Reply to the comments by Referee #1

To Referee #1,

The paper assesses biases in satellite-retrieved CO_2 concentrations at the lower and middle troposphere from GOSAT/TANSO-FTS TIR V1 product by comparing them with precise aircraft measurements by CONTRAIL CME, followed by global comparisons of bias-corrected CO_2 concentrations with model-simulated CO_2 by NICAM-TM. The authors found that the TIR data had negative biases of 1-1.5% against the aircraft measurements and bias-corrected TIR data showed generally good agreement with the NICAM-TM CO_2 data, which demonstrated the validity of the bias-correction values.

Observational CO_2 data in the free troposphere is still limited, and CO_2 profiles from high-resolution GOSAT TIR spectra will help to elucidate CO_2 variations in the free troposphere with its global coverage. Bias estimation of satellite-based CO_2 products is highly important for data users and further analysis of CO_2 fluxes by atmospheric inversion/data-assimilation studies. The paper is generally well written, and I recommend accepting it for publication after the comments listed below have been addressed.

We appreciate you reading our paper carefully and giving valuable comments and suggestions. We have considered your recommendations for revisions and made the necessary changes. The major points that we deal with in the revised manuscript are as follows:

1. Following your advice, we have added Table 2 to present bias values of GOSAT/TANSO-FTS thermal infrared (TIR) version 1 (V1) Level 2 (L2) CO_2 data against CONTRAIL CME CO_2 data to which TIR CO_2 averaging kernel functions were applied. It could help readers see Figure 6.
2. Following your advice, we have added Table 3 to present mode values of frequency distributions of differences in monthly averaged CO_2 concentrations between original or bias-corrected TIR and NICAM-TM CO_2 data and numbers of data categorized into the mode values and all 2.5° gridded data used for comparisons. It could help readers see Figure 7.

Individual responses to the Referee's comments are listed below.

General comments:

1 . Results section: The paper presents comparisons between the original TIR data and CONTRAIL CME data and between bias-corrected TIR data and NICAM-TM data. But the expressions of the

evaluations are often qualitative, such as "relatively low", "tend to be larger", "slightly increase", "nearly identical", "close to zero" without any supporting numbers. Although one can see tendencies on the plots, I would recommend illustrating the point with some numbers and add a table with quantitative values to explain the results clearly. The authors do not need to write all related numbers, but at least it would be better to write statistic values related to Figure 7, one of the main plots, to show the validity of the bias-correction values quantitatively. Statistic values in a table or the main text may help readers to follow the discussion. They can be mode values (or medians), standard deviations, kurtoses and skewnesses of frequency distributions, the total number of data pairs, or whatever the authors need to describe Figure 7.

Reply:

We totally agree with you. As described above, we have added Table 2 and Table 3 to present specific values of what we focused on in Figure 6 and Figure 7, respectively. In the revised manuscript, we have referred to Table 2 and Table 3 to clarify points of discussions related to Figure 6 and Figure 7. We have also referred to specific values presented in Table 2 and Table 3 in the main text of the revised manuscript. We appreciate your comment.

2. "East Asia" in abstract and discussion section: The authors conclude that one of the reasons of the overcorrection in JJA/low latitudes (0S-20N)/upper MT region is that the correction values were determined by using the data over East Asian airports. Since the authors write this finding to the abstract, this conclusion is thought to be important for the paper. But the explanation (p.10, L34 - L11, L8) is not clear enough to understand why data in the East Asia region strongly affects to the 0-20N bias correction. Usually, Asia in 20S-20N is called Southeast Asia (or part of South India). Do the authors mean "Southeast Asia" rather than "East Asia"? Or if the East Asian data truly affects the 0-20N bias-correction values via atmospheric transport, please give more explanation and references.

Reply:

We greatly appreciate you pointing out this. We wrote "East Asia" incorrectly in the sentences where we should have written "Southeast Asia" in the manuscript. We intended to say that the bias-correction values in low latitudes (20°S−20°N) in the JJA season in 2010 were determined on the basis of comparisons over the three airports over Southeast Asia: BKK (Bangkok), SIN (Singapore), and CGK (Jakarta). In the revised manuscript, we have replaced "East Asia" with "Southeast Asia" throughout the text and described these specific airports in the discussion part.

Specific comments:

Page 3, Section 2, TIR data: Does the TIR product include nighttime data as well as daytime data? I suggest writing time of the observations briefly somewhere in this section.

Reply:

The TIR products of GOSAT/TANSO-FTS include data obtained both in daytime and nighttime. Following your suggestion, we have stated this clearly in the revised manuscript as follows:

"The TIR band of TANSO-FTS makes observations both in daytime and nighttime, unlike the SWIR band."

Page 4, Section 3, NICAM-TM data: NICAM-TM inversion with CONTRAIL data was conducted for the period 2006-2008 (Niwa et al., 2012). It should be explained briefly how the 2010-2012 CO_2 data was calculated by NICAM-TM.

Reply:

We agree with you. As you pointed out, the NICAM-TM inversion simulation that was conducted in Niwa et al. (2012) used CONTRAIL and surface CO_2 data in 2006–2008 to estimate the natural flux of CO_2. The NICAM-TM CO_2 data used here were generated by using the estimated CO_2 natural flux (fixed for 2010−2012) and year-dependent CO_2 fluxes from fossil fuel and biomass burning emissions (considering their yearly trends). Following your suggestion, we have added more explanation of the NICAM-TM CO_2 inversion as follows:

"In this study, simulation of NICAM-TM used inter-annually varying flux data of fossil fuel emissions (Andres et al., 2013) and biomass burnings (van der Werf et al., 2010), and the residual natural fluxes from the inversion of Niwa et al. (2012), which mostly represent fluxes from the terrestrial biosphere and oceans. The inversion analysis of Niwa et al. (2012) was performed for 2006−2008 and the three-year-mean fluxes were used in this study."

We appreciate your comment.

Page 5, line 24, "the number of pairs": Could the authors show the number of pairs which finally used for the comparisons for each latitude bands?

Reply:

Following your suggestion, we have described the numbers of coincident pairs of TIR and CME_AK CO_2 profiles for each of the four latitude bands in the fourth paragraph of Chapter 4.1 in the revised manuscript:

"The numbers of coincident pairs of TIR and CME_AK CO_2 profiles varied depending on latitude band and season. The largest number of coincident pairs was obtained in the latitude band of 20°N−40°N including Narita airport, where 506−2501 pairs were obtained. 63−310 and 77−472 coincident pairs were obtained at 40°S−20°S and 40°N−60°N, respectively. The comparison area for low latitudes was extended to a band of 20°S−20°N, because the number of coincident pairs in that region was smaller (0−341) than in other latitude bands; nevertheless, there were no coincident pairs at 20°S−20°N in the JJA seasons of 2011 and 2012. The number of

coincident pairs was smallest (0−30) at 20°S−0° and no data were collected there after September 2010. Thus, all bias-correction values for 20°S−20°N after the SON season of 2010 were determined based on data from 0°−20°N."

The below-attached table shows the numbers of the coincident pairs for each season for each latitude band.

		40ºS−20ºS		20ºS−0º/0º−20ºN		20ºN−40ºN		40ºN−60ºN	
2010, MAM	2010, JJA	63	75	27/114	30/95	1305	2501	472	161
2010, SON	2010, DJF	128	114	0/172	6/155	2133	1588	454	132
2011, MAM	2011, JJA	209	183	0/49	0/0	506	1255	77	227
2011, SON	2011, DJF	179	78	0/137	0/234	1529	1049	199	253
2012, MAM	2012, JJA	310	105	0/49	0/0	748	1815	418	406
2012, SON	2012, DJF	145	166	0/31	0/341	2045	1664	326	119

Page 7, line 10, "On a global scale, the seasonality of negative biases was not clear, given the relatively large 1-σ standard deviations, although these biases tended to be larger in the spring hemisphere than in the fall hemisphere.": The sentence is not clear. Does this mean the negative biases had measurable spring-fall seasonality, but it was not statistically significant due to the large standard deviations? Or actually, the biases had no seasonality?

Reply:

In northern middle latitudes (20°N−40°N), negative biases in TIR CO_2 data were larger in spring (MAM) and summer (JJA) than in fall (SON) and winter (DJF). On a global scale from 40°S−60°N, any statistically significant seasonality was not found in negative biases in TIR CO_2 data against CONTRAIL CME_AK CO_2 data. In Table 2 of the revised manuscript, we have presented bias values of TIR CO_2 data against CME_AK CO_2 data in each season at 541−464 hPa and 464−398 hPa (corresponding to layers 5−6) to make readers refer to specific values that we focused on.

Page 7, line 26, "negative biases of TIR CO_2 data against NICAM-TM CO_2 data in all seasons slightly increased over time": Is there no possibility that small trend error in NICAM-TM CO_2 could attribute the bias increase in Fig.7? The NICAM-TM natural fluxes were estimated for the period 2006-2008, which is different from the target period of this article. In other words, does the NICAM CO_2 have no bias in trends against CONTRAIL CME data? The authors can confirm it by plotting NICAM-TM CO_2 data against CONTRAIL CME data like Fig.6.

Reply:

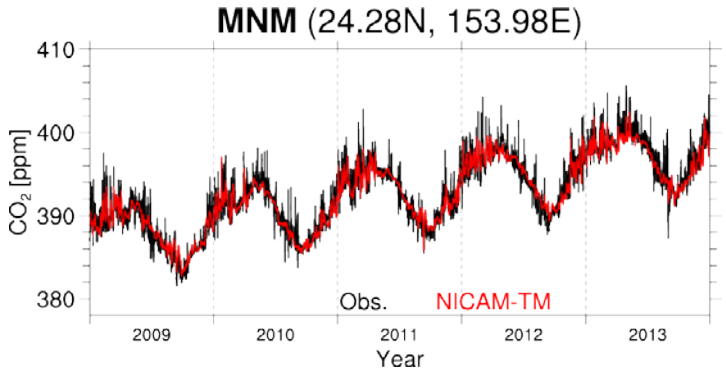
As explained above, the NICAM-TM inversion simulation that was conducted in Niwa et al. (2012) used CONTRAIL and surface CO_2 data in 2006–2008 to estimate the natural flux of CO_2. When calculating CO_2 concentrations in 2010–2012, the mean inversion fluxes were cyclically used, but fossil fuel and biomass burning CO_2 fluxes used were varied inter-annually. We

confirmed that the growth rate of the calculated NICAM-TM CO_2 concentrations for 2010–2012 is reasonable (2.4 ppm/yr) judging from an observation-based growth rate (2.2 ppm/yr), which is partly contributed by the fact that there were no major El Nino events for both the periods. The below-attached figure shows comparison between the NICAM-TM CO_2 simulations and observations at the surface station at Minamitorishima, which demonstrates the validity of the NICAM-TM CO_2 simulations. As Figure 6 is based on one-by-one coincident GOSAT−CME_AK CO_2 comparisons over airports selected by applying the thresholds of a 300-km distance and a 72-h time difference, we think that it is inappropriate to plot comparisons between 2.5°-gridded NICAM-TM and CME CO_2 data on the same figure. Alternatively, we have described the specific comparison in CO_2 growth rates between NICAM-TM simulation and surface observation data as follows:

"Furthermore, the CO_2 forward simulation of NICAM-TM for 2010−2012 showed a good agreement with in-situ CO_2 observations not only in seasonal cycles but also in trends in spite of using the fluxes optimized for 2006−2008; the simulated growth rate at the Minamitorishima station (e.g., Wada et al., 2011), which is one of the global stations of the Global Atmospheric Watch (GAW), was 2.4 ppm/yr for 2010−2012, while the growth rate based on in-situ observations was 2.2 ppm/yr."

"In addition, negative biases of TIR CO_2 data against NICAM-TM CO_2 data in all seasons slightly increased over time, judging from the mode values presented in the top left boxes of Table 3, although the increase in negative biases was not much evident as in the comparisons over airports shown in Figure 6; this may be partly because of slightly high growth rate of NICAM-TM simulations (2.4 ppm/yr) compared to in-situ observations (2.2 ppm/yr)."

We greatly appreciate your comment.



Reference figure. Time-series of observed (black) and simulated (red) CO_2 concentrations at the surface station at Minamitorishima. The observation data presented here were taken from the World Data Center for Greenhouse Gases (WDCGG). The observations have been conducted by JMA under the program of WMO/GAW. We would like to acknowledge the staff that supports the observations.

Page 9, line 5, other sources of negative biases: I'm not familiar with retrieval algorithms, but would any errors in cloud detection process cause retrieval errors in the low latitudes with enhanced convective activity? And H_2O or O_3 do not affect the CO_2 retrieval results?

Reply:

We appreciate your comment. As you pointed out, uncertainties in H_2O and O_3 data could also affect CO_2 retrievals, as shown in Figure 7(b) and (c) of Saitoh et al. (2009). The TIR V1 CO_2 retrieval algorithm (Saitoh et al., 2016) simultaneously retrieves H_2O and O_3 with CO_2, which could decrease the effect of their uncertainties on CO_2 retrieval results. However, water vapor is abundant in the tropics, so that we cannot completely deny the possibility of the effect of H_2O uncertainty on CO_2 retrieval results. Similarly, error in the judgement of cloud contamination may affect CO_2 retrieval results. We have added this point to the discussion part of the revised manuscript as follows:

"Although the effect of uncertainty in H_2O data on CO_2 retrieval results could be also decreased by simultaneous retrieval of H_2O with CO_2 in the TIR V1 algorithm, water vapor is abundant in the tropics, so that we cannot deny the possibility of its effect on CO_2 retrieval results. Similarly, error in the judgement of cloud contamination in low latitudes with high cloud occurrence frequency may affect CO_2 retrieval results."

Page 10, lines 29-30, "The CME data that determined the bias-correction values of the $20^{\circ}S-20^{\circ}N$ latitude band were concentrated in East Asia": I was confused with this sentence. Please see my general comment #2.

Reply:

As described above, we have replaced "East Asia" with "Southeast Asia" throughout the text. In the revised manuscript, we have listed specific airports (BKK, SIN, and CGK) where most CME data were obtained in the latitude band of 20°S−20°N as follows:

"The CME data that determined the bias-correction values of the 20°S−20°N latitude band were concentrated in Southeast Asia, as illustrated in Figure 1: BKK (Bangkok), SIN (Singapore), and CGK (Jakarta)."

We appreciate your comment.

Page 10, line 34 – page 11, line 1, "in most areas at $0^{\circ}-20^{\circ}N$, and the negative biases were largest near airport locations in East Asia.": Same as above. Please see my general comment #2.

Reply:

As described above, we have replaced "East Asia" with "Southeast Asia" throughout the text. We

appreciate your comment.

Page 11, lines 12-13, "More in-situ CO_2 data in the upper atmosphere in low latitudes": Hiaper Pole-to-Pole Observations (HIPPO) project observed latitudinal distributions of CO_2 concentrations in the free troposphere over the Pacific Ocean where mostly clean during 2009 to 2011 (e.g. Wofsy et al., 2011). The dataset has been used for transport model or satellite data validation (e.g. Wecht et al., 2012; Kulawik et al., 2013). The comparison with HIPPO data is out of the scope of this paper, but if the authors found some problems in using HIPPO data for validation, please write it in the discussion section or the introduction section.

Wofsy, S. C. et al.: HIAPER Pole-to-Pole Observations (HIPPO): fine-grained, global-scale measurements of climatically important atmospheric gases and aerosols, Phil. Trans. Roy. Soc. A: Math. Phys. Eng. Sci., 369, 2073–2086, doi:10.1098/rsta.2010.0313, 2011.

Reply:

We agree with you. The reason why we did not use HIPPO data in this study is that HIPPO campaign observations were conducted for limited periods (October−November in 2009, March−April in 2010, June−July in 2011, and August−September in 2011, after starting the regular operation of GOSAT) in limited areas (mainly over the Pacific Ocean), so that they are not suitable for evaluating season- and latitude-dependent biases in GOSAT/TANSO-FTS TIR CO_2 data. As you pointed out, however, HIPPO data themselves are useful to validate CO_2 vertical profiles observed by satellite-borne sensors and simulated in models. Following your advice, we have touched on HIPPO data in the discussion part of the revised manuscript as follows:

"Although HIAPER Pole-to-Pole Observations (HIPPO) data (Wofsy et al., 2011) are not suitable for a comprehensive validation study as in this study due to their limited observation periods, HIPPO CO_2 data are useful to validate CO_2 vertical profiles observed by satellite-borne sensors and simulated in models (Kulawik et al., 2013)."

We appreciate your comment.

Page 11, line 17, "Reconsideration of the setting of retrieval grid layers …": Why do the authors think the current setting of retrieval grid layers might not be suitable for retrievals and reconsideration might solve it?

Reply:

Total degree of freedom (defined as the trace of averaging kernel matrix) does not depend on the setting of retrieval grid layers theoretically. In this situation, partial degree of freedom for each retrieval grid layer (defined here as the diagonal element of averaging kernel matrix corresponding to each retrieval grid layer, see Saitoh et al. (2016)) should decrease as the number

of retrieval grid layers increases. As illustrated in reference figure attached in Authors' reply to Referee #2, the total degrees of freedom of GOSAT/TANSO-FTS TIR V1 CO_2 data are on average 1.1−2.2 (depending on latitude and season), which means that we can derive information on CO_2 concentrations in more than 1−2 vertical layers independently from observations by the TIR band. In the TIR V1 Level 2 CO_2 retrieval algorithm, we have set 28 vertical grid layers. Judging from the total degree of freedom of the TIR CO_2 data and the relatively small partial degree of freedom for each vertical grid layer, we think we should reconsider the setting of retrieval grid layers.

Page 11, line 20, "during the JJA seasons of 2011 and 2011": Does this mean "2011 and 2012"?

Reply:
We have modified the sentence. We appreciate you pointing out our mistake.

Figs.3: The Y axis is described in altitude, not in pressure as seen in the following plots. For easy reference, I would suggest adding a 2nd Y axis in pressure or adding a column in Table 1 to show altitude [km] for each pressure levels. (Rough altitudes from International Standard Atmosphere or the same kind might be enough for this purpose.

Reply:
Following your suggestion, we have added a second vertical axis (y-axis) in pressure in Figure 3 of the revised manuscript. Here, we have taken pressure levels corresponding to the measurement location of GOSAT/TANSO-FTS TIR data shown in the figure.

Fig.4: Please replace "Altitude [km] in Y axis label with "Pressure [hPa]".

Reply:
We have corrected the label of the vertical axis (y-axis) of Figure 4 of the revised manuscript. We appreciate you pointing out our mistake.

Fig.7: I think drawing zero lines (i.e. no bias) in each panel makes the bias correction validity more visible.

Reply:
Following your advice, we have drawn zero lines in each of the four panels of Figure 7 of the revised manuscript. We have also drawn zero lines in Figure 8 and 9 to show differences between each histogram clearly. We appreciate your suggestion.

Fig.7 caption "Thick and dashed lines indicate the biases of the original TIR CO_2 data (no bias correction) and bias-corrected TIR CO_2 data, respectively.":

1. On my screen, all lines in each panel seem to have same line thickness. Do the authors mean "solid and dashed lines"?

2. This sentence does not match the main text which says that thick lines are bias-corrected values.

Reply:

We appreciate you pointing out our mistake.

1. We have replaced "thick lines" with "solid lines" and exchanged "solid" for "dashed" in the caption for Figure 7 of the revised manuscript as follows:

"Dashed and solid lines indicate the biases of the original TIR CO_2 data (no bias correction) and bias-corrected TIR CO_2 data, respectively."

2. We have replaced "thick lines" with "solid lines" in the sentences related to Figure 7 in the revised manuscript.

Fig.11, gray shade: Could the authors explain what gray zones in the figure are? (No data or out of color scale?)

Reply:

Gray color in Figure 11 means no GOSAT/TANSO-FTS TIR CO_2 data in a 2.5° grid area. Following your advice, we have explained the meaning of gray color in the caption for Figure 11 of the revised manuscript as follows:

"There are no GOSAT/TANSO-FTS TIR CO_2 data in gray-shaded areas."