We would like to thank our anonymous referee#2 for the valuable comments on our manuscript. We have answered point-by-point each of the questions raised by the referee as follows: the original referee's inputs are shown in blue text, while our answer is in black.

The network COCCON focusses on column measurements of greenhouse gases (GHGs). Sub-percent variations are of interest for the GHG columns. Hence the measurements need to resolve these small variations and small biases between different instruments could lead to erroneous conclusions. For this reason the calibration of the instruments is of high importance. The manuscript describes very detailed how the calibrations are performed and compares the new and old calibration methods on the technical level (e.g. modulation efficiency amplitudes). This is certainly very important for documenting the COCCON calibration and also for readers working with FTIR spectrometry. However, currently the paper does not discuss the impact of the improved calibration on the GHG retrievals.

The propagation of ILS errors into the GHG retrievals has been discussed in depth in the PhD thesis of Qiansi Tu, which is available here: https://publikationen.bibliothek.kit.edu/1000095901. We have added this reference to the paper and added in the introduction the statement "The effect of ILS parameter errors on the retrieved column-averaged abundances has been discussed in detail by Q. Tu, 2019 (see chapter 3.4 and figure 3.6 in this work)."

In my opinion it would be good to address the following questions: Does the improved calibration improve the consistency within COCCON and if yes, how much for each gas. I understand that there might be no improvement, because of the instrument specific correction factors. In that case it should be discussed how much the improved calibration procedure influence the instrument specific correction factors.

This is a very interesting question (and a very demanding task). Unfortunately, we do not know the true atmospheric state and the best available reference measurements (provided by the collocated TCCON spectrometer and the COCCON reference) are exploited for the determination of the empirical instrument-specific calibration factors for each gas.

However, there is still a possible argument supporting the assumption that the refined analysis is superior. Ideally, the explicit description of the instrumental deviations from the nominal would remove any discrepancy between the spectrometer under test and the reference spectrometer. In this case, gas-specific calibration factors would not be needed at all (they all would adopt the same value across different spectrometers). This opens up a way to test for progress in the calibration procedures, even if we do not know the true state in the atmosphere. For detecting an improvement, XCO_2 is the best candidate. (The quality of the XCO calibration is limited by other factors not treated in this work. A main impact factor on XCO is the presence of weak channelling in the spectra, because CO is a very weak absorber. In case of XCH_4, we unfortunately face the drift of the reference spectrometer during the early years (so affecting the XCH_4 calibration results by Frey et al. (2019))). The table below provides the scatter of the gas-specific calibration factor for XCO_2 following each recipe (assuming ILS as: nominal, the previous Frey et al. (2019) results and by using the results obtained in this study). The numbers indicate that both methods deliver more consistent calibration factors than the "nominal ILS" assumption and the XCO_2 test even suggests an advantage of the refined calibration approach.

Procedure	Empirical STDEV of XCO_2 calibration factor between different spectrometers
Nominal ILS	9.49839E-4
ILS results from Frey et al, 2019	8.56409E-4
ILS results from this work	7.16057E-4

With respect to atmospheric impact factors, water content should be of secondary importance. The atmospheric humidity and temperature might affect slightly the GHG retrieval due to various inaccuracies of the model description, which connects the atmospheric state to a modelled spectrum. However, the side-by-side measurements of the spectrometer under test and the reference unit are recorded under identical atmospheric conditions covering the same airmass range, so the small impact of water vapour or temperature should cancel out.

A separate question is whether the parameterisation of the residual mismatch between two spectrometers as a gas-specific calibration factor is adequate. As we do not know the mechanism triggering the discrepancy, this is hard to answer. If the reason of the discrepancy is, e.g., mainly a spectral baseline offset, and then our approach of using a calibration factor might be suboptimal.

From the instrumental viewpoint, observed air mass is the most obvious variable, which might influence the calibration factor (as it affects the amount of radiation accepted by the spectrometer). The figure below shows the XCO_2 time series of a side-by-side measurement, shown as function of solar zenith angle SZA. As can be seen, the approach of using a single calibration factor seems to work reasonably well.
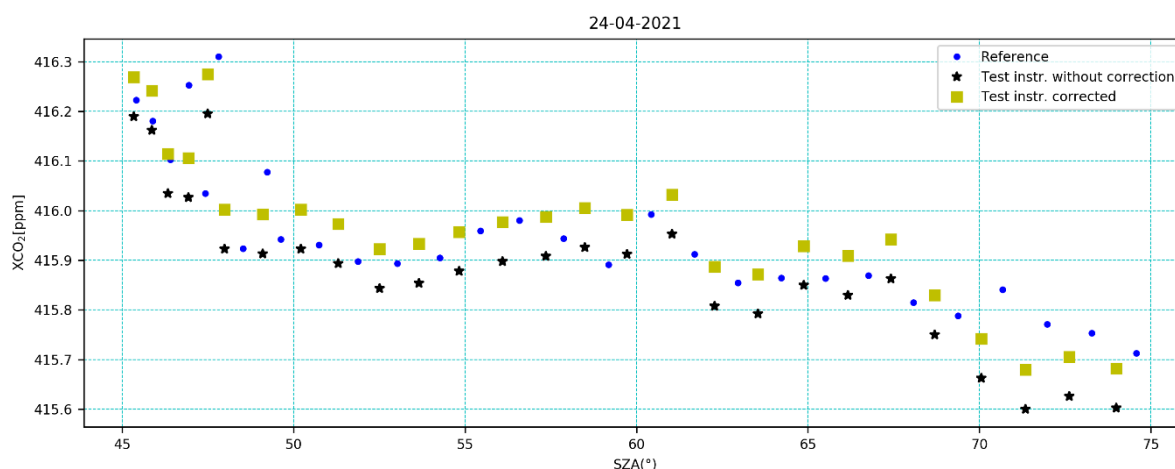

Figure 1: XCO_2 of reference and spectrometer under test, as function of SZA, show uncorrected and corrected XCO_2 of test spectrometer

Finally, again from the instrumental viewpoint, we need to address your question whether the EM27/SUN spectrometer shows some sensitivity to local temperature. Possibly the best handle on this is the retrieved XAIR. The reference spectrometer is operated outside and experiences ambient temperatures in the range 0°C to 35°C during the year.

While the PROFFAST Ver. 1 analysis showed a little annual cycle in XAIR values (this is seen in the lowest panel of figure 19), this spurious effect is gone using the latest HITRAN spectroscopy for the retrievals (PROFFAST ver. 2, this new code is currently in beta release). The annual variation of XAIR (if there is any) is within ± 0.001 according to PROFFAST Ver. 2. This is an excellent level of stability,

which does not leave much room for instrumental characteristics depending on temperature. The figure below shows the PROFFAST ver. 2 XAIR results for the whole year 2019 as function of SZA.
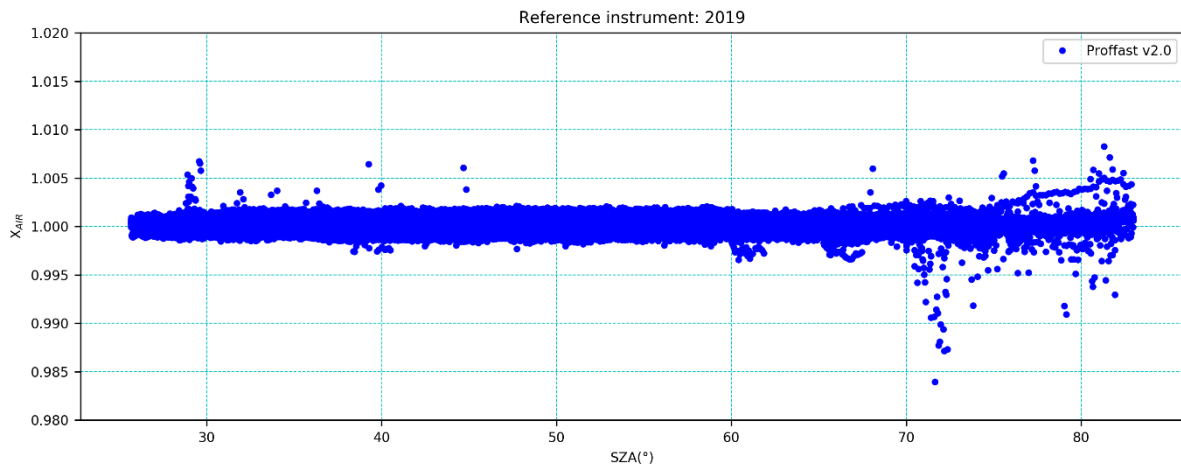

Figure 2: XAir results obtained with Proffast version 2.0 for the reference unit during 2019.

Technical comment: please note that we meanwhile discovered that our revised results for spectrometers SN29, SN32, SN50, SN52 and SN53 were suspicious outliers in the resulting instrument specific gas calibration factors when compared to the previous results by Frey et al. (2019). We identified the reason of this problem (we used outdated and not the latest sets of laboratory spectra available for those spectrometers as used by Frey et al. (2019)). We repeated the analysis for these spectrometers and updated accordingly all dependent figures and tables in the paper.