Black: referee's comments red: authors' answers
First of all, we want to thank the two referees for the detailed analysis of our paper.
For the details, please look into the paper with keeping track of changes.

Anonymous Referee #2

The study by Zhou et al. employs the full-physics retrieval code SFIT4, which is used by the Network for the Detection of Atmospheric Composition Change (NDACC) in order to retrieve vertical profile information on atmospheric methane from solar absorption spectra measured in the near infrared (NIR) by spectrometers within the Total Carbon Column Observing Network (TCCON). Comparisons of retrieval codes lead to improvements in the codes and therefore, this study is a contribution to remote sensing measurements of atmospheric CH4. I recommend its publication in AMT after the questions, issues and comments outlined below have been addressed.

Major Comments:
The authors state that the ILS parameters are retrieved simultaneously by the code. How does the retrieved instrument line shape look like and how constant is it for all the sites involved? The Bruker 125HR spectrometers exhibit excellent ILS stability, so the retrieved values should reflect this. Therefore, it would be beneficial if the authors could show a time series of the ILS and the parameters.
Thanks for the suggestions. We plot the time series of the retrieved amplitude error of the modulation efficiency at the maximum optimal path difference (MOPD) at the six sites (see Figure 1). As expected, the retrieved ILS are very constant, and the retrieved amplitude errors are within 2%. In the revised version, more information of the retrieved ILS is added in the retrieval strategy Section.

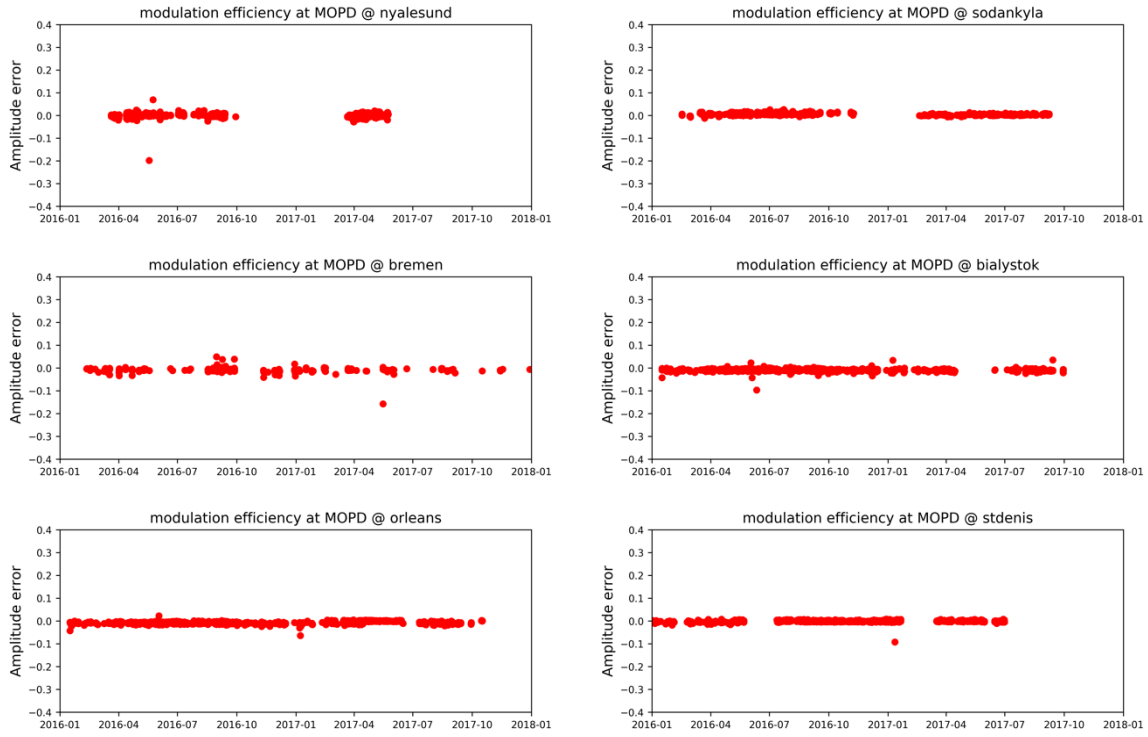

Figure 1. the time series of the retrieved amplitude of the modulation efficiency at maximum path difference (45 cm) at six sites.

The profile retrieval relies on the Alpha values, as discussed in Sec 2.2.3, but could the authors please explain the physical significance of the Alpha value?
The physical significance of the Alpha value is the correlation among layers. With a larger Alpha value, the stronger relationship among layers are constrained. Some explanations are added in the revised version.

Also, it seems to me, as shown in Fig. 3, that the retrieved profile just approaches a scaled value of the a-priori profile at Alpha values of 1,000 and 10,000. Optional addition to Fig 3: Could the authors add a plot of retrieved VMR profile divided by the a-priori VMR profile with altitude as y-axis or something similar? This is to show how much the a-priori is scaled and the altitude dependence of this value.
Following the referee's suggestion, we add one ratio profile in the revised version, along with the original plot.

In TCCON, the Xair value and its time series are indicative of instrument stability, I think a comparison of the SFIT retrieved Xair and the TCCON Xair for the sites is warranted for this study.
The Xair in TCCON data is defined as: Xair = $VC_{dryair}/(VC_{O2}/0.2095)$, where VC_{dryair} is calculated from the surface pressure and total column of water vapor. The Xair is often used by TCCON data user as an index to check the time error and instrument stability. However, in SFIT4TCCON retrieval, we do not retrieval O_2 and the dry air column is derived from the surface pressure and total column of water vapor directly. Therefore, there is no Xair for SFIT4TCCON data.

In situ measurements: In its current state, I do not see the full usefulness of the comparison between the in situ ground-based measurements and the tropospheric product of SFIT4TCCON (Sec 3.2). Both measurements have completely different sensitivities, as the authors mentioned, and I think comparing the time-series alone does not sufficiently provide information to say that "The SFIT4TCCON tropospheric and stratospheric XCH4 can observe the CH4 seasonal variation very well, which has been confirmed by the ground-based in situ measurements. . ." in the conclusions. For example, the agreement between SFIT4TCCON tropospheric CH4 and in situ looks to be closer during the winter months and farther during the summer months both at Orleans and St. Denis. But it is difficult to see from the scattered, overlapping data points. I recommend that the authors derive a seasonal fit and/or trend line to both FTS and in situ data. Alternatively, a correlation/scatter plot for the FTS vs. in situ would be useful, like in Fig 6.
Thanks for the suggestion. We now compare the monthly means from the in situ measurements and tropospheric product of SFIT4TCCON at Orleans and St Denis. The seasonal cycles from two measurements have similar phases and amplitudes, which can support our conclusion.

Finally, it wasn't clear to me until the last sentence of Sec 2.2.1 that the SFIT4TCCON retrieval does not actually use all the CH4 bands used by TCCON, so the term "SFIT4TCCON" is potentially misleading to the readers and data users. I recommend that this term be changed. Moreover, the authors state in the abstract that "the SFIT4 retrieval code is applied to retrieve CH4 mole fraction vertical profile using TCCON spectra". First, this is not entirely true because the retrieval only uses a subset of the NIR spectra used for retrieval of CH4 in the TCCON network. Second, the term "TCCON spectra" is a term that is not officially used by TCCON and this is also not a standard TCCON product, so I do not think that this term is appropriate to be used to describe the spectra used in this study.

Thanks for the suggestions. We agree with you that the SFIT4TCCON is not appropriate. So we use “SFIT4NIR” as the new name for our retrievals.

Technical/Minor Comments:
P1, Line 7: two distinct species -> two distinct pieces
Corrected

P1, Line 19: spectrometer -> spectrometers
Corrected

P2, Line 16: the atmosphere chemistry -> atmospheric chemistry
Corrected

P2, Line 23: started to increasing -> started to increase
Corrected

P2, Line 24: remove "the" in: partly by the getting -> partly by getting
Corrected

P3, Line 15: It seems that the spectra is converted to SFIT4 readable format then corrected for SIV, please arrange sentences if not the case.
Corrected

Figure 2. How do the retrieved columns for each band look like for the standard TCCON product? Does the TCCON CH4 at band 1 also exhibit the same curve?
We checked the TCCON CH4 (GGG2014) retrieved total columns from the three bands, and no such large curve is observed. However, the retrievals from Band 1 also show opposite diurnal variation compared to the results from Band 2 and 3.

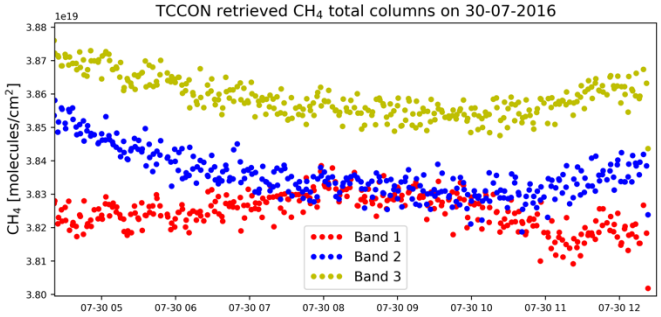

Figure 2. the retrieved CH4 total columns from three bands in the GGG2014 code at St Denis on 30 July 2016.

Although the TCCON ATM spectroscopy is used in our retrieval (SFIT4NIR), we notice that the spectrum in Band1 is not well fitted with SFIT4 code. More investigations are needed in the future to understand the reason. So far, we only choose the Band2 to do the CH4 profile retrieval.

P6, Line 3: The DOFS definition is not entirely accurate; moreover it is not consistent with line 7 of the Abstract.
Corrected. According to Rodgers (2000) Page 30, the DOFS is defined as “degree of freedom for signal”.

P6, Line 5: The sentence about S_(epsilon) needs to be checked, it seems wrong. Corrected

P6, Line 5: change "to constraint" to "to constrain" Corrected

P6, Line 5: "to determine whether the" or "to determine if the" Corrected

P7, Line 5: change "penal" to "panel" Corrected

P8, Line 17: remove "have" in "parameters have do not" Corrected

P8, Line 19: change "error" to "errors" Corrected

P10, Lines 25 and 26: use plural "measurements .... are "Corrected

P10, Line 28: Either quantify how well they are calibrated or change "well calibrated" to just "calibrated"
Corrected

Fig. 5: The panels are too crowded and the legend boxes partially cover the data. I think this should be improved. The TCCON and SFIT4TCCON data overlap and it is impossible to determine the data points. This figure needs to be revised. The same goes for Fig. 7. Fig. 7, Legend: change "Insitu" to "in situ". Title: fix "stdenis" and "orleans"
Corrected. Note that the number of the co-located hour means are slightly changed, because the official TCCON data have been updated at Caltech website.

Fig 8: The legends overlap with the actual data points, making it hard to read the figure. Additionally, Fig. 8 could be improved by adding a correlation plot for each panel because the data are too sparse and does not cover a long time series. In fact, the correlation plots could be a better representation.
The correlation plots are added.

P13, Line 9: There have been other validation activities after De Maziere et al., 2008 and the results for CH4 have improved since then, e.g. https://doi.org/10.4401/ag-6339
Added in the revised version.

P13, Line 19: "sits" to "sites" Corrected

Figure 9: I would like to know how the data points in Fig. 9 are treated. Daily means, hourly means? How is the filtering done and how are the errors in each retrieval taken into account? The answer to this could explain or support the statement "at St Denis (a moist site), the TCCON HF retrievals are strongly affected by H2O so that TCCON proxy method tropospheric and the stratospheric XCH4 data using HF have many outliers"
These are hourly means. Added in the revised version.

P13, Line 26-27: This "slight seasonal and site dependent bias" is not clear to me from the figure.
One sentence is added in the revised version.
“For example, the difference in tropospheric X$_{CH_4}$ between N$_2$O proxy method and HF proxy method is larger in summer than that in winter at Ny-\r{A}lesund. ”

P14, Line 3: "systematic larger" -> "systematically larger"
Corrected

P14, Line 5: The sentence starting from line 5 and ending in line 6 needs to be improved.
Corrected

Fig. 10 Caption: it is very hard to see the "scaled SFIT4TCCON a priori profile (dotted black line)"
Adapted.

Fig. 10 and Sec 3.5: Why and how is the AirCore profile smoothed? It seems that there is a lot of structure and information in the AirCore profile that is lost from the smoothing. Also, the sentence "The extended" AirCore profile is then smoothed with the closest SFIT4TCCON retrieval" is not very clear.
According to Rodgers et al., (2003), the in situ vertical profile with a fine vertical resolution must be smoothed to compare with the optimal estimation method retrieval from the remote sensing technique with a coarse vertical resolution. Therefore, the AirCore profile is smoothed to consider the vertical sensitivity of the FTIR retrievals. The method to do the smoothing is added in the revised version.

The plot is adapted.

Fig. 11: The authors need to provide an error estimate on the slopes of the lines.
Added

P18, Lines 13-15: I think measurements from a single tower compared to a single TCCON site are not sufficient to arrive at this conclusion. Moreover, the authors should quantify what they mean by "very close".
Thanks. To avoid the confusion, the sentence is removed.

P18, lines 17-20. These sentences seem contradicting. On one hand the authors state that "there is almost no systematic bias between the SFIT4TCCON and AirCore XCH4", but on the other hand the next sentence state "An overestimation of 1.2% in the SFIT4TCCON tropospheric XCH4 and an underestimation of 4.0% in the SFIT4TCCON stratospheric XCH4 is seen by comparing with AirCore measurements"
Adapted in the revised version.

Reference:
Rodgers, C. D.: Intercomparison of remote sounding instruments, J. Geophys. Res., 108, 46–48, https://doi.org/10.1029/2002JD002299, 2003.