

We are grateful to Reviewer 2 for taking the time to review our manuscript and for the reviewer's considered and constructive comments. They have helped improve the manuscript.

Below we have included the full text of the review (black text), interspersed with our responses (in italicised blue font) addressing their specific comments and changes to the manuscript.

Reviewer 2 Specific comments

It appears that the Authors have not smoothed the XCO₂ retrievals from the various instruments to account for differences in the averaging kernels and priors (if these differ between the retrieval schemes) per the method of Rogers and Connor (2003) as implemented by Wunch et. al. (2010). It may be that the effect of this process would be negligible, however this should still be discussed and quantified.

Thank you for this important comment. Under Section 4.1, we've added:

"We did not consider the averaging kernels in this work. The averaging kernels of the EM27 have been previously presented and compared to TCCON in a study by Hedelius et al. (2016). In their study, they found that although there are differences in the TCCON and EM27 averaging kernels, the effect of the differences in averaging kernels from the top of the atmosphere cancel out the effect of differences at the bottom. Further work and more measurements may be necessary to better understand the cause of this phenomenon but this is beyond the scope of this study."

Introduction and Fig. 1. Is there a reason why there appear to be more v2.72 M-Gain retrievals?

Yes, this is a good question, thanks. So under Section 2 "Alice Springs Australia Site Description", paragraph 1, we've added:

"Note that for Ver. 02.72 FTS SWIR L2 retrievals, upgraded input and reference products were used. For example, there was an improvement in the spatial resolution of the cloud flagging procedure, which employs the CAI L2 (Cloud and Aerosol Imager, Level 2) data. This improvement resulted in better screening of the data and may have resulted in an increased number of soundings that were passed for the Ver. 02.72 FTS SWIR L2 retrievals."

These changes are described in the release notes:

https://data2.gosat.nies.go.jp/doc/documents/ReleaseNote_FTSSWIRL2_V02.72_en.pdf

There should be some commentary on how the GOSAT specific point observations differ from normal GOSAT observations in Section 3.3.

Good point, therefore under Section 3.3, we wrote:

"The TANSO-FTS on GOSAT has a two-axis pointing system. Normally, TANSO-FTS follows an M shaped grid on a 5-point cross-track scan mode. By using this pointing system to vary the observation geometry, it is able to observe specific points, i.e. it can view targets with angles up to $\pm 20^\circ$ along the satellite track and by $\pm 30^\circ$ across the track. Specific point observations over Alice Springs were requested from July 2016, in preparation for the campaign in September. Five locations within 100 km of the center of Alice Springs were targeted (see Fig. 3).

In section 4.1.1 comparing the retrievals of Xair between the EM27 and TCCON instruments, it is noted that there are both airmass and seasonal variations in Xair. It seems likely that at least some of the seasonal dependence might be caused by the differing ranges of solar zenith angles that are observed throughout the year. It would be interesting to plot Fig 5 for a limited range of solar zenith angles to identify if the airmass dependence is the only reason for the seasonal variation.

Thanks, this is a good point. We revised Fig. 5 to show X_{air} values at $SZA > 45^\circ$, $SZA < 45^\circ$ and all SZA and revised the section accordingly. The SZA dependence is small and within approximately 1%.

In any case, it would be useful to have a few more details about the year of inter-comparison measurements in the introductory paragraph of Sect. 4.1 e.g. number of days of measurements, total number of measurements from each instrument and whether they were operated for the same periods on each day.

Under Section 3.2 we added in the last paragraph:

“In Wollongong, both the TCCON and EM27 solar tracker covers were opened and closed by the same pneumatic system. The same weather station provided the meteorological data that were used to pre-filter the data (e.g. fractional variation in solar intensity, wind speed and direction, pressure, etc). We did not filter the data according to solar zenith angles anymore, instead in addition to the pre-filter, we filtered out noisy retrievals by selecting only those with X_{air} values within 0.5 and 1.5 because anything beyond that would be unrealistic in the atmosphere but most likely be the cause of an interference or obstruction.”

And in Section 4.1, first paragraph, third sentence, we wrote:

“Here, we focus on comparisons of X_{air} and X_{CO_2} from measurements spanning almost one year (Nov. 2015 to Sep. 2016)” under varied environmental conditions. Both instruments normally measure at the same time, apart from interruptions due to occasional mid-infrared measurements with the 125HR or rare software glitches (e.g. JAVA issues).”

Section 4.1.2 and Fig. 6, what time averaging is applied to the data in Fig. 6? The colour scale and the size of the 1σ uncertainties suggest daily averages, but this is not made explicit.

Actually, these are hourly averages, so in the text we wrote:

*“Therefore, we fit hourly mean data from TCCON and EM27 employing linear least squares and force a zero intercept. The standard errors of the weighted means are used as weights in the fit. From this exercise, we arrive at a scaling factor of $EM27=0.9954*TCCON$ ”*

There also appears to be quite a large variety in the magnitude of the 1σ values in Fig. 6. which would bear explanation. Several times during the manuscript calibration factors are presented in the form $INST1 = F * INST2$ it would be useful to have some indication of the uncertainty of the F value, or the goodness of the regression fit used to derive it. Similarly, for the drift mentioned at P9L17.

We’ve added error bars on the inset plot for the drift and a Pearson R correlation coefficient ($R=1$) for the least squares fit in Fig. 6. The uncertainty in the F value arising from the linear fit, while forcing the intercept to zero, is very small ($<1 \times 10^{-6}$), so we did not include it anymore.

Section 4.4. describing the length of campaign required to reduce the error in the bias estimates is interesting but would benefit from some discussion of what the target for accuracy and precision in the bias estimate should be.

Thanks, this is a good point. We wrote 0.1 ppm as reference and we cite the work of Miller et al., 2007 and added on Sec 4.4, 2nd paragraph:

“Miller (2007) showed that a comparison of surface CO_2 concentration data and X_{CO_2} data flux inversions clearly reveals a land-ocean bias in the X_{CO_2} retrievals, even when the bias is only 0.1 ppm”.

Conclusions: this section should have some comment relating to the presented dataset specifically. Also, as an interested reader it would be good to know if there are any plans to repeat the field campaign and build the dataset.

We included in the conclusion, 2nd to last sentence:

“Subject to funding and interest from the community, a repeat of the campaign is possible.”

P15L3, the sentence starting “With the exception of Lauder...” should be qualified.

Thanks, we reworded this whole sentence to:

“With the exception of Lauder, New Zealand, there are no permanently dedicated ground targets for satellite calibration and validation in the Southern Hemisphere that is far enough from water (to avoid a possible land-ocean bias in retrievals) and surrounded by homogeneous topography (to avoid a possible altitude bias)”

Data availability: what about the GOSAT data?

We included in the section “Data Availability”

GOSAT data can be obtained from the GOSAT Data Archive Service (GDAS) after registration for access via: https://data2.gosat.nies.go.jp/index_en.html

Technical corrections

P2L24 “atmospheric conditions that can serve as a calibration point...”

Yes, this is better, thanks.

P4 Fig. 1 caption: Start with "Location of.." to indicate that these do not represent the retrieved values, for consistency, use XCO₂ and XCH₄.

Good point, done. Thanks.

P4L7 “below 5000 cm⁻¹ which allows for...”

We improved the sentences related to this to:

“For total column measurements of CO₂, CH₄, H₂O and O₂ spectra in the near infrared, the instrument is fitted with an Indium Gallium Arsenide (InGaAs) detector dedicated to 5,000–12,000 cm⁻¹. This also enables measurements of spectra covering the O₂ bands necessary to derive column-averaged dry-air mole fractions of CO₂ and CH₄ similar to the method used by TCCON”

P5L20 The last sentence on this page is disjointed and difficult to read, consider revising.

Thanks, we agree and revised these sentences.

P6L4 “However, to construct the time-series...”

Done, thanks

P6L4 should refer to Fig. 8, not Fig. 7.

Well spotted, thanks.

P8 Fig. 4 caption: it should be sufficient to say month number and delete “in a year”.

Done, thanks

P9L12 Sentence can be finished after “campaign” to avoid repetition.

Agreed, we’ve done this. Thanks.

P10L7 "...retrievals results in lower..."

Done, thank you.

P11 Fig. 7 caption: use XCO₂ for consistency.

Done, thanks.

P11 Fig. 7 caption: "days on the x-axis are not"

Corrected, thank you.

P11 Fig. 7 caption: pluralise interruption

Corrected, thank you.

P12L2 Start the sentence with "The number of..." to clarify that the retrieved values are not reducing.

Thanks for the suggestion, we corrected it.

P16L14 Explain what RA means.

Done, thanks.

P18L4 correct the rendering of the subscript

We tried many times but there seems to be a difficulty in the subscripting between Overleaf and Mendeley. We will bring this up during the typesetting.

P18L33 The DOI is repeated

Corrected, thanks

P19L6 insert a space between inverse and models

Done, thanks.