Authors' answer to the interactive comments of anonymous referee #1 on "Building the COllaborative Carbon Column Observing Network (COCCON): Long term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer" by Frey et al., Atmos. Meas. Tech. Discuss., amt-2018-146

First of all, we would like to thank the anonymous referee #1 for the help in further improving the current presentment by a thorough assessment with regards of content and the careful technical proofreading resulting in the identification of several imprecisions and typos.

***Referee**: "The manuscript "Building the COllaborative Carbon Column Observing Network (COCCON): Long term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer" by Frey et al. assesses the stability of Bruker EM27/SUN spectrometers and evaluates their use for greenhouse gas (GHG) observations. EM27/SUN spectrometers are the spectrometers used in the recently founded Collaborative Carbon Column Observing Network (COCCON). The EM27/SUN spectrometers are portable and easy to operate. They have been used to quantify GHG emission sources (cities, exploration sites, ..) by column budgeting and are also used at permanent sites to complement the well-established Total Carbon Column Observing Network (TCCON). Therefore, COCCON will increase the number of sites, which perform groundbased column observations of GHGs. This is of high importance for the validation of GHG satellite retrievals. Since a high stability and small bias are vital for a network of GHG observations, this paper is scientifically of very high interest. The manuscript is well written and in my opinion it should be published after minor revisions. Below are several point that should be addressed prior acceptance.*

Abstract: The abstract should be more quantitative and clearer. All the information I am requesting is contained somewhere in the manuscript but in my opinion it should be mentioned in the abstract, e.g. "the EM27/SUN is stable on timescales of several years" How stable, how many years? "average bias across the ensemble" Is the average a good measure? What about min/max? "the application of these empirical factors is expected to further improve" The abstract should contain a number for the bias after all corrections and one before. Future papers will cite this number and quite often this is only taken from the abstract and not from the text."

Authors: As suggested by the referee, we underpin the qualitative statements of the previous version ("stable", "very uniform") with the quantitative results given in sections 3.4, 3.5, and 4.2 of the paper. In addition to the robust metric based on absolute differences, we calculate the standard deviation among the empirical calibration factors of all spectrometers (this metric assigns higher weight to far min / max outliers) and report the resulting 2 sigma uncertainty in the abstract.

Concerning our (rephrased) claim that "the application of these empirical factors is expected to further improve … the network conformity … beyond the scatter among the empirical calibration factors" we cannot provide an explicit proof, as this would require to repeat the calibration for each spectrometer several times. However, in our opinion it is evident that the application of these factors will further improve the conformity, given that we also have demonstrated the long-term stability of instrumental characteristics for one selected spectrometer. Please note, that the calibration factor for several spectrometers is derived from different calibration runs (sometimes even a year apart), without indication of variable outcome.

***Referee**: "p.2, line 11: "Furthermore these measurements can be directly used to evaluate emissions reductions as demanded by international treaties:" I do not know a study, where such measurements have been used to evaluate emission reductions. Please give a reference or mention that this is a future plan."*

Authors: Correct, we changed the wording accordingly: "Furthermore, these measurements offer the prospect of being usable for the evaluation of emission reductions as demanded …"

***Referee**: "p.6, 3.2 It is mentioned that the EM27/SUN spectrometers are operated at a significantly different temperatures. It would be interesting to have a separate assessment of the temperature on the EM27 retrievals."*

Authors: Our major concern in the context of this paper is the effect of ambient temperature on the spectrometer itself. The demonstrated long term comparison with collocated low-resolution 125HR measurements (this spectrometer is operated in an air-conditioned enforced laboratory container) would reveal any drifts due to ambient temperature. Both setups share the same retrieval setup, so any possible effects of the tropospheric temperature on trace gas results cancel out – as desired, because our study aims at evaluating the hardware performance of COCCON.

***Referee**: "p.8, line 32: Why do HR125 LR and EM27/SUN have a different bias compared to TCCON?"*

Authors: The instrument-specific bias of an EM27/SUN spectrometer is the result of instrumental imperfections. We assume that a well-maintained HR125 spectrometer operated at the same spectral resolution as the EM27/SUN is the best realizable approximation of an ideal EM27/SUN spectrometer. The relative difference between the EM27/SUN and the HR125 LR is not at all conspicuous for XCH4. It is higher for XCO2 (0.0014), but still does not exceed the 2 sigma scatter of the calibration factors among the EM27/SUN spectrometers. Please note that all EM27/SUN spectrometers might share a common design feature invoking a common bias with respect to the ideal HR125 LR reference, so we would expect a slightly higher bias when comparing a typical EM27/SUN to the HR125 LR.

***Referee**: "p.9, line 4: "The offset between EM27/SUN and TCCON shows a seasonal variability. Reasons for this are mainly the differences in airmass correction, averaging kernels and retrieval algorithm." Maybe this is discussed in the papers mentioned. However, it is highly important for the network. Therefore the reasons should be (re-) discussed here."*

Authors: We added an additional citation [Kiel et al., 2016] on this topic discussing in depth a comparable finding of differing seasonal variations due to differing sensitivities of the sensors involved. Please note, that the task of characterization of the emerging network is twofold. The first item are instrumental issues (long-term stability of instrumental characteristics, instrument-to-instrument variations of instrumental characteristics, stability of the spectrometer when operated under different ambient conditions, especially temperature), the second item is related to the data analysis (preprocessing algorithms, approximations in the radiative transfer code, retrieval strategy, spectroscopic issues, empirical postprocessing steps, etc.). The second – as you mention correctly also important- item is under study in the framework of the ESA project FRM4GHG (http://frm4ghg.aeronomie.be/), for COCCON and other kinds of remote sensing devices and a paper on this topic is in preparation. In the paper under consideration here, we are focusing on the instrumental aspects.

Referee*: "p.11/12, 4.2: The instrument dependent calibration factors: It would be good to elaborate the discussion about them and have the numbers summarised in a table, e.g. overall biases a) with instrument dependent cal factor (including uncertainties on the cal factor) and b) w/o instrument dependent cal factor."*

Authors: Table 6 summarizes the biases of the empirical instrument-specific calibration factors and their estimated uncertainties. In our feeling, the provided estimates of the uncertainties as the average absolute deviation are quite conservative (application of the usual Gaussian statistics would assign an additional factor of 1/(SQRT(number of measurement pairs)) when determining the confidence level of the calibration factor). We have expanded the discussion in 4.2 by adding the empirical scatter of the calibration factors from Gaussian statistics (adding more weight to far outliers) and histograms for demonstrating that the statistical distribution of the sample is well-behaved.

Referee*: "p.13 line 9: define how you quantify "scatter""*

Authors: We have extended the discussion and now offer two different metrics for the scatter of the empirical instrument-specific calibration factors, based on either the average absolute deviations from the reference spectrometer or on the scatter derived from Gaussian statistics.

Referee*: "Figure 1: How much information is coming from the apriori in the ILS retrieval? If the constraints are high one can always get a stable ILS. Is it possible to add averaging kernels or similar for the ILS retrieval? At least some explanation should be included in section 3.1."*

Authors: The problem of an ILS shape not fully determined by the measured spectral scene mentioned by the referee typically occurs in the case of high-resolution spectrometers. The spectral scene used here for the EM27/SUN ILS retrieval is comprised of a large number of water vapor lines generated at ambient pressure. It contains plenty of spectral detail beyond the resolution capability of the EM27/SUN. Therefore, even the extended 40-parameter ILS model offered by LINEFIT could be used without adding significant a-priori information to the ILS solution. However, as the deviations from the nominal ILS of the spectrometers were found to be very small and as we would like to come up with a concise ILS characterization, we applied the simple 2 parameter ILS model.

We added some further information on this in the text: "Due to the fact that the EM27/SUN is equipped with a circular fieldstop aperture, the ILS is nearly nominal.

Therefore, for keeping the treatment concise, we use the simple 2 parameter ILS model offered by LINEFIT."

***Referee**: "Figure 2: I am not sure if this figure is really needed."*

Authors: We agree that it might not be of high importance, but in our feeling a figure providing an overview of the complete raw long term dataset is useful, as all the Xgas values discussed in the following are derived from these column data. For this reason, we would like to keep figure 2.

***Referee**: "Figure 4 and 6: The difference in the left figures is difficult to see and contains the same information than the ratio in the figures on the right. Therefore I would leave the difference out in the figures on the left."*

Authors: Ok, we have updated figures 4 and 6 accordingly.

***Referee**: "Figure 10: I would delete this."*

Authors: We agree that this figure is of marginal importance. However, as referee #2 points out, the figure provides information on the instrument size and illustrates the practical configuration of the side-by-side arrangement. Therefore, we would like to keep it, but we will check in the proof version of the article that its size is appropriate.

Reference:

Kiel, M., Hase, F., Blumenstock, T., and Kirner, O. (2016): Comparison of XCO abundances from the Total Carbon Column Observing Network and the Network for the Detection of Atmospheric Composition Change measured in Karlsruhe, Atmospheric Measurement Techniques, 9, 2223–2239, https://doi.org/10.5194/amt-9-2223-2016

Authors’ answer to the interactive comments of anonymous referee #2 on “Building the COllaborative Carbon Column Observing Network (COCCON): Long term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer” by Frey et al., Atmos. Meas. Tech. Discuss., amt-2018-146

First of all, we would like to thank the anonymous referee #2 for the help in further improving the current presentment by a thorough assessment with regards of content and the careful technical proofreading resulting in the identification of several imprecisions and typos.

***Referee**: “Overall comments:*

This paper analyzes multi-year analysis of EM27/SUN results compared to TCCON. The long-term performance and stability of the EM27/SUN systems is important to use EM27/SUN results for science analysis and satellite validation. The EM27/SUN systems have potential as lower cost stationary instruments, and for use in shorter term field campaigns since EM27/SUN are easier to move.

I agree with reviewer 1 that column averaging kernels should be shown and compared to TCCON and LR TCCON.”

Authors: The averaging kernels of the EM27/SUN have already been presented and discussed in comparison to TCCON in the literature [Hedelius et al., 2016]. We have added this information explicitly in the revised version. Please note that the scope of this paper is the characterization of the instrumental performance of the spectrometers used for COCCON.

***Referee**: “The assessment of EM27/SUN results relies on comparisons to a specially processed, modified TCCON dataset, called LR TCCON. LR TCCON is reduced resolution TCCON, with a differently derived ILS, and processed with the PROFFIT software. However, LR TCCON has not itself been validated.”*

Authors: This is true, we are fully aware of this limitation. However, the TCCON LR and EM27/SUN data products have been generated by applying exactly the same processing scheme. The idea behind this approach is to use the TCCON LR dataset to quantify instrument-to-instrument biases and possible instrumental drifts. This approach offers a much higher sensitivity in this regard than a direct comparison with official TCCON products derived from high-res interferograms, because the sensitivities of the reference and device under test are perfectly matched. The fact, that the TCCON LR data product is unvalidated is not harmful in our context, as we are aiming at only a highly sensitive relative comparison (between comparable sensors, comparable in the sense that we expect identical trace gas results if the same atmospheric state is observed).

***Referee**: “Significant differences are seen between EM27/SUN and the full resolution TCCON (shown in Figures 4 and 6) for XCO2 and XCH4. These errors should be quantified in the paper. The errors are seasonally dependent and look to have peak-*

to-peak seasonal errors of about 1 ppm for XCO2 and 20 ppb for XCH4, larger than the TCCON errors compared to aircraft validation (0.4 ppm for XCO2 and 5 ppb for XCO2 for GGG2014 (Wunch, 2015)). Comparisons of EM27/SUN results to LR TCCON are very good. However, LR TCCON has NOT been validated and comparisons of EM27/SUN versus LR TCCON is NOT validation of the EM27/SUN results and does NOT tie EM27/SUN to WMO."

Authors: An exhaustive comparison with TCCON will be given in a paper under preparation by the FRM4GHG consortium. We agree that the tying to WMO suffers from a significantly larger uncertainty than the instrument-to-instrument calibration within COCCON. The instrument-to-instrument calibration should be based on the comparison with TCCON LR, only the tying to WMO needs to be done via the official TCCON data products.
Unfortunately, due to higher spectral resolution the TCCON observations have different sensitivity characteristics than COCCON. If the a-priori profile shape assumed by TCCON differs from the truth (its quality might depend on season, and on the current meteorological situation, as demonstrated in the paper for the situation of polar air intrusion), it will give rise to e.g., seasonal differences between TCCON and COCCON of the observed size. Proof of this is given in section 3.5, where a period of polar air intrusion is discussed and in an upcoming FRM4GHG paper.

Following the suggestion of the referee, we have added a short discussion concerning the level of uncertainty with respect to WMO tying of COCCON, which is significantly higher than the internal consistency. This discussion is based on the results provided in tables 3 and 4, which clearly indicate the higher scatter in the EM27/SUN versus TCCON residuals, suggesting a current calibration uncertainty of 0.15% for XCO2 and 0.24% for XCH4 with respect to TCCON.

***Referee**: "In summary, if LR TCCON can be validated versus aircraft/AirCore with similar errors as the standard TCCON, then this paper will set useful limits on EM27/SUN errors. As the paper stands, validation that must be considered is versus the standard TCCON product, which is marginal for satellite validation and on the high side for other uses."*

Authors: We do not agree to this statement. The choice of TCCON LR product is fully appropriate for demonstrating the level of internal consistency achievable by COCCON. The paper under consideration does not claim to solve the problem of tying the COCCON data to WMO, as the title says, it aims at demonstrating the long-term stability of the EM27/SUN spectrometer, and it investigates the ensemble performance. The construction of a TCCON LR data set is in our opinion the best possible approach for achieving this.
Note that the vertical sensitivity offered by COCCON differs slightly from TCCON, but is not systematically poorer than TCCON (see Hedelius et al., 2016). Therefore, when a similar dataset of in-situ measurements will be exploited for COCCON, the tying to WMO is expected to be of similar quality as for TCCON. Work in this direction, also including AirCore observations, is under progress in the FRM4GHG project.

***Referee**: "Specific comments*

Introduction:

The COCCON project should be introduced in the introduction, with the objectives of the COCCON, and who is participating in COCCON, the length of the project (for example).”

Authors: We added the following paragraph in the introduction:

“COCCON is intended to be a lasting framework for creating and maintaining a greenhouse gas observing network based on common instrumental standards and data analysis procedures. Currently, about 18 working groups operating EM27/SUN spectrometers are contributing. We expect that COCCON will become an important supplement of TCCON, as the logistic requirements are low and the spectrometers are simple to operate. It will increase the global density of column-averaged greenhouse gas observations and due to the fact that the spectrometers are portable will especially contribute to the quantification of local sources.”

Referee*: “In the introduction, add in the importance of TCCON for OCO-2 and GOSAT validation, adding a sentence after line 23 something like: "TCCON stations are also the primary validation for OCO-2 (cite https://oco.jpl.nasa.gov/files/ocov2/OCOC22_SciValPlan_111005_ver1_0_revA_final_signed1.pdf) and validating the satellite observations at different locations is critical for the validation effort (Wunch et al., 2017)."”*

Authors: Ok, done

Referee*: “A figure showing the TCCON (original and degraded resolution) and EM27 spectral range and radiance would be helpful for the reader, or a reference to a previous paper showing this.”*

Authors: “We added a reference [Hedelius et al., 2016] in section 2.1. The figure contains a TCCON (original resolution) and EM27/SUN spectrum. We refrain from adding a figure in this paper because we think that the additional information from TCCON (degraded resolution) is marginal.”

Referee*: “The spectral ranges and approximate resolution should be given in wavelength in addition to wavenumber. Some scientists are used to wavelength and the translation is not immediately obvious.”*

Authors: “We included this information in section 2.1. For the sake of readability, in the other sections, only the wavenumber notation is given.”

Referee*: “Section 2.2 The description of the HR125 low resolution data set should include the software used to analyze it. I infer it is PROFFIT, but should be stated.”*

Authors: Ok, done

Referee*: “Page 5, line 15. Define ILS, modulation efficiency, phase error.”*

Authors: The paper includes a reference to a paper where the procedure of instrumental line shape measurements is explained in detail [Frey et al., 2015].

Additionally we added a reference for a more general description of the used ILS model [Hase et al., 2012].

Referee*: "Page 5, line 22. How is the phase error calculated – describe or cite a reference. Why is phase error important? What does it affect?"*

Authors: We included an additional sentence in the manuscript with a reference to the original LINEFIT paper [Hase et al., 1999]. Figure 1 and 2 of this reference illustrates the effect of differing modulation efficiency amplitudes and phase errors on a spectral line.

Referee*: "The statement on line 7, page 7, "The remaining difference can be attributed to the different measurement heights of the HR125 (112 m) and EM27/SUN (133 m)." This needs to be further explained and quantified. Is it the total column? It would be useful to the reader to have a calculation accounting for the offset."*

Authors: In this section total columns are discussed. So here it is expected that the total columns differ for instruments at slightly different heights. For an estimate of the ratio the barometric height formula can be utilized. As for this study the main interest lies in the analysis of XCO2 and XCH4, where the height dependency is expected to largely cancel out, we chose not to dwell on the small differences observed in the total columns at different heights.

Referee*: "Table 2, it would be useful to show the effect on XCO2, etc, which is the key result. The reader looks between columns and thinks it will probably cancel for XCO2 but is not sure."*

Authors: We agree that this information is vital. We now include the information not only in the text, but also in the caption of the table. Including the information in the table would enlarge the table too far, and we feel it is important to keep the basic information of the total columns.

Referee*: "Page 8, line 11, "From this higher variability it can be concluded that the airmass dependency in the official TCCON O2 retrieval is higher than for the PROFFIT retrieval, a finding also observed by Gisi et al. (2012)." This statement needs to be modified for clarity to "...higher than for the PROFFIT retrieval on reduced resolution TCCON measurements.""*

Authors: We changed the wording accordingly.

Referee*: "Page 8, line 25. "There are no obvious steps between the EM27/SUN and the HR125 LR data sets so that it can be concluded that the EM27/SUN is stable." The offset versus time needs to be quantified as well. Step functions and slower drift are both important to quantify."*

Authors: We changed the wording accordingly: "There are no obvious steps and there is no significant drift between…"

Referee*: "Page 9 line 7. The green line on Fig. 4 shows significant differences between TCCON and EM27, on the order of 1 ppm it looks like. This seasonal cycle* *amplitude difference should be quantified. The pink difference (comparison to LR TCCON to EM27) looks very good. As stated in the overall comments, if the difference of EM27/SUN vs. TCCON is larger than the reported TCCON error, then it is important to determine the cause of this difference. PROFFIT should be applied to the full resolution TCCON data, OR GFIT should be applied to the low resolution TCCON data to separate out the PROFFIT/GFIT differences vs. ILS/truncation differences to determine the source of the difference between full-resolution TCCON and LR TCCON. LR TCCON needs to be validated versus aircraft/AirCore before it can be used to validate EM27/SUN."*

Authors: As we discussed before in this reply and have illustrated in the paper exemplary on the intrusion event seen in March 2016, the differences are mostly due to different sensitivities. In our context of demonstrating the level of long-term stability and ensemble consistency, it is just important to use a common choice for the EM27/SUN and TCCON LR analyses. A comparison of GFIT with PROFFIT for both high- and low-resolution spectra is well beyond the scope of this paper.

***Referee**: "Similar comment for XCH4. In Fig. 6, differences for XCH4 between EM27 and full resolution TCCON look to have seasonal differences of about 20 ppb, which is higher than the TCCON estimated XCH4 error of 5 ppb."*

Authors: As explained before, this discrepancy simply reveals the smoothing errors inherent in both time series (TCCON and COCCON). The occurrence of larger differences during an episode of a polar air intrusion mentioned in the paper is clearly demonstrating the mechanism.

Only in simple situations – e.g. if one can assume that a certain excess signal is due to a nearby source generating enhanced values in the boundary layer, one can approximately correct for the differing sensitivity characteristics [Wunch et. al, 2011, Hedelius et al. , 2017], but in general, when differences in the seasonal cycle are observed, it is not possible to remove the smoothing error without knowledge of the real mixing ratio profile in the atmosphere.

***Referee**: "Wording/formatting suggestions:*

Line 11, suggestion: change "as demanded by" to "as specified by""

Authors: Ok, done

***Referee**: "Line 16, word suggestion: "Nonetheless" change to "However""*

Authors: Ok, done

***Referee**: "Line 20: "However, recently OCO-2 data was used for estimating the source strength of power plants (Nassar et al., 2017)", would reword to emphasize coverage issues, "Recently OCO-2 data was used for estimating the source strength of power plants (Nassar et al., 2017). However, this can only be done for power plants that lie directly under the OCO-2 overpass locations.""*

Authors: We rephrased the sentences as suggested.

***Referee**: “Make the dots bigger on the Fig 2-7 legends. It is very hard to tell which dot is blue and which is black in the legend.”*

Authors: We will change the size of the dots in the legend.

***Referee**: “Page 7 line 11, "Before, a sensitivity study is provided demonstrating the effect of changes in the ILS on the gas retrieval." I think change "Before" to "First".”*

Authors: Ok, done

***Referee**: “I see reviewer 1 suggests deleting Fig. 10. However I think Fig. 10 is useful to show the size of the instrument. Perhaps make this figure small.”*

Authors: As stated in the reply to reviewer 1, we will check that the size of the figure is appropriate in the final version of the paper.

References:

Hase, F., Blumenstock, T., and Paton-Walsh, C. (1999): 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

Wunch, D., Wennberg, P. O., Osterman, G., Fisher, B., Naylor, B., Roehl, C. M., O’Dell, C., Mandrake, L., Viatte, C., Kiel, M., Griffith, D. W. T., 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., 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., García, O. E., Te, Y., Jeseck, P., Gunson, M. R., Crisp, D., and Eldering, A. (2017): Comparisons of the Orbiting Carbon Observatory-2 (OCO-2) XCO2 measurements with TCCON, Atmospheric Measurement Techniques, 10, 2209–2238, https://doi.org/10.5194/amt-10-2209-2017

Authors' answer to the interactive comments of Jacob Hedelius on "Building the COllaborative Carbon Column Observing Network (COCCON): Long term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer" by Frey et al., Atmos. Meas. Tech. Discuss., amt-2018-146

First of all, we would like to thank Jacob Hedelius for the help in further improving the current presentment by a thorough assessment with regards of content and the careful technical proofreading resulting in the identification of several imprecisions and typos.

J. Hedelius*: "The authors may want to consider these comments in preparation for their final submission of this paper.*

Abstract: It may be helpful to emphasize that the instrument is solar-viewing."

Authors: Ok, done

J. Hedelius*: "P1L3 – The word "stable" is used throughout. The authors use the term to both refer to 1) mechanical stability of the instrument, and 2) comparability of the retrievals to another product. Because this is a subject term (e.g., one person may say 0.5% accuracy is stable, and another may say 0.05%) it would be useful if the author's metric of stability was defined numerically. In the future, requirements for "stability" may change as well."*

Authors: We agree and have added specific numbers (which also was requested by referee #1).

J. Hedelius*: "P1L5 – It may be useful to list the QA measures here, as the authors use several."*

Authors: The QA measures are explained in detail in the paper, we think that it would be too specific to provide these details in the abstract.

J. Hedelius*: "P2L2 – "Very uniform" is also subjective. It would be helpful to mention indicators of uniformness here in case readers only see the abstract."*

Authors: We have added further information (see reply to referee #1).

J. Hedelius*: "P2L13 – Numerically, what is the reference precision of the TCCON?"*

Authors: The performance of TCCON has been demonstrated and discussed in many papers, e.g. the cited paper by Wunch et al., 2011. We will provide the information in the manuscript for the readers.

J. Hedelius*: "P2L14 – Not only do 125HR instruments require more frequent maintenance than EM27/SUN instruments, it also needs to be done on site."*

Authors: True, we have added this information.

***J. Hedelius**: "P2L20 –Ye et al. (ACPD, 2017) https://doi.org/10.5194/acp-2017-1022 recently estimated city/urban emissions using satellite observations. Data from other current and future satellites may be used to estimate emissions from more localized sources, but that remains to be seen."*

Authors: Thanks, we added the reference.

***J. Hedelius**: "P2L25 – "Low-cost" is subjective, but I would actually say the EM27/SUN spectrometer is quite expensive, and cost-prohibitive for many institutions to own. The authors may consider stating the 2018 price range for these instruments."*

Authors: "Low-cost" here is meant in comparison with operating a TCCON spectrometer, which requires not only a more expensive spectrometer, but also the provision of a controlled environment, e.g., by operation in a laboratory or special container. We will add the current price range, around 100000 Euro, in the manuscript.

***J. Hedelius**: "P2L29 & throughout – The authors often use the world "calibrated" when "compared" or "scaled to" would be a better choice in this context. Calibrated is usually reserved for values more directly measured and compared to a standard."*

Authors: Thanks, we revised the text accordingly.

***J. Hedelius**: "P3L15 – Define IMECC"*

Authors: Ok, done

***J. Hedelius**: "P3L24 – Define/describe the NCEP data"*

Authors: Ok, done

***J. Hedelius**: "P4L4 – What does "nominal" mean here and throughout?"*

Authors: "Nominal" is matching the theoretical expectation.

***J. Hedelius**: "Sect. 3.1 – The ME at MOPD is consistently around 0.985, so what should users running PROFFIT use for the ILS? Should the input ME at MOPD be 1.0? What was used in this study (e.g., what does "real ILS" mean on P12L17)?"*

Authors: Our current choice is to accept the bias in the method and to use the ME as it results from the LINEFIT analysis. Renormalization to 1.0 would slightly change the calibration values for XCO2 and XCH4 (in a systematic way, the instrument-to-instrument relation will not be affected).

***J. Hedelius**: "P8L23 – Given that Dragos Ene is a coauthor of this study it seems strange to use "they" instead of "we." The authors may consider removing the private communications citation and instead put in an author contribution section at the end: see Manuscript composition -> 14. Author contribution under https://publications.copernicus.org/for_authors/manuscript_preparation.html"*

Authors: We added an author contribution section and revised the text accordingly.

***J. Hedelius**: "P9L7 – I agree with Reviewer #2 that the focus on comparing with a LR rather than an HR dataset from the 125 HR instruments is dissatisfying. I would expect the additional information in HR data should at least make it possible to construct a dataset with smaller absolute errors and biases. If 2 Xgas measurements have large, but equal errors or biases they will agree well."*

Authors: In our reply to referee #2 we tried to make clearer why the use of LR data is the most sensitive way to quantify the small instrument-to-instrument biases which we need to detect. The generation of an LR dataset from the high-resolution spectrometer allows to generate comparable observation systems. The smoothing error which occurs due to use of an imperfect a-priori trace gas profile during a side-by-side observation period is essentially a systematic error (as clarified by the annual variations seen between TCCON and COCCON), it can only be removed from the intercomaprison by matching sensitivities (this, in turn, by matching spectral resolution).

***J. Hedelius**: "P13L19 – I would disagree that no maintenance is ever required. In my experience at least 6 of 9 EM27/SUN instruments I have been on campaigns with required some form of maintenance within their first two years. Even the reference spectrometer in this study needed maintenance in 2016. However, an advantage is they do not need to be maintained on-site, but rather can be shipped back to Bruker or KIT."*

Authors: Correct, we have updated the text accordingly.

***J. Hedelius**: "P13L21 – From here and the TCCON meeting the COCCON PROCEEDS sounds like a very exciting upcoming development. I think this project deserves a more complete description earlier on in the paper. I also agree that a more concrete description of COCCON will be useful."*

Authors: We have added some more information on COCCON in the introduction.

***J. Hedelius**: "P13L29 – Perhaps the authors may want to check with the editor, but there may be some conflicts of interest that should be declared (https://www.atmospheric-measurement-techniques.net/about/competing_interests_policy.html). For example, receiving research funding from, or working for a commercial company could be considered a conflict of interest per the Copernicus policy."*

Authors: Thanks for pointing this out, we discussed the point with the Editor.

***J. Hedelius**: "Figure 4 – The authors may consider changing the y-axis scale. Scales of 15 ppm, and 5% (_20 ppm) are, in my opinion, quite large and make it difficult to judge comparability of the retrievals on shorter timescales. Especially as the satellite community is pushing towards accuracy of 1 ppm (_0.25%) or better for XCO2."*

Authors: We have revised figures 4 and 6, as was also suggested by referee #1.

J. Hedelius: *"Metrics of stability in the Xgas retrievals in addition to the linear fit over the full time series may be useful in the text. For example, on different timescales such as months or seasons – especially since differences on these timescales are quite noticeable. This will help if the COCCON is used in satellite validation to know if comparisons should only be done over multi-annual scales to get an overall bias as high and low values will cancel out, or if shorter time-scales are plausible. Seasonal or month-to-month biases would also lead to artificial cycles in global assimilation models."*

Authors: The linear fit is performed for quantifying instrumental drifts. The discussion of seasonal changes is work in progress in the FRM4GHG consortium. These variations are mainly driven by variations of differing smoothing error contributions between TCCON and EM27/SUN. It is not related to the questions of concern in the publication under consideration: investigation of long-term instrumental stability and ensemble performance.

J. Hedelius: *"Table 1 – It would be helpful to have a caption as to why some uncertainties always propagate to negative on ME."*

Authors: Thanks, we corrected this inconsistency.

J. Hedelius: *"Table 5 - Would all the authors advise that regular ILS monitoring is unnecessary and other EM27/SUN operators just use the values in this Table?"*

Authors: If regular atmospheric measurements are performed with a spectrometer, a drift or step change in XAIR will be a sensitive early indicator of any instrumental instability (assuming the availability of a reliable pressure record). If a change in the XAIR timeseries is detected, we would strongly recommend ILS measurements as a measure of diagnosis.

J. Hedelius: *"Table 6 - Would the authors recommend instrument operators not make their own side-by-side comparison at the beginnings and ends of instrument campaigns, and instead use these scaling factors?"*

Authors: We definitely would recommend as a measure of precaution to perform side-by-side comparisons before and after campaigns, if the campaign schedule allows. If e.g. one of the participating spectrometers received a mechanical shock during overseas transport due to mishandling, it could after recalibration still contribute (then with a slightly changed calibration factor) to the campaign dataset. Ideally, this spectrometer should be resend to the central calibration facility afterwards for recalibration (a change in instrumental characteristics might in addition indicate an instrumental damage).

Reference:

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., 5 Lin, J. C., Yang, E., and Wu, D. (2017): Constraining fossil fuel CO2 emissions from urban area using OCO-2 observations of total column CO2, Atmospheric Chemistry and Physics Discussions, 2017, 1–30, https://doi.org/10.5194/acp-2017-1022

Building the COllaborative Carbon Column Observing Network (COCCON): Long term stability and ensemble performance of the EM27/SUN Fourier transform spectrometer

Matthias Frey[1], Mahesh Kumar Sha[2,a], Frank Hase[1], Matthäus Kiel[3,a], Thomas Blumenstock[1], Roland Harig[4], Gregor Surawicz[4], Nicholas M. Deutscher[5], Kei Shiomi[6], Jonathan E. Franklin[7], Hartmut Bösch[8,9], Jia Chen[10], Michel Grutter[11], Hirofumi Ohyama[12], Youwen Sun[13], André Butz[14,b], Gizaw Mengistu Tsidu[15], Dragos Ene[16], Debra Wunch[17], Zhensong Cao[13], Omaira Garcia[18], Michel Ramonet[19], Felix Vogel[20], and Johannes Orphal[1]

[1]Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany
[2]Royal Belgian Institute for Space Aeronomy, Brussels, Belgium
[3]Division of Geological and Planetary Sciences, California Institute of Technology, Pasadena, CA, USA
[4]Bruker Optics GmbH, Ettlingen, Germany
[5]Centre for Atmospheric Chemistry, School of Chemistry, University of Wollongong, Wollongong, Australia
[6]Japan Aerospace Exploration Agency, Tsukuba, Japan
[7]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA
[8]Department of Physics and Astronomy, University of Leicester, UK
[9]National Centre for Earth Observation (NCEO), University of Leicester, UK
[10]Environmental Sensing and Modeling, Technische Universität München, Munich, Germany
[11]Universidad National Autonoma de Mexico, Mexico City, Mexico
[12]National institute for Environmental Studies, Tsukuba, Japan
[13]Anhui Institute of Optics and Fine Mechanics, Hefei, China
[14]Institut für Umweltphysik, Universität Heidelberg, Germany
[15]Botswana International University of Science and Technology, Gaborone, Botswana
[16]National Institute for Research and Development in Optoelectronics (INOE), Magurele, Romania
[17]University of Toronto, Toronto, Canada
[18]Izaña Atmospheric Research Centre (IARC), Meteorological State Agency of Spain (AEMET), Tenerife, Spain
[19]Laboratoire des sciences du climat et de l'environment, Gif-Sur-Yvette, France
[20]Environment and Climate Change Canada, Toronto, Canada
[a]before at: Karlsruhe Institute of Technology (KIT), Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe, Germany
[b]before at: Institut für Physik der Atmosphäre,Deutsches Zentrum für Luft- und Raumfahrt e. V., Oberpfaffenhofen, Germany

Correspondence: M. Frey (m.frey@kit.edu)

Abstract. In a 3.5 year long study, the long term performance of a mobile, solar absorption Bruker EM27/SUN spectrometer, used for greenhouse gases observations, is checked with respect to a co-located reference Bruker IFS 125HR spectrometer, which is part of the Total Carbon Column Observing Network (TCCON). We find that the EM27/SUN is stable on timescales of several years, the drift per year between the EM27/SUN and the official TCCON product is 0.02 ppmv for XCO_2 and 0.9 ppbv for XCH_4, which is within the 1σ precision of the comparison, 0.6 ppmv for XCO_2 and 4.3 ppbv for XCH_4. The bias between the two datasets is 3.9 ppmv for XCO_2 and 13.0 ppbv for XCH_4. In order to avoid sensitivity dependent artefacts, the EM27/SUN is also compared to a truncated IFS 125HR dataset derived from full resolution TCCON interferograms. The drift is 0.02 ppmv for XCO_2 and 0.2 ppbv for XCH_4 per year, with 1σ precisions of 0.4 ppmv for XCO_2 and 1.4 ppbv for XCH_4, respectively. The bias between the two datasets is 0.6 ppmv for XCO_2 and 0.5 ppbv for XCH_4. ~~qualifying it~~ With the presented long term stability, the EM27/SUN qualifies as an useful supplement for the existing TCCON network in remote areas. For achieving consistent performance, such an extension requires careful testing of any spectrometers involved by application of common quality assurance measures. One major aim of the COllaborative Carbon Column Observing Network (COCCON) infrastructure is to provide these services to all EM27/SUN operators. In the framework of COCCON development, the performance of an ensemble of 30 EM27/SUN spectrometers was tested and found to be very uniform, enhanced by the centralized inspection performed at the Karlsruhe Institute of Technology prior to deployment. Taking into account measured instrumental line shape parameters for each spectrometer, the resulting average bias across the ensemble with respect to the reference EM27/SUN used in the long term study in XCO_2 is 0.20 ppmv, while it is 0.8 ppbv for XCH_4. The average standard deviation of the ensemble is 0.13 ppmv for XCO_2 and 0.6 ppbv for XCH_4. In addition to the robust metric based on absolute differences, we calculate the standard deviation among the empirical calibration factors. The resulting 2σ uncertainty is 0.6 ppmv for XCO_2 and 2.2 ppbv for XCH_4. As indicated by the executed long-term study on one device presented here, the remaining empirical calibration factor deduced for each individual instrument can be assumed constant over time. Therefore the application of these empirical factors is expected to further improve the EM27/SUN network conformity beyond the ~~raw residual bias~~ scatter among the empirical calibration factors reported above.

1 Introduction

Precise measurements of atmospheric abundances of greenhouse gases (GHGs), especially carbon dioxide (CO_2) and methane (CH_4), are of utmost importance for the estimation of emission strengths and flux changes (Olsen and Randerson, 2004). ~~Furthermore these measurements can be directly used to evaluate emissions reductions~~ Furthermore, these measurements offer the prospect of being usable for the evaluation of emission reductions as ~~demanded~~ specified by international treaties, e.g. the Paris COP21 agreement (http://unfccc.int/resource/docs/2015/cop21/eng/l09r01.pdf/). The Total Carbon Column Observing Network (TCCON) (Wunch et al., 2011) measures total columns of CO_2 and CH_4 with reference ~~precision~~ quality. TCCON achieves a calibration accuracy with a 1σ error of 0.2 ppmv for XCO_2 and 2 ppbv for XCH_4 and a total uncertainty budget of below 1 ppmv for XCO_2 and below 5 ppbv for XCH_4, respectively.(Wunch et al., 2010, 2015)~~, however~~ However, the instruments used by this network are rather expensive and need large infrastructure to be set up and expert maintenance, which has to be performed on site. Therefore TCCON stations have sparse global coverage, especially in Africa, South America and large parts of Asia (Wunch et al., 2015). Current satellites like the Orbiting Carbon Observatory-2 (OCO-2) (Frankenberg et al., 2015) and the Greenhouse Gases Observing Satellite (GOSAT) (Morino et al., 2011) on the other hand offer global coverage. Nonetheless, they suffer from coarse temporal resolution (the repeat cycle of OCO-2 is 16 days), and in the case of GOSAT from sparse spatial sampling as well as limited precision of a single measurement. These limitations mostly inhibit a straight-forward estimation of the emission strength of localised sources of CO_2 and CH_4 like cities, landfills, swamps or fracking and mining areas from satellite observations. ~~However, r~~Recently OCO-2 data was used for estimating the source strength of power plants (Nassar et al., 2017) and urban emissions (Ye et al., 2017). However, this can only be done for powerplants and urban areas that lie directly under the OCO-2 overpass locations. TCCON stations are also the primary validation for OCO-2 (https://oco.jpl.nasa.gov/files/ocov2/OCOC22_SciValPlan_111005_ver1_0_revA_final_signed1.pdf) and validating the satellite observations at different locations is critical for the validation effort (Wunch et al., 2017).

The previously described Bruker EM27/SUN portable FTIR spectrometer (Gisi et al., 2011; Frey et al., 2015; Hedelius et al., 2016) is a promising instrument to overcome the above mentioned shortcomings as it is a mobile, reliable, easy to deploy and low-cost supplement to the Bruker IFS 125HR spectrometer used in the TCCON network. So far the EM27/SUN was mainly used in campaigns for the quantification of local sinks and sources (Hase et al., 2015; Chen et al., 2016). In this work the long term performance of the EM27/SUN with respect to a reference high resolution TCCON instrument is investigated. Additionally, the ensemble performance of several EM27/SUN spectrometers is tested. During 2014-2018, 30 EM27/SUN were ~~calibrated~~ tested at the Karlsruhe Institute of Technology (KIT) before being shipped to the customers. Several instruments that were distributed before this calibration routine at KIT was established, were upgraded with a second channel for CO observations at Bruker Optics^{TM} and after this also ~~calibrated~~ checked at KIT. This results in a unique data set as all EM27/SUN are directly ~~calibrated with respect~~ compared to a reference EM27/SUN, continuously operated at KIT, as well as a co-located TCCON instrument. From this data set an EM27/SUN network precision and accuracy can be estimated.

The COllaborative Carbon Column Observing Network (COCCON) is intended to be a lasting framework for creating and maintaining a greenhouse gas observing network based on common instrumental standards and data analysis procedures. Currently, about 18 working groups operating EM27/SUN spectrometers are contributing. We expect that COCCON will become an important supplement of TCCON, as the logistic requirements are low and the spectrometers are easy to operate. It will increase the global density of column-averaged greenhouse gas observations and due to the fact that the spectrometers are portable will especially contribute to the quantification of local sources.

2 Methodology

2.1 TCCON data set

As part of the TCCON, the Karlsruhe Institute of Technology (KIT) operates a high resolution ground based spectrometer at KIT, Campus North (CN) near Karlsruhe (49.100°N, 8.439°E, 112 m a.s.l.). Standard TCCON instruments have been described in great detail elsewhere (Washenfelder et al., 2006; Wunch et al., 2011). The Karlsruhe instrument, in the following called HR125, is the first demonstration of synchronized recordings of TCCON near infrared (NIR) and NDACC mid infrared (MIR) spectra using a dedicated dichroic beamsplitter (BS) arrangement (Optics Balzers Jena GmbH, Germany) with a cut-off wavenumber of 5250 cm^{-1}. It uses an InGaAs (indium gallium arsenide) detector in conjunction with an InSb (indium antimonide) detector, details can be found in Kiel et al. (2016b). By the TCCON measurements, the relevant wavenumber region 4000 - 11000 cm^{-1}, corresponding to wavelengths λ between 0.9 μm and 2.5 μm, is covered so that, among other species, O_2, CO_2, CH_4, CO and H_2O can be retrieved. A figure showing the spectral range of TCCON and the EM27/SUN can be found in Hedelius et al. (2016), figure 1. The TCCON measurements were chosen as reference measurements because these gases are also measured by the EM27/SUN spectrometer. For TCCON measurements in the NIR the HR125 records single sided interferograms with a resolution of 0.014 cm^{-1} ($\Delta\lambda = 3.5$ pm) or 0.0075 cm^{-1} ($\Delta\lambda = 1.9$ pm), corresponding to a maximum optical path difference (MOPD) of 64 cm and 120 cm. The recording time for a typical measurement consisting of two forward and two backward scans is 212 s, and 388 s, respectively. The applied scanner velocity is 20 kHz. The TCCON site Karlsruhe participated in the Infrastructure for Measurement of the European Carbon Cycle (IMECC) aircraft campaign (Messerschmidt et al., 2011; Geibel et al., 2012). The spectrometer has been used for calibrating all gas cells used by TCCON for instrumental line shape (ILS) monitoring (Hase et al., 2013).

TCCON data processing is performed using the GGG Suite software package (Wunch et al., 2011). In this study, the current release version, GGG 2014 is used (Wunch et al., 2015). The software package includes a pre-processor correcting for solar brightness fluctuations (Keppel-Aleks et al., 2007) and performing a fast Fourier transform including phase error correction routine to convert recorded interferograms into solar absorption spectra. Note that forward and backward scans are split by the preprocessing software and analyzed separately. The central part of the software package is the non-linear least-squares retrieval algorithm GFIT. It performs a scaling retrieval with respect to an a priori profile, then integrates the scaled profile over height to calculate the total column of the gas of interest. The software package additionally uses meteorological data ~~(NCEP)~~ from the National Center for Environmental Protection and National Center for Atmospheric Research (NCEP/NCAR) (Kalnay et al., 1996) and provides daily a priori gas profiles. TCCON converts the retrieved total column abundances VC_{gas} of the measured gases into column-averaged dry air mole fractions (DMFs), where the DMF of a gas is denoted $\mathrm{X}_{gas} = \frac{VC_{gas}}{VC_{O_2}} \times 0.2095$. In this representation several errors cancel out that affect both the target gas and O_2. However, residual bias with respect to in situ measurements still persists, as well as a residual spurious dependence of retrieval results on the apparent airmass. Therefore the GGG suite also includes a post-processing routine applying an empirical airmass-dependent correction factor (ADCF) and airmass-independent correction factor (AICF). The AICF are deduced from comparisons with in situ instrumentation on aircrafts (Wunch et al., 2010).

2.2 HR125 low resolution data set

In addition to the afore mentioned TCCON data product, a second data product from the HR125 will be used in this work, in the following called HR125 LR. For this product the raw interferograms are first truncated to the resolution of the EM27/SUN, 0.5 cm^{-1}. At 0.5 cm^{-1} resolution, the ILS of the HR125 is expected to be nearly nominal. However, to avoid any systematic bias of the HR125 LR data with respect to the EM27/SUN results, the same procedure for ILS determination from H_2O signatures in open path lab air spectra was applied and the resulting ILS parameters adopted for the trace gas analysis. The analysis procedure will be explained in detail in section 2.3, the retrieval software used for this dataset is PROFFIT Version 9.6 (Hase et al., 2004). The reason for the construction of this HR125 LR data set is that with this approach the analysis for the two instruments can be performed exactly the same way. The resolution is harmonized, the averaging kernels for a given airmass are nearly identical. Differences betweens the EM27/SUN and the HR125 LR data set can then be attributed to instrumental features alone and do not need to be disentangled from retrieval software, resolution and airmass dependency differences. Note that for the low resolution data set, forward and backward scans are averaged and then analysed whereas they are analysed separately for the TCCON data set. Therefore number of coincident measurements with the EM27/SUN data set compared to the TCCON data set is lower.

2.3 EM27/SUN data set

The EM27/SUN spectrometer, which was developed by KIT in collaboration with Bruker Optics^{TM}, is utilized for the acquisition of solar spectra. The instrument has been described in great detail in Gisi et al. (2012), in the following a short overview is given. Central part of this Fourier transform spectrometer (FTS) is a RockSolid^{TM} pendulum interferometer with two cube corner mirrors and a CaF_2 beamsplitter. The EM27/SUN routinely records double sided interferograms, the compensated BS design minimizes the curvature in the phase spectrum. This setup achieves high stability against thermal influences and vibrations. The retroreflectors are gimbal-mounted, which results in frictionless and wear-free movement. In this aspect the EM27/SUN is more stable than the HR125 high resolution FTS, which suffers from wear because of the use of friction bearings on the moving retroreflector. Over time this leads to shear misalignment and requires regular realignment (Hase, 2012). The gimbal-mounted retroreflectors move a geometrical distance of 0.45 cm leading to an optical path difference of 1.8 cm which corresponds to a spectral resolution of 0.5 cm^{-1}.

In a first pre-processing step, a solar brightness fluctuation correction is performed similar to Keppel-Aleks et al. (2007). Furthermore, the recorded interferograms are Fourier transformed using the Norton-Beer-Medium apodisation function (Davis et al., 2010). This apodisation is useful for reducing sidelobes around the spectral lines, an undesired feature in low resolution spectra, which would complicate the further analysis. A quality control, which filters interferograms with intensity fluctuations above 10 % and intensities below 10 % of the maximal signal range, is also applied.

In this work, spectra were analyzed utilizing the PROFFIT Version 9.6, a non-linear least-squares spectral fitting algorithm, which gives the user the opportunity to provide the measured ILS as an input parameter, an option chosen for this study (Hase et al., 2004). This code is in wide use and has been thoroughly tested in the past for the HR125 as well as the EM27/SUN, e.g. Schneider and Hase (2009); Sepúlveda et al. (2012); Kiel et al. (2016a); Chen et al. (2016). Due to the low resolution of the EM27/SUN, the atmospheric spectra were fitted by scaling of a priori trace gas profiles, although PROFFIT has the ability to perform a full profile retrieval (Dohe, 2013). As source of the a priori profiles, the TCCON daily profiles introduced in section 2.1 are utilized to be consistent with the TCCON analysis. Also for the daily temperature and pressure profiles, the approach from TCCON was adopted, using NCEP model data together with on site ground pressure data from a meteorological tall tower (www.imk.kit.edu/messmast/).

For the evaluation of the O_2 column the 7765 - 8005 cm^{-1} spectral region is used, which is also applied in the TCCON analysis (Wunch et al., 2010). For CO_2 we combine the two spectral windows used by TCCON to one larger window ranging from 6173 to 6390 cm^{-1}. CH_4 is evaluated in the 5897 - 6145 cm^{-1} spectral domain. For H_2O the 8353 - 8463 cm^{-1} region is used. This differs from TCCON, which deploys several narrow spectral windows, a strategy which is more in line with high resolution spectral observations. For consistency reasons, and to reference the results to the WMO scale, the EM27/SUN retrieval also performs a post-processing. The AICFs from TCCON are adopted, and similar to Wunch et al. (2010), an airmass dependency correction is performed, although other numerical values for the correction parameters are used. Details can be found in Frey et al. (2015); Klappenbach et al. (2015).

3 Long term performance

3.1 ILS analysis

Accurate knowledge of the real ILS of a spectrometer is extremely important because errors in the ILS lead to systematic errors in the trace gas retrieval. For this reason regular ILS measurements were performed from the beginning of this study four years ago to detect possible misalignments and alignment drifts. Source of a de-adjustment is mostly mechanical shock, due to e.g. impacts or vibrations especially due to transportation of the instruments. For the analysis of the measured data version 14.5 of the retrieval software LINEFIT (Hase et al., 1999; Hase, 2012) is used. Due to the fact that the EM27/SUN is equipped with a circular field stop aperture, the ILS is nearly nominal. Therefore, for keeping the treatment concise, we use the simple two parameter ILS model offered by LINEFIT. A detailed description of the ILS analysis is given in Frey et al. (2015). The time series of the ILS measurements is shown in Fig. 1, the modulation efficiency (ME) at maximum optical path difference (MOPD) ranges between 0.9835 and 0.9896, with a mean value of 0.9862 and a standard deviation of 0.0015. The phase error is close to zero for the whole time series with a mean value of 0.0019 $\pm$ 0.0018. This modulation efficiency is significantly different from nominal, which is surprising, as great care was taken to align the instrument. Therefore open path measurements were also performed for the HR125 at a resolution of 0.5 cm^{-1} to investigate whether this method shows a bias. For this small optical path difference, the alignment of the HR125 should be very close to nominal. However, the LINEFIT analysis shows a ME of 0.9824 at MOPD. From this result it is concluded that this method shows an overall low bias of around 1.5 - 2 %, probably due to a slight underestimate of the pressure broadening parameters of H_2O in the selected spectral region.

There is no overall trend apparent in the time series, the remaining differences in the modulation efficiency are probably due to the remaining uncertainty of the measurement technique. As is indicated by the more frequent measurements in 2017, there is also no seasonality in the results of the open path measurements. It should be noted that the measurement routine was refined in the course of this work. In particular, in the beginning (2014) it was assumed that the inside of the EM27/SUN is free of water vapor, so the instrument was not vented during the lamp measurements. However, sensitivity studies as presented in Frey et al. (2015) revealed that the influence of the water vapor column inside of the spectrometer can not always be neglected. After this discovery the instrument was vented during the open path measurements. This is why the 2014 calculations show larger scatter, as here the amount of water vapor inside the spectrometer is not known. For this analysis it was assumed that also for the 2014 measurements the total pressure inside the spectrometer is the same as of the surrounding air, which is a sensible assumption as the spectrometer is not evacuated. This explains also that the deviations become smaller in 2017. A further test to verify the stability of the instrument is the X_{air} parameter, which is the surface pressure divided by the measured column of air. This test will be shown in section 3.3.

The grey lines in Fig. 1 denote transportation of the spectrometer over longer distances for field campaigns in Berlin (North-Eastern Germany), Oldenburg (Northern Germany) and Paris (France) and for maintenance at Bruker Optics. Note that no realignment of the interferometer was performed during this maintenance. Only the reference HeNe laser was exchanged due to sampling instabilities during interferogram recordings. More specifically, the laser wavelength was unstable resulting in a corruption of parts of the measured spectra. Later in 2016 and 2017 this instrument was not used for campaigns since it has been chosen as the reference EM27/SUN for comparison measurements next to the HR125 spectrometer in order to take measurements at Karlsruhe as continuously as possible. The instrument was not realigned during the whole comparison study. An error estimation for the open path measurements is given in Table 1. For the temperature and pressure error, the stated accuracies of the data logger manufacturer were used. For the other potential error sources reasonable estimates were made. The total error, given by the root-squares-sum of the individual errors, is 0.29 % in ME amplitude, consisting of several errors of approximately the same magnitude.

3.2 Total column time series

In this section the total column measurements from the EM27/SUN are compared to the reference HR125 spectrometer. For the measurements, the EM27/SUN was moved to a terrace on the top floor of the IMK-ASF, building 435 KIT CN (49.094 °N, 8.436 °E, 133 m a.s.l.), on a daily basis if weather conditions were favourable. The spectrometer was moved from the lab on the fourth floor to the roof terrace on the seventh floor thus being exposed to mechanical stress. The instrument was coarsely oriented north, without effort for levelling. If further orientation was needed, the spectrometer was manually rotated so that the solar beam was centered onto the entrance window. The CamTracker program was then able to track the sun. The spectrometer was operated at ambient temperatures. During summer, the spectrometer heated up to temperatures above 40 °C. In order to protect the electronics from the heat, a sun cover for the EM27/SUN was built, which reduced the temperatures inside the spectrometer by about 10 °C. In winter the temperatures were as low as -4 °C at the start of measurements. Double-sided interferograms with 0.5 cm^{-1} resolution were recorded. With 10 scans and a scanner velocity of 10 kHz, one measurement takes about 58 s. For precise time recording, a GPS receiver was used.

The full time series from March 2014 to November 2017 is shown in Figure 2 for the three data sets. For better visibility only coincident data points measured within one minute between EM27/SUN and the other data sets are shown. There are 8349 paired measurements between EM27/SUN and TCCON and 4624 between EM27/SUN and HR125 LR, in total there are 50550 EM27/SUN and 25361 TCCON measurements.

All gases show a pronounced seasonal cycle, where the variability in water vapour is strongest with values below 1×10^{26} molecules m^{-2} in winter and up to 14×10^{26} molecules m^{-2} in summer. Furthermore, the seasonal cycle of water vapour is shifted with respect to the other species. Another feature seen is that there is an offset in the EM27/SUN (red squares) and HR125 LR (blue squares) total column data with respect to the TCCON data (black squares). The occurence of a systematic bias when reducing the spectral resolution has been observed by several investigators (Petri et al., 2012; Gisi et al., 2012). The observed offset between EM27/SUN and HR125 LR measurements is smaller. The remaining difference can be attributed to the different measurement heights of the HR125 (112 m) and EM27/SUN (133 m). For a quantitative analysis we do not utilize the total column measurements but rather use the X_{Gas}, as in this representation systematic errors, e.g. ILS errors, timing errors, tracking errors and nonlinearities mostly cancel out. Furthermore the height dependence largely cancels out in this representation. The comparison will be presented in the following sections.

~~Before~~ First, a sensitivity study is provided demonstrating the effect of changes in the ILS on the gas retrieval. For this one hour of measurements around solar noon on 01 August 2016 and 15 February 2017, corresponding to solar elevation angles (SEA) of 60° and 30°, were analysed with artificially altered ILS values. The results are shown in Table 2. An increase of 1 % in the modulation efficiency leads to a decrease of 0.35 % (0.37 %) on the retrieved O_2 column, 0.31 % (0.31 %) on H_2O, 0.26 % (0.28 %) on CH_4 and 0.50 % (0.57 %) on CO_2 for an SEA of 60° (30°). So the change in the retrieved total column is not alike, but a unique characteristic of each species, and also slightly airmass dependent. As the decrease in the CO_2 column is larger than the decrease in the O_2 column, XCO_2 decreases with an increasing ME, 0.16 % (0.19 %) for 1 % ILS increase, whereas XCH_4 increases 0.10 % (0.09 %). This is opposed to prior studies (Gisi et al., 2012; Hedelius et al., 2016), reporting an increase of XCO_2 and decrease of XCH_4 for an increase of the modulation efficiency, albeit in agreement with the findings from Hase et al. (2013) for the HR125 spectrometer, reporting that a change in the modulation efficiency results in a larger relative decrease in the CO_2 column than in the O_2 column.

3.3 $\mathbf{X}_{air}$

In this section the column averaged amount of dry air (X_{air}) is investigated. This quantity is a sensitive test of the stability of a spectrometer because for X_{air} there is no compensation of possible instrumental problems, in contrast to the DMFs, where errors can partially cancel out. X_{air} compares the measured oxygen column (VC_{O_2}) with surface pressure measurements (P_S):

$$X_{air} = \frac{0.2095}{VC_{O_2} \cdot \overline{\mu}} \cdot \left(\frac{P_S}{g} - VC_{H_2O} \cdot \mu_{H_2O} \right) \tag{1}$$

Here $\overline{\mu}$ and μ_{H_2O} denote the molecular masses of dry air and water vapour, respectively, g is the column averaged gravitational acceleration and VC_{H_2O} is the total column of water vapour. The correction with VC_{H_2O} is necessary as the surface pressure instruments measure the pressure of the total air column, including water vapour. For an ideal measurement and retrieval with accurate O_2 and H_2O spectroscopy, as well as accurate surface pressure, X_{air} would be 1. However, due to insufficiencies in the oxygen spectroscopy, this value is not obtained. For TCCON measurements X_{air} is typically $\sim$ 0.98 (Wunch et al., 2015). For the EM27/SUN prior studies showed a factor of $\sim$ 0.97 (Frey et al., 2015; Hase et al., 2015; Klappenbach et al., 2015). Large deviations ($\sim$ 1 %) from these values indicate severe problems, e.g. errors with the surface pressure, pointing errors, timing errors or changes in the optical alignment of the instrument. As mentioned in section 3.1, here X_{air} is used to check whether the small changes in the modulation efficiency indicated by the open path measurements are due to actual alterations in the alignment of the EM27/SUN or due to the residual uncertainty of the calibration method.

The left panel of Figure 3 shows the X_{air} time series of TCCON, the EM27/SUN and HR125 LR. For clarity, only coincident data points that were measured within one minute between the different data sets are shown. Grey areas denote periods where the EM27/SUN was moved over long distances for campaigns or maintenance. The absolute values of X_{air} differ for the data sets, with 0.9805 $\pm$ 0.0012 for TCCON, 0.9669 $\pm$ 0.0010 for the EM27/SUN and 0.9670 $\pm$ 0.0011 for HR125 LR. The difference between the EM27/SUN and the HR125 LR is within 1σ precision. The difference between the EM27/SUN and TCCON data set, which is commonly observed as previously noted, is a consequence of the different resolution together with the different retrieval algorithm (Gisi et al., 2012). It can be seen that all data sets exhibit a seasonal variability, which is more prominent in the TCCON data as can also be seen from the higher standard deviation. From this higher variability it can be concluded that the airmass dependency in the official TCCON O_2 retrieval is higher than for the PROFFIT retrieval on reduced resolution TCCON measurements, a finding also observed by Gisi et al. (2012) between the TCCON retrieval and the PROFFIT retrieval at full resolution. For the PROFFIT retrieval, it is suspected that part of the variability stems from insufficiencies in the utilized HITRAN 2008 H_2O linelist. It was reported by Tallis et al. (2011) that in the 8000-9200 cm^{-1} region, line intensities are low by up to 20 % compared to other wavenumber regions. This in return will lead to a systematic overestimation of the water column, which also affects X_{air}. To test the sensitivity of X_{air} with respect to the measured H_2O column, in the right panel of Figure 3 the original EM27/SUN time series is compared to a data set where the H_2O column is artificially reduced by 20 %. This approach is further justified by a study from the Romanian National Institute for Research and Development in Optoelectronics (INOE) conducted in 2017, where ~~they~~ we compared total column amounts of water vapor from an EM27/SUN and a radiometer. ~~They~~We found that the EM27/SUN values were systematically higher by 20 % ~~(Dragos Ene, priv. comm.)~~. And indeed, the standard deviation, which is here used as a measure for the seasonal variability, of the modified time series (0.0009) is lower when compared to the original time series (0.0010).

There are no obvious steps and there is no significant drift between the EM27/SUN and the HR125 LR data sets so that it can be concluded that the EM27/SUN is stable during the complete course of the over three year long comparison and differences seen in the modulation efficiency are introduced by the remaining uncertainty in the calibration method.

3.4 $\mathbf{XCO_2}$

In Figure 4 XCO_2 time series of the three data sets are shown together with the offsets between the data sets. The general characteristics of the data sets are similar. The yearly increase of XCO_2 due to anthropogenic emissions of about 2 ppmv can be seen as well as the seasonal cycle with a decrease of XCO_2 of approximately 10 ppmv during summer due to photosynthesis, characteristic for mid latitude stations. Despite these agreements in the general trend, there are also differences between the data sets. Relative to the TCCON data the EM27/SUN and the HR125 LR data sets are biased high (0.98 % and 0.84 % respectively). The scaling factors are calculated by taking the mean of all individual coincident point ratios (EM27/SUN / TCCON and EM27/SUN / HR125 LR), together with these ratios also a standard deviation is derived, see Table 3. A high bias was also observed by Gisi et al. (2012); Frey et al. (2015), albeit with smaller absolute differences. This is due to the fact that (1) in the Gisi et al. paper the TCCON data was retrieved with an earlier version of GFIT (GGG2012) and (2) after the publication of the Frey et al. paper the Karlsruhe TCCON data was reprocessed with a customized GFIT retrieval accounting for baseline variations (Kiel et al., 2016b). The offset between EM27/SUN and TCCON shows a seasonal variability. Reasons for this are mainly the differences in airmass correction, averaging kernels and retrieval algorithm. These effects have been investigated before (Gisi et al., 2012; Frey et al., 2015; Klappenbach et al., 2015; Hedelius et al., 2017), (Kiel et al., 2016a). The averaging kernels of the EM27/SUN have been previously presented and compared to TCCON in a study by Hedelius et al. (2016).

It has to be noted that the level of uncertainty for XCO_2 is significantly higher between COCCON and TCCON compared to the internal EM27/SUN consistency. According to table 3, a current calibration uncertainty with respect to TCCON of 0.6 ppmv is estimated.

For the long term stability of the EM27/SUN the focus lies on the comparison with the HR125 LR data set, where the above mentioned differences cancel out. There is a small offset between the two data sets, resulting in a calibration factor of 1.0014, which is constant over time in the analyzed time period. To test this assumption a linear fit was applied to the XCO_2 ratios, see right panel of Figure 4. In Table 3 the slope coefficient is depicted. For both comparisons the yearly trend in the ratio is well within the 1σ precision (0.44 ppmv) of the data set. In absolute numbers the slope per year is $\approx$ - 0.02 ppmv for both ratios, or a drift smaller than 0.1 ppmv over the whole comparison period of around three and a half years.

Figure 5 shows the data sets in a different representation. In the left panel the EM27/SUN is compared to the HR125 LR, the colorbar indicates the date of measurement and the dashed line is the 1 : 1 line. It can be seen that there is no trend in the data apart from the overall increase in time due to anthropogenic emissions. In the right panel the EM27/SUN is compared to the TCCON data set, the colorbar shows the solar elevation angle (SEA). This representation is chosen so that the remaining airmass dependency of the ratio can be seen. It is also interesting to note that omitting the TCCON airmass independent correction factor (AICF) for our analysis would move the data set significantly closer to the 1 : 1 line. The scaling factor would change from 1.0098 to 0.9995. As this finding is not true for XCH_4 and is probably coincidental, we maintain the AICF.

3.5 XCH_4

Figure 6 shows the XCH_4 time series of the different data sets. As for XCO_2, the general features are in agreement for all data sets. There is a slight annual increase of about 10 ppbv. Also there is a seasonal cycle with a variability of $\approx$ 30 ppbv; however, compared to XCO_2 the interannual seasonality strength and phase varies significantly between the years due to the many different variable sinks and sources of methane, e.g. Dlugokencky et al. (1997). The differences between the data sets largely resemble the differences observed for XCO_2. The bias between EM27/SUN and TCCON is 0.72 %, see Table 4. This bias is close to the bias observed by Hedelius et al. (2016), 0.75 %, where they used the GGG software package for the analysis of EM27/SUN spectra. Although a single bias is reported, as was observed for XCO_2 the offset is not constant, but rather shows a seasonality. The calibration uncertainty between COCCON and TCCON is estimated to amount to 5 ppbv for XCH_4, see table 4. The retrievals between EM27/SUN and HR125 LR agree within 1σ precision (0.9997 ± 0.0008). The left panel of Figure 7 shows the ratio between EM27/SUN and HR125 LR color coded with the observation date. As for XCO_2, no trend is apparent. An explicit linear fit to the XCH_4 ratio produces a slope coefficient of 0.0001, one order of magnitude smaller than the 1σ precision of the ratio (0.0008).

An interesting feature is observed in the ratio between EM27/SUN and TCCON data sets, see right panel of Figure 7. In general the pattern is similar to that of XCO_2, with a slight dependence on the SEA. The ratio in the figure is color coded with the date of observation rather than the SEA. It can be seen that for 01 March 2016 and 14 March 2016 (shaded area in Figure 7) the XCH_4 ratio significantly differs from the other observations. Previous work by Ostler et al. (2014) has shown that stratospheric intrusion, caused for example by the subsidence of the polar vortex, has a different effect on MIR and NIR retrievals, even when using the same a priori profile. This is due to the differing sensitivity of the retrievals with respect to altitude. Therefore, differences between the true atmospheric profile and the assumed a priori profiles on these days could cause the differences seen. This effect will also lead to larger differences between EM27/SUN and TCCON XCH_4 because of the different impact on the retrieved columns due to differing sensitivities. A spread of the polar vortex to mid-latitudes could lead to significantly altered CH_4 profiles compared to the a priori profiles, explaining the observed differences in the XCH_4 ratio.

The left panel of Figure 8 shows N_2O data from the Microwave Limb Sounder (MLS) on the Aura satellite for several days in February and March 2016 on the 490 K potential temperature level, corresponding to a height of approximately 18 km. N_2O is chosen because it serves as a tracer for the position of the polar vortex. Indeed it seems that beginning of March 2016 the polar vortex stretches out to mid-latitudes. To further test this hypothesis in the right panel of Figure 8 independent NDACC CH_4 profiles from the Jungfraujoch station in 2016 are shown. The station is situated approximately 270 km south of Karlsruhe with a station height of 3580 m. For dates without measurements, the data was interpolated using a weighted average. The dotted black lines denote 1 March 2016 and 14 March 2016, the dates on which the XCH_4 ratio between EM27/SUN and TCCON shows an anomaly. The changed profile shape during that period is clearly visible. As this station is south of Karlsruhe, it is expected that also for Karlsruhe the CH_4 profile shows considerable downwelling, explaining the observed anomaly in the XCH_4 ratio.

4 Ensemble performance

Having investigated the long term stability of the EM27/SUN with respect to a reference spectrometer in the previous section, here the level of agreement of an ensemble of EM27/SUN spectrometers is presented. The procedure is the same as for the comparison between the reference EM27/SUN and the HR125. First, the ILS is analysed, followed by calibration factors for XCO_2 and XCH_4.

4.1 ILS measurements and instrumental examination

The measurement of the ILS is a valuable diagnostic for detecting misalignments of spectrometers. Differences in the ILS of the EM27/SUN spectrometers due to misalignment can lead to biases in the data products between the instruments. Here the spread of ILS values of all EM27/SUN spectrometers that were checked at KIT in the past four years is estimated. Numerical values are given in Table 5, the results are shown in Figure 9. The black square denotes an ILS measurement of the HR125 spectrometer, also with 1.8 cm MOPD. This test was done to check for an absolute offset of our method. The HR125 would be expected to show an ideal ILS for short optical path differences, but a value of 0.9824 was obtained. From this measurement it is concluded that our method shows an absolute offset and that values between 0.98 and 0.99 are desired.

In general, the agreement between the 30 tested EM27/SUN is good with an ensemble mean of 0.9851 ± 0.0078, which is not differing significantly from the value obtained for the HR125, but there are exceptions. Instrument SN 44 was checked at KIT only after an upgrade with the second channel was performed at Bruker Optics. Before realignment, the instrument showed a very low ME value of 0.9374. A realignment of the instrument enhanced the ME to 0.9714. This is still significantly low compared to the EM27/SUN ensemble mean, but the difference was drastically reduced. The second instrument showing strong deviations from the ensemble mean is SN76 with an ILS of 1.0160, the only instrument showing overmodulation. The ILS was even higher (1.0350) when first ILS measurements were performed. Due to our findings, the manufacturer exchanged the beamsplitter which reduced the overmodulation, but it partly remained. In the meantime it was recognized as the cause of error that the manufacturer during assembling of the instrument forgot to insert the foreseen spacer to achieve the correct detector position with respect to the beamsplitter. The beamsplitter is coated, and the coating is applied on both sides of the beamsplitter over half the surface area. If the optical axis of the detector element coinciding with the transition region of the two coating areas, detrimental effects occur. For this reason the detector element needs to be raised with respect to the interferometer. This problem occured for instrument SN 77 but there it was diagnosed and corrected by KIT (ILS before lifting: 1.0340, ILS after correction: 0.9855).

The above mentioned problems show the benefit of the calibration routine at KIT. Imperfections from nonideal alignments were diagnosed and corrected. Also other detrimental effects, e.g. double-passing, channeling, nonlinearity issues, solar tracker problems, inaccurate positioning of the second detector or camera issues, were corrected or minimized for a number of instruments. Finally, it was checked whether the linear interpolation method suppressing sampling ghosts was activated.

4.2 XCO_2 and XCH_4 comparison measurements

After checking the alignment and performing lamp measurements, side-by-side solar calibration measurements were performed on the terrace on top of the KIT-IMK office building with each spectrometer with respect to the reference EM27/SUN and also a co-located HR125 spectrometer. Calibration measurements started in June 2014 and are ongoing, if new spectrometers arrive for testing. The aim is to have at least one day of comparison measurements so that the spectrometers can be scaled to TCCON via the reference EM27/SUN. TCCON is extensively compared to measurements on the WMO scale. Dates of the comparison measurements for the different spectrometers as well as number of coincident measurements are given in Table 6. On January 21 2016, our reference spectrometer suffered from laser sampling errors after approximately one hour of measurements. Therefore the number of coincident measurements for SN62 and 63 that were ~~calibrated~~ checked on this date are sparse. A typical calibration day is depicted in Fig. 10.

The calibration factors and standard deviations for all instruments with respect to the reference spectrometer are also depicted in Table 6. Calibration factors and standard deviations were obtained using the methods described in section 3.4. The calibration factors are close to nominal for all species and instruments. For XCO_2 the ensemble mean is high compared to the reference EM27/SUN, with a mean calibration factor of 0.9993 and a standard deviation of 0.0007. In Fig. 11 histograms of the calibration factor distributions are depicted for XCO_2, XCH_4 and O_2, respectively. The histograms are not conspicuous.

Applying ~~this~~ the mean calibration factor to all calculated calibration factors centers the data around the ensemble mean. As an estimate for the spread of the calibration factors $\frac{1}{n}\Sigma|X\text{Gas factor} - 1|$, we arrive at an average bias between the instruments of 0.20 ppmv. From Table 6 we can also calculate an average standard deviation $\frac{1}{n}\Sigma|\sigma|$ of 0.13 ppmv. For XCH_4 the ensemble mean is closer to the reference EM27/SUN (0.9997 ± 0.0006) as compared to XCO_2. From this results an average bias of 0.8 ppbv. The average standard deviation is 0.6 ppbv. These values are comparable to results obtained in a study from Hedelius et al. (2017). They checked the intercomparability of the 4 United States TCCON sites using an EM27/SUN as a traveling standard. They report average biases of 0.11 ppmv for XCO_2 and 1.2 ppbv for XCH_4, for the average standard deviations they obtain 0.34 ppmv (XCO_2) and 1.8 ppbv (XCH_4). It has to be noted that for the Hedelius et al. (2017) study only data within ± 2 h local noon was taken into account whereas here no constraints regarding the time of measurement were applied. As another sensitive test the O_2 total column calibration factors are given. In contrast to XCO_2 and XCH_4, there is no canceling of errors in this quantity. The ensemble mean is slightly high compared to the reference EM27/SUN (0.9999 ± 0.0014). The average bias is 0.11 % O_2 with an average standard deviation of 0.04 % O_2.

Note that for our setup this average bias is a worst case scenario. The bias only applies if no calibration factor is used in the subsequent analysis. The strength of this calibration routine is that the computed calibration factors can be used, thereby significantly lowering the bias between different EM27/SUN spectrometers. The remaining bias is then given by the long term drift of the individual instrument, see section 3.4 and 3.5, and sudden alignment drifts due to mechanical strain from e.g. transport, campaign use. To estimate this drift, we utilize the calibration factors before and after the Berlin campaign performed in 2014. There the drifts between five instruments were below 0.005 % XCO_2 and 0.035 % XCH_4 (Frey et al., 2015).

Ideally, we would expect identical calibration factors as we took the real ILS of the instruments into account. As this is not the case, we investigate whether the remaining differences can be attributed to the uncertainties of the open path measurements, which are summarized in Table 1. The results are incorporated in Fig. 12. The left panel shows the correlation between O_2 and XCO_2 calibration factors. Black squares denote the empirical calibration factors derived from the side-by-side measurements. The red squares show calculated calibration factors based on the ME uncertainty budget. The dashed red line is a linear fit through the calculated factors. About half the measured empirical factors are within the bounds of the factors derived from the ME error budget. Furthermore the slopes of the calculated and empirical factors are in good agreement, confirming that the ME uncertainty is contributing to the uncertainty of the calibration factors. The other contributions for this uncertainty are due to a superposition of various small device-specific imperfections. The right panel of Fig. 12 shows the correlation between O_2 and XCH_4 calibration factors. The findings mentioned above for the O_2 and XCO_2 correlation also hold true here.

5 Conclusions and Outlook

Based on a long-term intercomparison of column-averaged greenhouse gas abundances measured with an EM27/SUN FTIR spectrometer and with a co-located 125HR spectrometer, respectively, we conclude that the EM27/SUN offers highly stable instrument characteristics on timescales of several years. The drifts on shorter timescales reported by Hedelius et al. (2016) were probably exclusively - as conjectured by the authors of the study - due to a deviation from the instrumental design as originally recommended. The application of a wideband detector suffering from nonlinearity together with steadily decreasing signal levels due to ageing of the tracker mirrors seem to be the reason for the observed drifts.

The favourable instrument stability which is preserved even during transport events and operation under ambient conditions suggests that the EM27/SUN spectrometer is well suited for campaign use and long-term deployment at very remote locations as a supplement of the TCCON. A deployment at remote sites is further facilitated by the recent development of an automated enclosure for the EM27/SUN, which enables unattended remote operation (Heinle and Chen, 2018; Dietrich and Chen, 2018). An annual to biannual check of the instrument performance by performing a side-by-side intercomparison with a TCCON spectrometer seems adequate for quality monitoring. For separating out instrumental drifts from atmospheric signals, the addition of low-resolution spectra derived from the TCCON measurements is highly useful, because in this kind of comparison, the smoothing error and any possible resolution-dependent biases of the analysis software cancel out. The ensemble performance of 30 EM27/SUN spectrometers turns out to be very uniform, supported by a centralized acceptance inspection performed at KIT before the spectrometers are deployed. When using the empirical ILS parameters derived for each spectrometer, the scatter in XCO_2 amounts 0.13 ppmv, while it is 0.6 ppbv for XCH_4. The standard deviation of the oxygen columns is 0.04%. We expect that the conformity of measurement results will be even better than indicated by this scatter, if the remaining empirical calibration factors are taken into account. These empirical calibration factors are likely composed of several small device-specific error contributions, a major contribution was identified to stem from the uncertainty of the ILS measurements.

Continuation and further development of the COCCON activities seem highly desirable for achieving the optimal performance of the growing EM27/SUN spectrometer network. The implemented pre-deployment procedures of testing, optimizing, and calibrating each device - executed by experts at a central facility – help to ensure consistent results from EM27/SUN spectrometers operated in any part of the world. This approach is corroborated by the proven excellent long-term stability of instrumental characteristics, and the proven high degree of stability under thermal and mechanical burdens as they occur during transport. In order to maintain the reliability of the EM27/SUN spectrometers, we propose to investigators to send the instrument to KIT for a biennial inspection. The EM27/SUN spectrometer does not require continuing expert maintenance and it is very simple to operate, we therefore expect that many investigators world-wide who are not keen to become FTIR experts will be attracted by this measurement device, operating it as a side activity. Current COCCON work supported by ESA in the framework of the COCCON PROCEEDS project will result in an easy-to-handle preprocessing tool optimized for the EM27/SUN spectrometer. This tool will generate quality-checked spectra from raw interferograms, which then are forwarded to a central data analysis facility. A demonstration setup of the central facility will be part of COCCON PROCEEDS. When finally implemented on an operational level, the facility will remove the whole burden of the quantitative trace gas analysis from the operator and ensure the consistency of the trace gas analysis chain to the utmost degree. Furthermore it will enable a timely reanalysis of all submitted spectra after upgrades of the retrieval procedures and minimize the risk of data loss if operators for some reason are stopping their activity. Finally, this centralized facility will serve as a unique contact point for the data users.

Author contributions. Matthias Frey: performed measurements, data analysis, paper writing
Mahesh Kumar Sha: performed measurements, contributed to data analysis
Frank Hase: performed measurements, data analysis, paper writing
Matthäus Kiel: contributed to data analysis

Thomas Blumenstock: performed measurements, contributed to calibration efforts

Roland Harig: contributed to calibration efforts

Gregor Surawicz: contributed to calibration efforts

Nicholas M. Deutscher: contributed to calibration efforts

Kei Shiomi: contributed to calibration efforts

Jonathan Franklin: contributed to calibration efforts

Hartmut Bösch: contributed to calibration efforts

Jia Chen: contributed to calibration efforts

Michel Grutter: contributed to calibration efforts

Hirofumi Ohyama: contributed to calibration efforts

Youwen Sun: contributed to calibration efforts

André Butz: contributed to calibration efforts

Gizaw Mengistu Tsidu: contributed to calibration efforts

Dragos Ene: contributed to calibration efforts, provided evidence for XH_2O bias

Debra Wunch: contributed to calibration efforts

Zhesong Cao: contributed to calibration efforts

Omaira Garcia: contributed to calibration efforts

Michel Ramonet: contributed to calibration efforts

Felix Vogel: contributed to calibration efforts

Johannes Orphal: supported the advance of the project, contributed to calibration efforts

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

Acknowledgements. We acknowledge support by the ACROSS research infrastructure of the Helmholtz Association of German Research Centres (HGF).

~~We acknowledge support by the MOSES research infrastructure of the HGF.~~ This work was supported by funding from the Helmholtz Association in the framework of MOSES (Modular Observation Solutions for Earth Systems).

We thank the National Center for Environmental Prediction (NCEP) for providing atmospheric temperature profiles.

We thank the NASA science team for providing MLS data from the Aura satellite.

We thank the Jungfraujoch NDACC team for providing Jungfraujoch FTIR data.

I. Morino and A. Hori contributed by procuring the NIES instrument and performing additional instrumental line shape measurements in Tsukuba.

We acknowledge funding from the Australian Space Research Program - Greenhouse Gas Monitoring Project, the Australian Research Council project DE140100178, and the Centre for Atmospheric Chemistry (CAC) Research Cluster supported by the University of Wollongong Faculty of Science, Medicine and Health.

We thank Minqiang Zhou (BIRA-IASB) for his contribution to the tool which was used to truncate the IFS 125HR interferograms.
We acknowledge support from the ESA project 4000118115/16/NL/FF/gp: Technical Assistance for a Romanian Atmospheric Mobile Observation System (RAMOS).

References

Chen, J., Viatte, C., Hedelius, J. K., Jones, T., Franklin, J. E., Parker, H., Gottlieb, E. W., Wennberg, P. O., Dubey, M. K., and Wofsy, S. C.: Differential column measurements using compact solar-tracking spectrometers, Atmospheric Chemistry and Physics, 16, 8479–8498, https://doi.org/10.5194/acp-16-8479-2016, https://www.atmos-chem-phys.net/16/8479/2016/, 2016.

Davis, S., Abrams, M. C., and Brault, J. W.: Fourier Transform Spectrometry, Academic press, 2010.

Dietrich, F. and Chen, J.: Portable Automated Enclosure for a Spectrometer Measuring Greenhouse Gases, in: Geophysical Research Abstracts Vol. 20, EGU2018-16281-1, 2018.

Dlugokencky, E. J., Masarie, K. A., Tans, P. P., Conway, T. J., and Xiong, X.: Is the amplitude of the methane seasonal cycle changing?, Atmospheric Environment, 31, 21–26, https://doi.org/10.1016/S1352-2310(96)00174-4, http://www.sciencedirect.com/science/article/pii/S1352231096001744, 1997.

Dohe, S.: Measurements of atmospheric CO_2 columns using ground-based FTIR spectra, Ph.D. thesis, Karlsruhe Institute of Technology, 2013.

Frankenberg, C., Pollock, R., Lee, R. A. M., Rosenberg, R., Blavier, J.-F., Crisp, D., O'Dell, C. W., Osterman, G. B., Roehl, C., Wennberg, P. O., and Wunch, D.: The Orbiting Carbon Observatory (OCO-2): spectrometer performance evaluation using pre-launch direct sun measurements, Atmospheric Measurement Techniques, 8, 301–313, https://doi.org/10.5194/amt-8-301-2015, http://www.atmos-meas-tech.net/8/301/2015/, 2015.

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.

Geibel, M. C., Messerschmidt, J., Gerbig, C., Blumenstock, T., Chen, H., Hase, F., Kolle, O., Lavrič, J. V., Notholt, J., Palm, M., Rettinger, M., Schmidt, M., Sussmann, R., Warneke, T., and Feist, D. G.: Calibration of column-averaged CH_4 over European TCCON FTS sites with airborne in-situ measurements, Atmospheric Chemistry and Physics, 12, 8763–8775, https://doi.org/10.5194/acp-12-8763-2012, https://www.atmos-chem-phys.net/12/8763/2012/, 2012.

Gisi, M., Hase, F., Dohe, S., and Blumenstock, T.: Camtracker: a new camera controlled high precision solar tracker system for FTIR-spectrometers, Atmospheric Measurement Techniques, 4, 47–54, https://doi.org/10.5194/amt-4-47-2011, https://www.atmos-meas-tech.net/4/47/2011/, 2011.

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, http://www.atmos-meas-tech.net/5/2969/2012/, 2012.

Hase, F.: Improved instrumental line shape monitoring for the ground-based, high-resolution FTIR spectrometers of the Network for the Detection of Atmospheric Composition Change, Atmospheric Measurement Techniques, 5, 603–610, https://doi.org/10.5194/amt-5-603-2012, https://www.atmos-meas-tech.net/5/603/2012/, 2012.

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., Hannigan, J., T. Coffey, M., Goldman, A., Höpfner, M., Jones, N., P. Rinsland, C., and Wood, S.: Intercomparison of retrieval codes used for the analysis of high-resolution, ground-based FTIR measurements, Journal of Quantitative Spectroscopy & Radiative Transfer, 87, 25–52, 2004.

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., Blumenstock, T., Groß, J., Kiel, M., Kohlhepp, R., Mengistu Tsidu, G., Schäfer, K., Sha, M. K., and Orphal, J.: Application of portable FTIR spectrometers for detecting greenhouse gas emissions of the major city Berlin, Atmospheric Measurement Techniques, 8, 3059–3068, https://doi.org/10.5194/amt-8-3059-2015, https://www.atmos-meas-tech.net/8/3059/2015/, 2015.

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.

Heinle, L. and Chen, J.: Automated enclosure and protection system for compact solar-tracking spectrometers, Atmospheric Measurement Techniques, 11, 2173–2185, https://doi.org/10.5194/amt-11-2173-2018, https://www.atmos-meas-tech.net/11/2173/2018/, 2018.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Leetmaa, A., Reynolds, R., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–471, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

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.

Kiel, M., Hase, F., Blumenstock, T., and Kirner, O.: Comparison of XCO abundances from the Total Carbon Column Observing Network and the Network for the Detection of Atmospheric Composition Change measured in Karlsruhe, Atmospheric Measurement Techniques, 9, 2223–2239, https://doi.org/10.5194/amt-9-2223-2016, https://www.atmos-meas-tech.net/9/2223/2016/, 2016a.

Kiel, M., Wunch, D., Wennberg, P. O., Toon, G. C., Hase, F., and Blumenstock, T.: Improved retrieval of gas abundances from near-infrared solar FTIR spectra measured at the Karlsruhe TCCON station, Atmospheric Measurement Techniques, 9, 669–682, https://doi.org/10.5194/amt-9-669-2016, https://www.atmos-meas-tech.net/9/669/2016/, 2016b.

Klappenbach, F., Bertleff, M., Kostinek, J., Hase, F., Blumenstock, T., Agusti-Panareda, A., Razinger, M., and Butz, A.: Accurate mobile remote sensing of XCO_2 and XCH_4 latitudinal transects from aboard a research vessel, Atmospheric Measurement Techniques, 8, 5023–5038, https://doi.org/10.5194/amt-8-5023-2015, https://www.atmos-meas-tech.net/8/5023/2015/, 2015.

Messerschmidt, J., Geibel, M. C., Blumenstock, T., Chen, H., Deutscher, N. M., Engel, A., Feist, D. G., Gerbig, C., Gisi, M., Hase, F., Katrynski, K., Kolle, O., Lavrič, J. V., Notholt, J., Palm, M., Ramonet, M., Rettinger, M., Schmidt, M., Sussmann, R., Toon, G. C., Truong, F., Warneke, T., Wennberg, P. O., Wunch, D., and Xueref-Remy, I.: Calibration of TCCON column-averaged CO_2: the first aircraft campaign over European TCCON sites, Atmospheric Chemistry and Physics, 11, 10 765–10 777, https://doi.org/10.5194/acp-11-10765-2011, https://www.atmos-chem-phys.net/11/10765/2011/, 2011.

Morino, I., Uchino, O., Inoue, M., Yoshida, Y., Yokota, T., Wennberg, P. O., Toon, G. C., Wunch, D., Roehl, C. M., Notholt, J., Warneke, T., Messerschmidt, J., Griffith, D. W. T., Deutscher, N. M., Sherlock, V., Connor, B., Robinson, J., Sussmann, R., and Rettinger, M.: Preliminary validation of column-averaged volume mixing ratios of carbon dioxide and methane retrieved from GOSAT short-wavelength infrared spectra, Atmospheric Measurement Techniques, 4, 1061–1076, https://doi.org/10.5194/amt-4-1061-2011, http://www.atmos-meas-tech.net/4/1061/2011/, 2011.

Nassar, R., Hill, T. G., McLinden, C. A., Wunch, D., Jones, D. B. A., and Crisp, D.: Quantifying CO2 Emissions From Individual Power Plants From Space, Geophysical Research Letters, 44, 10,045–10,053, https://doi.org/10.1002/2017GL074702, https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL074702, 2017.

Olsen, S. C. and Randerson, J. T.: Differences between surface and column atmospheric CO_2 and implications for carbon cycle research, Journal of Geophysical Research: Atmospheres, 109, https://doi.org/10.1029/2003JD003968, http://dx.doi.org/10.1029/2003JD003968, 2004.

Ostler, A., Sussmann, R., Rettinger, M., Deutscher, N. M., Dohe, S., Hase, F., Jones, N., Palm, M., and Sinnhuber, B.-M.: Multistation intercomparison of column-averaged methane from NDACC and TCCON: impact of dynamical variability, Atmospheric Measurement Techniques, 7, 4081–4101, https://doi.org/10.5194/amt-7-4081-2014, https://www.atmos-meas-tech.net/7/4081/2014/, 2014.

Petri, C., Warneke, T., Jones, N., Ridder, T., Messerschmidt, J., Weinzierl, T., Geibel, M., and Notholt, J.: Remote sensing of CO_2 and CH_4 using solar absorption spectrometry with a low resolution spectrometer, Atmospheric Measurement Techniques, 5, 1627–1635, https://doi.org/10.5194/amt-5-1627-2012, https://www.atmos-meas-tech.net/5/1627/2012/, 2012.

Schneider, M. and Hase, F.: Ground-based FTIR water vapour profile analyses, Atmospheric Measurement Techniques, 2, 609–619, https://doi.org/10.5194/amt-2-609-2009, http://www.atmos-meas-tech.net/2/609/2009/, 2009.

Sepúlveda, E., Schneider, M., Hase, F., García, O. E., Gomez-Pelaez, A., Dohe, S., Blumenstock, T., and Guerra, J. C.: Long-term validation of tropospheric column-averaged CH_4 mole fractions obtained by mid-infrared ground-based FTIR spectrometry, Atmospheric Measurement Techniques, 5, 1425–1441, https://doi.org/10.5194/amt-5-1425-2012, http://www.atmos-meas-tech.net/5/1425/2012/, 2012.

Tallis, L., Coleman, M., Gardiner, T., Ptashnik, I., and Shine, K.: Assessment of the consistency of H_2O line intensities over the near-infrared using sun-pointing ground-based Fourier transform spectroscopy, Journal of Quantitative Spectroscopy and Radiative Transfer, 112, 2268–2280, https://doi.org/10.1016/j.jqsrt.2011.06.007, 2011.

Washenfelder, R. A., Toon, G. C., Blavier, J.-F., Yang, Z., Allen, N. T., Wennberg, P. O., Vay, S. A., Matross, D. M., and Daube, B. C.: Carbon dioxide column abundances at the Wisconsin Tall Tower site, Journal of Geophysical Research: Atmospheres, 111, https://doi.org/10.1029/2006JD007154, http://dx.doi.org/10.1029/2006JD007154, 2006.

Wunch, D., Toon, G. C., Wennberg, P. O., Wofsy, S. C., Stephens, B. B., Fischer, M. L., Uchino, O., Abshire, J. B., Bernath, P., Biraud, S. C., Blavier, J.-F. L., Boone, C., Bowman, K. P., Browell, E. V., Campos, T., Connor, B. J., Daube, B. C., Deutscher, N. M., Diao, M., Elkins, J. W., Gerbig, C., Gottlieb, E., Griffith, D. W. T., Hurst, D. F., Jiménez, R., Keppel-Aleks, G., Kort, E. A., Macatangay, R., Machida, T., Matsueda, H., Moore, F., Morino, I., Park, S., Robinson, J., Roehl, C. M., Sawa, Y., Sherlock, V., Sweeney, C., Tanaka, T., and Zondlo, M. A.: Calibration of the Total Carbon Column Observing Network using aircraft profile data, Atmospheric Measurement Techniques, 3, 1351–1362, https://doi.org/10.5194/amt-3-1351-2010, http://www.atmos-meas-tech.net/3/1351/2010/, 2010.

Wunch, D., C Toon, G., L Blavier, J.-F., A Washenfelder, R., Notholt, J., J Connor, B., W T Griffith, D., Sherlock, V., and Wennberg, P.: The Total Carbon Column Observing Network, 369, 2087–112, 2011.

Wunch, D., Toon, G. C., Sherlock, V., Deutscher, N. M., Liu, X., Feist, D. G., and Wennberg, P. O.: The Total Carbon Column Observing Network's GGG2014 Data Version, 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., Kiel, M., Griffith, D. W. T., 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., 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., García, O. E., 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, 10, 2209–2238, https://doi.org/10.5194/amt-10-2209-2017, https://www.atmos-meas-tech.net/10/2209/2017/, 2017.

Ye, X., Lauvaux, T., Kort, E. A., Oda, T., Feng, S., Lin, J. C., Yang, E., and Wu, D.: Constraining fossil fuel CO_2 emissions from urban area using OCO-2 observations of total column CO_2, Atmospheric Chemistry and Physics Discussions, 2017, 1–30, https://doi.org/10.5194/acp-2017-1022, https://www.atmos-chem-phys-discuss.net/acp-2017-1022/, 2017.

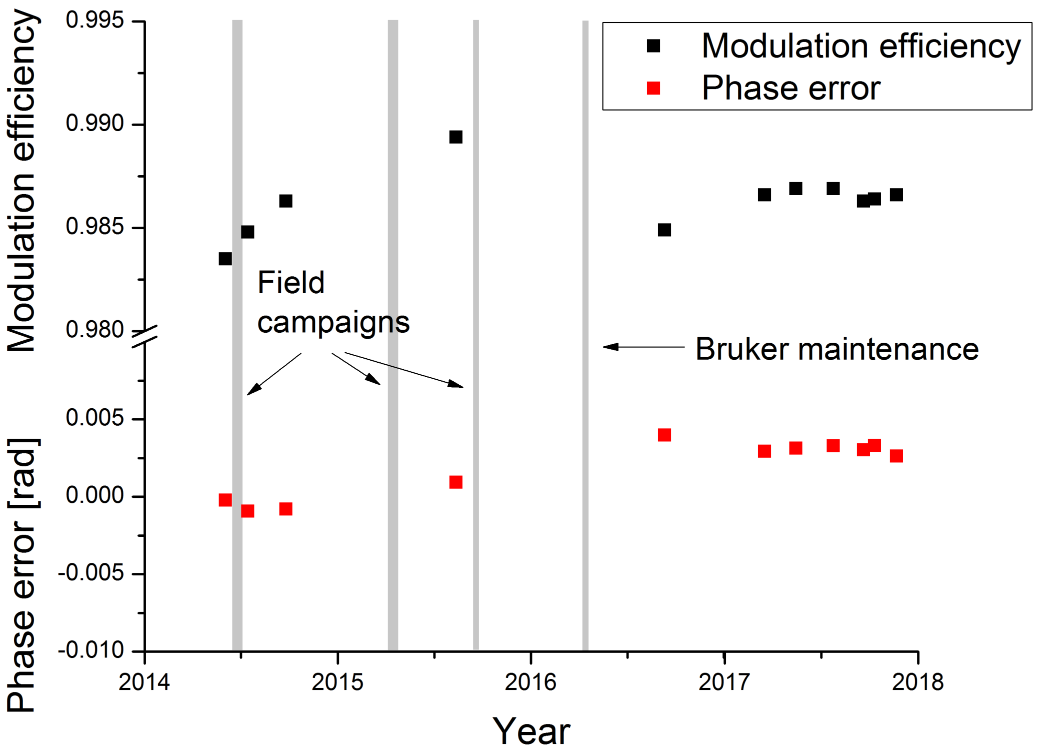

Figure 1. ILS time series of the reference EM27/SUN. Results for modulation efficiency and phase error were obtained with LINEFIT 14.5. The mean value of the modulation efficiency is 0.9862 with a standard deviation of 0.0015. For the phase error an average value of 0.0019 ± 0.0018 is observed. As can be seen from the closely spaced measurements in 2017, there is no seasonality in the ILS values. Grey areas denote periods of transportation of the instrument.

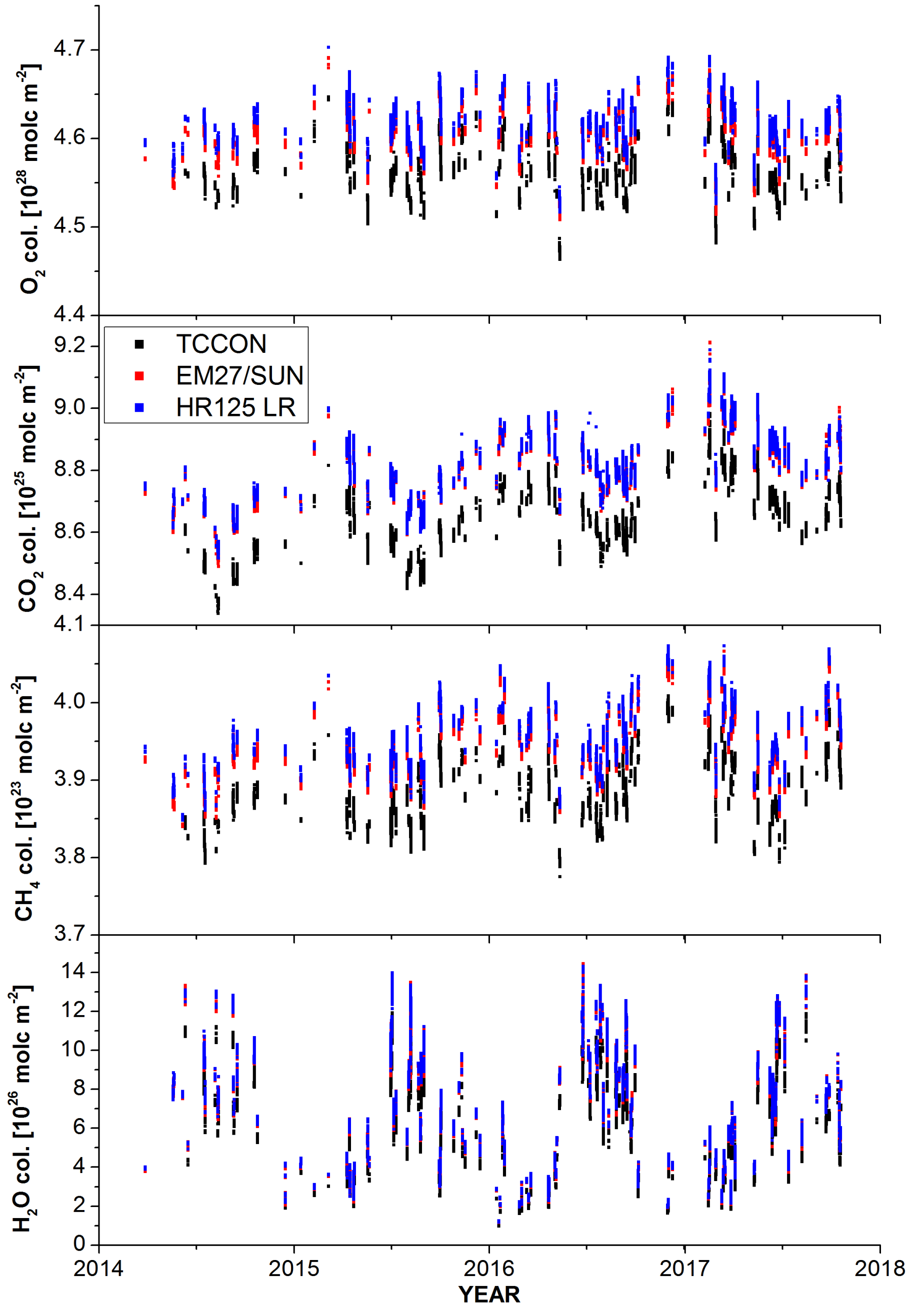

Figure 2. Total column time series for O_2, CO_2, CH_4 and H_2O measured at KIT in Karlsruhe from March 2014 until October 2017. The number of interferograms and recording time for the different data types are the following: TCCON: 2 IFGs, 114 s; EM27/SUN: 10 IFGs, 58 s; HR125 LR: 4 IFGs, 152 s. Only coincident measurement points (within one minute) are depicted.

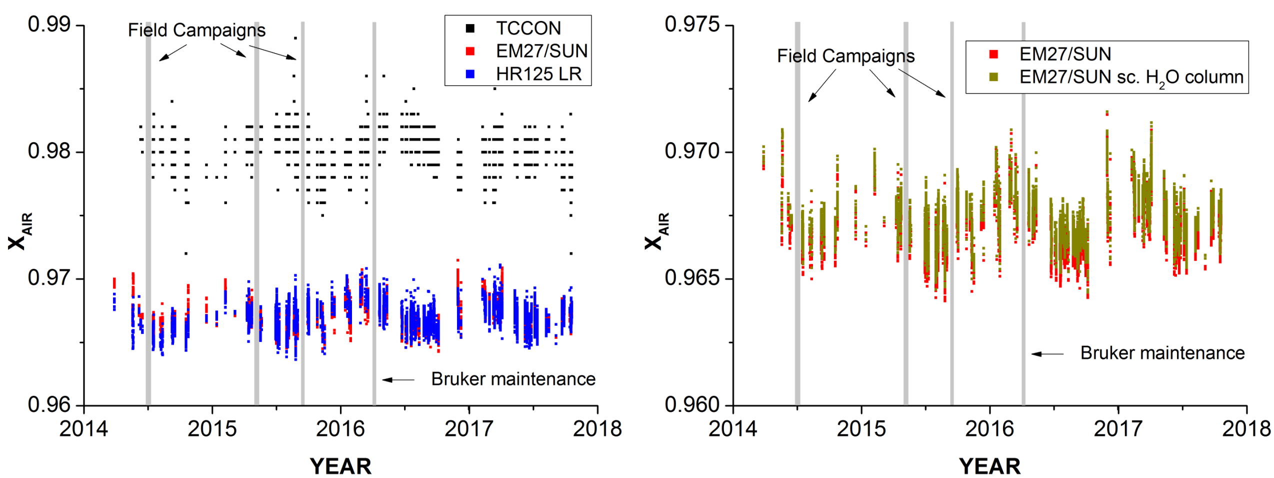

Figure 3. The left panel shows the X_{air} time series measured at KIT in Karlsruhe for the TCCON, EM27/SUN and HR125 LR data sets. For clarity, only coincident measurements (within one minute) of the data sets are plotted. Grey areas denote periods where the EM27/SUN was moved over long distances. The right panel shows a comparison of the original EM27/SUN time series with a modified version, where a scaling factor of 0.8 was applied to the H_2O total column.

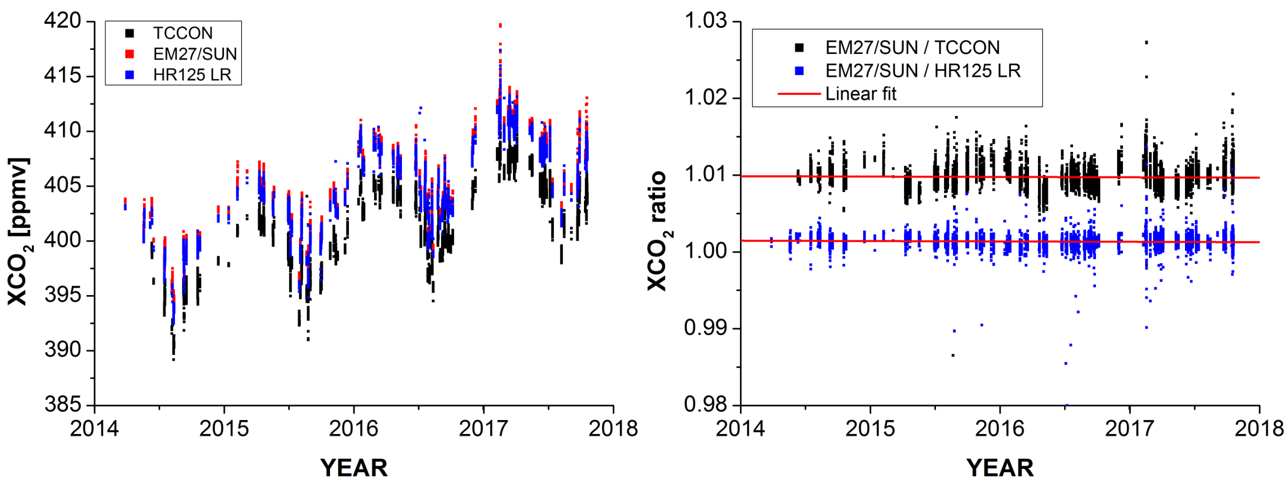

Figure 4. The left panel shows the XCO_2 time series measured at KIT in Karlsruhe for the three data sets from March 2014 to October 2017.~~Additionally the absolute offsets between the EM27/SUN and the two other data sets are shown.~~ For clarity, only coincident measurements (within one minute) of the data sets are plotted. The right panel shows the XCO_2 ratio between the EM27/SUN and the two HR125 data sets. A linear fit was applied to investigate a possible trend in the ratios.

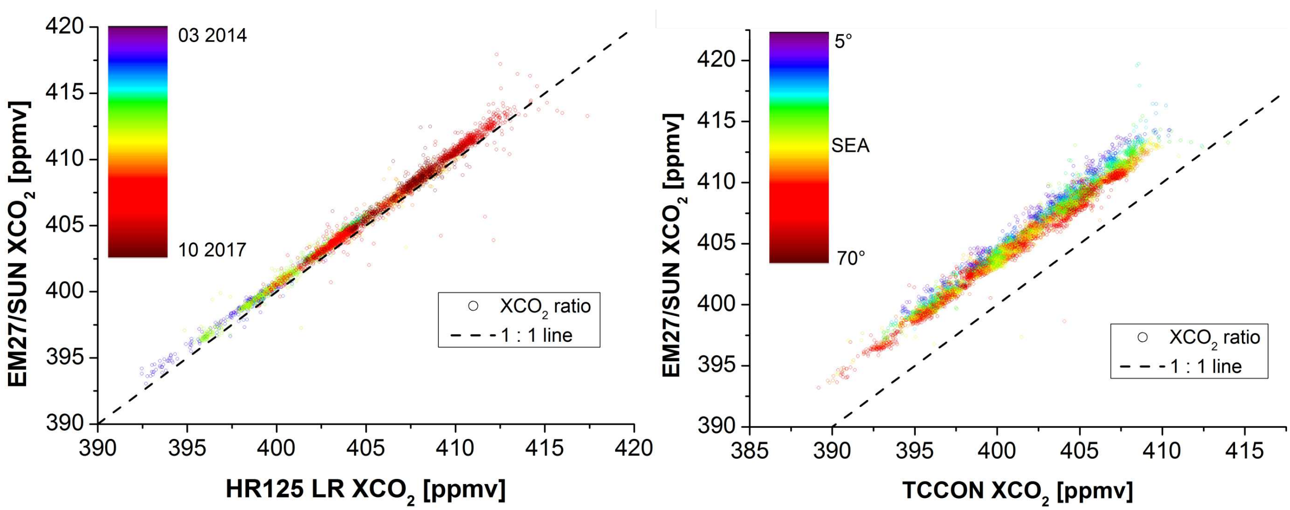

Figure 5. The left panel shows the XCO_2 comparison between EM27/SUN and HR125 LR. The colorbar denotes the date of the measurement, the dashed line is the $1:1$ line. In the right panel the comparison with TCCON is shown. Note that here the colorbar shows the solar elevation angle.

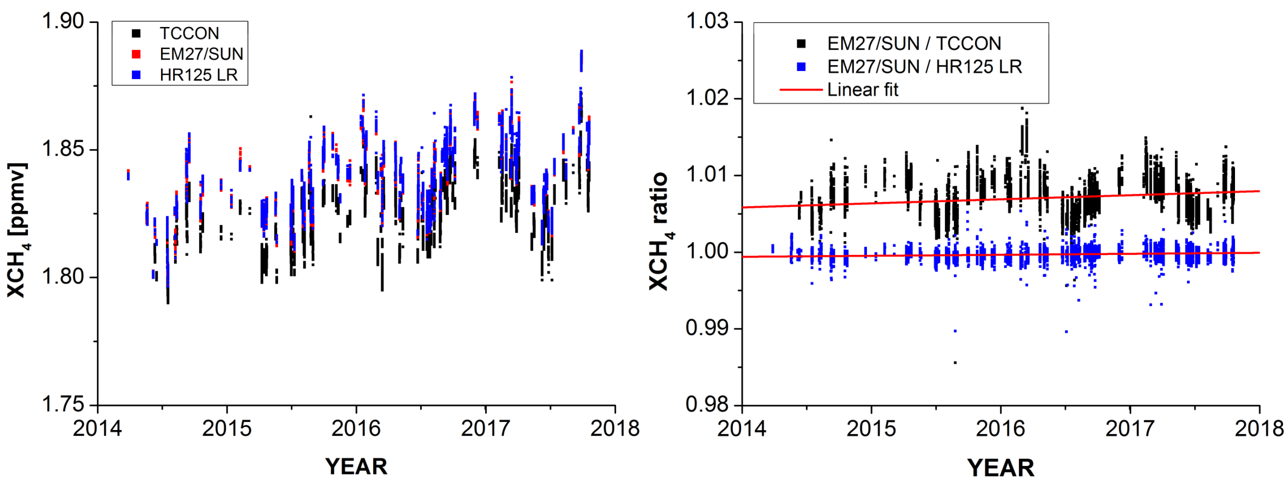

Figure 6. The left panel shows the XCH_4 time series measured at KIT in Karlsruhe for the three data sets from March 2014 to October 2017.~~Additionally the absolute offsets between the EM27/SUN and the two other data sets are shown.~~ For clarity, only coincident measurements (within one minute) of the data sets are plotted. The right panel shows the XCH_4 ratio between the EM27/SUN and the two HR125 data sets. A linear fit was applied to investigate a possible trend in the ratios.

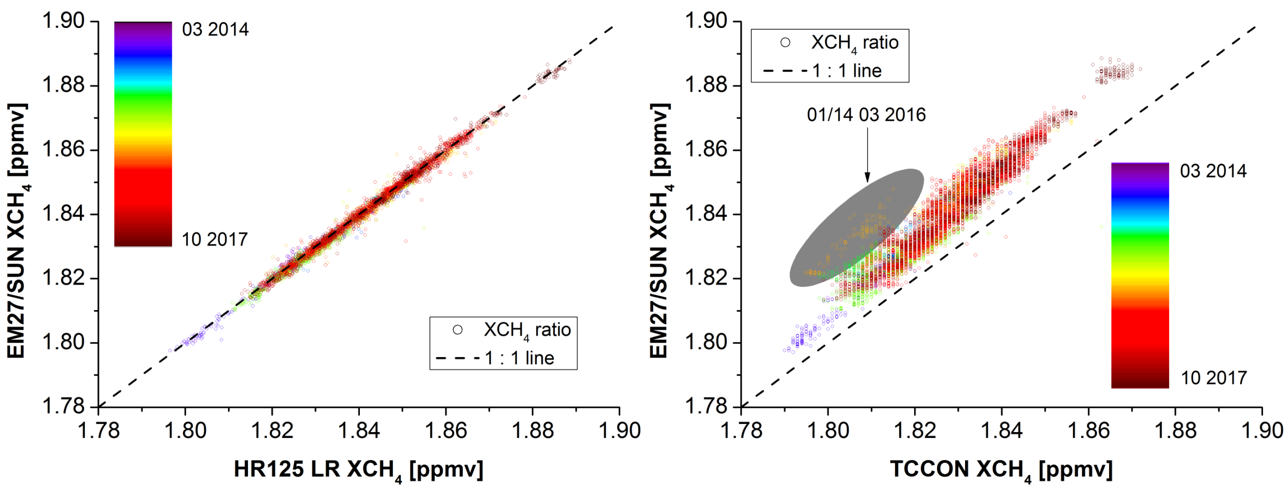

Figure 7. The left panel shows the $\mathrm{XCH_4}$ comparison between EM27/SUN and HR125 LR. The colorbar denotes the date of the measurement, the dashed line is the $1:1$ line. In the right panel the comparison with TCCON is shown. The shaded area encloses measurements from 01 and 14 March 2016. For these days the ratio is significantly different with respect to the remaining data set (see text for discussion).

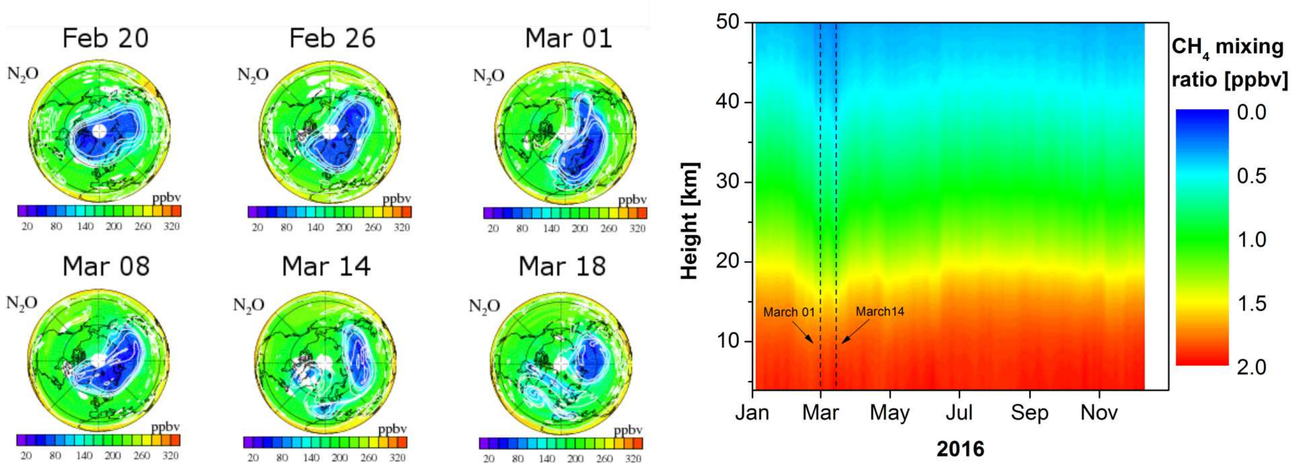

Figure 8. In the left panel N_2O MLS data from the Aura satellite is shown as a tracer for the position of the polar vortex for several days in February and March 2016. Data and plots courtesy of the NASA science team (https://mls.jpl.nasa.gov/). The right panel shows CH_4 mixing ratios from the NDACC FTIR station Jungfraujoch in Switzerland, downloaded from the NDACC archive (http://www.ndaccdemo.org/stations/jungfraujoch-switzerland/). For dates with no measurements the data has been interpolated using a weighted average. Dotted lines depict March 01 and 14 2016. For these dates, the XCH_4 data significantly differs from the remaining data set.

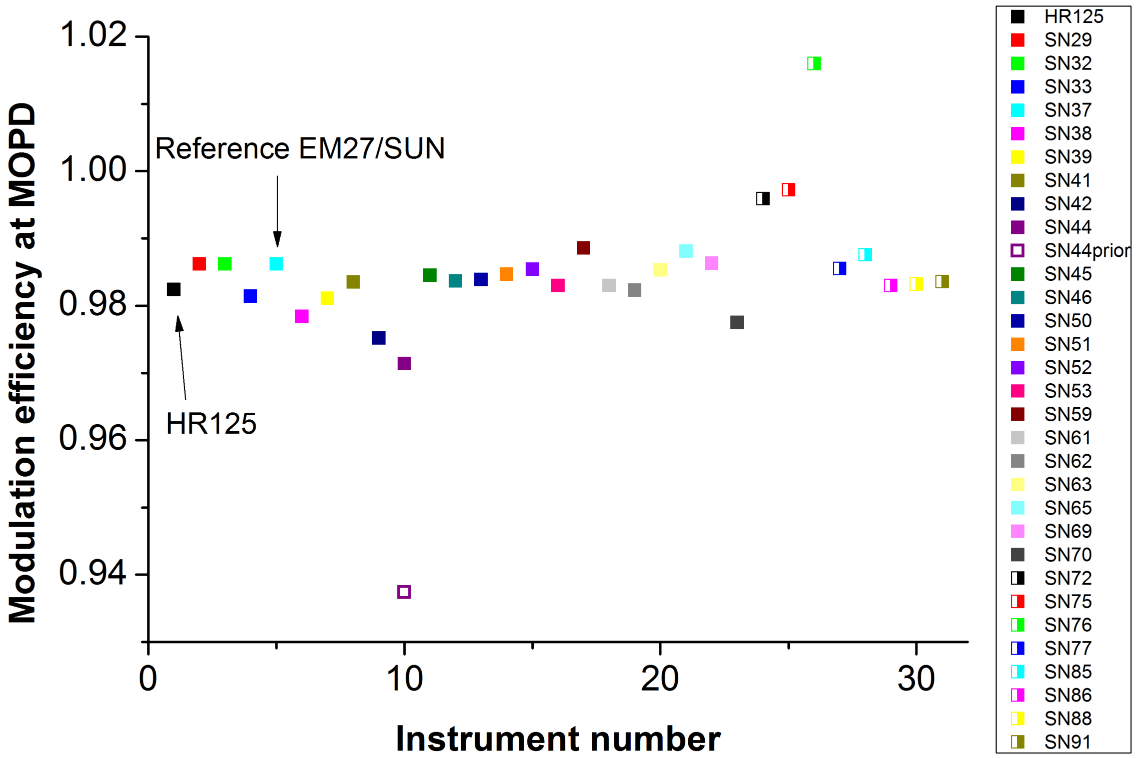

Figure 9. Modulation efficiencies at MOPD for all EM27/SUN ~~calibrated~~ tested in Karlsruhe. For SN44 prior, ILS measurements were taken before an alignment check and subsequent realignment of the instrument. For comparison reasons, also an ILS measurement for the HR125 was performed.

Figure 10. Calibration measurements performed on April 14 2015 on top of the KIT-IMK office building north of Karlsruhe.

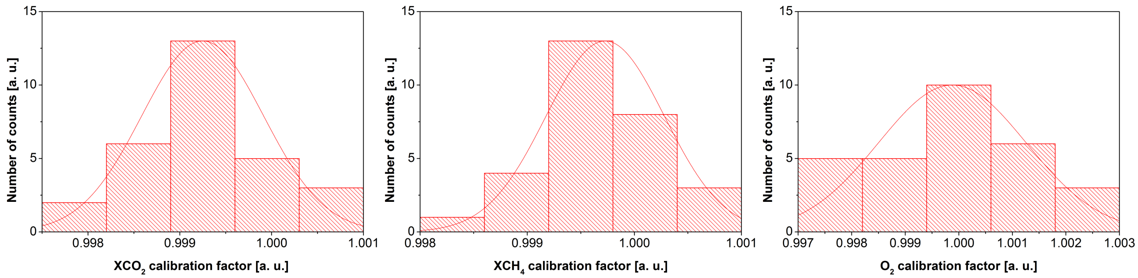

Figure 11. Histograms of the empirical XCO_2, XCH_4 and O_2 calibration factors for the different instruments with respect to the reference EM27/SUN. The red line overlying the histograms is a fit of a Gaussian function to the histogram. For the histograms, calibration measurements of 29 instruments were accumulated.

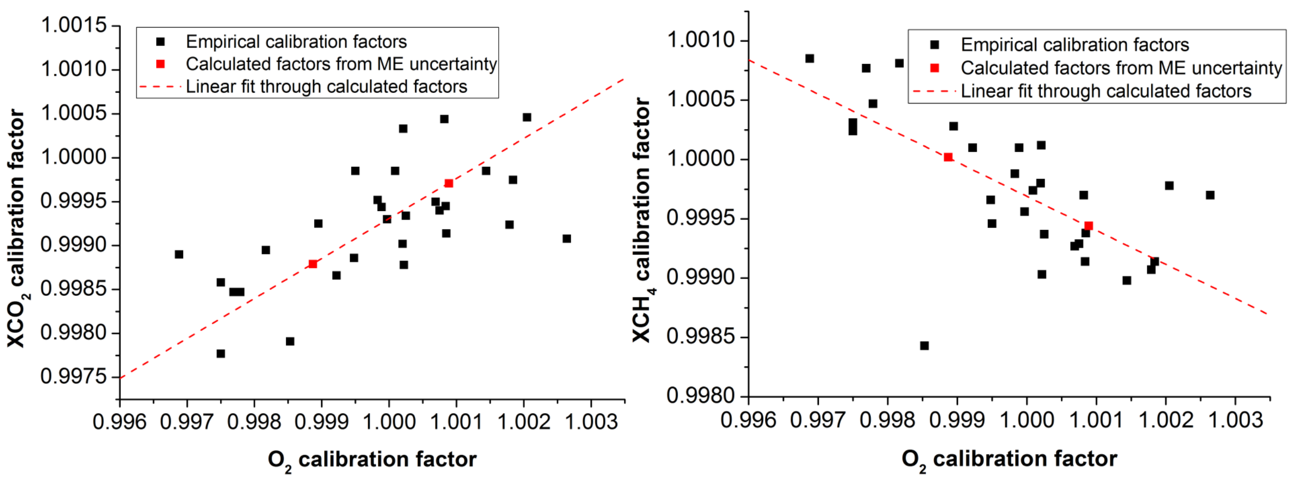

Figure 12. Correlation of O_2 calibration factors and XCO_2 (left panel) as well as XCH_4 (right panel) calibration factors. Black squares show the empirical calibration factors from the side-by-side measurements, red squares show calculated factors derived from the total ME uncertainty shown in Table 1, the dashed red line is a linear fit through the calculated factors. The slope of empirical and calculated factors is in good agreement.

Table 1. Estimated ME uncertainties for various error sources.

Error source	Uncertainty	Propagation on ME
Temperature	$\pm$0.8 K	-$\pm$0.16 %
Total pressure	$\pm$3 mbar	$\pm$0.19 %
Distance	$\pm$5 cm	-$\pm$0.04 %
Partial pressure H_2O	$\pm$0.5 mbar	$\pm$0.13 %
Measurement noise		$\pm$0.05 %
Total		$\pm$0.29 %

Table 2. Sensitivity study on the effect of ILS changes on the retrieval of the total gas columns. Depicted is hourly pooled data on 01 August 2016 and 15 February 2017 around solar noon, corresponding to a solar elevation angle of 60° and 30°. The resulting ILS dependency of XCO_2 is -0.16 % and -0.19 % for 60° and 30° SEA, for a 1 % ME increase. XCH_4 increases by 0.10 % (0.09 %).

ME	O_2 [10^{28} molc m^{-2}]	H_2O [10^{26} molc m^{-2}]	CH_4 [10^{23} molc m^{-2}]	CO_2 [10^{25} molc m^{-2}]
August 2016				
0.99	4.6097	7.4551	3.9457	8.7321
1.00	4.5936	7.4323	3.9356	8.6879
February 2017				
0.99	4.6718	3.7746	4.0261	9.0968
1.00	4.6545	3.7628	4.0148	9.0455

Table 3. XCO_2 biases between EM27/SUN and HR125 data sets.

XCO_2 ratio	No. coincidences	Mean (1σ)	Yearly trend in the ratio
EM27 / TCCON	8349	1.0098 (0.0015)	-5×10^{-5}
EM27 / HR125 LR	4624	1.0014 (0.0011)	-5×10^{-5}

Table 4. XCH_4 biases between EM27/SUN and HR125 data sets.

XCH_4 ratio	No. coincidences	Mean (1σ)	Yearly trend in the ratio
EM27 / TCCON	8349	1.0072 (0.0024)	0.0005
EM27 / HR125 LR	4624	0.9997 (0.0008)	0.0001

Table 5. Summary of the modulation efficiencies at MOPD and phase errors for all EM27/SUN calibrated in Karlsruhe. "ref" denotes the reference EM27/SUN and "prior" denotes an ILS measurement with instrument SN44 prior to calibration at KIT.

Instrument SN	ME at MOPD	Phase error [rad]
29	0.9862	0.0014
32	0.9862	0.0034
33	0.9814	-0.0017
37 (ref)	0.9862	0.0019
38	0.9784	0.0009
39	0.9811	-0.0005
41	0.9835	0.0001
42	0.9752	0.0039
44	0.9714	-0.0019
44 (prior)	0.9374	-0.0074
45	0.9845	0.0034
46	0.9837	0.0024
50	0.9839	0.0023
51	0.9847	0.0017
52	0.9854	0.0048
53	0.9830	0.0025
59	0.9886	0.0029
61	0.9830	0.0013
62	0.9823	0.0053
63	0.9853	0.0011
65	0.9881	0.0024
69	0.9863	0.0030
70	0.9775	0.0056
72	0.9959	0.0030
75	0.9972	0.0041
76	1.0160	0.0007
77	0.9855	0.0016
85	0.9876	0.0025
86	0.9830	0.0031
88	0.9832	0.0007
91	0.9836	0.0021

Table 6. Calibration factors for XCO_2, XCH_4 and O_2 for all investigated instruments with respect to the reference EM27/SUN spectrometer (SN37) as well as calibration dates and number of coincident measurements. Values in brackets denote percent standard deviations.

Instr. SN	Dates	No. co.	XCO_2 factor	XCH_4 factor	O_2 factor
29	140606, 140718	490	1.0004 (0.02)	0.9997 (0.03)	1.0008 (0.03)
32	150414 - 150422	1548	0.9997 (0.03)	0.9997 (0.03)	1.0004 (0.03)
33	170807, 170815	339	0.9991 (0.03)	0.9994 (0.04)	1.0009 (0.05)
38	150410 - 150421, 160121	1609	0.9989 (0.03)	0.9997 (0.04)	0.9988 (0.04)
39	140717, 150414, 150415	1210	0.9992 (0.04)	0.9994 (0.04)	1.0003 (0.04)
41	140717, 150414 - 150422	1877	0.9999 (0.03)	1.0002 (0.03)	0.9991 (0.03)
42	160730, 160801	368	0.9978 (0.04)	1.0003 (0.04)	0.9975 (0.03)
44	170227	286	0.9979 (0.03)	0.9984 (0.03)	0.9985 (0.03)
45	170807, 170815	382	0.9995 (0.03)	0.9991 (0.04)	1.0008 (0.02)
46	170808, 170815	503	0.9993 (0.03)	0.9994 (0.03)	1.0003 (0.03)
50	150421, 150422	699	0.9999 (0.03)	0.9995 (0.03)	0.9995 (0.03)
51	160126, 160129	256	0.9995 (0.03)	0.9993 (0.03)	1.0007 (0.05)
52	150421, 150422	727	0.9990 (0.04)	0.9998 (0.05)	1.0002 (0.05)
53	150421, 150422	729	0.9987 (0.03)	1.0001 (0.03)	0.9992 (0.04)
59	160318	273	0.9998 (0.03)	0.9991 (0.03)	1.0019 (0.04)
61	151002, 170713	618	0.9993 (0.03)	0.9996 (0.04)	1.0000 (0.04)
62	160121	18	0.9988 (0.04)	0.9990 (0.02)	1.0002 (0.02)
63	160121	15	1.0003 (0.05)	1.0001 (0.05)	1.0002 (0.07)
65	160511	234	1.0005 (0.04)	0.9998 (0.05)	1.0020 (0.03)
69	160908, 170713	636	0.9994 (0.03)	0.9993 (0.03)	1.0008 (0.03)
70	160831, 160906	522	0.9985 (0.02)	1.0005 (0.03)	0.9978 (0.03)
72	170215, 170216	433	0.9994 (0.05)	1.0001 (0.03)	0.9999 (0.04)
75	170516, 170517	852	0.9993 (0.03)	0.9991 (0.03)	1.0018 (0.05)
76	170608	365	0.9991 (0.04)	0.9997 (0.04)	1.0026 (0.06)
77	170927	389	0.9999 (0.03)	0.9997 (0.03)	1.0001 (0.04)
85	180213, 180214	371	0.9993 (0.03)	1.0003 (0.03)	0.9990 (0.03)
86	180213, 180214	464	0.9986 (0.03)	1.0002 (0.03)	0.9975 (0.05)
88	180314	154	0.9990 (0.03)	1.0008 (0.03)	0.9982 (0.03)
91	180228	148	0.9985 (0.03)	1.0008 (0.03)	0.9977 (0.04)