Authors' Response to Reviews of

Using portable low-resolution spectrometers to evaluate TC-CON biases in North America

Nasrin Mostafavi Pak, Jacob Hedelius, Sébastien Roche, Liz Cunningham, Bianca Baier, Colm Sweeney, Coleen Roehl, Joshua Laughner, Geoffrey Toon, Paul Wennberg, Harrison Parker, Colin Arrowsmith, Joseph Mendonca, Pierre Fogal, Tyler Wizenberg, Beatriz Herrera , Kimberly Strong, Kaley A. Walker, Paul Wennberg, Felix Vogel, Debra Wunch
Atmospheric Measurement Techniques,

RC: ***Reviewers' Comment***, AR: Authors' Response, ☐ Manuscript Text

We would like to thank the reviewer for their thoughtful comments that have helped to improved our manuscript substantially. The detailed response and corrections are given below.

Reviewer #1

General Comments

1. General Comment #1

RC: ***In Section 2.3, you are dealing with ILS measurements. In Figure 2 you show the ILS measurements of the TCCON stations but you are not classifying them (are they good/bad?). They are not used for a further data analysis, I do not see a strong need to show them here. Is there an acceptance level? The variation of 1% found for the fluctuations in EM27/SUN modulation efficiency (ascribed to variable humidity and room temperature) seems way too high, Alberti et al., 2022 demonstrated a significantly better reproducibility (see Fig. 15 in this work). It is important to take the variable partial pressure of H2O into account when doing the analysis (which can be calculated from total pressure, path length and H2O column) and to do the analysis repeatedly until a self-consistent solution (for ILS, column, and partial pressure) is found. If this has been taken into account, I would suspect that there was a problem in coupling the light source to the spectrometer in a reproducible manner.***

AR: ILS results from TCCON and EM27/SUNs are used as a qualitative measure to ensure the stability of the instruments over the course of the road trip and ILS values are not currently used in the retrieval procedure. Regarding ILS measurements for TCCON, guidelines require that the modulation efficiency deviates less than 5% from 1.0, over 0 to 45 cm OPD. which is true for all the TCCON sites during our visit.

Regarding the EM27/SUN ILS measurements, we have followed the same procedure using partial pressure of H2O. We agree that the variability observed in Alberti et al., 2022 is significantly smaller, however there are multiple factors that could have caused a larger range for our measurements. In Alberti et al, all measurements are performed in one room using a single distance between the lamp and the FTS mirror. However during our road trip, we have performed ILS measurements in three different regions in North America with different laboratory conditions. We have also used 3 different lengths between the lamp and the FTS mirror and took the average of the 3 trials. This could contribute to the variability in the calculated ME. Further, these instruments underwent thousands of kilometers of travel, compared to the KIT reference instrument which by design is not subjugated to as significant of movements. This movement could affect the ILS of the instruments.

We have added additional explanation in this section to clarify these points:

> Instrumental Line Shape (ILS) is a measure of the optical alignment of the instrument and imperfections in this alignment can cause biases in the retrievals. The ILS of an FTS can be described by two parameters: Phase Error (PE) and Modulation Efficiency (ME). For high resolution FTS instruments (TCCON) ILS is typically reported as function of optical path difference (OPD) and for the low resolution EM27/SUNs these are typically reported at the maximum optical path difference. ILS values are not implemented into the GGG2014 and GGG2020 retrieval algorithms and are only used to evaluate the instruments' alignment qualitatively. TCCON guidelines require that the modulation efficiency deviates less than 5% from 1.0, over 0 to 45 cm OPD. A modulation loss of 1% in the EM27/SUN causes a bias of 0.1% in XCO_2 and 0.15% in XCH_4 (Hedelius et al., 2016), therefore we aim to ensure that the EM27/SUN ME variations remain less than 1%.
>
> Because the EM27/SUNs were moved from one site to another, we evaluated their optical alignment by measuring the ILS of all the EM27/SUNs at various sites (Table 1). For EM27/SUNs, we use a method introduced by Frey et al. (2015). and further developed by Alberti et al. (2022), in which we collect spectra from an external lamp (Quartz Tungsten-Halogen Lamp, Thorlabs) in the laboratory and use LINEFIT (version 14.0) to derive the ILS parameters from H_2O lines in the 7000 and 7400 cm $^{-1}$ spectral region (Frey et al., 2015). Our method differs slightly because we use three different distances between the lamp and the EM27/SUN, and we compute the standard deviation of the calculated ME across the three distances to evaluate the variability in the calculated ME.

2. General Comment #2

RC: ***You derive the airmass-depended correction factors by comparing the measurements in the course of the day to the daily median, thereby assuming the variation is solely due to an airmass dependency. I am in doubt if this assumption is valid since it ignores intraday variability of GHGs, which, especially in a rural area like Toronto, seem reasonable for me to occur.***

AR: Although intraday variability in CO_2 is expected to some extent in urban areas such as Toronto due to peak traffic hours in the mornings and the afternoons, it's not regularly observed in our daily measurements. Based on measurements in multiple locations in Toronto, positive daily anomalies of larger than the standard deviation (0.6 ppm) occur mostly around solar zenith angles of 40–50 and around 14-16 UTC that is equivalent to 9 am- 11 am local time and are not symmetric around noon (See left panel of Figure 1 below). Therefore wouldn't interfere with the negative anomalies that occur at high solar zenith angles and are symmetric around noon (See right panel of Figure 1 below) and didn't have a significant effect on the final ADCF calculations. We collect fewer measurements in the early morning than in the late afternoon which skews the distribution. As can be seen in Figure 2, after applying the airmass correction, the negative anomalies have been eliminated effectively (right panel of Figure 2) and the enhancements due to urban emissions are not affected as strongly.

Figure 1: Before airmass correction

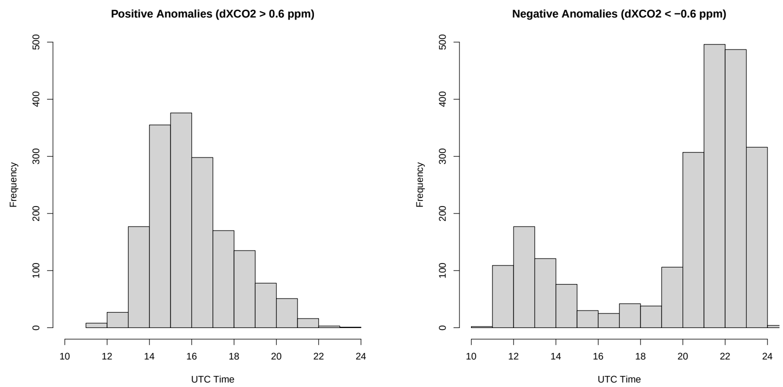
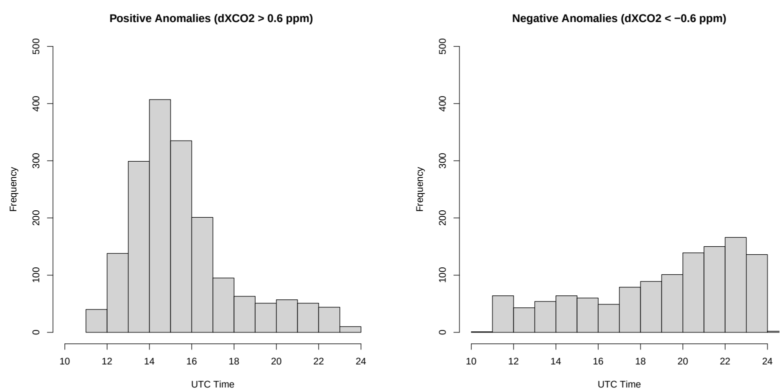


Figure 2: After airmass correction



We have added additional explanation in the text to clarify this point:

> To derive the airmass-dependent correction factors (ADCFs) for the EM27/SUNs, we use the long term record of measurements in Toronto from 2018 to 2021 with four EM27/SUNs ("ta", "tb", "tc" and "td") to calculate an average ADCF value for each gas. Although Toronto is under the influence of urban emissions, our analysis showed that the enhancements due to traffic emissions are not symmetric around noon and therefore would not interfere with the airmass dependent calculations.

3. General Comment #3

RC: ***Next you are writing in line 281 you are using the same method as GGG2014 to apply the correction factors. However, there is missing an explanation or at least a specific citation on how this is done in GGG2014.***

AR: We have made necessary changes in the text to make this more clear:

> TCCON applies ADCFs for GGG2020 differently than in GGG2014. In the TCCON post-processing procedure, ADCF values are calculated for each retrieval window *before* averaging them (Laughner et al., 2020), whereas in GGG2014, ADCFs are calculated and applied for each gas after averaging different retrieval windows. Additionally, in GGG2020, θ_0 and p vary from window to window in order to best capture the airmass dependence, whereas in GGG2014 all gases used $\theta_0 = 13°$ and $p = 3$. For EM27/SUNs, we follow the same method as in GGG2014 and derive and apply ADCFs for each gas *after* averaging individual windows. It is therefore not possible to directly compare TCCON and EM27/SUN ADCF values for gases with multiple windows (i.e., CO_2 and CH_4); it is expected that they will be different.

We have also added the formula used to apply the airmass dependance correction:

$$y_c = \frac{y_i}{[1 + \beta S(\theta_i)]} \tag{1}$$

> where β is the airmass dependent correction factor (ADCF) and y_c is the airmass corrected Xgas value.

4. General Comment #4

RC: ***You are using the median for the comparison of the 10-minute bins of the different spectrometer but you are using the mean value to calculate the 10-minute bins. What are the reasons for choosing the mean or the median for the different situations?***

AR: Over 10 minutes we expect the data from each instrument to be smooth, therefore the average and the median are not significantly different. To compare the two instruments, we use median over the entire data set to eliminate the possible outliers in the biases. The results wouldn't be greatly altered if we used the average.

5. General Comment #5

RC: ***In chapter 3.1.2 you are describing how the maximum biases are calculated. However, this procedure is not clear to me. I understand you are taking the 10-minutes bins of the different spectrometers and then calculate the difference of the minimal and maximal bin each, even though they are not temporal coincident. This however, would include the variation of the XGas value to the maximum bias. Please clarify what is done there.***

AR: The aim of this section is to estimate biases between pair of EM27/SUNs at different measurement locations and compare them as we move from one site to the other to ensure it remains consistent. We agree that presenting the absolute bias is not as critical as the changes in the bias over time. Therefore we updated the text and the table 4 to present the changes in the biases. We took the minimum bias and maximum bias in each pair and took the difference to account for the worst case scenario, ie. what would be the maximum variation in the biases and show even in the worst case the variations in the bias are smaller than the medium absolute deviation calculated for the average biases calculated in each case.

6. General Comment #6

RC: ***Lastly, you are writing in the introduction that you were taking a pressure sensor to the road trip to compare with the pressure measurement done on site. However, no detailed comparison is included in the paper. It would be nice to at least say a few words to the pressure comparison or better to show some results (maybe in a table?).***

AR: The digiquartz pressure sensor accompanied the EM27/SUNs and performed the measurements at the same height as the EM27/SUNs. As mentioned in Line 169, the slight differences in height of the EM27/SUNs and 125HRs in some of the TCCON sites led to minor differences in pressure. The offset was calculated based on the bias between the digiquartz pressure sensor and TCCON's pressure sensors that were previously calibrated against a similar pressure sensor. Therefore, we don't see the need to present the pressure comparison other than the cases we observed an offset due to height difference. We have also added an appendix to present the comparisons between the Digiquartz standard and TCCON pressure sensors at each site.

We made some changes to the corresponding paragraph to make it more clear how the standard pressure sensor was used.

> For the EM27/SUN retrievals, we use the pressure measurements recorded by the local TCCON weather stations which have already been calibrated against a Digiquartz pressure sensor, with the exception of Park Falls where we used the Vaisala WXT536 pressure data for both TCCON and EM27/SUNs, since the pressure measurements made by the TCCON pressure sensor were not stable during the campaign. In addition, at AFRC and Lamont we applied additional corrections to the EM27/SUN pressures as they were deployed at a slightly different altitude than the 125HR tracking mirror at the TCCON site. In these cases, we used the Digiquartz sensor pressure standard that was measuring at the same level as the EM27/SUNs to calculate the difference in surface pressure and added an offset of +0.1 hPa at AFRC and +0.3 hPa at Lamont to the original pressure value.

Specific comments

RC: ***L48: I am not fully aware of TCCON doing calibration of surface pressure measurements. If they do, please provide a citation or an explanation how this is done.***

AR: As part of the TCCON guidelines the primary investigators at each station sites calibrate the pressure sensors against a pressure standard. There is an additive offset added to the measured pressures before running the retrievals. The pressure calibration data are not included in a publication.

RC: ***L93: The abbreviations used for the EM27/SUNs and the TCCON station seem to be chosen randomly (e.g. it is unclear to me why the Armstrong Flight Research Center is abbreviated with "df"). Furthermore, for a reader it is quite confusing which abbreviation is a TCCON station and which is an EM27/SUN. Maybe add TCCON-xx to the TCCON sites or vice versa.***

AR: We recognise that there can be misunderstandings if the reader is unfamiliar with the two-letter TCCON station ID naming convention. We carefully reviewed the manuscript and made sure we clearly point out explicitly if we are referring to an EM27/SUN or a TCCON site.

RC: ***Table 1: It would help to reduce the reader's confusion with the abbreviations if you would add the abbreviations of the TCCON sites in the "Site" column.***

AR: We made the changes in Table 1 as suggested:

Table 1: This table lists the TCCON site locations and the dates the EM27/SUNs were on-site, the number of days with successful measurements, the average number of spectra collected by each EM27/SUN, and the total number of spectra collected by the TCCON instrument during the visit. The ILS column indicates whether EM27/SUN ILS measurements were collected at that location. The number of AirCore launches performed near the TCCON station during the dates listed is included in the final column. *instrument in brackets was not operational

Site	Latitude °N	Longitude °W	Elevation (masl)	Dates	Days	EM27/SUNs*	Spectra count EM27/SUN	Spectra count TCCON	ILS	AirCore Launches
Caltech (ci)	34.136	118.127	237	2018-07-06 – 2018-07-12	7	ta,tb,tc,dn	21356	1268	Yes	-
AFRC (df)	34.960	117.881	699	2018-07-13 – 2018-07-19	7	ta,tb,tc,dn	22667	2177	No	6
Lamont (oc)	36.605	97.486	320	2018-07-21 – 2018-07-29	5	ta,tb,tc,dn	17744	921	Yes	9
Park Falls (pa)	45.945	90.273	442	2018-07-31 – 2018-08-07	4	(ta),tb,tc	4436	406	No	4
East Trout Lake (et)	54.354	104.987	517	2018-08-09 – 2018-08-18	6	ta,tb,tc	14910	770	Yes	-
Eureka (eu)	80.053	86.417	610	2020-03-04 – 2020-08-31	61	tb	131713	5166	No	-

RC: ***L169 -- L171: I am not sure if I understand correctly what you are doing. This is what I understood: You are using the Digiquartz data as a „standard" measuring at height a. For AFRC and Lamont you are adding a factor to bring the pressure of the TCCON station to the level of the "standard" to correct for height. Have you ever compared the pressure measurements of the "standard" with the pressure of the TCCON station? Because otherwise, you cannot be sure if the correction is only due to height or also compensating sensor biases.***

AR: We are assuming TCCON pressure sensors are already calibrated. We also tend not to make any changes to TCCON data that is used by a lot of users. Digiquartz is used to ensure the consistency which was observed at all sites except the differences due to the differences in height of measurements which was consistent with theoretical calculation.

RC: ***L290: In the list of citations Alberti et al. 2022 would be good do mention, too.***

AR: We made the changes in Table 1 as suggested:

> Small biases in Xgas are expected to exist between the EM27/SUNs due to the differences in the instrument alignment (i.e., ILS) (Alberti et al., 2022). As long as these biases remain constant in time, including after shipping, a simple additive correction can place all the EM27/SUNs onto the same scale .

RC: ***Caption Figure 6: I was confused by this plot first, since I thought you are comparing something with the TCCON data and not that you only recorded the measurements at these sides. Maybe it is worth to write this explicitly.***

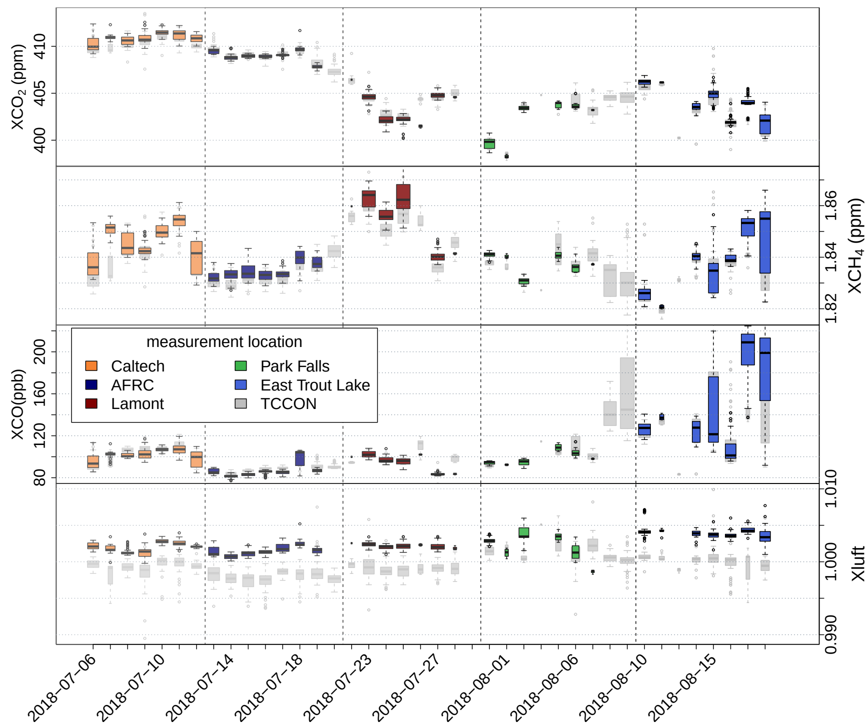
AR: To make it more clear we made the change to the x-axis in figure 6 to represent different measurement locations instead of TCCON abbreviations to avoid further confusions. We have also updated the caption as follows:

> These figures show the median differences between the 10-minute averaged EM27/SUN XCO_2 (in ppm), XCH_4 and XCO (both in ppb) from coincident EM27/SUN measurements collected throughout the

2018 field campaign at each TCCON site before applying the instrument-to-instrument bias correction. The colour bars show the median biases relative to the "tb" EM27/SUN instrument at each stop during the field trip. Top panel shows GGG2014 and the bottom panel GGG2020 results respectively. The error bars on each colour bar represent the median absolute deviation in the 10-minute averaged biases.

RC: ***Figure 8: Adding vertical lines separating the measurements of the different sites could help to make the plot clearer.***

AR: We have added vertical lines to the plot:



RC: ***L365: Add explicitly which is the reference EM27/SUN (is it tb?). This could help the reader for making fast comparisons with the figures.***

AR: We made the changes in the text as suggested:

Figure 9 presents the median bias between the 10-minute-average Xgas values from the reference EM27/SUN ("tb") and each TCCON instrument.

RC: ***L408: This sentence is more appropriate to section 4, not 4.1. Because it sums up the results of all the other stations than eureka, it appears misplaced to me in a section treating the peculiarities of the Eureka station.***

AR: We have moved the discussion about Eureka to section 4.1.