Dear anonymous referee #2,

Thank you very much for your careful reading of our manuscript and valuable suggestions. We hope that revised manuscript looks better.

We do not avoid using the GOSAT SWIR CAK. We consider that it is necessary to apply the GOSAT SWIR CAK to the aircraft measurement data for a meaningful comparison between the two measurements. In this study, we applied the GOSAT CAK to aircraft-based XCH$_4$ calculation when comparing the GOSAT data with temporally matched aircraft data (Sect. 3.3.1). On the other hand, we could not apply the GOSAT SWIR CAK to the fitted aircraft-based XCH$_4$ due to the absence of

the vertical information for all aircraft measurements when comparing of GOSAT SWIR XCH_4 with the gap-filling time series of the aircraft-based XCH_4 through curve fitting (Sect. 3.3.2). Therefore, we first evaluated the impact of GOSAT SWIR CAK on the aircraft-based XCH_4 calculation (Sect. 3.1).

- in my opinion, the proposed method of extending the aircraft measurements with the ACE/HALOE climatology is flawed (check Geibel et al. 2012 for the reasons). Retrieval theory tells us that the GOSAT a priori profile is the most reasonable choice because that is what the retrieval falls back to when there is no other information. Any other choice of profile will only introduce an additional bias and never improve anything. Just imagine the extreme case where the aircraft coverage would be close to zero: with ACE/HALOE you would still get a bias despite the fact that there is no information from the aircraft measurement! If you think otherwise, you should explain much more why you think your choice is better.

You mentioned "retrieval falls back to when there is no other information", but exactly we believe that this must be rephrased as "the retrieval value is identical to a priori value when the observed spectrum has no sensitivity to the target state." This is not related to the presence or absence of more probable profile (e.g., ACE/corrected HALOE in this study) other than a priori profile. In addition, Geibel et al. (2012) did not use the a priori profile itself for the stratospheric part of the column. They used the GFIT a priori profile multiplied by the retrieval scaling factor, which corresponds to the retrieval value rather than the a priori value.

The aim of Geibel et al. (2012) is to derive a better calibration factor (TCCON-to-aircraft ratio) by using a new method they developed, whereas the aim of our study is to validate the GOSAT XCH_4. To prepare the aircraft-based XCH_4 as the validation dataset, we should use the most probable data for the part of the column that was not measured by aircraft. If the GOSAT a priori profile data were the most probable data, we would use them for the stratospheric profile. However, the GOSAT a priori profile was calculated by NIES TM and the stratospheric part of the model was nudged to the HALOE data (Saeki et al., 2013, GMD). As noted in Sect. 2.2.3, the HALOE CH_4 data were underestimated compared to the data provided by the ACE-FTS whose observing period (February 2004 to February 2009) was relatively close to that of GOSAT. Therefore, in this study, the ACE data or the HALOE data corrected by the ACE data were used as the most probable stratospheric and mesospheric profiles.

We understand that error contribution comes from the part of the column that was not measured by aircraft should be discussed. However, "the 2-7 km altitude" you mentioned is not typical observing altitude in this study. As described in Sect. 2.2.1, typical observing altitudes of NOAA, DOE, NIES, and NIES-JAXA campaign were from 0.5 km up to about 6 or 7 km, and the HIPPO missions were able to provide atmospheric measurements covering altitudes from 0.3 km up to 14 km. We found it interesting that Geibel et al. (2012) mentioned incomplete vertical coverage of aircraft profiles can lead to a bias in the calibration factor (TCCON-to-aircraft ratio), and developed a new approach to derive a calibration factor without biases even the aircraft profiles with incomplete vertical coverage. By using the HIPPO profiles with higher altitude observation than other aircraft platforms, we investigated how aircraft-based XCH_4 at HIPPO sites differed when calculated using all aircraft profiles and only aircraft profiles below 7 km altitude (Sect. 3.2).

We added a figure and a table to support our results (Fig. R2-1 and Table R2-1) in the manuscript. In addition, we revised the sentences in Sect. 3.3.3 as follows.

"To clarify the cause of this difference, we compared the aircraft measurement data with TCCON data at several sites. The results show that on average aircraft-based XCH_4 is approximately 6–8 ppb (SD = ~10 ppb) smaller than TCCON XCH_4."
--->

"To clarify the cause of this difference, we compared the ground-based FTS data (GGG2012 release) obtained from four TCCON sites – Park Falls, Lamont (USA), Tsukuba (Japan), and Wollongong (Australia) with aircraft measurement data at four aircraft sites (LEF, SGP, TKB, and HPC) which were obtained within ±5° boxes of each TCCON site. TCCON data are the mean values of XCH_4 data obtained within ±30 min of GOSAT overpass time. Figure 12 and Table 7 describe the comparisons at four sites. The results show that on average aircraft-based XCH_4 is 8.6 ppb (SD = 10.4 ppb) smaller than TCCON XCH_4."
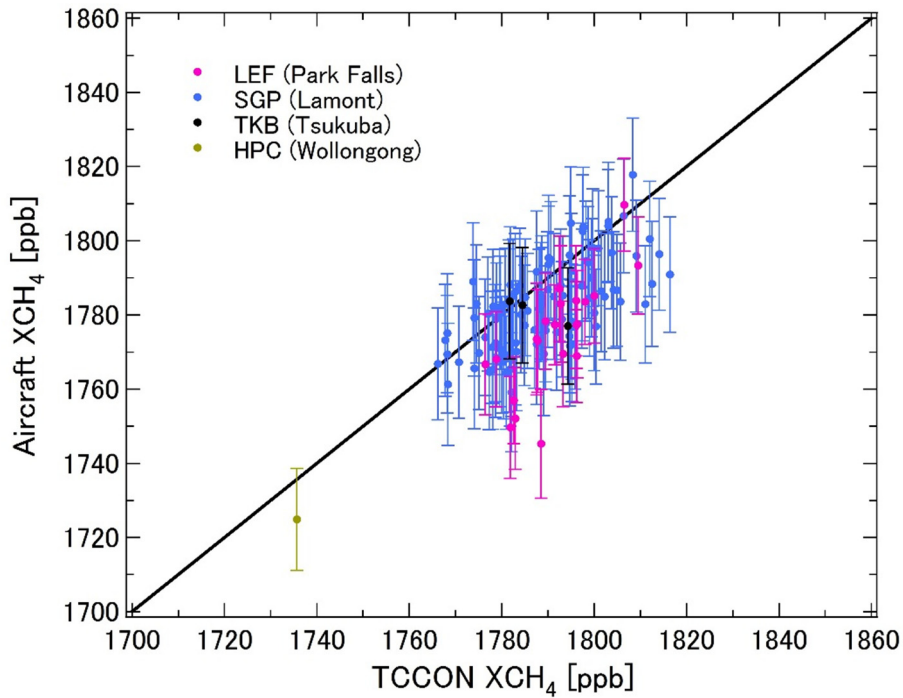

Fig. R2-1. Scatter diagram between aircraft-based XCH_4 and TCCON XCH_4 (GGG2012 release) on the same day as aircraft measurement at each site. The one-to-one line is plotted as a black line.

Table R2-1. The average, maximum, minimum, and 1 standard deviation of differences of TCCON XCH_4 (GGG2012 release) and aircraft-based XCH_4 at each observation site.

	LEF (Park Falls)	SGP (Lamont)	TKB (Tsukuba)	HPC (Wollongong)	All sites
Number of dataset	21	98	3	1	123
average [ppb]	17.4	6.8	5.7	10.8	8.6
1σ [ppb]	10.3	9.6	10.2	—	10.4
maximum [ppb]	43.2	28.2	17.2	10.8	43.2
minimum [ppb]	-3.2	-15.2	-2.0	10.8	-15.2

Specific comments:
- The title is somewhat misleading. A proper validation should provide more than just a comparison of two datasets. It should rather be called "Intercomparison of ...".

We used all of the aircraft vertical profile data available, and used the most probable data (ACE data / corrected HALOE data) for the part of the column that was not measured by aircraft to calculate aircraft-based XCH_4. We would like to use the term "validation".

p. 4739, l. 6-10: the use of "above" and "below" is ambiguous when you use pressure as a vertical coordinate. Please rephrase. A figure might help to show which part of the profile was taken from which source. This was partly done in Fig. 8. However, this figure - along with all others - uses geometric altitude as the vertical coordinate.

We replaced "above 10 hPa" by "above the 10 hPa level (above ~30 km)", and "below 10 hPa" by "below the 10 hPa level". In addition, vertical coordinate of Fig. 3 was represented by geometric height (left axis) and pressure (right axis). We added those descriptions in Sect. 2.2.5 and caption of Fig. 3.

p. 4741: sorry, from the description in the text it is not clear to me how Fig. 4 was derived. Was this derived from aircraft measurements at SGP? If so, what was the altitude coverage of the aircraft measurements? This number is not provided in any

As shown in Sect. 2.2.6, Fig. 4 shows the temporal variations of the partial XCH_4 calculated in the three domains (I), (II), and (III) over SGP. The time series of partial XCH_4 were calculated from aircraft profiles or assumed profiles below the PBL height for domain (I) (Fig. 4a), from aircraft profiles between the PBL height and the tropopause for domain (II) (Fig. 4b), and from ACE/corrected HALOE data in the stratosphere and mesosphere for domain (III) (Fig. 4c).
We revised a description in Sect. 2.2.6 as follows.

"As an example, we show the temporal variations of the partial XCH_4 in the three domains over SGP (Fig. 4)."
--->

"Figure 4 shows the temporal variations of the partial XCH_4 calculated in the three domains (I), (II), and (III) over SGP. The time series of partial XCH_4 were calculated from aircraft profiles or assumed profiles below the PBL height for domain (I) (Fig. 4a), from aircraft profiles between the PBL height and the tropopause for domain (II) (Fig. 4b), and from ACE/corrected HALOE data in the stratosphere and mesosphere for domain (III) (Fig. 4c)."

p. 4742, l. 11-18: I find it somehwat unusual to consider values outside a 1-sigma range to be outliers and remove them. That is a very strong filtering criterion which leaves you with a very smooth dataset with very limited variability. Was that really necessary?

The purpose of this study is to validate the GOSAT XCH_4. We consider that this data screening is needed to use aircraft-based data as validation data.

Sec 3.3.3: "Comparison with validation by ground-based FTS data"? Either comparison or validation!

We replaced "Comparison with validation by ground-based FTS data" by "Comparison of validation results between aircraft-based data and ground-based FTS data".

p. 4748, l. 5-8: the idea by Geibel et al. was to minimize biases introduced by filling the domain not covered by the aircraft measurements (which turned out to be the

largest error contribution). I would not be surprised if your biases were the result of the climatological profiles that you used to extend your aircraft profiles (see my arguments above).

Thank you for your comment.