and modified AirCore a priori for the measurements performed with the TCCON and other low resolution FTIR spectrometers are shown in Fig. **??**. The left and right columns show the results for measurements performed on 15 May and on 28 August 2017, respectively. The difference between the TCCON a priori and the modified AirCore a priori profiles is relatively small on 15 May as compared to the high difference of the profiles on 28 August (see Fig. **??**). This implies that the TCCON a priori is closer to the true atmospheric state on 15 May than on 28 August. As a result the difference between the standard retrievals from each instrument using TCCON a priori and the retrievals using the modified AirCore a priori is smaller on 15 May compared to the difference on 28 August. The retrieval results for all instruments for the measurements on 28 August (see Fig. **??**) show a bias between the TCCON a priori and the modified AirCore a priori retrievals. The bias shows a strong dependency of the retrieval on the SZA of the measurements. This is due to the TCCON CO_2 AK dependence on the SZA as seen in the left panel plot of the bottom-row of Fig. **??**. With these AK the a priori information is very relevant. The AirCore a priori is in principle the closest a priori to the truth. When applying the AirCore a priori and doing the retrieval we see that the airmass dependence is much reduced.

For example the AK values for CO_2 for lower altitudes are >1 for higher SZA measurements 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, 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. Therefore we can see in Fig. **??** that the TCCON a priori underestimates the lower partial columns during the summer months and that the TCCON columns are too high compared to the TCCONAC. The plots also show the scatter of the retrieval results from the individual instruments for two days. With these AK the a priori information is very relevant. The AirCore a priori is the closest a priori to the truth. When applying the AirCore a priori and doing the retrieval we see that the airmass dependence is much reduced. The EM27/SUN shows a lower scatter as compared to the TCCON. The Vertex70 measurements on 15 May were performed before the instrument modifications. As a result a high bias relative to the TCCON 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 TCCON.~~**~~XCH_4 intercomparison results~~**~~The intercomparison results for XCH_4 retrieved using TCCON a priori and modified AirCore a priori for the measurements performed with TCCON and other low resolution FTIR spectrometers are shown in Fig. **??**. The left and right columns show the results for measurements performed on 15 May and on 28 August 2017. The difference between the TCCON a priori and the modified AirCore a priori profiles of CH_4 is the highest for 24 April followed by 15 May and the smallest for 28 August (see Fig. **??**). The vertical distribution of the CH_4 concentration during the winter and spring period is poorly modelled by the TCCON a priori tool. The a priori during the summer is in better agreement with the AirCore measurements as seen for the 28 August profiles. As a result the difference between the standard retrievals from each instrument using TCCON a priori and the retrievals using the modified AirCore as a priori is smaller for 28 August than for 15 May.~~

~~The middle panel plot of the bottom-row of Fig. **??** shows the TCCON CH_4 AK dependence as a function of the SZA. The AK values are >1 for measurements at a lower SZA which means that the retrieval overestimates the contribution from all layers above 10 km. However, the AK values are <1 for measurements with SZA >65° which means that the retrieval underestimates the contribution from all layers above 10 km. The TCCONAC results are higher than the TCCON results for the lower SZA values and vice versa. This effect is stronger for the retrieval results of 15 May compared to the results of 28 August where the TCCON a priori is closer to the AirCore a priori. The retrieval results for 15 May measurements for all instruments show a bias between the TCCON a priori and the modified AirCore a priori. The bias shows a strong dependency of the retrieval on the SZA. The EM27/SUNAC results show a small bias compared to the EM27/SUN. The difference plot shows that the change of the retrieved XCH_4 values with the modified AirCore a priori has the same sign as compared to the TCCON. The same feature is also seen by the Vertex70 and IRcube results. The bias for the 28 August is largely reduced compared to that of the 15 May. The small remaining bias is due to the difference in the a priori and the AK of the instruments.~~**~~XCO intercomparison results~~**~~The intercomparison results for XCO retrieved using TCCON a priori and modified AirCore a priori for the measurements performed with the TCCON and the other low resolution FTIR spectrometers are shown in Fig. **??**. The left and right columns show the results for measurements performed on 15 May and on 28 August 2017. The TCCON a priori and modified AirCore a priori profiles of CO for three days in 2017 are shown in Fig. **??**. The AirCore measured CO profiles are provided for altitudes up to 17 km and in some cases as high as 19 km. 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 difference in the profiles in the stratosphere is the largest for 24 April followed by 15 May and the difference is the smallest for 28 August. As a result the difference between~~

the standard retrievals using TCCON a priori and the retrievals using the modified AirCore as a priori is slightly higher for 15 May than for 28 August. The right panel plot of the bottom-row of Fig. **??** shows the TCCON CO AK dependence as a function of the SZA. The AK contribution to the retrieval results are underestimated (AK values <1) for layers below 5 km and overestimated for layers above 5 km with AK values >1, even increasing up to above 2 for higher layers. The bias dependence on SZA is significant for measurements performed only at high SZA. The TCCONAC XCO retrievals show a constant bias relative to the TCCON XCO retrievals for most of the SZA and the deviation is seen only for measurements performed at the high SZAs. The EM27/SUN and the Vertex70 results also show a slight dependency of the XCO retrieved using the TCCON a priori and the modified AirCore a priori on the measurements performed at a high SZA and a constant bias for measurements performed at a low SZA.

F4.1 Direct intercomparison results of the Xgas calculated from AirCore relative to the TCCON and non-linearity corrected TCCON data sets

In this section we discuss the intercomparison results of the Xgas values calculated directly from the modified AirCore profiles by using the TCCON AKs in relation to the Xgas values retrieved from the standard TCCON and the non-linearity corrected TCCON data sets. Any difference in the intercomparison results is a direct reflection of the difference between the measured AirCore profile and the ground-based in-situ data relative to the TCCON a priori for the same altitude coverage. The time corresponding to 90% of the profile (starting at the top of the atmosphere) acquisition time is taken as the AirCore timestamp for the intercomparison of the Xgas values. A 3h time window around the AirCore measurement time was used as the co-incidence limit. All Xgas values from TCCON data and TCCONmod data in this time window were averaged and taken as the coincident data sets for the intercomparison. The 3hr time window was selected for the remote sensing measurement as it is a good representation of the AirCore measurements. Reducing the time window resulted in the reduction of co-located measurement days and increasing the time window introduced the true variability of the atmospheric state in the remote sensing data.

The mean bias, the standard deviation of the difference and the correlation coefficient of the Xgas values calculated from the AirCore relative to the TCCON and the TCCONmod are shown in Table 5. The XCO_2 mean bias between AirCore and TCCON is 0.47 ppm with a standard deviation of 0.66 ppm and a correlation coefficient of 0.994. The mean bias is reduced significantly to -0.03 ppm for the intercomparison between the AirCore and the TCCONmod. The standard deviation of the difference is very similar, however the correlation coefficient improved slightly for the TCCONmod. This shows that the XCO_2 values from the TCCONmod data set are a better representation of the true atmospheric state.

The XCH_4 mean bias between the AirCore and the TCCONmod increases to -0.007 ppm as compared to the mean bias of -0.004 ppm between AirCore and TCCON. The scatter remains the same with an improvement in the correlation for the TCCONmod. The improvement in the correlation indicates that the TCCONmod data is a better representation of the true atmospheric state. The increase in the mean bias is due to the difference in the TCCON a priori profiles used for the retrieval relative to the true atmospheric profiles. The top-panel of Fig. 3 shows the time series of a 30 min averaged TCCONmod XCH_4 data set and XCH_4 calculated from the AirCore measurements. The middle-panel plot shows the difference in the XCH_4 bias.

The large difference between the two data sets in April is due to the difference between the a priori from the true atmospheric state. The bias reduces significantly for all later AirCore measurement days.

The XCO mean bias between AirCore and TCCONmod is slightly reduced to 6.25 ppb as compared to the mean bias of 6.4 ppb between AirCore and TCCON. The scatter is almost the same with very similar correlation coefficients. The CO retrieval from the AirCore has a large uncertainty. As a result the impact due to the change of the data set from the TCCON to the TCCONmod is within the uncertainty budget of the AirCore measurements.

The direct intercomparison results of the Xgas calculated from AirCore relative to the TCCON and non-linearity corrected TCCON data sets indicate clearly that the non-linearity corrected data set gives Xgas amounts which are closer to the AirCore amounts and hence closer to our best estimate of the true atmospheric conditions.

Appendix G: Summary and outlook

The FRM4GHG campaign was successfully executed by comparing four portable remote sensing instruments against the reference TCCON instrument at the Sodankyläsite during the year 2017. The EM27/SUN was set up every day at the ambient temperature and pressure and was operated without configuration changes during the whole campaign. The other low resolution FTIR and the LHR were operated from inside a dedicated temperature controlled container. The instruments needed optimisation and behaved better with a low bias and a high correlation relative to the TCCON instrument afterwards.

The EM27/SUN Xgas biases relative to the TCCON data were low for the target species except for the high XCH_4 bias during the March–May period which is due to the difference in the sensitivity of the high and low resolution instruments and the a prioris not matching well with the actual profile shape. The EM27/SUN measurements include a bias correction determined independently prior to this study.

The IReube Xgas values show relatively high biases which are related to the possible dependence of the signal level on the extended InGaAs detector known to have non-linearity characteristics. The ILS of the IReube is also less ideal compared to other larger instruments due to the compact short focal length optics. The impact of the ILS on the biases is being further investigated. However the comparison shows low scatter and a good correlation relative to the TCCON data.

The Vertex70 was equipped with an extended InGaAs detector which led to identifiable non-linearity effects. The optical path was modified by introducing an aperture stop to avoid saturation and operate in the linear region of the detector to improve the ILS. The bias of the Xgas values, the standard deviation of the difference and the correlation of the modified Vertex70 instrument relative to the TCCON data were significantly lower after this instrument modification and comparable to the EM27/SUN results relative to the TCCON data.

The LHR was a new instrument deployed under test during this campaign. It showed large scatter and large biases with a strong diurnal variation relative to the TCCON and other FTS instruments. The LHR data for the 2017 campaign are not yet able to provide meaningful geophysical information. However this comparison has proven to be invaluable to characterise and understand the instrumental biases and possibly the retrieval biases. Both aspects are currently under investigation and improvements are being developed.

The intercomparison results of the low resolution measurements performed with the Bruker IFS 125HR relative to the standard TCCON and other low resolution instruments provided useful analysis of the resolution dependent effects on the Xgas retrieved for the target gases. The low resolution measurements performed with the Bruker IFS 125HR also helped to identify that the high bias in the XCH_4 during the March–May period was caused by the resolution difference and the corresponding different sensitivities to the vertical profile shape as seen in the averaging kernels.

The airmass dependence of the retrievals is an effect of the software. The EM27/SUN and the HR125LR results retrieved with PROFFAST do not show SZA dependence for species where an airmass correction factor was applied. All other results show SZA dependence to some degree. The correction for the SZA dependence is a long-standing and ongoing issue for TCCON and relevant to all instruments. In order to minimise the effect of the SZA, measurements with SZA <75° were used for the intercomparison of the different data sets.

The bias dependence on the humidity along the measurement line-of-sight was investigated for each target species. However no dependence was found.

In the course of the campaign not only the Vertex70, IReube and LHR instruments have been improved but also the TCCON instrument by detecting and correcting non-linearity of the detector response. Detecting this issue by comparison with the EM27/SUN shows the potential of this instrument as a traveling standard for TCCON. The intercomparison results showed that the non-linearity corrected TCCON data gave a better match with the low resolution instruments. The standard deviation of the bias and the correlation coefficient was similar for the target species for the non-linearity corrected TCCON data relative to the standard TCCON data.

The intercomparison results using AirCore profiles as a priori showed interesting insights to the FTS retrievals, its sensitivity to the resolution and the averaging kernels. The AirCore profiles also showed the differences relative to the TCCON a prioris and the resulting biases in the retrievals of the target species. The Xgas calculated from AirCore and compared to the TCCON and the non-linearity corrected TCCON data sets show that the latter data set is a better representation of the true atmospheric state.

The EM27/SUN, the IReube, the modified Vertex70 and the HR125LR provided stable and precise measurements of the target gases during the campaign. The portable low resolution instruments can be used for campaigns or long-term measurements from any site, and complement the TCCON network. The Xgas measurements from these instruments will be of similar quality as the TCCON Xgas data.

Appendix G: Non-linearity of the Sodankylä TCCON InGaAs detector

The TCCON measurements performed during the campaign in 2017 are found to be affected by the non-linearity of the InGaAs detector used for the measurements. A correction method has been developed based on the method described in Hase (2000, chap. 5) to correct the signal in the interferogram domain using the following equation:

$$I' = I + a \times I^2 + b \times I^3 + c \times I^4$$

where I is the original interferogram and I′ is the non-linearity corrected interferogram. The coefficients a, b and c are determined for each measurement by minimising the cost function. The cost function is defined as the ratio of the signal sum in the out-of-band regions covering 100–3600 cm^{-1} and 14200–15750 cm^{-1} to the signal sum for the in-band spectral region covering 4100–9700 cm^{-1}. The individual correction functions are plotted in Fig. A1 as red curves, their mean is plotted as blue open circles and the black line is the fourth order polynomial fit of the mean with the coefficient values of a = 0.00138, b = 0.00302 and c = 0.0166. The coefficients of the fitted curve representing the averaged correction function are used to correct the non-linearity of all TCCON measurements. The top-left and bottom-left panel plots of Fig. A2 show the original and the non-linearity corrected spectra for the measurement day of 06 September 2017. The colours of the spectrum depend on the interferogram maximum signal at the centre burst. The highest values represented by the dark red colour are for measurements performed with the highest solar intensity during noon time and the measurements performed with the lowest solar intensity are represented by blue as the minimum. The top-right and bottom-right panel plots of Fig. A2 show the original and the non-linearity corrected spectra for the out-of-band region between 100 and 3600 cm^{-1}. The signal reduction in the out-of-band spectral region clearly shows that the non-linearity correction worked for the spectra. In order to quantify the effect of the non-linearity correction on the Xgas values, explicit results are shown in detail for the 06 September 2017. The top panel plot of Fig. A3 shows the retrieved XCO_2 values from the original spectra (red) and the non-linearity corrected spectra (black). The middle and bottom row panels show the difference in ppm and the relative difference in percentage for the individual measurements. The non-linearity correction is dependent on the signal intensity of the recorded interferograms. The interferograms with less signal intensity are affected less as compared to the signal with high intensity. The maximum of the correction is therefore applied for the measurements with the highest signal during the noon time and the minimum correction is needed for measurements with the lowest signal when the sun is near the horizon.

Author contributions. MKS, MDM and JN designed the study. MKS wrote the paper and produced the intercomparison analysis and results with input from all authors. All authors contributed in the generation of data used for this study. All authors read and provided comments on the paper.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The FRM4GHG project received financial support from the European Space Agency under the grant agreement number ESA-IPL-POE-LG-cl-LE-2015-1129. The authors wish to thank the instrument operators as well as the scientists doing the retrieval for the delivery of the data which was used for the intercomparison and discussed in this paper. We thank Tuomas Laurila and Juha Hatakka from FMI for providing in-situ measurements from the 50 m tower at the site and for providing critical instrumentation needed for the AirCore analysis, including gas analyser, calibration gases and fill gases. We thank Nicholas Kumps for checking on the automatic operation and data recording of the Vertex70 instrument. KIT would like to thank the FMI staff and their student Elina Paulus for supporting the operation of the

EM27/SUN spectrometer. The UK team also wishes to acknowledge support from the Centre of Earth Observation Instrumentation (CEOI) and the UK Space Agency (UKSA). We thank Paolo Castracane for his contribution during the project meetings and discussions.

References

Algorithm Theoretical Baseline Document for Sentinel-5 Precursor Methane Retrieval, https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-Methane-retrieval, a.

Algorithm Theoretical Baseline Document for Sentinel-5 Precursor: Carbon Monoxide Total Column Retrieval, https://sentinel.esa.int/documents/247904/2476257/Sentinel-5P-TROPOMI-ATBD-Carbon-Monoxide-Total-Column-Retrieval, b.

Borsdorff, T., aan de Brugh, J., Hu, H., Hasekamp, O., Sussmann, R., Rettinger, M., Hase, F., Gross, J., Schneider, M., Garcia, O., Stremme, W., Grutter, M., Feist, D. G., Arnold, S. G., De Mazière, M., Kumar Sha, M., Pollard, D. F., Kiel, M., Roehl, C., Wennberg, P. O., Toon, G. C., and Landgraf, J.: Mapping carbon monoxide pollution from space down to city scales with daily global coverage, Atmospheric Measurement Techniques, 11, 5507–5518, https://doi.org/10.5194/amt-11-5507-2018, https://www.atmos-meas-tech.net/11/5507/2018/, 2018.

Deutscher, N. M., Griffith, D. W. T., Bryant, G. W., Wennberg, P. O., Toon, G. C., Washenfelder, R. A., Keppel-Aleks, G., Wunch, D., Yavin, Y., Allen, N. T., Blavier, J.-F., Jiménez, R., Daube, B. C., Bright, A. V., Matross, D. M., Wofsy, S. C., and Park, S.: Total column CO_2 measurements at Darwin, Australia – site description and calibration against in situ aircraft profiles, Atmospheric Measurement Techniques, 3, 947–958, https://doi.org/10.5194/amt-3-947-2010, https://www.atmos-meas-tech.net/3/947/2010/, 2010.

Dirksen, R. J., Sommer, M., Immler, F. J., Hurst, D. F., Kivi, R., and Vömel, H.: Reference quality upper-air measurements: GRUAN data processing for the Vaisala RS92 radiosonde, Atmospheric Measurement Techniques, 7, 4463–4490, https://doi.org/10.5194/amt-7-4463-2014, https://www.atmos-meas-tech.net/7/4463/2014/, 2014.

Dlugokencky, E. and Tans, P.: Trends in Atmospheric Carbon Dioxide and Methane, www.esrl.noaa.gov/gmd/ccgg/trends/.

Dudhia, A.: The Reference Forward Model (RFM), Journal of Quantitative Spectroscopy and Radiative Transfer, 186, 243 – 253, https://doi.org/https://doi.org/10.1016/j.jqsrt.2016.06.018, http://www.sciencedirect.com/science/article/pii/S0022407316301029, satellite Remote Sensing and Spectroscopy: Joint ACE-Odin Meeting, October 2015, 2017.

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.

Frey, M., Sha, M. K., Hase, F., Kiel, M., Blumenstock, T., Harig, R., Surawicz, G., Deutscher, N. M., Shiomi, K., Franklin, J. E., Bösch, H., Chen, J., Grutter, M., Ohyama, H., Sun, Y., Butz, A., Mengistu Tsidu, G., Ene, D., Wunch, D., Cao, Z., Garcia, O., Ramonet, M., Vogel, F., and Orphal, J.: Building the COllaborative Carbon Column Observing Network (COCCON): long-term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer, Atmospheric Measurement Techniques, 12, 1513–1530, https://doi.org/10.5194/amt-12-1513-2019, https://www.atmos-meas-tech.net/12/1513/2019/, 2019.

Gisi, M., Hase, F., Dohe, S., Blumenstock, T., Simon, A., and Keens, A.: XCO_2-measurements with a tabletop FTS using solar absorption spectroscopy, Atmospheric Measurement Techniques, 5, 2969–2980, https://doi.org/10.5194/amt-5-2969-2012, https://www.atmos-meas-tech.net/5/2969/2012/, 2012.

Hase, F.: Inversion von Spurengasprofilen aus hochaufgelöstenbodengebundenen FTIR-Messungen in Absorption, Ph.D. thesis, http://www.imk-asf.kit.edu/downloads/bod/disshase.pdf, 2000.

Hase, F., Blumenstock, T., and Paton-Walsh, C.: Analysis of the instrumental line shape of high-resolution Fourier transform IR spectrometers with gas cell measurements and new retrieval software, Appl. Opt., 38, 3417–3422, https://doi.org/10.1364/AO.38.003417, http://ao.osa.org/abstract.cfm?URI=ao-38-15-3417, 1999.

Hase, F., Drouin, B. J., Roehl, C. M., Toon, G. C., Wennberg, P. O., Wunch, D., Blumenstock, T., Desmet, F., Feist, D. G., Heikkinen, P., De Mazière, M., Rettinger, M., Robinson, J., Schneider, M., Sherlock, V., Sussmann, R., Té, Y., Warneke, T., and Weinzierl, C.: Calibration of sealed HCl cells used for TCCON instrumental line shape monitoring, Atmospheric Measurement Techniques, 6, 3527–3537, https://doi.org/10.5194/amt-6-3527-2013, https://www.atmos-meas-tech.net/6/3527/2013/, 2013.

Hase, F., Frey, M., Kiel, M., Blumenstock, T., Harig, R., Keens, A., and Orphal, J.: Addition of a channel for XCO observations to a portable FTIR spectrometer for greenhouse gas measurements, Atmospheric Measurement Techniques, 9, 2303–2313, https://doi.org/10.5194/amt-9-2303-2016, https://www.atmos-meas-tech.net/9/2303/2016/, 2016.

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_{CO_2}, X_{CH_4}, X_{CO}, and X_{N_2O} from a $0.5\,cm^{-1}$ resolution solar-viewing spectrometer, Atmospheric Measurement Techniques, 9, 3527–3546, https://doi.org/10.5194/amt-9-3527-2016, https://www.atmos-meas-tech.net/9/3527/2016/, 2016.

Hedelius, J. K., Parker, H., Wunch, D., Roehl, C. M., Viatte, C., Newman, S., Toon, G. C., Podolske, J. R., Hillyard, P. W., Iraci, L. T., Dubey, M. K., and Wennberg, P. O.: Intercomparability of X_{CO_2} and X_{CH_4} from the United States TCCON sites, Atmospheric Measurement Techniques, 10, 1481–1493, https://doi.org/10.5194/amt-10-1481-2017, https://www.atmos-meas-tech.net/10/1481/2017/, 2017.

Hoffmann, A., Macleod, N. A., Huebner, M., and Weidmann, D.: Thermal infrared laser heterodyne spectroradiometry for solar occultation atmospheric CO_2 measurements, Atmospheric Measurement Techniques, 9, 5975–5996, https://doi.org/10.5194/amt-9-5975-2016, https://www.atmos-meas-tech.net/9/5975/2016/, 2016.

Hoffmann, A., Macleod, N. A., Huebner, M., and Weidmann, D.: LHR ILS Validation for FRM4GHG campaign, Tech. rep., Spectroscopy Group, Earth Observation & Technology Department, Space Science & Technology Department, Rutherford Appleton Laboratory, http://frm4ghg.aeronomie.be/ProjectDir/documents/ILS_parameters/LHR_ILS_validation.pdf, 2017.

Inoue, M., Morino, I., Uchino, O., Nakatsuru, T., Yoshida, Y., Yokota, T., Wunch, D., Wennberg, P. O., Roehl, C. M., Griffith, D. W. T., Velazco, V. A., Deutscher, N. M., Warneke, T., Notholt, J., Robinson, J., Sherlock, V., Hase, F., Blumenstock, T., Rettinger, M., Sussmann, R., Kyrö, E., Kivi, R., Shiomi, K., Kawakami, S., De Mazière, M., Arnold, S. G., Feist, D. G., Barrow, E. A., Barney, J., Dubey, M., Schneider, M., Iraci, L. T., Podolske, J. R., Hillyard, P. W., Machida, T., Sawa, Y., Tsuboi, K., Matsueda, H., Sweeney, C., Tans, P. P., Andrews, A. E., Biraud, S. C., Fukuyama, Y., Pittman, J. V., Kort, E. A., and Tanaka, T.: Bias corrections of GOSAT SWIR XCO_2 and XCH_4 with TCCON data and their evaluation using aircraft measurement data, Atmospheric Measurement Techniques, 9, 3491–3512, https://doi.org/10.5194/amt-9-3491-2016, https://www.atmos-meas-tech.net/9/3491/2016/, 2016.

Jing, Y., Wang, T., Zhang, P., Chen, L., Xu, N., and Ma, Y.: Global Atmospheric CO2 Concentrations Simulated by GEOS-Chem: Comparison with GOSAT, Carbon Tracker and Ground-Based Measurements, Atmosphere, 9, https://doi.org/10.3390/atmos9050175, https://www.mdpi.com/2073-4433/9/5/175, 2018.

Karion, A., Sweeney, C., Tans, P., and Newberger, T.: AirCore: An Innovative Atmospheric Sampling System, Journal of Atmospheric and Oceanic Technology, 27, 1839–1853, https://doi.org/10.1175/2010JTECHA1448.1, https://doi.org/10.1175/2010JTECHA1448.1, 2010.

Keppel-Aleks, G., Toon, G. C., Wennberg, P. O., and Deutscher, N. M.: Reducing the impact of source brightness fluctuations on spectra obtained by Fourier-transform spectrometry, Appl. Opt., 46, 4774–4779, https://doi.org/10.1364/AO.46.004774, http://ao.osa.org/abstract.cfm?URI=ao-46-21-4774, 2007.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quéré, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter,

B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. A., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E., and Zeng, G.: Three decades of global methane sources and sinks, Nat. Geosci., https://doi.org/doi:10.1038/ngeo1955, 2013.

Kivi, R. and Heikkinen, P.: Fourier transform spectrometer measurements of column CO_2 at Sodankylä, Finland, Geoscientific Instrumentation, Methods and Data Systems, 5, 271–279, https://doi.org/10.5194/gi-5-271-2016, https://www.geosci-instrum-method-data-syst.net/5/271/2016/, 2016.

Kivimäki, E., Lindqvist, H., Hakkarainen, J., Laine, M., Sussmann, R., Tsuruta, A., Detmers, R., Deutscher, N. M., Dlugokencky, E. J., Hase, F., Hasekamp, O., Kivi, R., Morino, I., Notholt, J., Pollard, D. F., Roehl, C., Schneider, M., Sha, M. K., Velazco, V. A., Warneke, T., Wunch, D., Yoshida, Y., and Tamminen, J.: Evaluation and Analysis of the Seasonal Cycle and Variability of the Trend from GOSAT Methane Retrievals, Remote Sensing, 11, https://doi.org/10.3390/rs11070882, https://www.mdpi.com/2072-4292/11/7/882, 2019.

Kong, Y., Chen, B., and Measho, S.: Spatio-Temporal Consistency Evaluation of XCO2 Retrievals from GOSAT and OCO-2 Based on TCCON and Model Data for Joint Utilization in Carbon Cycle Research, Atmosphere, 10, https://doi.org/10.3390/atmos10070354, https://www.mdpi.com/2073-4433/10/7/354, 2019.

Neefs, E., Mazière, M. D., Scolas, F., Hermans, C., and Hawat, T.: BARCOS, an automation and remote control system for atmospheric observations with a Bruker interferometer, Review of Scientific Instruments, 78, 035 109, https://doi.org/10.1063/1.2437144, https://doi.org/10.1063/1.2437144, 2007.

Novelli, P. C., Masarie, K. A., and Lang, P. M.: Distributions and recent changes of carbon monoxide in the lower troposphere, Journal of Geophysical Research: Atmospheres, 103, 19 015–19 033, https://doi.org/10.1029/98JD01366, https://doi.org/10.1029/98JD01366, 1998.

Ostler, A., Sussmann, R., Patra, P. K., Houweling, S., De Bruine, M., Stiller, G. P., Haenel, F. J., Plieninger, J., Bousquet, P., Yin, Y., Saunois, M., Walker, K. A., Deutscher, N. M., Griffith, D. W. T., Blumenstock, T., Hase, F., Warneke, T., Wang, Z., Kivi, R., and Robinson, J.: Evaluation of column-averaged methane in models and TCCON with a focus on the stratosphere, Atmospheric Measurement Techniques, 9, 4843–4859, https://doi.org/10.5194/amt-9-4843-2016, https://www.atmos-meas-tech.net/9/4843/2016/, 2016.

Sha, M. K., De Mazière, M., Notholt, J., Blumenstock, T., Chen, H., Griffith, D., Hase, F., Heikkinen, P., Hoffmann, A., Huebner, M., Jones, N., Kivi, R., Petri, C., Tu, Q., Warneke, T., and Weidmann, D.: FRM4GHG raw dataset from the Sodankylä campaign, Royal Belgian Institute for Space Aeronomy, https://doi.org/https://doi.org/10.18758/71021040, 2018.

Sha, M. K., De Mazière, M., Notholt, J., Blumenstock, T., Chen, H., Griffith, D., Hase, F., Heikkinen, P., Hoffmann, A., Huebner, M., Jones, N., Kivi, R., Petri, C., Tu, Q., Warneke, T., and Weidmann, D.: FRM4GHG level2 dataset from the Sodankylä campaign, Royal Belgian Institute for Space Aeronomy, https://doi.org/https://doi.org/10.18758/71021048, 2019.

Stocker, T. F., Qin, D., Plattner, G.-K., Alexander, L., Allen, S., Bindoff, N., Bréon, F.-M., Church, J., Cubasch, U., Emori, S., Forster, P., Friedlingstein, P., Gillett, N., Gregory, J., Hartmann, D., Jansen, E., Kirtman, B., Knutti, R., Kumar, K. K., Lemke, P., Marotzke, J., Masson-Delmotte, V., Meehl, G., Mokhov, I., S.Piao, Ramaswamy, V., Randall, D., Rhein, M., Rojas, M., Sabine, C., Shindell, D., Talley, L., Vaughan, D., and Xie, S.-P.: Technical Summary, in: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)], Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2013.

Toon, G. C. and Wunch, D.: A stand-alone a priori profile generation tool for GGG2014 release, https://doi.org/10.14291/TCCON.GGG2014.PRIORS.R0/1221661, 2017.

Tsai, T. R., Rose, R. A., Weidmann, D., and Wysocki, G.: Atmospheric vertical profiles of O3, N2O, CH4, CCl2F2, and H2O retrieved from external-cavity quantum-cascade laser heterodyne radiometer measurements, Appl. Opt., 51, 8779–8792, https://doi.org/10.1364/AO.51.008779, http://ao.osa.org/abstract.cfm?URI=ao-51-36-8779, 2012.

Weidmann, D., Reburn, W. J., and Smith, K. M.: Ground-based prototype quantum cascade laser heterodyne radiometer for atmospheric studies, Review of Scientific Instruments, 78, 073 107, https://doi.org/10.1063/1.2753141, https://doi.org/10.1063/1.2753141, 2007.

Wunch, D., Toon, G. C., Blavier, J.-F. L., Washenfelder, R. A., Notholt, J., Connor, B. J., Griffith, D. W. T., Sherlock, V., and Wennberg, P. O.: The Total Carbon Column Observing Network, 369, 2087–2112, https://doi.org/10.1098/rsta.2010.0240, 2011.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, C., Feist, D. G., and Wennberg, P. O.: The Total Carbon Column Observing Network's GGG2014 Data Version, Tech. rep., Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, Oak Ridge, Tennessee, U.S.A., https://doi.org/10.14291/tccon.ggg2014.documentation.R0/1221662, 2015.

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O'Dell, C., Mandrake, L., Viatte, C., Griffith, D. W., Deutscher, N. M., Velazco, V. A., Notholt, J., Warneke, T., Petri, C., De Maziere, M., Sha, M. K., Sussmann, R., Rettinger, M., Pollard, D., Robinson, J., Morino, I., Uchino, O., Hase, F., Blumenstock, T., Kiel, M., Feist, D. G., Arnold, S. G., Strong, K., Mendonca, J., Kivi, R., Heikkinen, P., Iraci, L., Podolske, J., Hillyard, P. W., Kawakami, S., Dubey, M. K., Parker, H. A., Sepulveda, E., Rodriguez, O. E. G., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A.: Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) X_{CO_2} measurements with TCCON, Atmospheric Measurement Techniques Discussions, 2016, 1–45, https://doi.org/10.5194/amt-2016-227, http://www.atmos-meas-tech-discuss.net/amt-2016-227/, 2016.

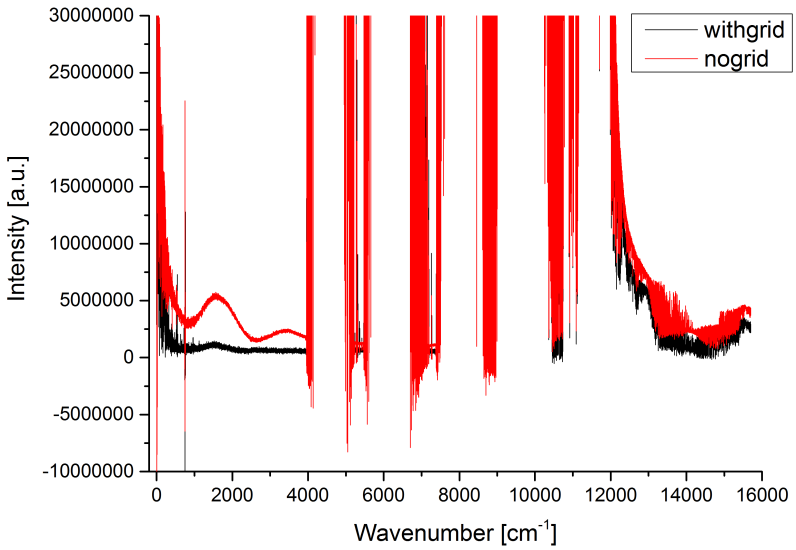

Figure 1. ~~Timeseries of XCO_2 retrievals for TCCON, LHR, Vertex70, IRcube and EM27/SUN using~~ Standard spectrum (red) recorded with the ~~standard procedure using TCCON a~~Bruker IFS ~~priori for measurements performed~~ 125HR at Sodankylä~~in 2017~~ä TCCON facility. Spectrum (~~top row panel~~black) ~~,~~ recorded with a grid placed in the ~~difference~~ parallel optical light path showing reduction of ~~XCO_2 time series for each instrument relative to~~ the ~~reference TCCON results (second row panel)~~non-linearity features in the out-of-band spectral regions. ~~The correlation plots of XCO_2 from LHR~~For comparison, ~~Vertex70, IRcube and EM27/SUN instruments vs TCCON for all measurements with SZA < 75° : Third row-left panel: LHR vs TCCON; third row-right panel: Vertex70 vs TCCON; bottom row-left panel: IRcube vs TCCON; bottom row-right panel: EM27/SUN vs TCCON measurements. The colours represent~~ the ~~measurements performed during the different months~~ maximum intensity of both spectra (not visible in the ~~year~~plot) have been normalized to the same value.

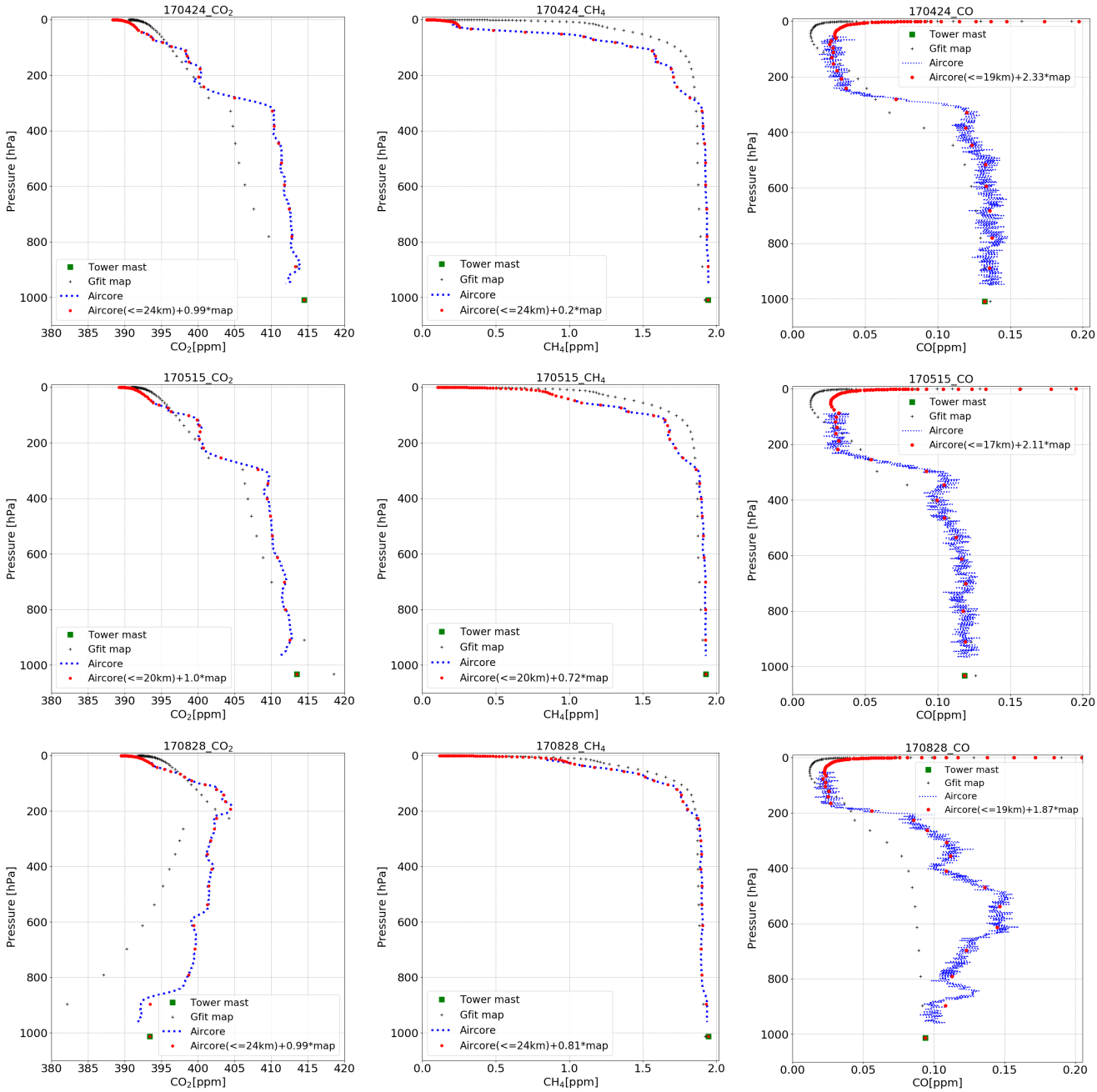

Figure 2. ~~Timeseries of XCH_4 retrievals for TCCON~~AirCore profile, ~~Vertex70~~GFIT map profile, ~~IReube~~ AirCore extended profile and ~~EM27/SUN using the standard procedure using TCCON a priori for~~ Tower mast measurements ~~performed at Sodankyläin 2017~~ are plotted for XCO_2 (~~top row panel~~left-column)~~and the difference of~~ , XCH_4 ~~time series for each instrument relative to the reference TCCON results~~ (~~middle row panel~~middle-column) ~~. The correlation plots~~ and XCO (right-column) as a function of ~~XCH_4 from Vertex70~~altitude for 24 April 2017 with launch time at 15:13:39 and landing time at 16:13:10 UTC (top-row), ~~IReube~~ 15 May 2017 with launch time at 09:33:22 UTC and ~~EM27/SUN instruments vs TCCON for all measurements~~ landing time at 10:25:32 UTC (middle-row) and 28 August 2017 with ~~SZA < 75°~~ launch time at 09:~~Bottom row-left panel~~13:~~Vertex70 vs TCCON; bottom row-middle panel~~15 UTC and landing time at 10:~~IReube vs TCCON; bottom row-right panel~~10:~~EM27/SUN vs TCCON measurements~~33 UTC (bottom-row), respectively.~~The colours represent the measurements performed during the different months of the year.~~

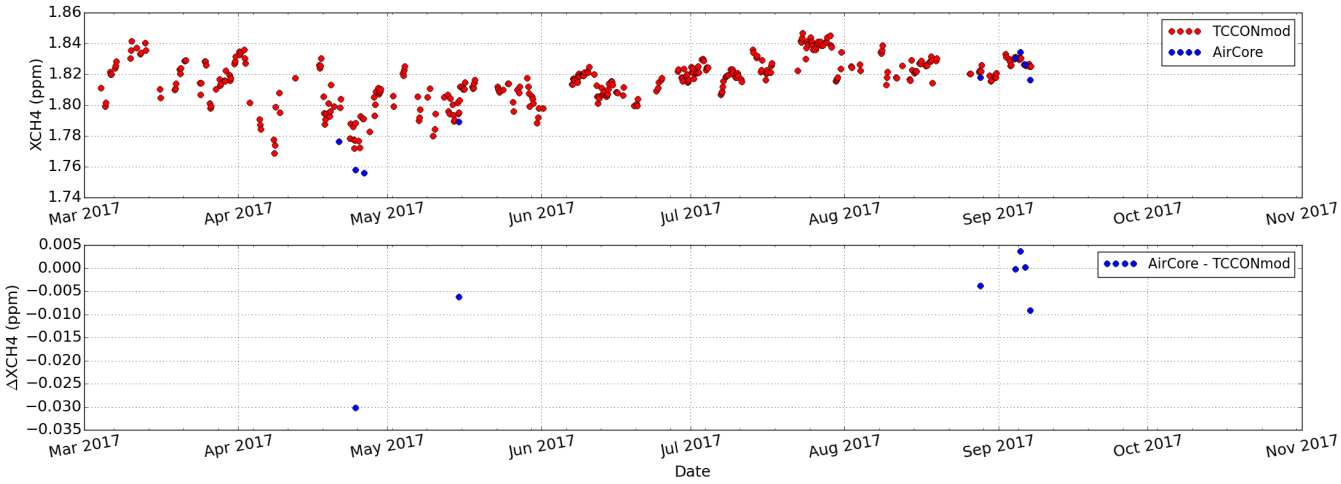

Figure 3. Timeseries of ~~XCO~~ XCH_4 retrievals ~~for TCCON, Vertex70 and EM27/SUN~~ using ~~the standard procedure using~~ non-linearity corrected TCCON ~~a priori for~~ and AirCore measurements ~~performed at Sodankyläin 2017~~ (top ~~row~~ panel) and ~~the difference of XCO time series for each instrument relative to the reference TCCON results~~ bias plot in absolute unit (~~middle row~~ bottom panel) ~~. The correlation plots of XCO from Vertex70 and EM27/SUN instruments vs TCCON~~ plotted for ~~all~~ measurements ~~with~~ performed in 2017 at SZA < 75° ~~: Bottom row-left panel: Vertex70 vs TCCON; bottom row-right panel: EM27/SUN vs TCCON measurements.The colours represent the measurements performed during the different months of the year.~~

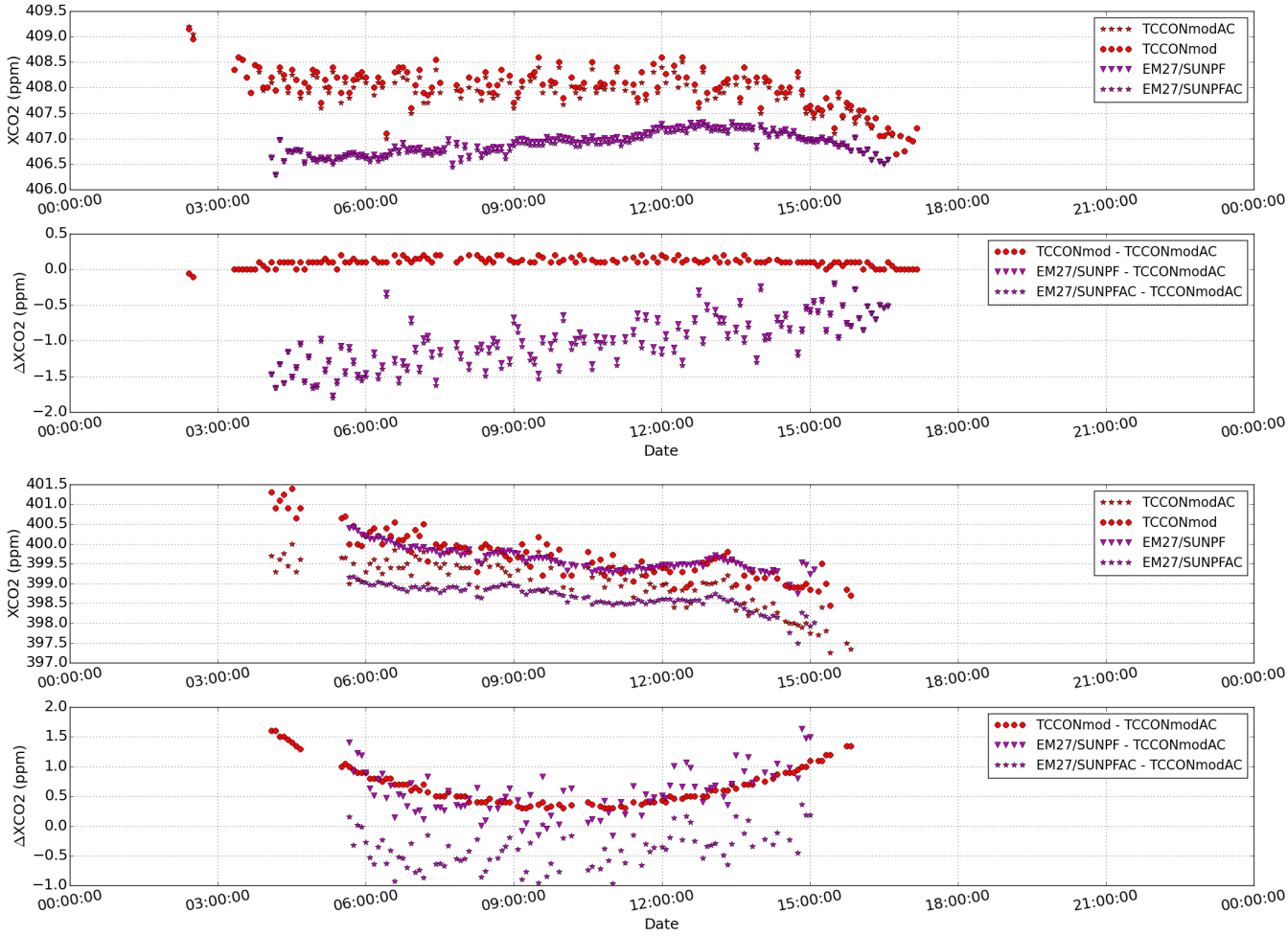

Figure 4. ~~Timeseries of Xair~~ Top figure - upper panel: XCO_2 plotted for ~~TCCON, Vertex70, IRcube~~ TCCONmod and EM27/SUN ~~using~~ retrievals with the ~~standard procedure using~~ TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) for measurements performed on 15 May 2017 at Sodankylä~~in 2017 (top row~~ ä. Top figure - lower panel~~) and~~ : shows the difference ~~of Xair time series for each instrument relative to~~ between the ~~reference TCCON results (middle row panel)~~two retrievals in absolute units. ~~The correlation~~ Bottom figure shows the same plots ~~of Xair from Vertex70, IRcube and EM27/SUN instruments vs TCCON~~ as mentioned above for ~~all~~ measurements ~~with SZA < 75° : Bottom row-left panel: Vertex70 vs TCCON; bottom row-middle panel: IRCUBE vs TCCON; bottom row-right panel: EM27/SUN vs TCCON measurements. The colours represent the measurements~~ performed ~~during the different months of the year~~on 28 August 2017 at Sodankylä.

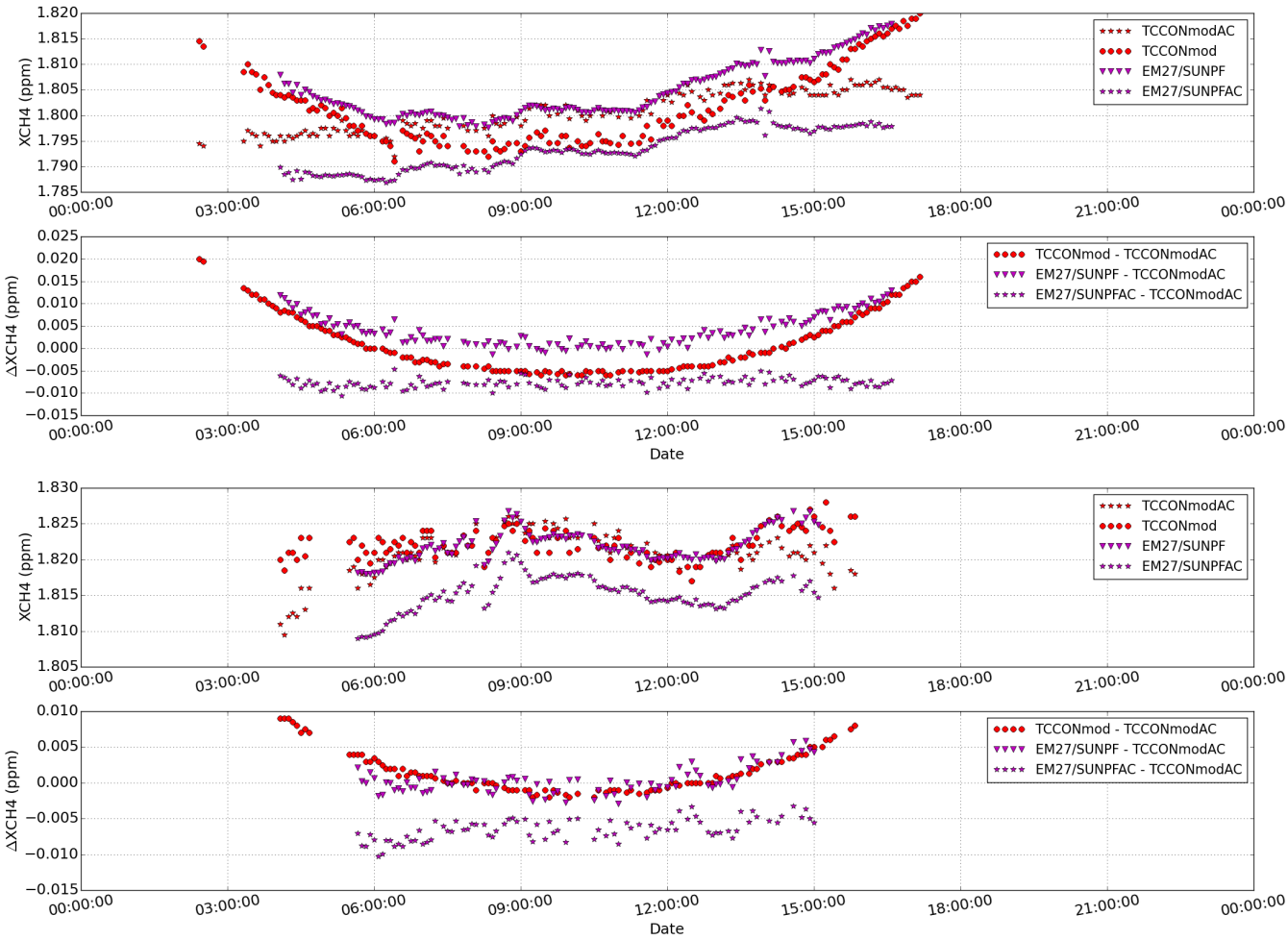

Figure 5. Top figure - upper panel: XCH_4 plotted for TCCONmod and EM27/SUN retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) for measurements performed on 15 May 2017 at Sodankylä. Top figure - lower panel: shows the difference between the two retrievals in absolute units. Bottom figure shows the same plots as mentioned above for measurements performed on 28 August 2017 at Sodankylä.

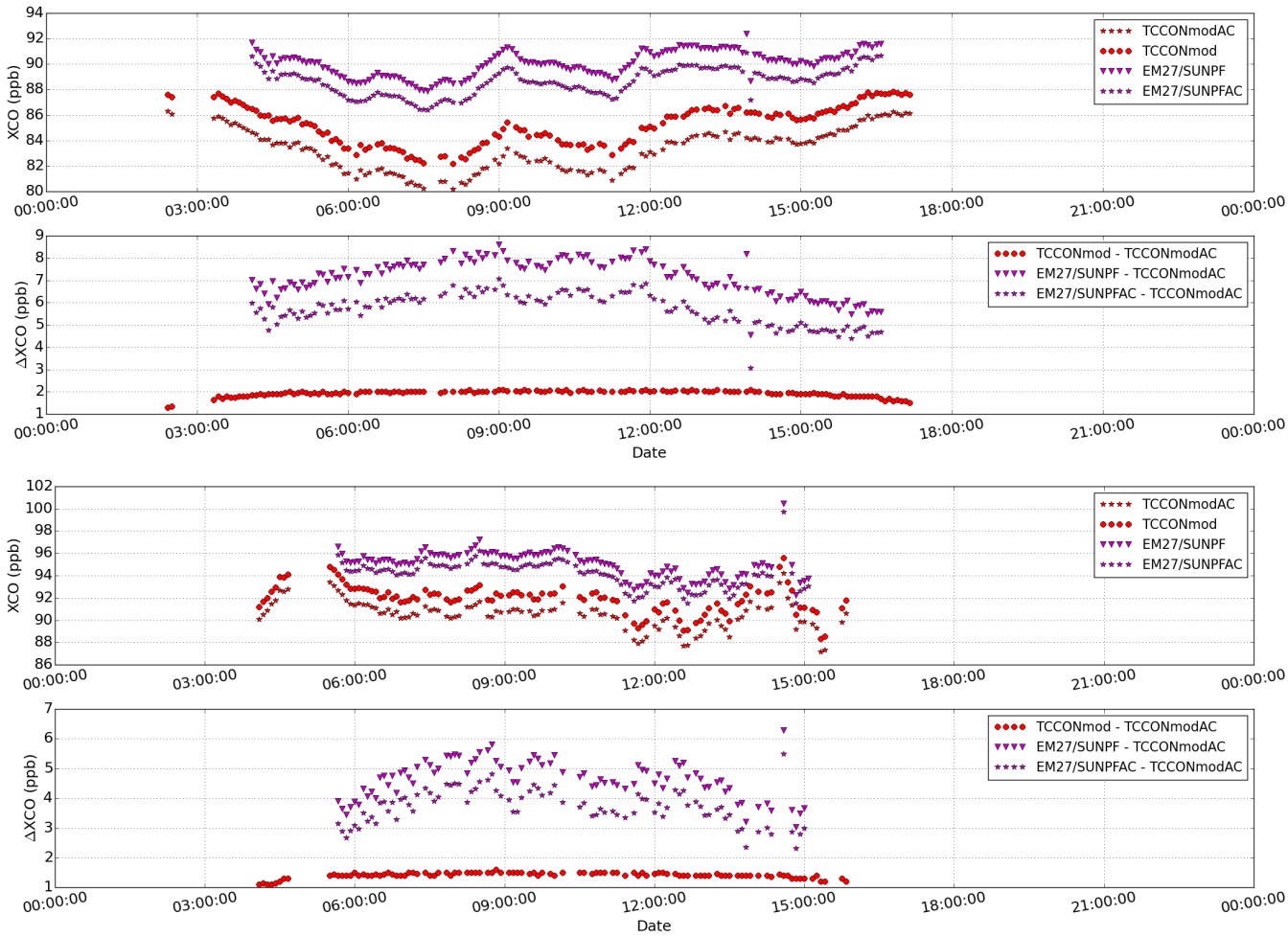

Figure 6. Top figure - upper panel: XCO plotted for TCCONmod and EM27/SUN retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) for measurements performed on 15 May 2017 at Sodankylä. Top figure - lower panel: shows the difference between the two retrievals in absolute units. Bottom figure shows the same plots as mentioned above for measurements performed on 28 August 2017 at Sodankylä.

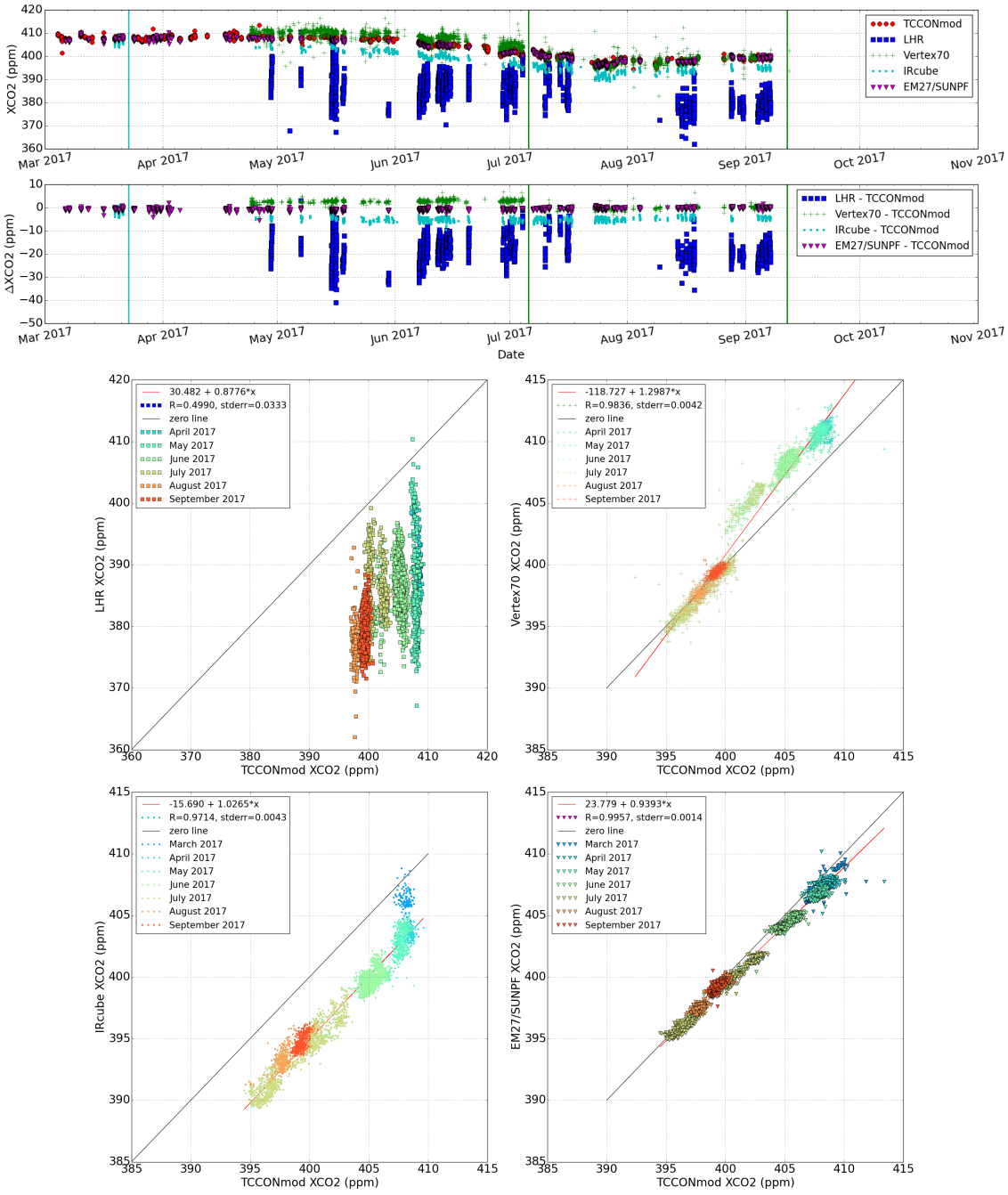

Figure 7. Timeseries of XCO_2 retrievals for TCCONmod, LHR, Vertex70, IRcube and EM27/SUN using the standard procedure using TCCON a priori for measurements performed at Sodankylä in 2017 (top row panel), the difference of XCO_2 time series for each instrument relative to the reference TCCONmod results (second row panel). The correlation plots of XCO_2 from LHR, Vertex70, IRcube and EM27/SUN instruments vs TCCONmod for all measurements with SZA < 75° : Third row-left panel: LHR vs TCCONmod; third row-right panel: Vertex70 vs TCCONmod; bottom row-left panel: IRcube vs TCCONmod; bottom row-right panel: EM27/SUN vs TCCONmod measurements. The colours represent the measurements performed during the different months of the year.

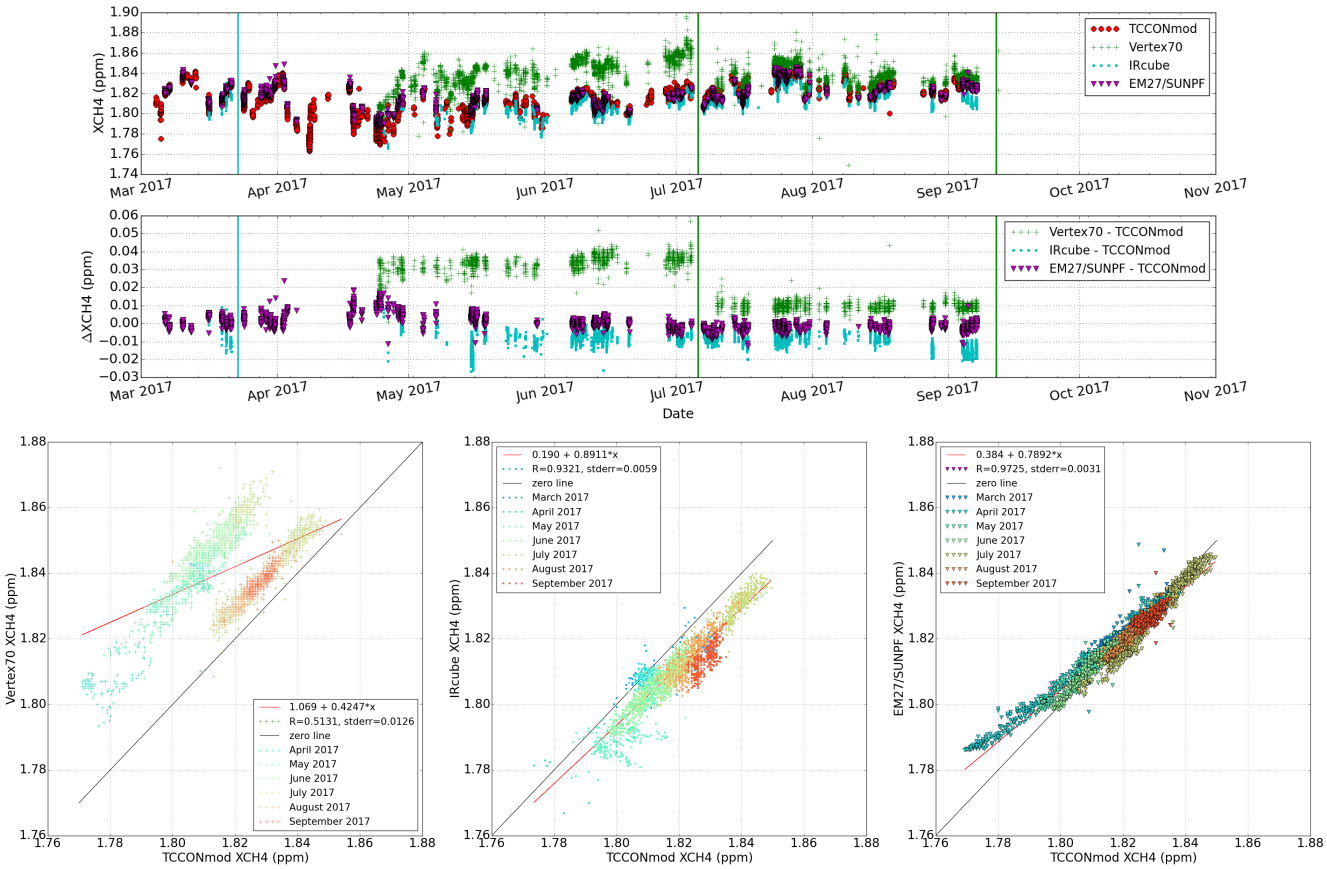

Figure 8. Timeseries of XCH_4 retrievals for TCCONmod, Vertex70, IRcube and EM27/SUN using the standard procedure using TCCON a priori for measurements performed at Sodankylä in 2017 (top row panel) and the difference of XCH_4 time series for each instrument relative to the reference TCCONmod results (middle row panel). The correlation plots of XCH_4 from Vertex70, IRcube and EM27/SUN instruments vs TCCONmod for all measurements with SZA < 75° : Bottom row-left panel: Vertex70 vs TCCONmod; bottom row-middle panel: IRcube vs TCCONmod; bottom row-right panel: EM27/SUN vs TCCONmod measurements. The colours represent the measurements performed during the different months of the year.

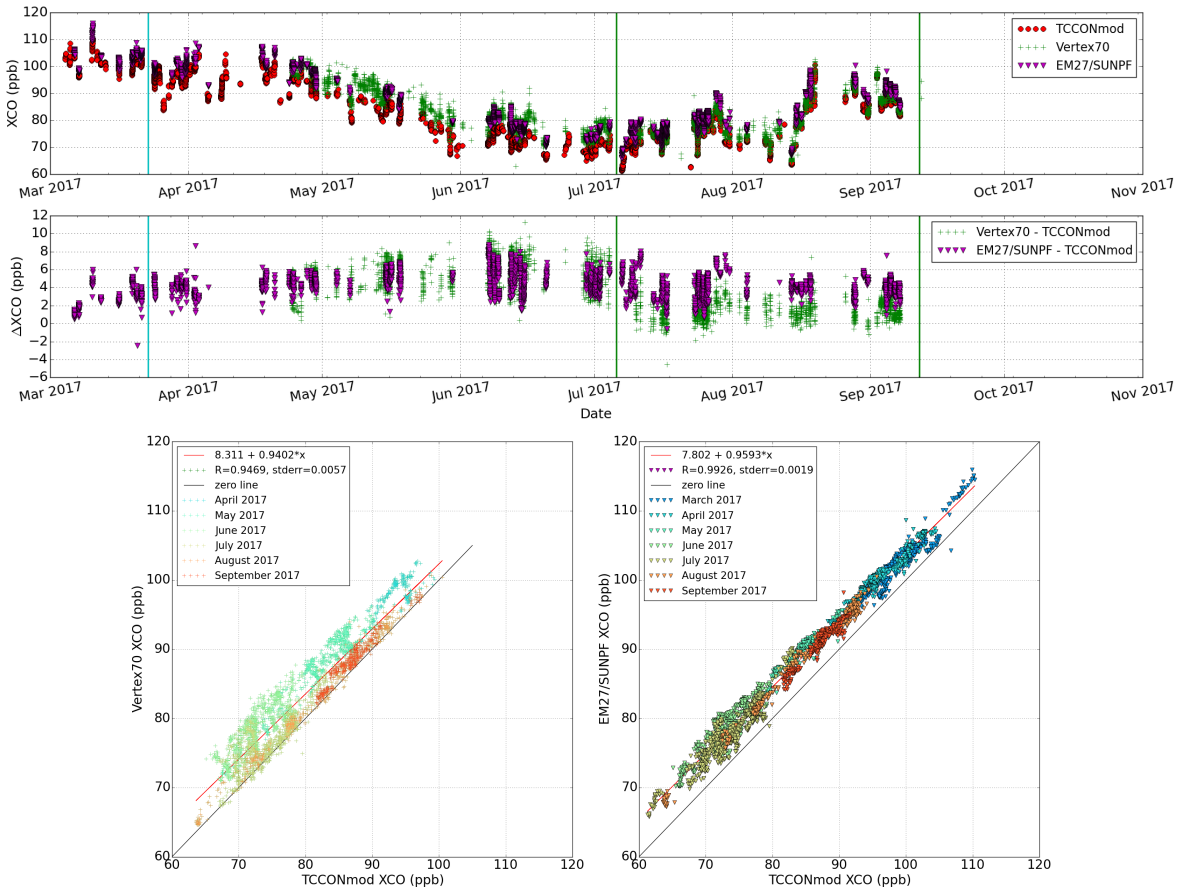

Figure 9. Timeseries of XCO retrievals for TCCONmod, Vertex70 and EM27/SUN using the standard procedure using TCCON a priori for measurements performed at Sodankylä in 2017 (top row panel) and the difference of XCO time series for each instrument relative to the reference TCCONmod results (middle row panel). The correlation plots of XCO from Vertex70 and EM27/SUN instruments vs TCCONmod for all measurements with SZA $< 75°$: Bottom row-left panel: Vertex70 vs TCCONmod; bottom row-right panel: EM27/SUN vs TCCONmod measurements. The colours represent the measurements performed during the different months of the year.

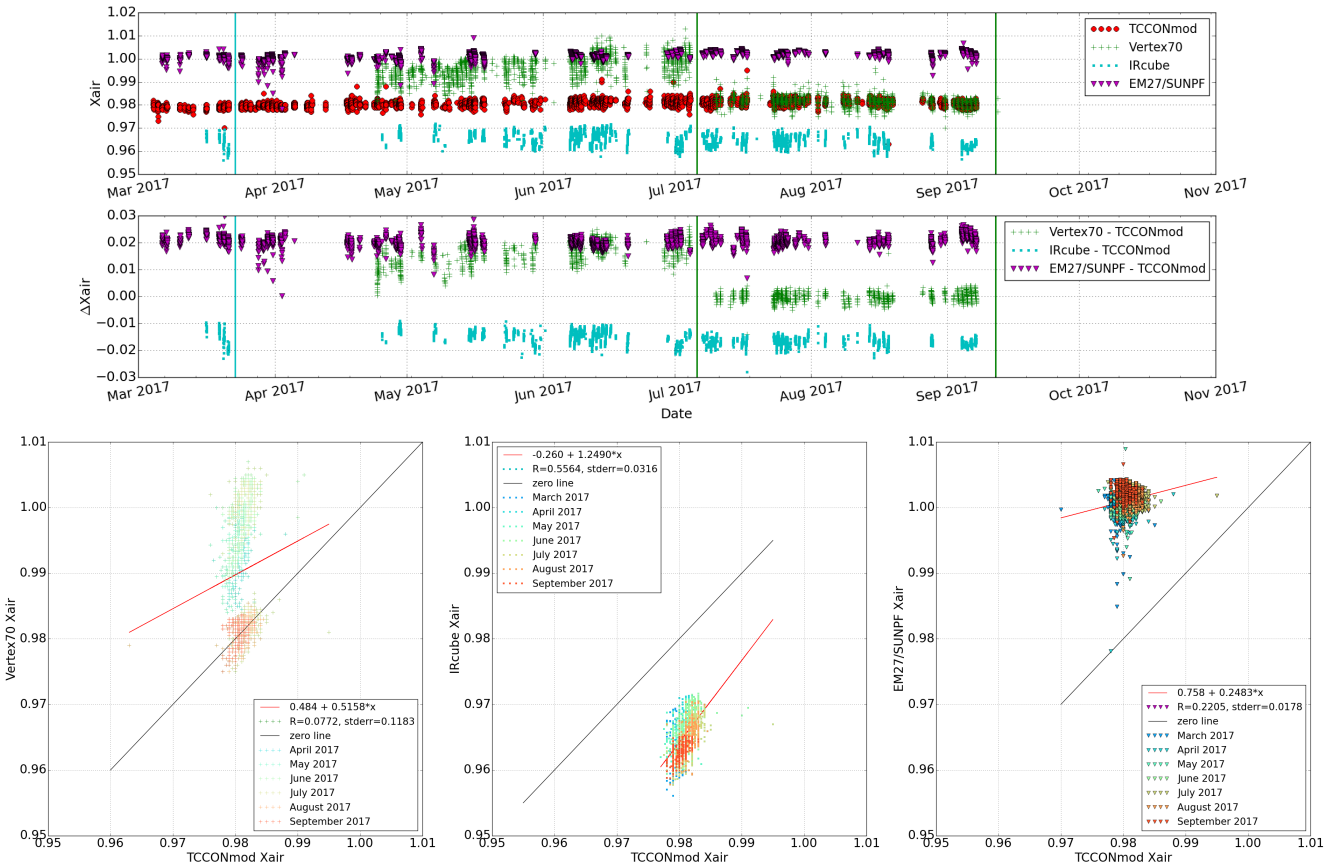

Figure 10. Timeseries of Xair for TCCONmod, Vertex70, IRcube and EM27/SUN using the standard procedure using TCCON a priori for measurements performed at Sodankylä in 2017 (top row panel) and the difference of Xair time series for each instrument relative to the reference TCCONmod results (middle row panel). The correlation plots of Xair from Vertex70, IRcube and EM27/SUN instruments vs TCCONmod for all measurements with SZA < 75° : Bottom row-left panel: Vertex70 vs TCCONmod; bottom row-middle panel: IRCUBE vs TCCONmod; bottom row-right panel: EM27/SUN vs TCCONmod measurements. The colours represent the measurements performed during the different months of the year.

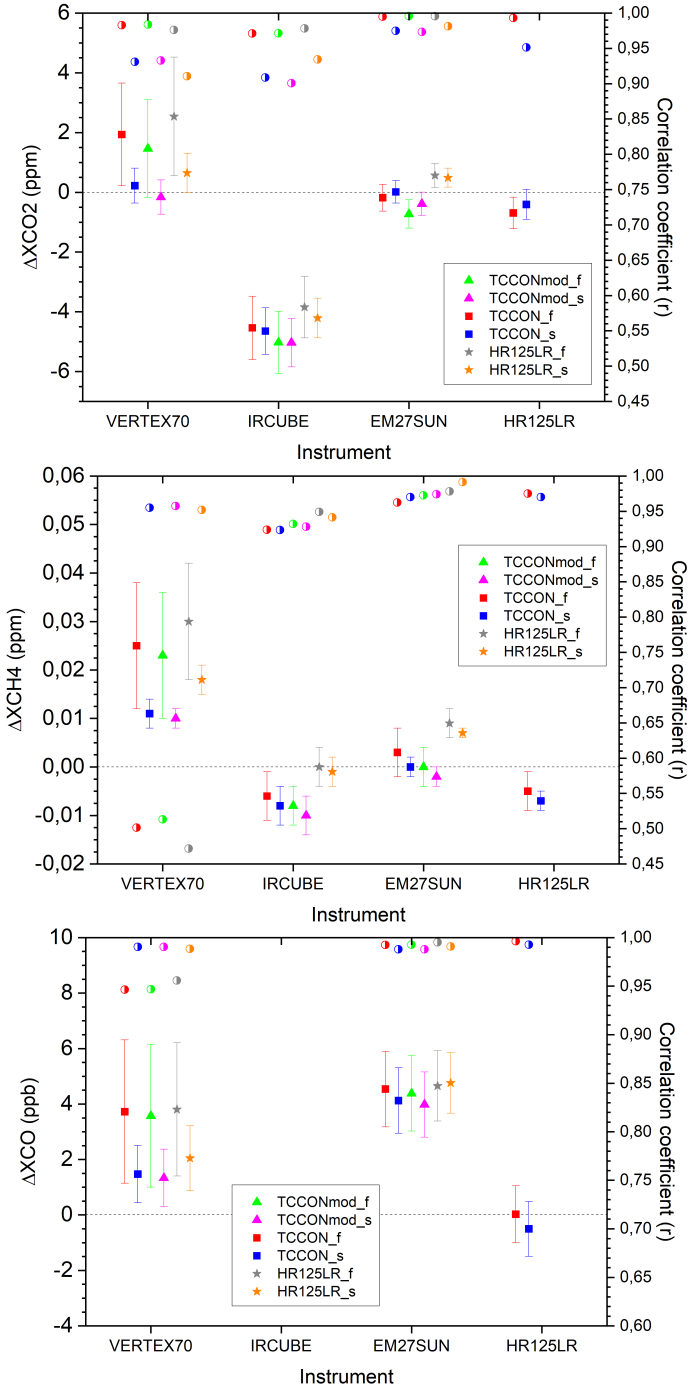

Figure 11. Top plot: XCO_2 bias plotted for each instrument relative to non-linearity corrected TCCON (full year–green triangle, short period–magenta triangle), relative to TCCON (full year–red box, short period–blue box) and relative to HR125LR (full year–grey star, short period–orange star). The correlation coefficient of the respective data set are plotted as half filled circles and correspond to the right side y-axis.The XCH_4 and XCO biases for each instrument are plotted in the middle and lower panel plots, respectively. A horizontal dashed line at zero is overlayed on each plot to help in the interpretation of the results.

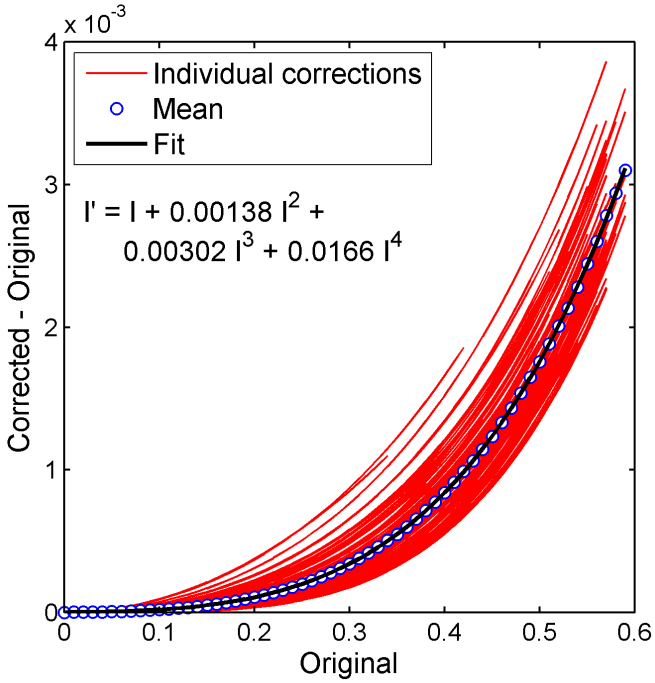

Figure A1. Plot showing the difference of the corrected interferograms - original interferograms vs the original interferograms. The individual corrections are plotted in red, the mean value is plotted as blue open circles and the black line is the fit.

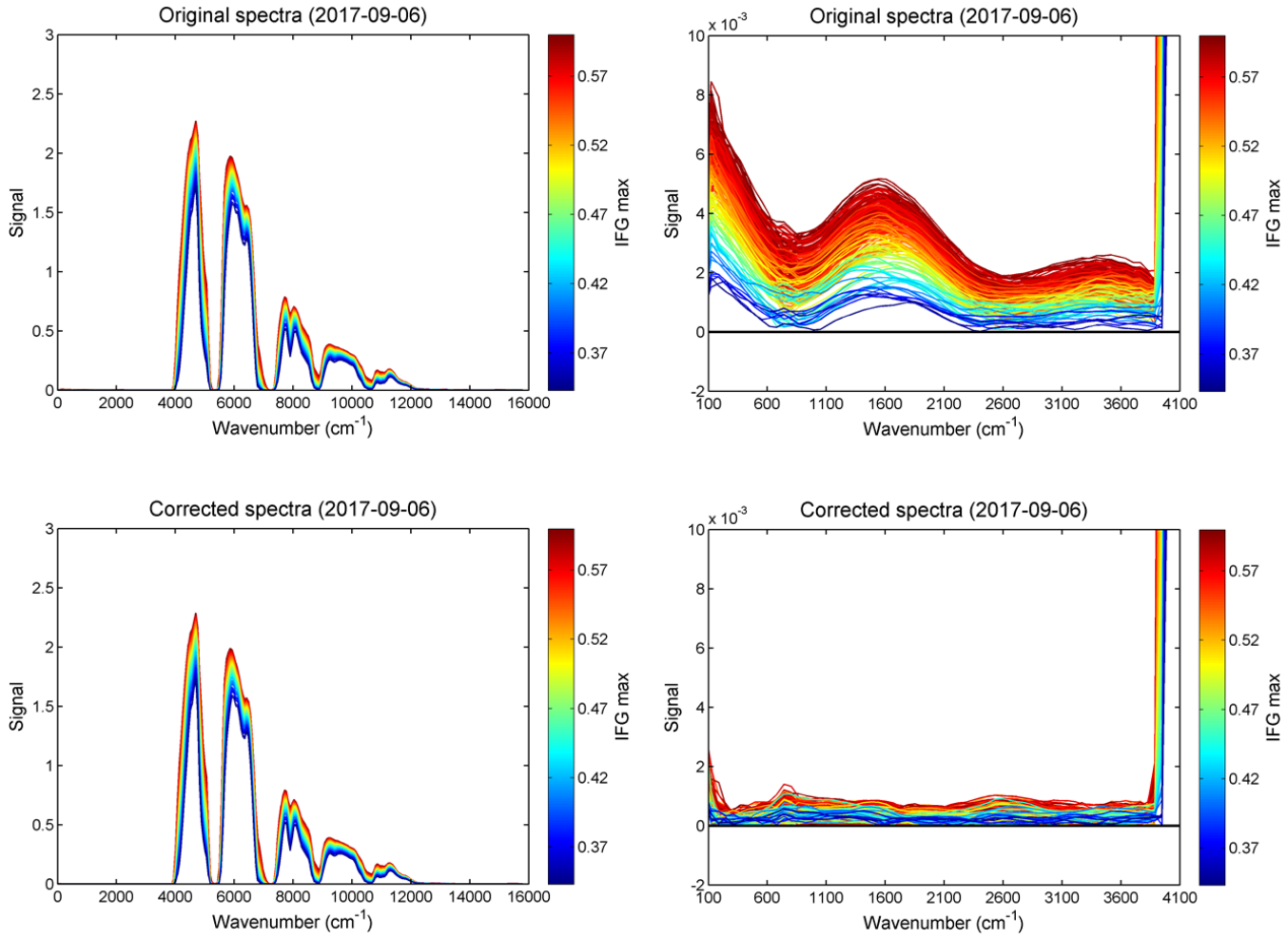

Figure A2. Original (top left) and non-linearity corrected (bottom left) spectra; zoom of the out-of-band spectral region (100–3600 cm^{-1}) with the original spectra (top right) and non-linearity corrected (bottom right) spectra from the Bruker IFS 125HR at Sodankylä TCCON facility. The colour of the spectrum depends on the interferogram maximum signal at the center burst. The highest values corresponding to the dark red colour are recorded during the noon time when the signal is the highest.

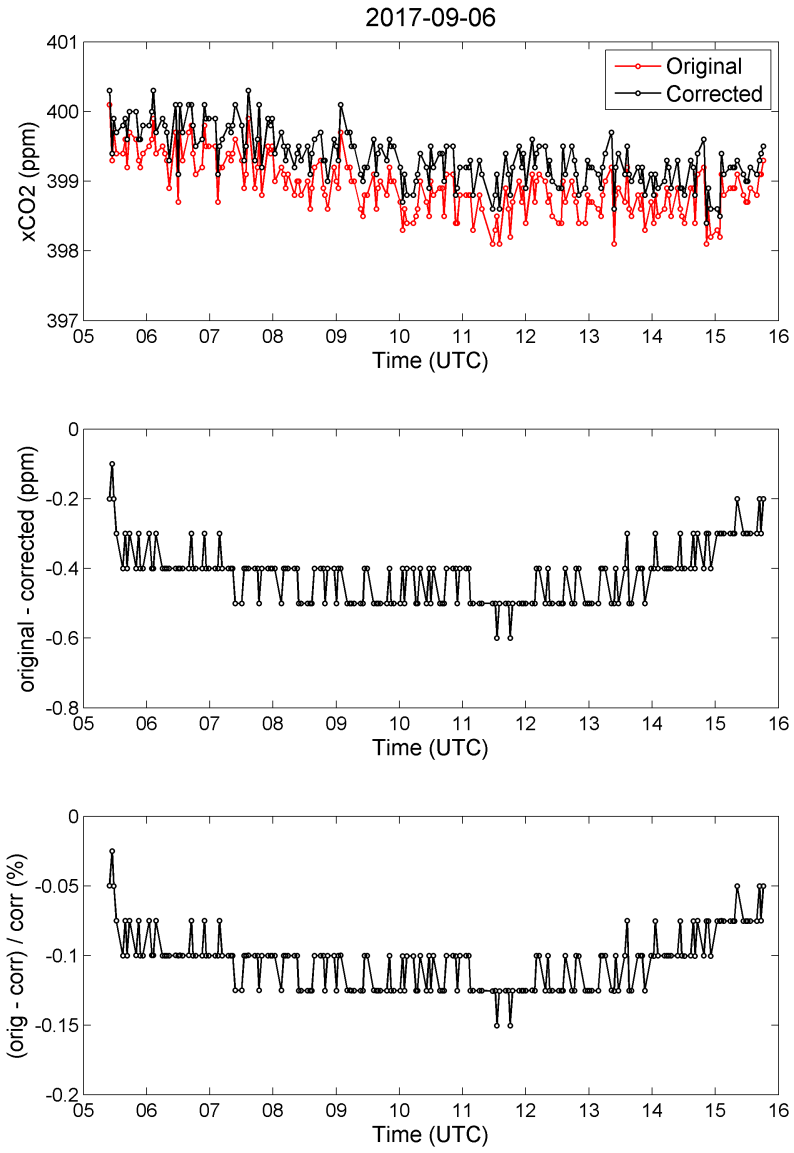

Figure A3. Top panel: plot showing the original (red) and non-linearity corrected (black) XCO2 values for one day of measurement performed on 06 September 2017 by the Bruker 125HR Sodankylä TCCON instrument. Middle-panel: shows the difference between the original and the corrected XCO2 values. Lower-panel: shows the relative difference (original - corrected)/corrected in percentage for the XCO2 values.

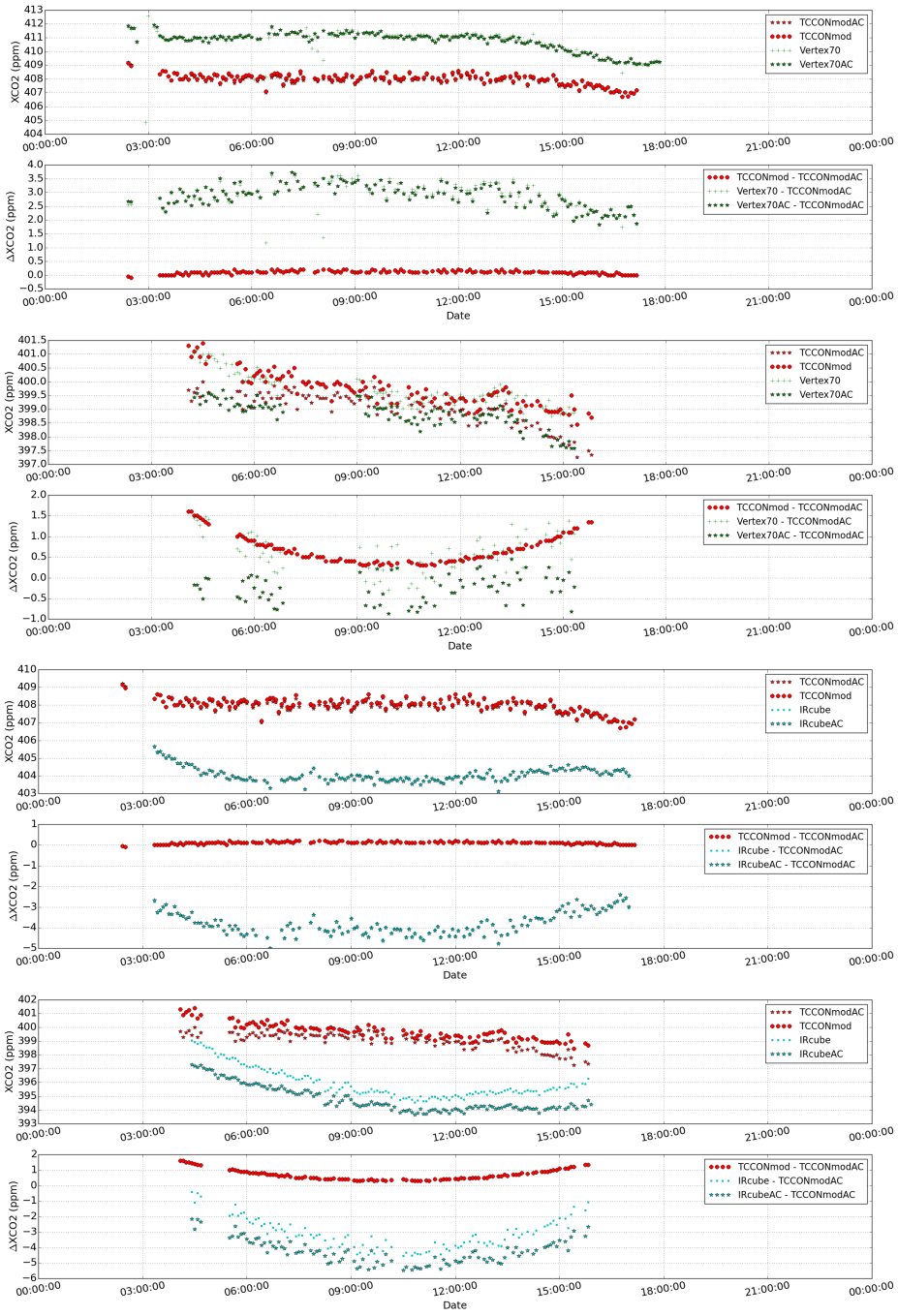

Figure A4. XCO_2 plotted for TCCONmod and Vertex70 retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) are shown in top row panel and their difference in second row panel for measurements performed on 15 May 2017 at Sodankylä. Third and fourth panel plots show the same plots for 28 August 2017. The same plots for TCCONmod and IRcube retrievals for measurements performed on 15 May 2017 are shown in fifth and sixth row panels and that for the 28 August 2017 are shown in seventh and eight panel plots.

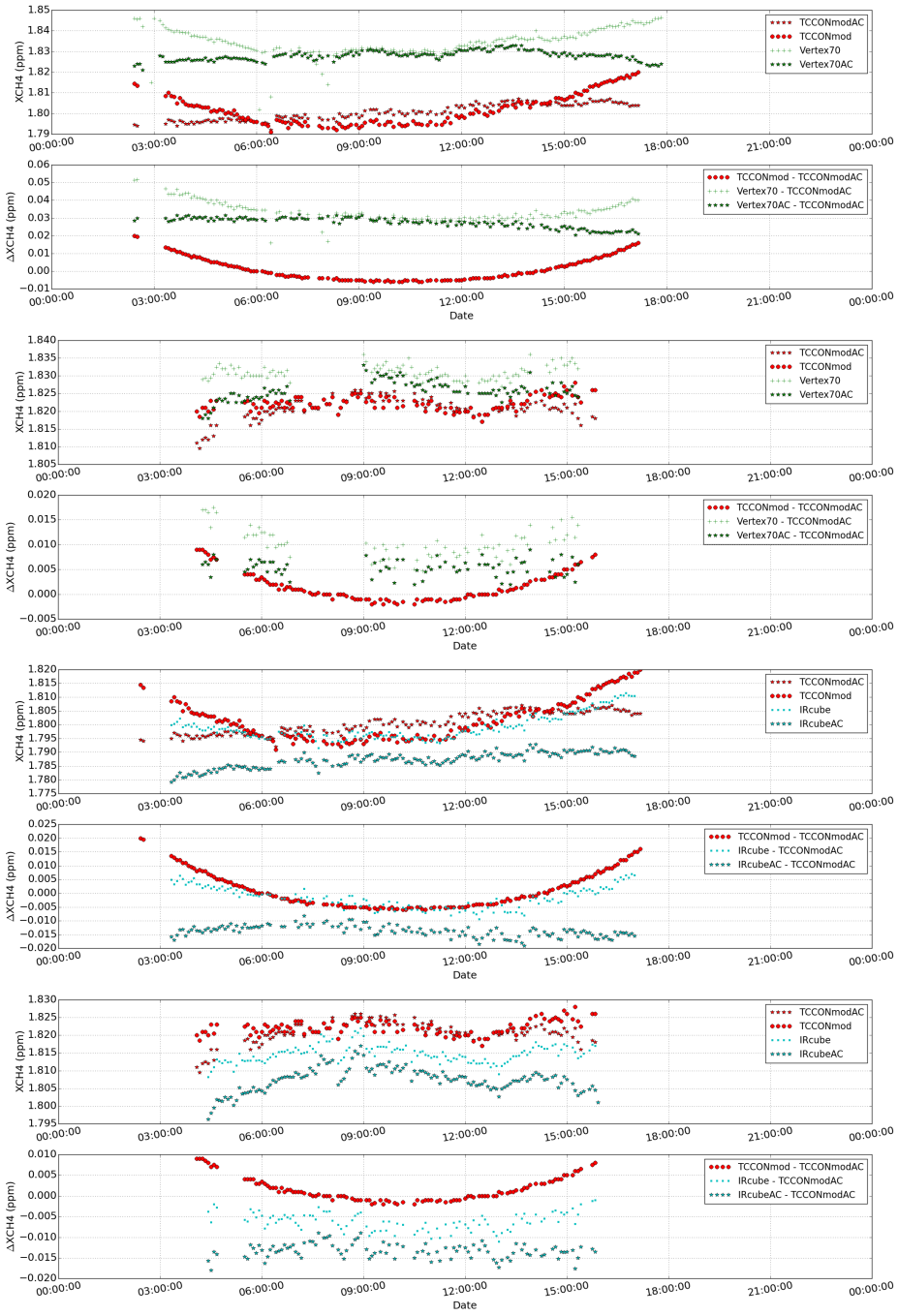

Figure A5. XCH_4 plotted for TCCONmod and Vertex70 retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) are shown in top row panel and their difference in second row panel for measurements performed on 15 May 2017 at Sodankylä. Third and fourth panel plots show the same plots for 28 August 2017. The same plots for TCCONmod and IRcube retrievals for measurements performed on 15 May 2017 are shown in fifth and sixth row panels and that for the 28 August 2017 are shown in seventh and eight panel plots.

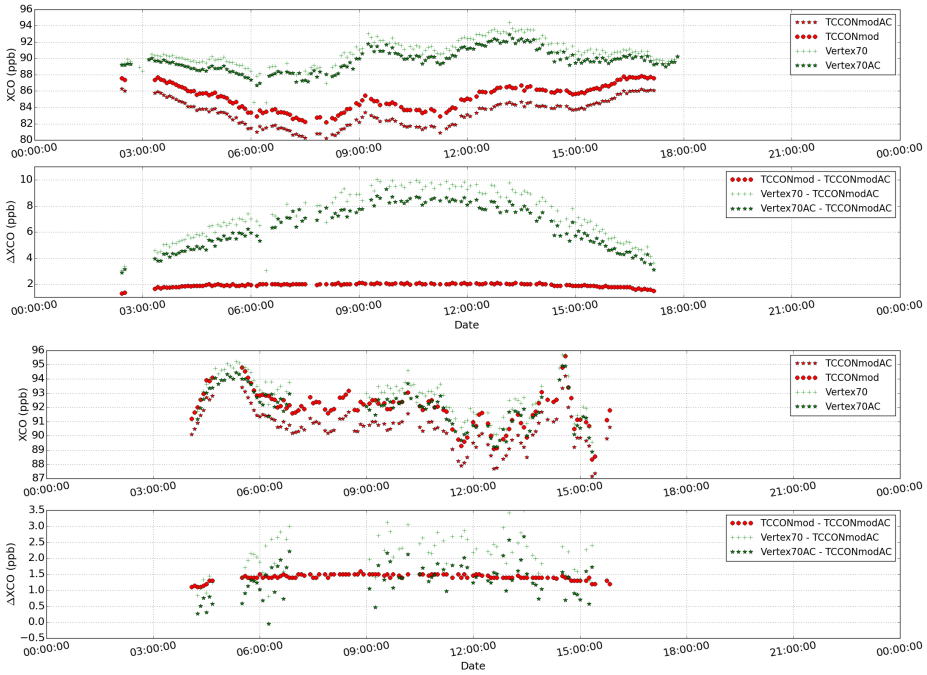

Figure A6. XCO plotted for TCCONmod and Vertex70 retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) are shown in top row panel and their difference in second row panel for measurements performed on 15 May 2017 at Sodankylä. Third and fourth panel plots show the same plots for 28 August 2017.

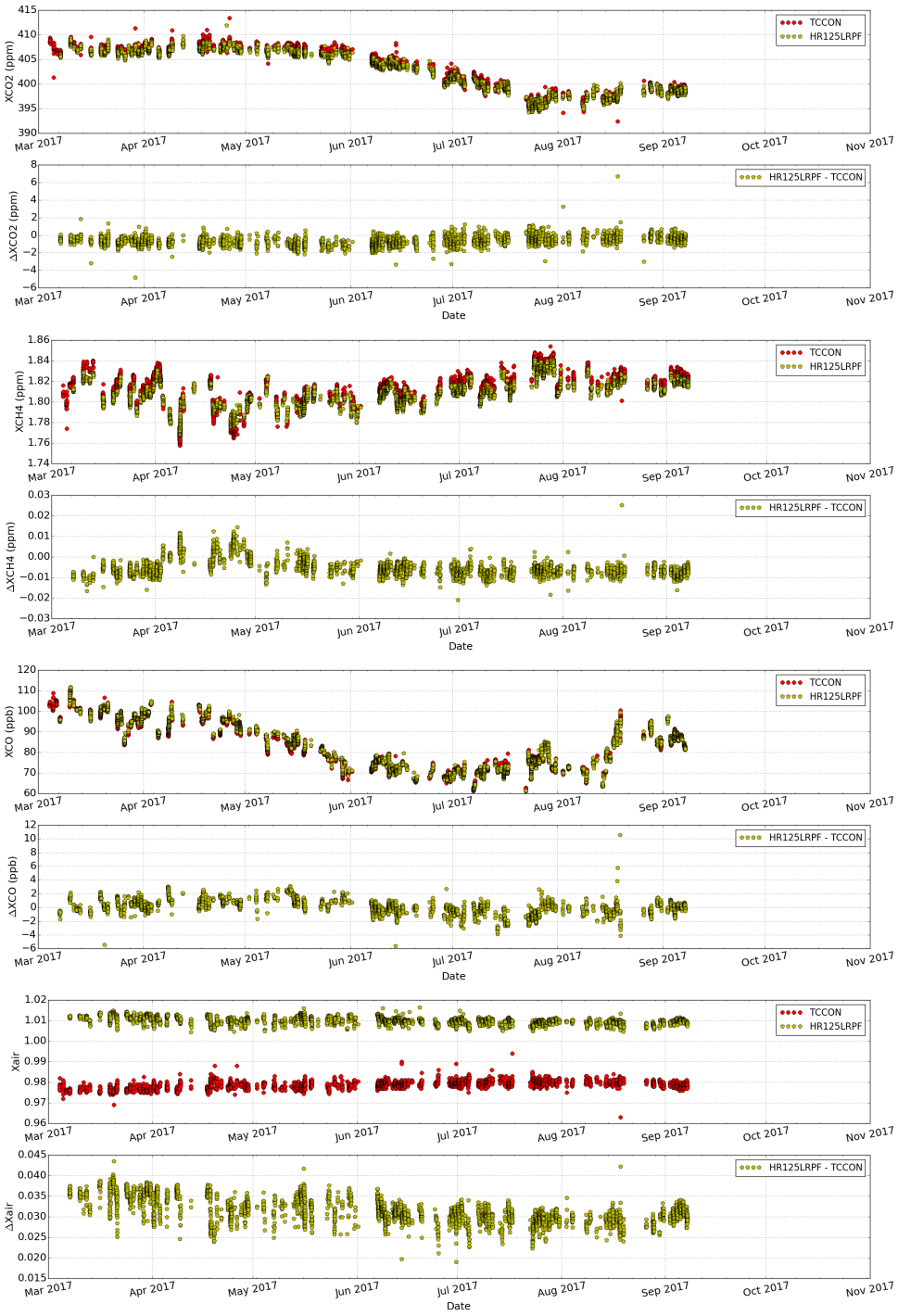

Figure A7. Timeseries of XCO_2 (first row panel), XCH_4 (third row panel), XCO (fifth row panel) and Xair (seventh row panel) retrievals for TCCON and HR125LR using the standard procedure using TCCON a priori for measurements performed at Sodankylä in 2017. The difference of XCO_2 (second row panel), XCH_4 (fourth row panel), XCO (sixth row panel) and Xair (eighth row panel) time series for HR125LR relative to the reference TCCON results.

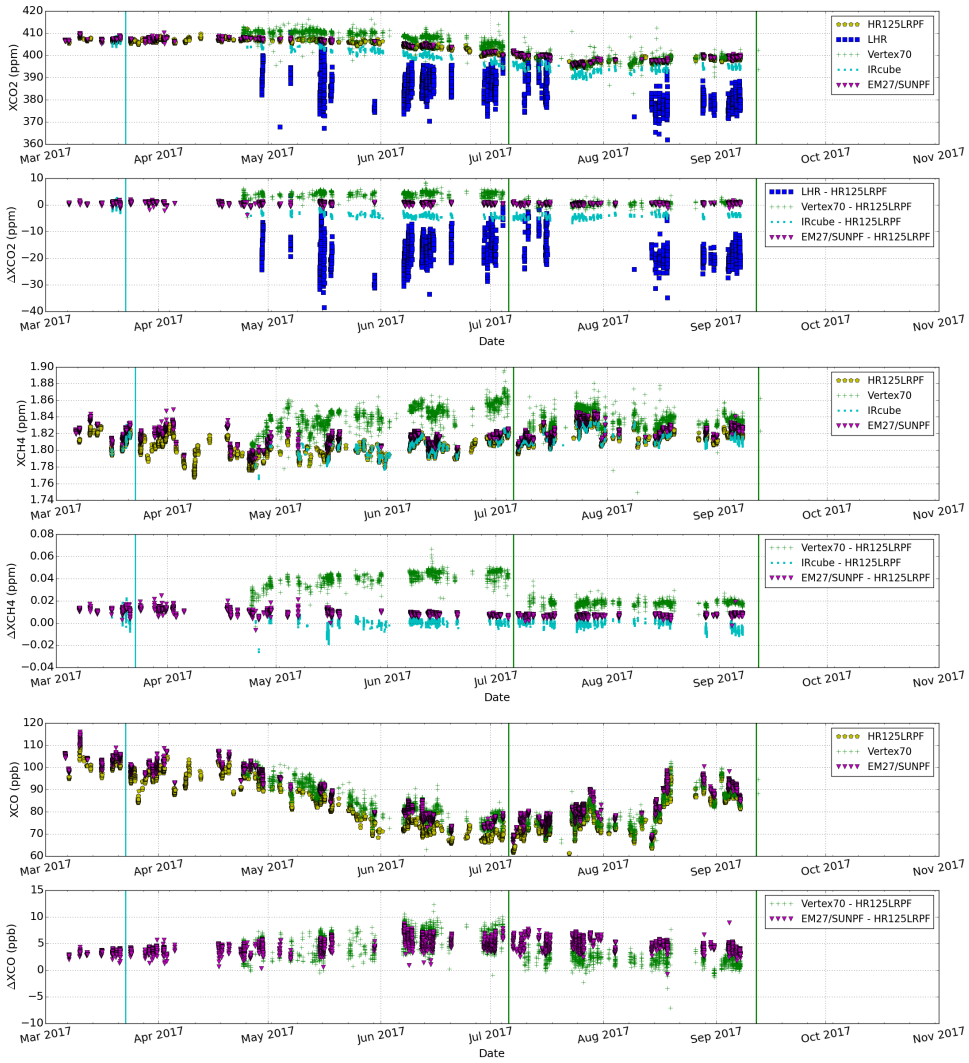

Figure A8. Timeseries of XCO_2 (first row panel), XCH_4 (third row panel) and XCO (fifth row panel) retrievals for HR125LR, LHR, Vertex70, IRcube and EM27/SUN using the standard procedure using TCCON a priori for measurements performed at Sodankylä in 2017. The suffix PF indicates that the retrieval were performed with PROFFAST code. The difference of XCO_2 (second row panel), XCH_4 (fourth row panel) and XCO (sixth row panel) time series for the test instruments relative to the HR125LR results.

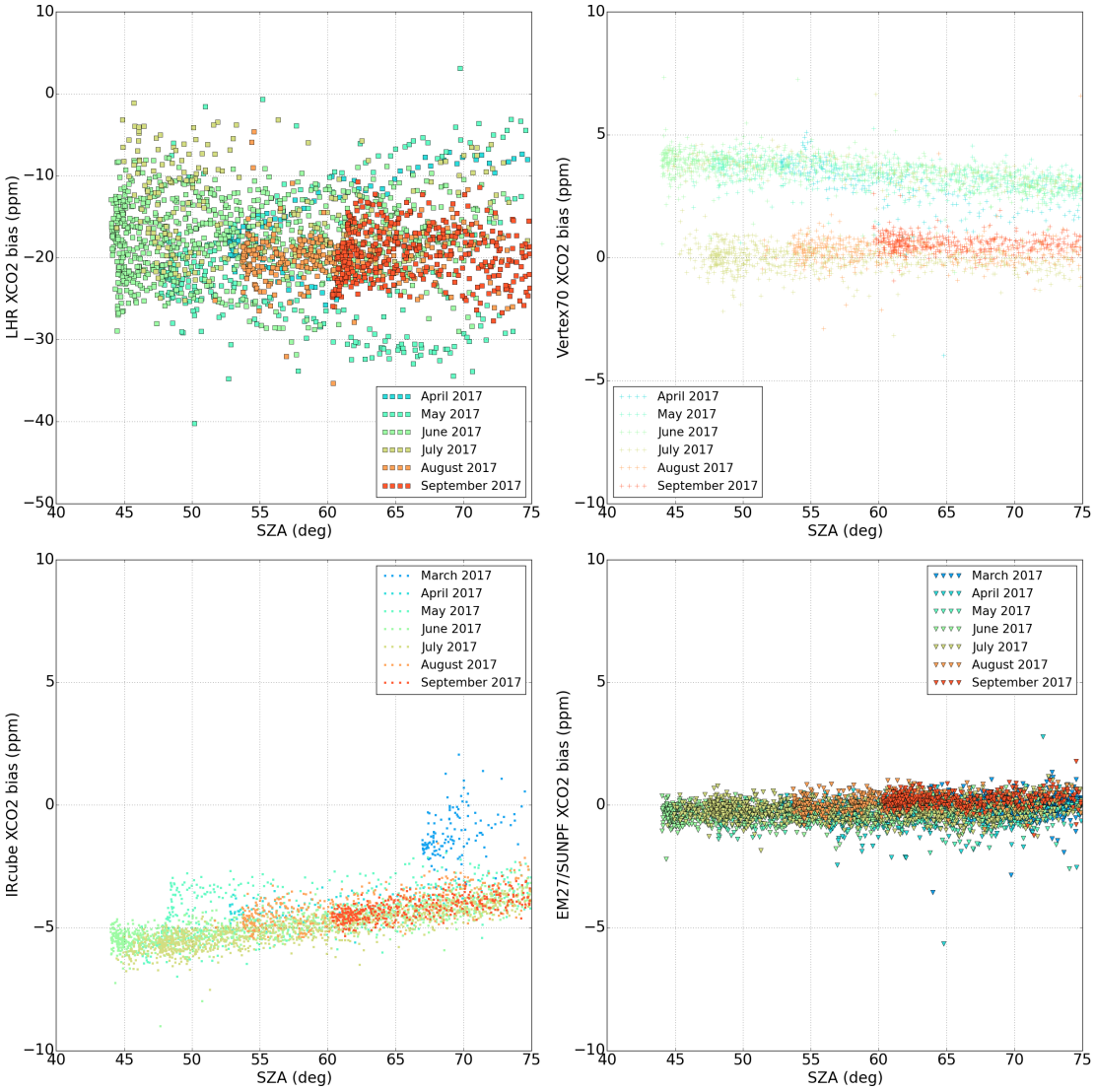

Figure A9. XCO_2 bias relative to TCCON for each instrument plotted w.r.t. the solar zenith angle: LHR (top left), Vertex70 (top right), IRcube (bottom left), EM27/SUN (bottom right). The colours represent the measurement performed during the different months of the year.

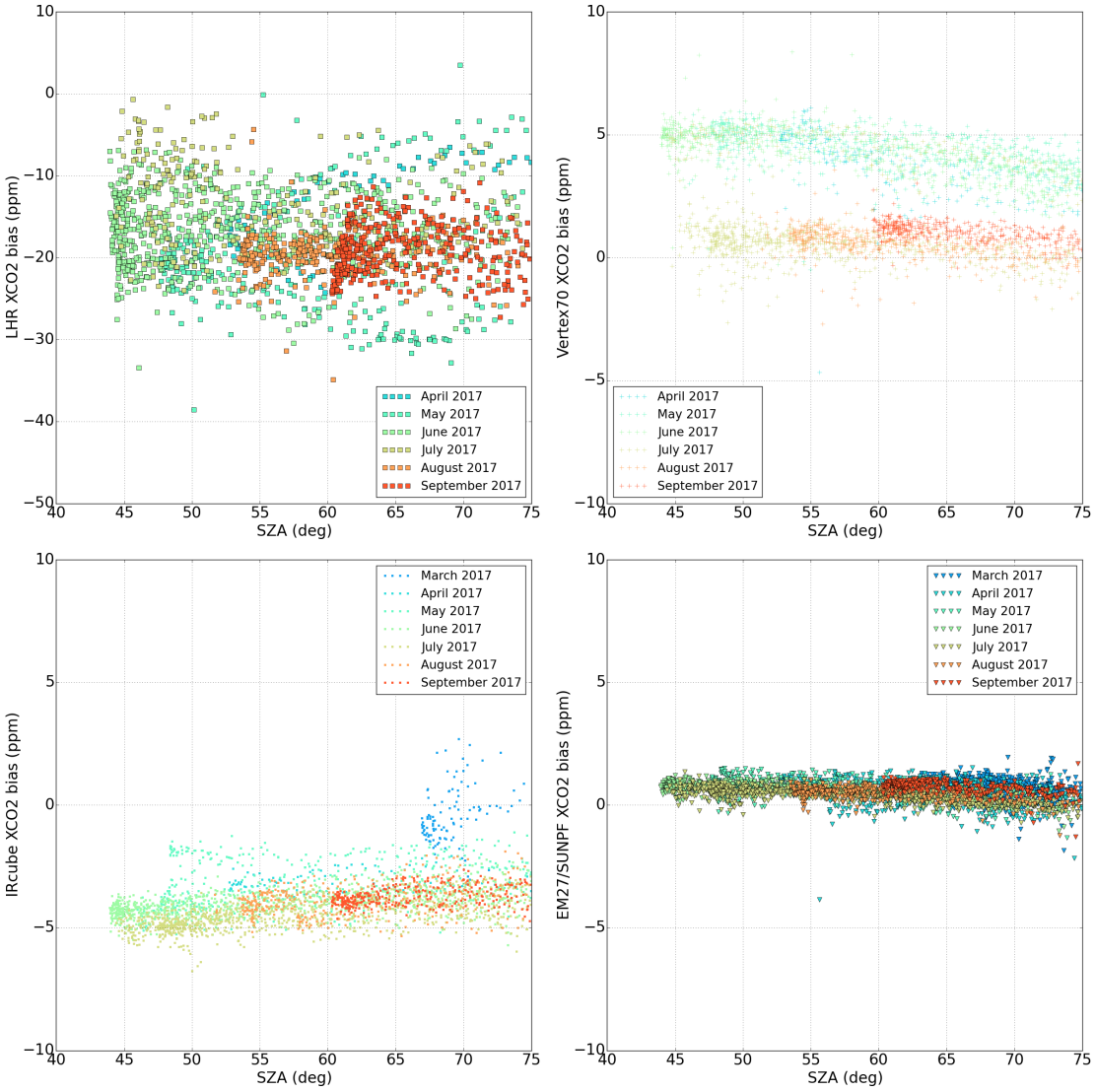

Figure A10. XCO_2 bias relative to HR125LR for each instrument plotted w.r.t. the solar zenith angle: LHR (top left), Vertex70 (top right), IRcube (bottom left), EM27/SUN (bottom right). The colours represent the measurement performed during the different months of the year.

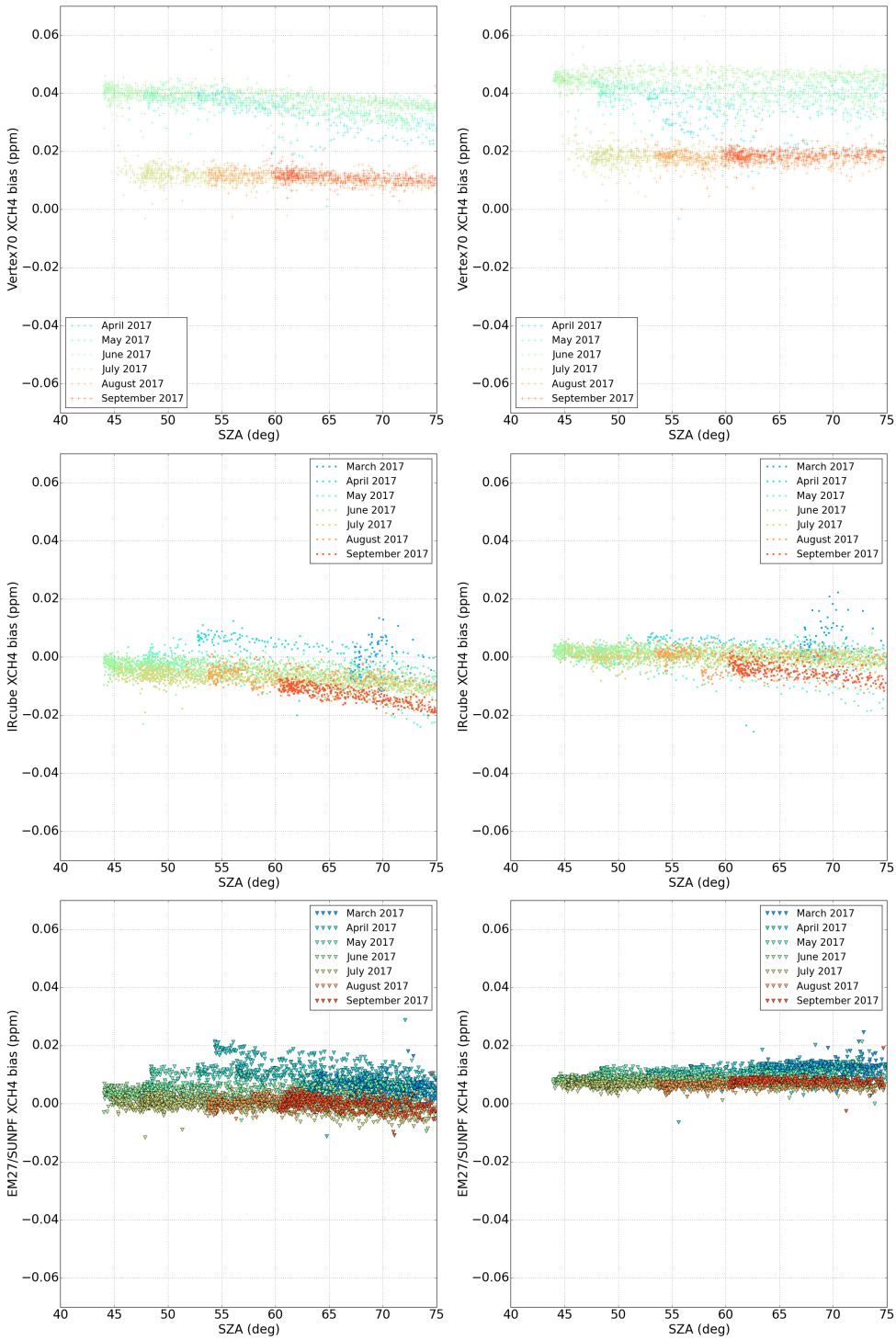

Figure A11. XCH_4 bias relative to TCCON (left column) and relative to HR125LR (right column) for each instrument plotted w.r.t. the solar zenith angle: Vertex70 (top row), IRcube (middle row), EM27/SUN (bottom row). The colours represent the measurement performed during the different months of the year.

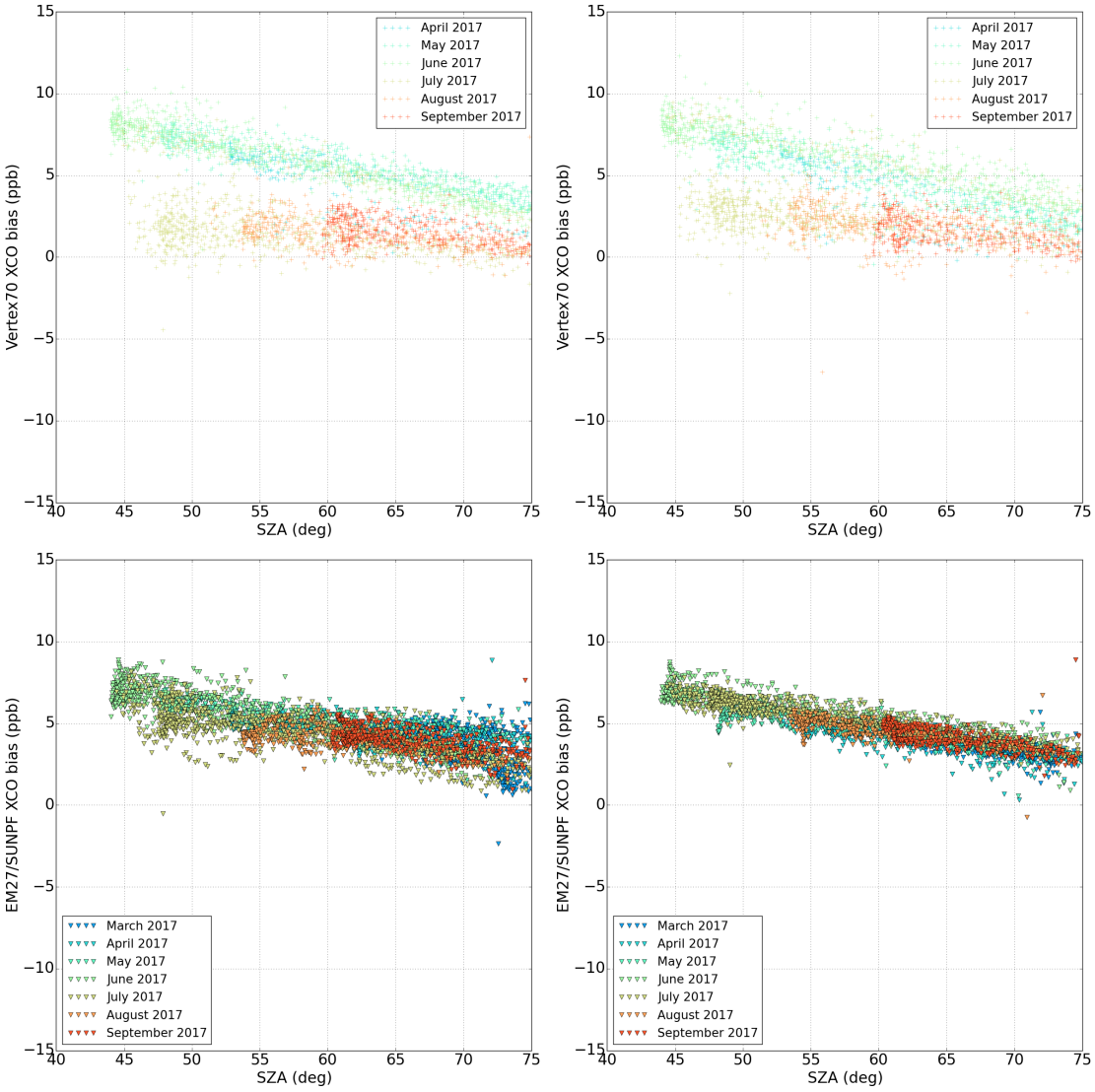

Figure A12. XCO bias relative to TCCON (left column) and relative to HR125LR (right column) for each instrument plotted w.r.t. the solar zenith angle: Vertex70 (top row) and EM27/SUN (bottom row). The colours represent the measurement performed during the different months of the year.

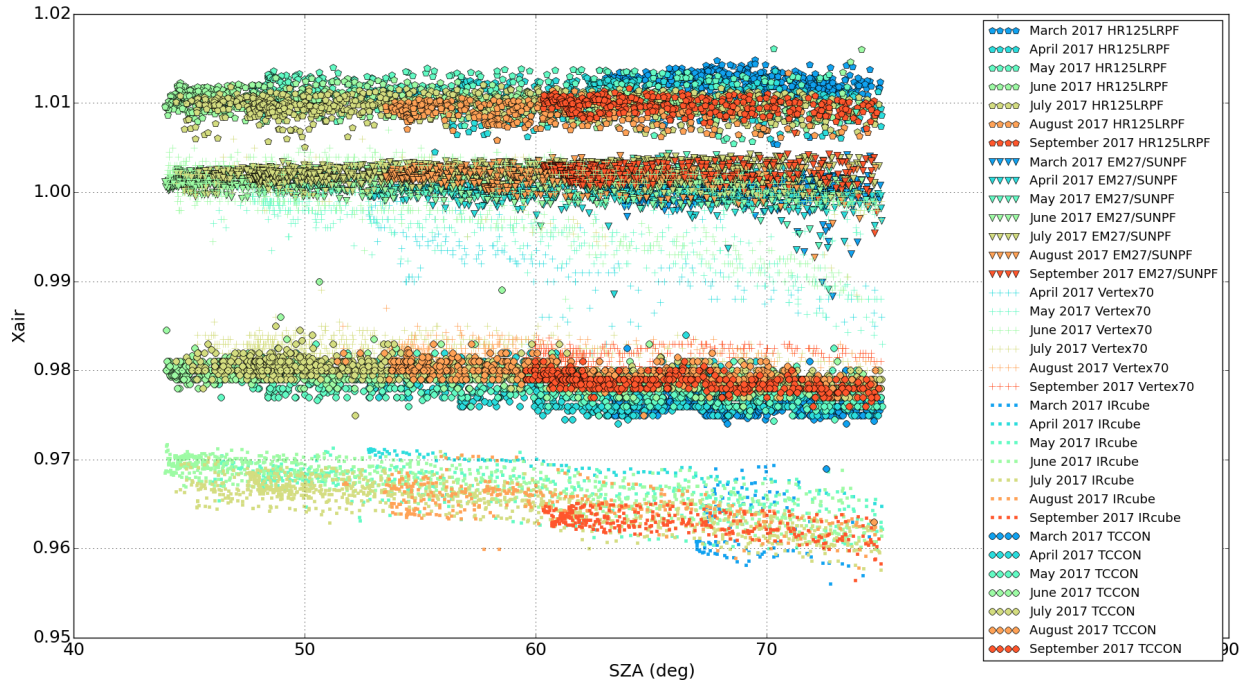

Figure A13. Xair plotted for HR125LR, EM27/SUN, Vertex70, IRcube and TCCON and w.r.t. the measurement solar zenith angle. The colours represent the measurement performed during the different months of the year.

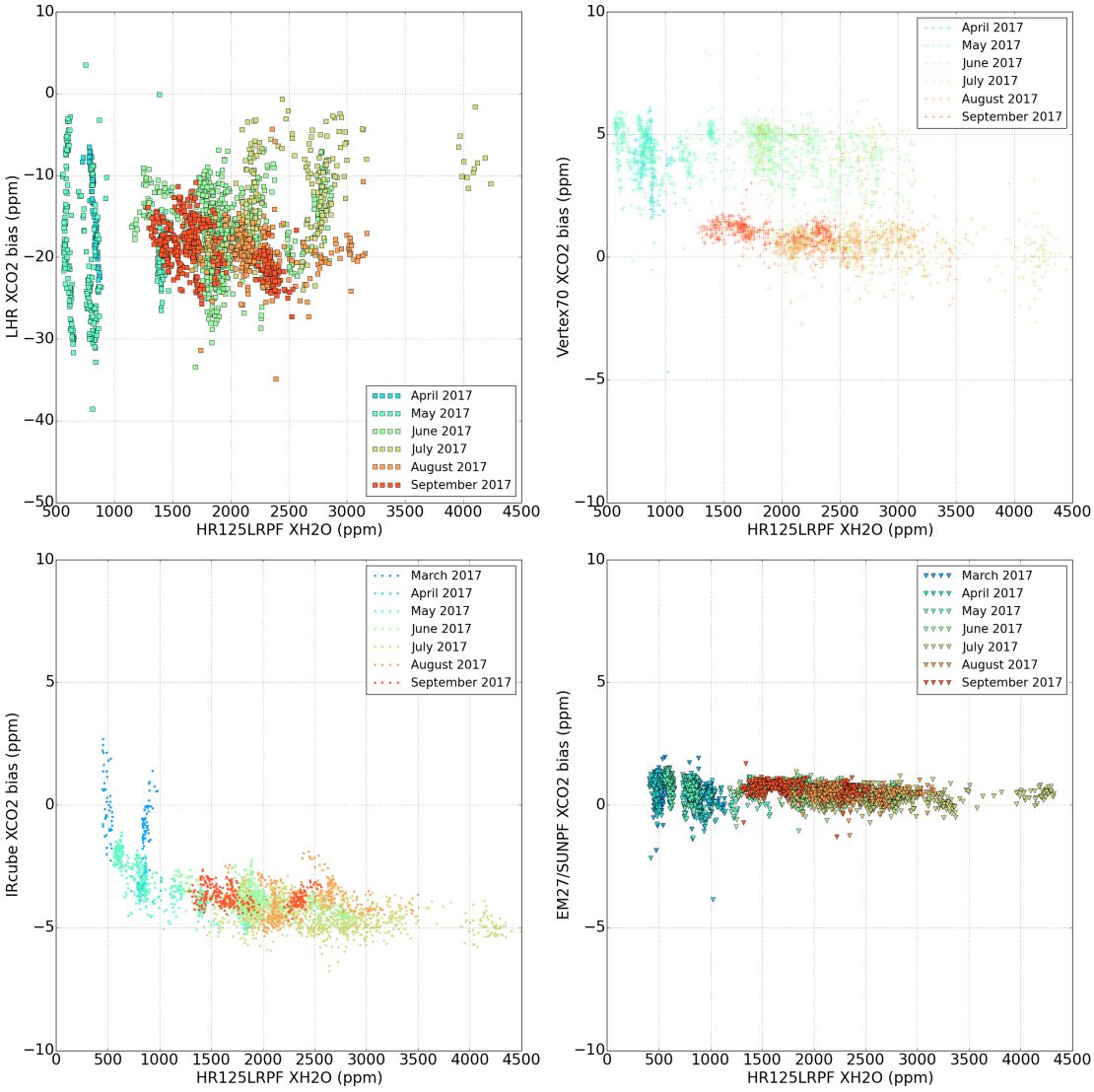

Figure A14. XCO_2 bias relative to HR125LR for each instrument plotted w.r.t. the total column water vapour retrieved by the reference HR125LR measurements: LHR (top left), Vertex70 (top right), IRcube (bottom left), EM27/SUN (bottom right). The colours represent the measurement performed during the different months of the year.

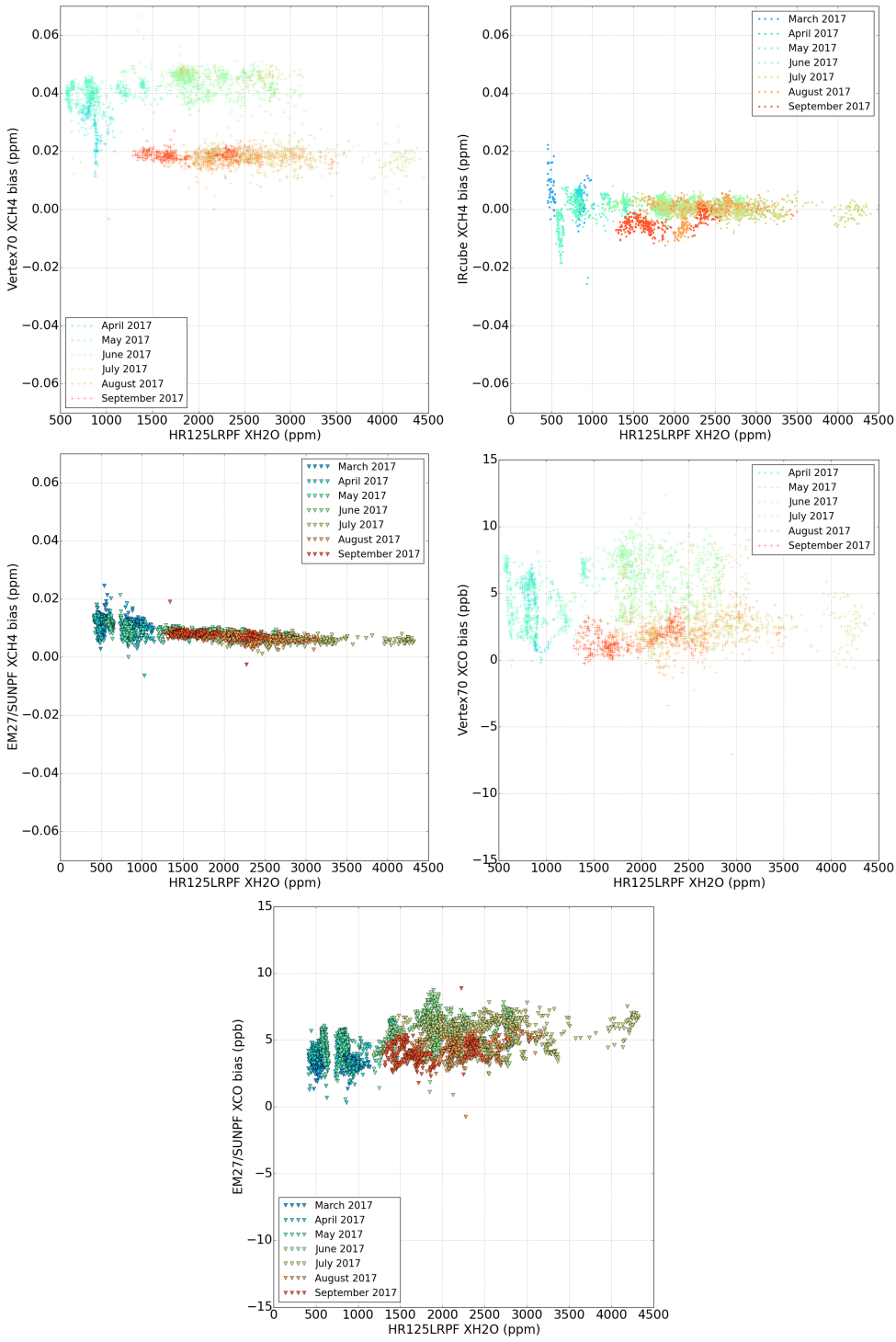

Figure A15. XCH_4 bias relative to HR125LR for Vertex70 (top-left panel), for IRcube (top-right panel) and for EM27/SUN (middle-left panel) and XCO bias relative to HR125LR for Vertex70 (middle-right panel) and for EM27/SUN (bottom panel~~panel~~) plotted w.r.t. the total column water vapour retrieved by the reference HR125LR measurements. The colours represent the measurement performed during the different months of the year.

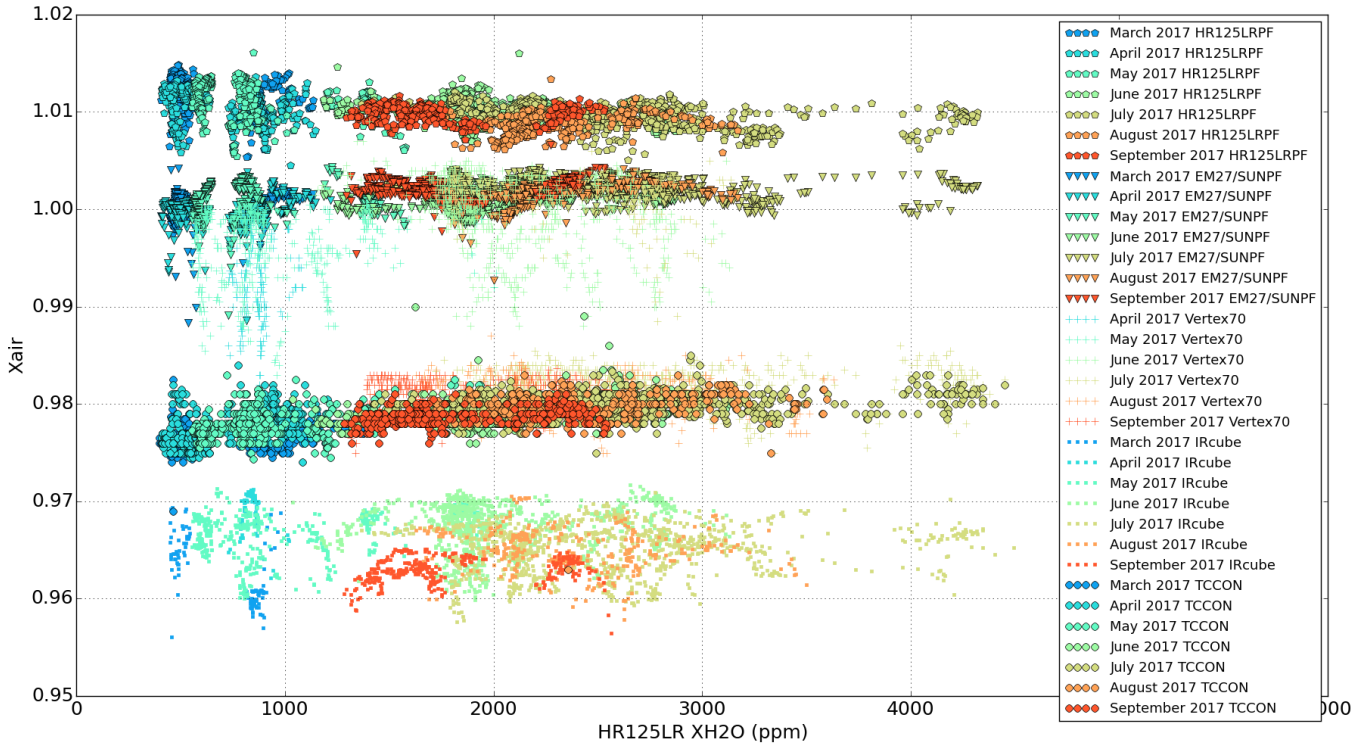

Figure A16. Xair plotted for HR125LR, EM27/SUN, Vertex70, IRcube and TCCON w.r.t. the total column water vapour retrieved by the reference HR125LR measurements. The colours represent the measurement performed during the different months of the year.

~~Standard spectrum (red) recorded with the Bruker IFS 125HR at Sodankylä TCCON facility. Spectrum (black) recorded with a grid placed in the parallel optical light path showing reduction of the non-linearity features in the out-of-band spectral regions. For comparison, the maximum intensity of both spectra (not visible in the plot) have been normalized to the same value.~~

~~Top plot: XCO_2 bias plotted for each instrument relative to TCCON (full year–red box, short period–blue box), relative to non-linearity corrected TCCON (full year–green triangle, short period–magenta triangle) and relative to HR125LR (full year–grey star, short period–orange star). The correlation coefficient of the respective data set are plotted as half filled circles and correspond to the right side y-axis.The XCH_4 and XCO biases for each instrument are plotted in the middle and lower panel plots, respectively.~~

~~AirCore profile, GFIT map profile, AirCore extended profile and Tower mast measurements are plotted for XCO_2 (left-column), XCH_4 (middle-column) and XCO (right-column) as a function of altitude for 24 April 2017 with launch time at 15:13:39 and landing time at 16:13:10 UTC (top-row), 15 May 2017 with launch time at 09:33:22 UTC and landing time at 10:25:32 UTC (2nd-row) and 28 August 2017 with launch time at 09:13:15 UTC and landing time at 10:10:33 UTC (3rd-row), respectively.~~

~~Bottom-row - left to right: Column averaging kernels of CO_2, CH_4 and CO retrievals for TCCON, plotted as a function of pressure and altitude. The different colours correspond to the varying solar zenith angle.~~

~~Top-left - upper panel: XCO_2 plotted for TCCON and EM27/SUN retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) for measurements performed on 15 May 2017 at Sodankylä. Top-left - lower panel: shows the difference between the two retrievals in absolute units. Middle-left and bottom-left figures show the same plots as mentioned above for Vertex70 and IRcube, respectively. Top-right, middle-right and bottom-right: show the above mentioned XCO_2 plots for measurements performed on 28 August 2017 at Sodankylä.~~

~~Top-left - upper panel: XCH_4 plotted for TCCON and EM27/SUN retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) for measurements performed on 15 May 2017 at Sodankylä. Top-left - lower panel: shows the difference between the two retrievals in absolute units. Middle-left and bottom-left figures show the same plots as mentioned above for Vertex70 and IRcube, respectively. Top-right, middle-right and bottom-right: show the above mentioned XCH_4 plots for measurements performed on 28 August 2017 at Sodankylä.~~

~~Top-left - upper panel: XCO plotted for TCCON and EM27/SUN retrievals with the TCCON a priori and with a modified a priori (calculated using in-situ, AirCore and TCCON map files; labelled with AC in the end) for measurements performed on 15 May 2017 at Sodankylä. Top-left - lower panel: show the difference between the two retrievals in absolute units. Bottom-left figures show the same plots as mentioned above for Vertex70. Top-right and bottom-right: show the above mentioned XCO plots for measurements performed on 28 August 2017 at Sodankylä.~~

~~Timeseries of XCH_4 retrievals using non-linearity corrected TCCON and AirCore measurements (top panel) and bias plot in absolute unit (middle panel) plotted for measurements performed in 2017 at SZA < 75°. The correlation of XCH_4 retrievals between non-linearity corrected TCCON and AirCore measurements are plotted in the lower panel. The colours represent the measurement performed during the different months of the year.~~

~~Plot showing the difference of the corrected interferograms - original interferograms vs the original interferograms. The individual corrections are plotted in red, the mean value is plotted as blue open circles and the black line is the fit.~~

~~Original (top left) and non-linearity corrected (bottom left) spectra; zoom of the out-of-band spectral region (100–3600 cm^{-1}) with the original spectra (top right) and non-linearity corrected (bottom right) spectra from the Bruker IFS 125HR at Sodankylä TCCON facility. The colour of the spectrum depends on the interferogram maximum signal at the center burst. The highest values corresponding to the dark red colour are recorded during the noon time when the signal is the highest.~~

~~Top panel: plot showing the original (red) and non-linearity corrected (black) XCO2 values for one day of measurement performed on 06 September 2017 by the Bruker 125HR Sodankylä TCCON instrument. Middle-panel: shows the difference between the original and the corrected XCO2 values. Lower-panel: shows the relative difference (original - corrected)/corrected in percentage for the XCO2 values.~~

Table 1. List of instruments participating in the FRM4GHG campaign in 2017 and their properties.

Item	Bruker IFS 125HR	Bruker Vertex70	Bruker IRcube	Bruker EM27/SUN	LHR
Beamsplitter	CaF_2	CaF_2	Quartz	Quartz (single plate)	ZnSe
Entrance window	CaF_2	CaF_2	CaF_2	Schott RG (IR transmitting filter glasses)	open
Aperture (mm)	1	0.25	0.5	0.6	0.3
Focal length (mm)	418	100	69	127	75
Scanner velocity (kHz)	10	10	10	10	N/A
Detectors	RT-Si Diode DC RT-InGaAs DC	RT-InGaAs DC LN2 cooled - InSb DC	RT-InGaAs DC	RT-InGaAs DC extended RT-InGaAs DC	Thermoelectrically cooled MCT
Acquisition mode	Single–sided forward–backward	Single–sided forward–backward	Single–sided forward–backward	Double–sided forward–backward	Sequential ~~LO~~ local oscillator scanning
Dimension (cm^3)		80x50x30	29x31x23.5	35x40x27	40x40x20
Weight (kg)		62 (without tracker)	14 (without tracker)	25 (with tracker)	∼10 (without tracker)
Vacuum	yes	no	no	no	no

Table 2. List of instruments, their measurement properties and retrieval strategy for the FRM4GHG campaign in 2017.

Instrument	Institute	Spectral range (cm^{-1})	Resolution (cm^{-1})	Measurement time (approx., min)	Sample scans	Main species	Dataset	Retrieval code
Bruker IFS 125HR (TCCON)	FMI	1800–15000	0.004	2.6	4	XCO_2, XCH_4, XCO @ 0.02 cm^{-1}	TCCON TCCONNLC	GFIT 2014 Posterior non-linearity correction
Bruker Vertex70	Uni Bremen & BIRA-IASB	2500–15000	0.16	2.5	18	XCO_2, XCH_4, XCO @ 0.2 cm^{-1}	VERTEX70	GFIT 2014
Bruker IRcube	Uni Wollongong	4500–15000	0.5	1.7	33	XCO_2, XCH_4	IRcube	GFIT 2014
Bruker EM27/SUN (COCCON)	KIT	4000–9000	0.5	1	10	XCO_2, XCH_4, XCO	EM27/SUN	PROFFAST
Bruker IFS 125HR (HR125LR)	FMI & KIT	1800–15000	0.004	1	10	XCO_2, XCH_4, XCO @ 0.5 cm^{-1}	HR125LR	PROFFAST
LHR	RAL	952–955	0.002 and 0.02	0.5	1	CO_2, H_2O @ 0.02 cm^{-1}	LHR	own code optimal estimation
AirCore	Uni Groningen & FMI	In-situ sampling	13.4 mbar (Amb.P. > 232 mbar) 3.9 mbar (Amb.P. < 232 mbar)			CO_2, CH_4, CO vertical profiles calibrated to WMO standards	AirCore	

Table 3. Instrumental line shape characteristics and modifications of the participating instruments.

Instrument	max OPD (cm)	Modulation efficiency	Phase error (mrad)	Modification periods in 2017	Modification comments
TCCON	45	<1.02	±2	begin–end	no modifications
EM27/SUN	1.8	1.02	-3 to +1	begin–end	no modifications
Vertex70					
– before blind phase	4.5	-0.935	-16 to -36	begin–6 July	parallel beam diameter 40 mm
– after blind phase		-0.973	-13	6 July–12 September	reduced aperture with parallel beam diameter 20 mm
IRcube	1.8	-0.95	-5 to +1.5	begin–23 March	old fibre cable
				~~23 March–end~~ 25 April–end	new fibre cable

Table 4. ~~Statistics of intercomparison results of LHR, Vertex70, IRcube and EM27/SUN vs TCCON for measurements~~ AirCore flight performed during the FRM4GHG campaign in 2017 ~~with SZA < 75°~~at the Sodankylä TCCON site.

Flights	Date	Start time of flight / UTC	End time of flight / UTC
1	21/04/2017	07:39:24	08:23:10
2	24/04/2017	15:13:39	16:13:10
3	26/04/2017	09:16:15	10:00:05
4	15/05/2017	09:33:22	10:25:32
5	28/08/2017	09:13:15	10:10:33
6	04/09/2017	09:15:58	10:04:15
7	05/09/2017	09:23:35	10:06:12
8	06/09/2017	09:10:20	09:49:10
9	07/09/2017	08:52:19	09:40:41
10	09/10/2017	09:49:48	10:50:14

Table 5. Statistics of intercomparison results of AirCore vs TCCON and non-linearity corrected TCCON data sets for measurements performed in 2017 with SZA < 75° for the TCCON measurements. The values provided are the mean bias ± the standard deviation and the correlation coefficient (r).

Species	XCO_2 / ppm	XCH_4 / ppm	XCO / ppb
AirCorevsTCCON	0.47±0.66 (0.994)	-0.004±0.011 (0.959)	6.40±1.88 (0.950)
AirCorevsTCCONmod	-0.03±0.71 (0.995)	-0.007±0.011 (0.969)	6.25±1.88 (0.951)

Table 6. Statistics of intercomparison results of LHR, Vertex70, IRcube and EM27/SUN vs non-linearity corrected TCCON for measurements performed in 2017 with SZA < 75°. The values provided are the mean bias ± the standard deviation and the correlation coefficient (r). The statistics of the first four rows cover the full year of measurements in 2017. While that of the last four rows are for the measurements performed between 06 July and 12 September 2017.

Species	XCO_2 / ppm	XCH_4 / ppm	XCO / ppb	Xair
LHR	-18.89±5.34 (0.499)	-	-	-
VERTEX70	1.46±1.63 (0.984)	0.023±0.013 (0.513)	3.57±2.57 (0.947)	-0.009±0.009 (0.077)
IRCUBE	-5.02±1.04 (0.971)	-0.008±0.004 (0.932)	-	-0.015±0.002 (0.556)
EM27/SUN	-0.73±0.47 (0.996)	0.000±0.004 (0.973)	4.38±1.36 (0.993)	0.02±0.002 (0.221)
LHR	-19.02±4.44 (0.462)	-	-	-
VERTEX70	-0.16±0.57 (0.933)	0.010±0.002 (0.958)	1.34±1.04 (0.991)	0.000±0.002 (0.454)
IRCUBE	-5.03±0.81 (0.901)	-0.010±0.004 (0.928)	-	-0.017±0.002 (0.739)
EM27/SUN	-0.38±0.39 (0.973)	-0.002±0.002 (0.974)	3.98±1.18 (0.988)	0.021±0.002 (-0.164)

Table 7. Statistics of intercomparison results of LHR, Vertex70, IRcube and EM27/SUN vs TCCON for measurements performed in 2017 with SZA < 75°. The values provided are the mean bias ± the standard deviation and the correlation coefficient (r). The statistics of the first four rows cover the full year of measurements in 2017. While that of the last four rows are for the measurements performed between 06 July and 12 September 2017.

Species	XCO_2 / ppm	XCH_4 / ppm	XCO / ppb	Xair
LHR	-18.37±5.32 (0.502)	-	-	-
VERTEX70	1.93±1.72 (0.983)	0.025±0.013 (0.502)	3.73±2.58 (0.946)	0.011±0.009 (-0.134)
IRCUBE	-4.54±1.06 (0.971)	-0.006±0.005 (0.924)	-	-0.013±0.003 (0.296)
EM27/SUN	-0.18±0.45 (0.995)	0.003±0.005 (0.962)	4.54±1.37 (0.993)	0.023±0.002 (0.336)
LHR	-18.62±4.43 (0.463)	-	-	-
VERTEX70	0.22±0.58 (0.931)	0.011±0.003 (0.955)	1.47±1.04 (0.991)	0.002±0.002 (0.361)
IRCUBE	-4.65±0.78 (0.909)	-0.008±0.004 (0.924)	-	-0.015±0.002 (0.629)
EM27/SUN	0.02±0.38 (0.975)	-0.000±0.002 (0.970)	4.12±1.18 (0.988)	0.022±0.002 (-0.158)

Table 8. Statistics of the intercomparison results of HR125LR vs TCCON for measurements performed in 2017 with SZA < 75°. The values provided are the mean bias ± the standard deviation and the correlation coefficient (r). The statistics of the first row cover the full year of measurements in 2017. While that of the second row cover the measurements performed between 06 July and 12 September 2017.

Species	XCO_2 / ppm	XCH_4 / ppm	XCO / ppb	Xair
HR125LR	-0.69±0.53 (0.993)	-0.005±0.004 (0.975)	0.03±1.02 (0.996)	0.032±0.003 (-0.596)
HR125LR	-0.40±0.50 (0.951)	-0.007±0.002 (0.970)	-0.50±0.99 (0.993)	0.030±0.002 (-0.189)

Table 9. Statistics of intercomparison results of LHR, Vertex70, IRcube and EM27/SUN vs HR125LR for measurements performed in 2017 with SZA < 75°. The values provided are the mean bias ± the standard deviation and the correlation coefficient (r). The statistics of the first four rows cover the full year of measurements in 2017. While that of the last four rows are for the measurements performed between 06 July and 12 September 2017.

Species	XCO_2 / ppm	XCH_4 / ppm	XCO / ppb
LHR	-17.52±5.27 (0.526)	-	-
VERTEX70	2.54±1.99 (0.976)	0.030±0.012 (0.472)	3.81±2.41 (0.956)
IRCUBE	-3.85±1.03 (0.978)	0.000±0.004 (0.949)	-
EM27/SUN	0.56±0.40 (0.996)	0.009±0.003 (0.978)	4.65±1.27 (0.995)
LHR	-18.00±4.61 (0.472)	-	-
VERTEX70	0.65±0.66 (0.911)	0.018±0.003 (0.952)	2.05±1.17 (0.988)
IRCUBE	-4.21±0.67 (0.934)	-0.001±0.003 (0.942)	-
EM27/SUN	0.49±0.32 (0.981)	0.007±0.001 (0.991)	4.76±1.09 (0.991)

~~Statistics of intercomparison results of LHR, Vertex70, IRcube and EM27/SUN vs non-linearity corrected TCCON for measurements performed in 2017 with SZA < 75°. Species XCO_2 / ppm XCH_4 / ppm XCO / ppbLHR -18.89±5.34 (0.499) - -VERTEX70 1.46±1.63 (0.984) 0.023±0.013 (0.513) 3.57±2.57 (0.947)IRCUBE -5.02±1.04 (0.971) -0.008±0.004 (0.932) -EM27/SUN -0.73±0.47 (0.996) 0.000±0.004 (0.973) 4.38±1.36 (0.993)LHR -19.02±4.44 (0.462) - -VERTEX70 -0.16±0.57 (0.933) 0.010±0.002 (0.958) 1.34±1.04 (0.991)IRCUBE -5.03±0.81 (0.901) -0.010±0.004 (0.928) -EM27/SUN -0.38±0.39 (0.973) -0.002±0.002 (0.974) 3.98±1.18 (0.988)~~

~~AirCore flight performed during the FRM4GHG campaign in 2017 at the SodankyläTCCON site. Flights Date Start time of flight / UTC End time of flight / UTC 1 21/04/2017 07:39:24 08:23:10 2 24/04/2017 15:13:39 16:13:10 3 26/04/2017 09:16:15 10:00:05 4 15/05/2017 09:33:22 10:25:32 5 28/08/2017 09:13:15 10:10:33 6 04/09/2017 09:15:58 10:04:15 7 05/09/2017 09:23:35 10:06:12 8 06/09/2017 09:10:20 09:49:10 9 07/09/2017 08:52:19 09:40:41 10 09/10/2017 09:49:48 10:50:14~

~~Statistics of intercomparison results of AirCore vs TCCON and non-linearity corrected TCCON data sets for measurements performed in 2017 with SZA < 75° for the TCCON measurements. Species XCO_2 / ppm XCH_4 / ppm XCO / ppbTCCONvsAirCore 0.47±0.66 (0.994) -0.004±0.011 (0.959) 6.40±1.88 (0.950)TCCONmodvsAirCore -0.03±0.71 (0.995) -0.007±0.011 (0.969) 6.25±1.88 (0.951)~~