

We sincerely thank the reviewer for taking the time to review our manuscript and for the reviewer's considered and constructive comments that certainly helped improve the manuscript.

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

Reviewer 1 Comments:

G1 – The word “calibrate” has been used throughout to describe reducing the mismatch between TCCON and EM27/SUN retrievals, when in most cases “scale” would be a better choice. Calibration is generally reserved for something directly observed of known accuracy, and XCO₂ is neither. In some situations “calibration” is okay so the wording is not too awkward (e.g., calibration curve), but should be changed in most instances.

We agree and have made the changes, for example in Section 4.1.2. we made changes to a few sentences:

“From the EM27 and TCCON measurements in 2016, we derived a ~~calibration~~-scaling factor for the EM27. We ~~calibrated~~ scaled the EM27 measurements to Wollongong TCCON data because the Wollongong XCO₂ has been calibrated against aircraft profile measurements that are traceable to the WMO in-situ scale (Wunch et al., 2010).”

*“From this exercise, we arrive at a ~~calibration~~ scaling factor of $EM27=0.9954*TCCON$.”*

“Therefore, we are confident that the ~~calibration~~ scaling factor relative to TCCON was consistent during the campaign with that derived relative to Wollongong TCCON.”

Also under Section 4.3

*“From here, we derived a ~~calibration~~ scaling factor of $0.9927*EM27$ for GOSAT H-gain and $0.9983*EM27$ for the GOSAT M-gain,...”*

G2 – It seems that differences in averaging kernels (AKs) and a priori profiles have not been considered in this study. E.g., on P2L35 it is stated measurements can be compared directly. It should be explicitly stated why AKs were not considered. Perhaps the a priori profiles are close enough to the true profiles *in this location* that accounting for these different sensitivities would make little difference? Also, it seems observations were first averaged rather than directly compared from both the EM27/SUN and GOSAT? It would also be interesting to know how much of the H to M gain bias is from AKs, if any.

We've added to Section 4.1.2, 2nd paragraph:

“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.”

We have also reworded the last sentence in Section 4th paragraph:

“The Alice Springs measurements are unique considering that they have been collected from a clean desert environment and can be directly compared to GOSAT retrievals using the M-gain measurements.”

To

The Alice Springs measurements are unique considering that they have been collected from a clean desert environment where GOSAT M-gain soundings are also abundant and close enough (<100 km) to compare with the EM27.

Specific comments

S1, P1L7: State length of campaign here

Done, thanks

S2, P1L_1: How exactly are the different gains chosen and used for GOSAT? Is gain chosen in real time by an onboard sensor, or are all gains recorded and the retrieval just picks the best one later?

We have revised the last sentences of the first paragraph in the Introduction:

“These gain settings are prespecified at certain locations because GOSAT does not observe both M and H gains simultaneously. For the majority of the soundings over land, H-gain is used. GOSAT M-gain retrievals over land are used over surfaces that are bright in the SWIR such as deserts and semi-arid regions. However, a bias between GOSAT M- and H-gain retrievals of XCO₂ has been reported, along with a lack of M-gain validation with TCCON (Yoshida et al., 2013). In Australia where surface reflectivity values are generally high, GOSAT was configured to observe using M gain for much of the land surface, at first to avoid detector saturation. However, it was found that some observations using H gain did not result in saturated signals and were still useful. Therefore from 2012-Feb-11, GOSAT started to perform alternate observations using both H and M gains in order to investigate the differences from retrievals between these two gain settings. Therefore from 2012-Feb-11, GOSAT started to perform alternate observations using both H and M gains in order to investigate the differences from retrievals between these two gain settings.”

How/why did the gains change in Fig. 1 going from v 2.60 to v 2.72?

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

S3, P2L16: What does this mean? There are still some large cities in deserts.

True, this sentence was poorly constructed, we now changed it to:

“The world’s deserts and semi-arid regions encompass large areas that are mostly undisturbed by anthropogenic emissions and are important for understanding the carbon cycle.”

S4, P2L20: This goes back to the question of how gain is chosen, but are M-gain regions always exclusively semi-arid? If so where does the classification of climate come from in the algorithm?

Thanks for the question. We believe we’ve addressed this above

S5, P2L20: Carbon cycle studies could be (and have been) carried out with biased satellite retrievals, even if the bias is “small.” I think the focus here though is if the goal is to continuously improve the accuracy of these studies, then accuracy of satellite retrievals needs to be improved as well. What would happen if all observations over high albedo were not available? Likely results would be biased, and would lead to misinterpretation.

This is correct! Thank you for this very important point. We would like to use it in the manuscript so we added:

“Recently, the importance of semi-arid regions in the carbon cycle inter-annual variability has been highlighted (Poulter et al., 2014). If the goal is to continuously improve the accuracy of carbon cycle

studies, then the accuracy of satellite retrievals needs to be improved as well because if observations over high albedo were not available or biased, flux estimates would likely be biased as well and would lead to misinterpretation. Therefore, measurements over M-gain regions are needed by the satellite community (Yoshida et al., 2013) and highlights the significance of this study.

S6, P2L25: Why is high reflectivity a challenge for satellite observations? How does the extra reflectance influence the retrievals negatively? (Same question on P15L12) S7, P2L25: How does having challenging observations naturally lead to their improvement? (Maybe the meaning here is it rather an opportunity?)

Thanks for these good questions we would like to address them at the same time. In the manuscript, we explained this in the text as below:

“The desert environment provides high surface reflectivity, which challenges satellite retrievals because aerosols, depending on type, can lead to an effect called optical path lengthening and this effect is dominant in regions with high albedo (Yoshida et al., 2013). Recently, Iwasaki et al. (2019) showed that there is an increase in the XCO₂ retrievals using their PPDF-S algorithm when the albedo at 1.6μm was high, implying that the retrieved XCO₂ is strongly related to the surface albedo. This challenge leads to the improvement of satellite retrievals.”

S8, P2L27: Why are such studies needed?

Thanks, we have expanded this sentence to:

“Benchmark measurements and pilot studies for desert sites are needed to assess the benefit and feasibility of such sites because setting up a TCCON site in remote deserts will be difficult logistically and financially.”

S9, P3L1-2: Suggest you pick a notation for EM27 or EM27/SUN early on and stay consistent throughout the entire paper.

Good point, we have done this early on in the Introduction by writing:

“We address this need by utilizing a well-established portable spectrometer, an EM27/SUN by Bruker Optics GmbH (Gisi et al., 2012; Frey et al., 2015; Hase et al., 2015; Hedelius et al., 2016; Frey et al., 2018), which was retrofitted with a protective fairing and automated solar tracker hatch for operations in a harsh environment. For brevity, we will refer to this instrument as EM27.”

S10, P3L2: Are v 2.60 data bias corrected? How?

Actually, both are NOT bias corrected, so we rewrote that part:

“We compare NIES GOSAT retrieval versions 2.60 and the new version 2.72, which are both not bias corrected.”

S11, P3L1-2: Suggest you move this sentence to later on in the paragraph where again it states EM27/SUN retrievals were compared with GOSAT.

Thanks, we moved this after mentioning Section 4.

S12, P3L9: How was the EM27/SUN retrofitted?

We added this in the introduction:

“...retrofitted with a protective fairing and automated solar tracker hatch for operations in a harsh environment.”

S13, P3L10: suggest you include 2 decimals on latitude. Also, include dates here.

Done, thanks.

S14, P4L4: Specify these are column measurements.

Well spotted, thanks.

S15, P4L7: Was the 2nd detector present for this study? Give full spectral range for this detector.

No, it wasn't present so we deleted that sentence, which takes care of the next issue

S16, P4L7-8: Suggest you move this sentence before the previous one. Currently it sounds like O₂ is measured on the secondary detector.

True indeed, thanks, we did this.

S17, P4L9: What is the spectral resolution of TCCON measurements?

TCCON uses a maximum optical path difference (MOPD) of 45 cm, corresponding to a spectral resolution of 0.02 cm⁻¹. We've added it now, thanks.

S18, P5L2: "greenhouse gas (CO₂ and CH₄) total columns" could simply be replaced with "XCO₂ and XCH₄."

Thanks, done.

S19, P5L₂₃: Suggest you describe how X_{air} can provide info on stability (e.g., as a measure of retrieved O₂, which is not particularly variable in dry air).

Thank you, this is a very good and important point. We improved the X_{air} description according to this suggestion.

"X_{air} is a good indicator of instrument stability and changes in spectrometer alignment because 1) VC_{air} is calculated using the surface pressure, which is independently measured by a pressure sensor to better than 0.3 hPa, keeping accuracy over long periods (Wunch et al., 2011); and 2) the atmospheric oxygen column is not particularly variable in dry air, hence the retrieved VC_{O2} by the spectrometer should be close to constant. From Eq. 3, it follows that a perfectly accurate measurement would lead to an X_{air} value of unity, however, due to spectroscopic limitations, the actual value is approximately 0.98 for all TCCON sites.

S20, P6L2: Clarify "specific" here. Is it only soundings within 100 km of the EM27? Are they compared individually or averaged together?

We elaborated this paragraph to:

"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 ±20° along the satellite track and by ±30° 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).

S21, P7L4: Frey et al., 2018 (<https://doi.org/10.5194/amt-2018-146>) would also be an appropriate reference here.

Thanks, done.

S22, P8: Was there any other alignment of the EM27 or TCCON instrument during this period?

Good question. In the revised version, we state:

“We note that there were no re-alignments done on the EM27 and TCCON instruments during this period, however the EM27 clamshell cover and fairing were fitted on the EM27 in February 2016, it is possible that this may have affected the alignment resulting in a shift in the Xair.”

S23, P10 Fig6: Specify what points represent in caption. Daily averages? Daily averages within certain sza?

In the Figure caption, we now write:

“Weighted hourly averaged XCO₂ from TCCON vs XCO₂ from EM27 in Wollongong for 2015-2016. To avoid noisy data, only measurements corresponding to TCCON with retrieved Xair values within 0.9783 and 0.9853 were selected (Xair values within the 2nd to 98th percentile)”

S24, P11L2: How were GOSAT data interpolated? (spatially? temporally? method?)

We changed this sentence. The word “interpolated” was removed because it can be confusing. Instead we write:

“To demonstrate the variability of XCO₂ in the region, we plot in Fig. 8 the daily mean time series of GOSAT soundings using these two coincidence criteria: 1) all soundings within a 1000-km radius centered at the BOM facility in Alice Springs and 2) all soundings measured within the same day (local time). Each data point considered in the calculation of the daily mean is weighted by the corresponding reported retrieval error for that particular GOSAT sounding.”

S25, P11L3: This is an unweighted mean?

No, as mentioned above, each data point considered in the calculation of the daily mean is weighted by the corresponding retrieval error for each GOSAT sounding falling within the spatial and temporal coincidence criteria. This associated error is reported in the GOSAT data.

S26, P11L12: Quantify approximate magnitude of annual increase here.

We have quantified from the slope of line in Fig. 8 the following XCO₂ annual increase:

H-gain: 2.2815 ppm/year; intercept: 132.9996

M-Gain: 2.2829 ppm/year; intercept: 134.9174

So, at the end of the first paragraph in Sec. 4.3, we wrote:

“We confirm that the M-gain retrievals are biased high (around 2 ppm) compared to the H-gain retrievals. Nevertheless, both least squares fitted lines to the H and M-gain retrievals show an increase of about 2.28 ppm/year with y-intercept values at 133.0 ppm (H-gain) and 134.92 ppm (M-gain).”

S27, P12L1: Did the number of H-gain observations decrease, or just the number of successful retrievals? If it's just the retrievals could it be an increased failure of convergence?

The number of successful retrievals increased. At the end of Sec 2, we added:

“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.”

S28, P12L2: Apr-Aug mean in absolute number of soundings would also be useful here, same on line 4.

We agree and included this (please refer to the reply below).

S29, P12L1-7: Seems a better sentence order could be: less rain -> less vegetation -> bright surface/more M-gain (rather than less vegetation -> bright surface/more M-gain -> less rain).

Thank you, we agree. This paragraph has now been changed to:

Rainfall records indicate that April-August 2011 had the least amount of rainfall on record in Alice Springs after 2002, which was only broken by 2015, then 2017 (<http://www.bom.gov.au/climate/data/>, station number 015590). This significant absence of rain in the region could result in stunted vegetation growth or mortality, which could have led to brighter surfaces. Bright surfaces mean more M-gain measurements and this may explain the 74% increase in M-gain soundings in April-August 2011 compared to the April-August 2010-2017 average of c. 220 soundings/month. In contrast, the number of H-gain retrievals seems to have diminished around 2011, coinciding with the dry months starting from April 2011. The number of H-gain soundings from April-August 2011 was 23% fewer compared to the April-August average from the years 2010-2017 (c. 450 soundings/month).

S30, P12Table1: What do the seconds mean on the measurement times? Start time? Central time? Seems the measurements could take up to 20 seconds. Also, the purpose of the first and second to last rows is not clear if no GOSAT soundings were acquired.

We removed the seconds and simply wrote:

The times correspond to the times when the satellite is directly above the site. The satellite normally performs 5 observations points across track, with an interferometric scan time lasting 4 seconds (Shiomi et al. 2006). But specific point observations deviate from this pattern by pointing and maximizing observations near the target.

S31, P13Fig9: I do not really like lines fit through single points. Granted the intercept is forced through zero, but I think this information would be better for a table.

Yes, the intercept is forced through zero because both instruments are expected to produce zero XCO₂ when there is no CO₂ absorption. We added the sentence to Sec. 4.1.2:

“The retrieval method is predicted to be both linear and have zero intercept (Wunch et al., 2010). Therefore, we fit hourly mean data from TCCON and EM27 employing linear least squares and force a zero intercept.”

We would like to keep the figure because for future campaigns and comparisons with more data points we would like to use this as reference.

S32, P14L2: Where did these values come from?

We clarified this item in the paper and also considered the standard errors of the weighted mean for version 2.72:

“The standard errors of the weighted hourly mean GOSAT specific point observations are: 0.25 ppm and 0.26 ppm for M-gains (versions 2.6 and 2.72, respectively) and 0.29 ppm and 0.28 ppm for H-Gains (versions 2.6 and 2.72, respectively). These values are taken from the averaged standard errors of the weighted hourly means from all specific point observation data falling within 100 km from the site for 2 Sep. 2015 - Feb. 2017”.

S33, P14L7: satellite measurements -> satellite and ground-based measurements

Done, thanks.

S34, P15L2: What about Wollongong? Darwin? Reunion? Showa? Arrival Heights? Ascension? (Fig. 1: <https://doi.org/10.5194/amt-9-2381-2016>)

We elaborated this 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).”

S35, P15L4-5: This sentence seems a bit redundant with the first sentence.

Agreed, we removed this sentence

S36, P16L_1: Also add a description on how GOSAT data can be acquired

Done. thanks

Technical

T1, P1L7: values, a -> values, another

done, thanks.

T2, P1L11: improve -> better understand (or estimated -> estimation of)

thanks, changed to “better understand”

T3, P2L4: along with -> and there has been

agreed, thanks.

T4, P2L6: version 2.72 -> version 2.72 Xgas retrieval algorithm (w/o sentence seems to be missing a subject)

Thanks, sentence modified.

T5, P2L8: precise -> precise and accurate (?)

Yes, that is right, change done.

T6, P2L17: by anthropogenic -> by recent anthropogenic

Thanks, this is better.

T7, P2L20: are -> is

Done.

T8, P2L25: provides -> has

Done, thanks.

T9, P2L25: which challenges -> which is a challenge for

Corrected, thank you.

T10, P2L27: could be -> is

Changed.

T11, P2L28: portable -> portable solar-viewing

Done, thanks.

T12, P3L8: “in the urban area” seems redundant, maybe omit?

Agreed, done.

T13, P3L17: average high -> average daily high

Done, thanks

T14, P3L18-19: omit parenthetical comment (already on page 2, and Fig 1 caption)

We agree and omission done.

T15, P3L21: maybe omit “reasonably accessible” as this is somewhat vague
We agree and did this.

T16, P5L19-20: suggest “s” on Ps and “a” on Na be subscripts
Changed, thanks.

T17, P5L21: can -> is (?)
Yes, thanks. The sentence has been changed.

T18, P6L6: retrievals -> retrievals separately.
This is better, thanks

T19, P8L6: should give -> gives
Done, thanks.

T20, P11L5: were -> are
Done, thanks.

T21, P12L1: H-gain -> the number of H-gain (or better, -> there were 23% fewer Hgain)

T22, P12L2: Could omit “derived”

Thanks, to accommodate both of these, we changed the sentences to:

“In contrast, the number of H-gain retrievals seems to have diminished around 2011, coinciding with the dry months starting from April 2011. H-gain soundings from April-August 2011 were 23% fewer compared to the April-August mean from the years 2010-2017.”

T23, P14L14: Could omit “in Alice Springs”
Thanks, done.

T24, P14L15: cover -> enclosure (x2)
Done, thanks.

T25, P14L17: omit “the assumption”
Omitted, thanks

T26, P15L8: suggest omission of “and on other future satellites”
Suggestion accepted, thanks.

T27, P15L11: provides -> has
Done.

T28, P15L12: to -> for
Done, thanks

T29, P15L12: this data is especially needed -> observations are especially needed here
This is an improvement, thanks

T30, P16L7: calculated -> characterized
Done, thanks.

T31, P16L9: co-funded -> co-acquired funding for (?)
Agreed, thanks

T32, P16L14: define RA

Defined, thanks

T33, P16L14: advise -> advice
Corrected, thanks.

Other notes/optional

O1, P2L15: The population of the greater LA area includes parts of other counties. <https://www.citypopulation.de/world/Agglomerations.html> lists the population as 17.7 million.

Thank you very much for this information. The agglomerated population count represents the region better, so we changed it according to this.

O2, P5L3: The authors may also consider publishing a description of the design and/or control software in the future (e.g., compare <https://doi.org/10.5194/amt-11-2173-2018>). Such a project/paper could be especially useful to the community if the control software were open source, and fully automated neither of which have been done yet. Example automation of OPUS: <https://doi.org/10.1364/AO.57.000689>, example of alternate solar tracking software: <http://hdl.handle.net/10222/64642>, Chapter 4.

Yes indeed, the EM27 enclosure has undergone several revisions, our aim is to have 3D drawings that we can share to the community. The Davis weather station has a built-in programmable logic; we are also planning to distribute the configuration.

O3, P6Fig2: More details on this schematic could be useful, such as the path of light, and parts that move/rotate or disassemble (it looks like a seam at the “v” part?) An actual picture at Alice Springs could be nice if available. A picture would provide the readers an idea if any precautions needed to be taken to prevent interference from unique Australian fauna, such as fencing or placement away from trees.

Yes, there is a seam at the “V” part for easy access to the solar beam inlet. However, the design has changed a little bit, so we are currently working on the engineering drawings that we can distribute. We have added a picture of the EM27 at the Alice Springs site as an inset to Fig 3. Fortunately, this site is just beside the airport and is fully fenced, so the noise and the fencing keep away large Australian fauna (e.g. Dingoes, Kangaroos, Sand goannas, etc.).

O4, P9L9-11: The ME at MOPD values seem particularly small, compared to typical values around 98-99% (Frey et al, 2018). It may be beneficial to realign the spectrometer. Though changes in Xair do not look large. . .

Yes, well noted. Unfortunately, unlike the 125HR, alignment of the EM27 is still not well documented in the community, therefore for a re-alignment we have to send the EM27 to Karlsruhe Institute of Technology (to Frey et al.) or directly to Bruker Germany. This was done in 2017.

O5, P10: Future measurements in Alice Springs may be useful to help derive an airmass correction for all EM27/SUN instruments. This dataset may not be ideal though since the ME at MOPD seems to differ from most other EM27/SUN instruments, and is only for part of one season.

Deriving an airmass correction is an excellent point. For future campaigns, we will make sure to have the EM27 re-aligned at Bruker or at Karlsruhe. A loan from the COCCON network (or others) may be a possibility as well.

O6, P14L2: A histogram of the standard deviations for the different gains could be useful.

This is a good idea. We followed this and added an inset to Fig 10. Also, we’ve added on Section 4.4, 2nd paragraph, sentences 1 to 3:

“Fig. 10 shows a plot of how the statistical errors in the estimated bias improve with the number of weeks in the campaign. The inset shows the normalized histogram of the reported GOSAT single sounding errors within 100 km of the site. Note that although H-gain retrievals have smaller errors on average, the amount of M-gain retrievals improves the standard error of the weighted mean.”