

Interactive comment on “Comparison of the GOSAT TANSO-FTS TIR CH₄ volume mixing ratio vertical profiles with those measured by ACE-FTS, ESA MIPAS, IMK-IAA MIPAS, and 16 NDACC stations” by Kevin S. Olsen et al.

Kevin S. Olsen et al.

kevin.olsen@latmos.ipsl.fr

Received and published: 27 July 2017

1 Anonymous Referee #2

1.1 General comments

Olsen et al. (2017) compares GOSAT TANSO-FTS profile retrievals and partial columns of methane (CH₄) to coincident data from ACE-FTS, ESA MIPAS, IMK-IAA MIPAS, and NDACC, focusing on the upper tropo-
C1

sphere/lower stratosphere (UTLS) over 2009-2013. This work expands an earlier TANSO inter-comparison in the Arctic to include global measurements from additional satellite and ground-based remote sensing instruments, aiming to identify possible zonal variability in the profile retrievals. Given the importance of understanding global CH₄ and the uncertainty in its recent trends, precise and accurate measurements with global coverage are very much needed. As GOSAT is one of the primary satellites in monitoring atmospheric CH₄ today, the assessment of the spatial biases of TANSO measurements would give confidence to the scientific community's use of an important data set. The approach described in the paper is comprehensive, and the methodology is thought-through. However, the exposition of the significance, conclusions, and limitations of this work require more development. The paper would benefit from a rebalancing of its structure, with the most critical changes being: (1) augmenting the discussion of previous and/or similar validation efforts of TANSO CH₄, (2) paring down of the instrument background sections to focus on details directly relevant to their results, and (3) discussing the reasons for and implications of the differences highlighted in the comparison. Therefore, I would recommend publication of this manuscript after these points are addressed.

We thank the reviewer for taking time to make such a thorough review of this manuscript. Their comments are about the general structure and language of the paper, and cover specific technical and grammatical points. We have worked very hard to address each comment with care and believe that the outcome is a more streamlined paper that better addresses the scientific merit of our study without distracting the reader with overly technical discussions. The three primary points made above are discussed in more detail below by the reviewer, asking pointed questions and offering examples. After careful consideration, we agree with the reviewer and have made several changes and additions to the manuscript,

A complete list of changes is included as a supplement to the authors' response to anonymous referee #1.

1.2 Scientific evaluation general comments

1. This paper would benefit from a more rigorous analysis (or if already done, a more comprehensive description) of the causes of the differences noted in the results. The authors are thorough in considering different parameters, but the text lacks a synthesis of how these differences relate to the results of the comparisons. For example, what component of the differences between TANSO and the other instruments can be explained by spectroscopy versus a priori profiles used? This can be addressed for the specific parameters already mentioned in the text; e.g. many of the retrievals incorporate the same line lists—do those profiles show better agreement to each other? The implications of these differences for the scientific community's use of the TANSO product based on these differences would also strengthen the paper.

Some of the primary comments from reviewer #1 are relevant to Sect. 3 where we look at the variability in each dataset. A discussion of these results in the context of our inter-instrumental comparison is lacking. We agree that the text lacks a synthesis and have made significant changes to Sect. 3 and added text to our discussion in Sect. 8. We think this reasonably addresses the reviewers' comments and strengthens our manuscript.

Some of the specific questions raised in this reviewer's comment are, we feel, beyond the scope of this research topic and manuscript. Such aspects as the TANSO-FTS retrieval's dependence on the a priori or line list can be insinuated from our work, but not properly quantified. We have made a better effort to comment on our results in Sect. 8, but, since we cannot vary these parameters, redo

C3

the retrievals, and test the effects of these parameters, thoroughly addressing these questions is best left to a future, specific manuscript about the TANSO-FTS TIR CH₄ vertical profile retrievals.

The impact of differences in spectroscopy and retrieval methods on our results is difficult to quantify, but these factors certainly make significant contributions. For example, the retrievals for the instruments we considered use different versions of the HITRAN line list. There are significant and much discussed differences between CH₄ line lists in each version and CH₄ spectroscopy may be a limiting factor in CH₄ retrieval accuracy. The next version of the TANSO-FTS TIR CH₄ data product may use a different version of HITRAN and the TANSO-FTS team will quantify the impact of the line lists.

We quantified the difference between the IMK-IAA and ESA MIPAS retrievals by comparing pairs of profiles retrieved from the same MIPAS observations. This shows the impact of the retrieval algorithm and spectral line lists. The discussion of this in the manuscript has been improved to highlight this. The largest effect on our study, however, is due to physics and instrument design. TANSO-FTS is a much lower resolution instrument than MIPAS, ACE-FTS and those used by NDACC. Viewing geometry is what led to the use of a retrieval algorithm that incorporates the a priori into its solution, and also results in smaller signal strength than the other instruments. Our discussion has been extended to more thoroughly discuss our findings in Sect. 3, and to address the differences in instrument capability and retrieval algorithms.

2. At several points throughout the paper, principally in Sections 2-5, the authors include technical details about the satellites/instruments that do not seem pertinent to the study presented (e.g. the inclination angles). The extraneous information was distracting and diluted the narrative of the instrument comparison. I recommend including only those details that the reader should know to understand and evalu-

C4

ate the results and conclusions of this comparison, especially if those details are published elsewhere. For necessary details where the connection to the present study is not clear, explicitly listing the relevance and/or implications would be useful. (e.g. The authors list measurement windows and spectral resolution but need to comment on how these inter-instrument differences might relate to the results.) The authors might also consider moving some of these details that are useful but not central to the paper into an appendix or supplement.

While we believe these details are necessary for the reader, we understand the drawbacks pointed out by the reviewer, especially thanks to several specific examples. We removed the ranges of the non-TIR TANSO-FTS channels, the details about MAESTRO on SCISAT, details about TANSO CAI, and the inclination of MIPAS. The inclination of SCISAT is important, since it affects the location of solar occultation measurements which leads to far more coincidences at high latitudes. The inclination has been moved to the coincidence discussion, Sect. 4, and follows a statement about the distribution of the ACE-FTS measurements. In the discussion, Sect. 8, we have addressed the different spectral ranges and resolutions, and their possible impact on our study.

3. Given that methane is generally provided in units of ppb, do the authors have a specific reason for using ppm? If not, I suggest changing the references and figures to ppb. Figure 5 in particular would be more clear without extraneous zeros and decimal places.

When discussing methane on Earth, the abundance in the troposphere is around 1.5–1.7 ppmv, and the literature commonly uses ppmv when discussing CH₄. Our work is consistent with other, related literature sources, such as the CH₄ validation studies for MIPAS and ACE-FTS (Payan et al., 2009; Pleininger et al., 2015; de Maziere et al., 2008) and the preceding work by Holl et al. (2015). That being said, we do agree with the reviewer about Fig. 5 and have changed its x -

C5

axis from ppmv to ppbv. We have left Figs. 1 and 6, since we would be adding zeros to the x -axes of those graphs. Furthermore, we agree that sticking to ppmv when discussing CH₄ *differences* (e.g., Sect. 8) is cumbersome, and we have made several changes from ppmv to ppbv in the text.

1.3 Specific comments

1. p. 2 l. 11-13: The redundancy in the list of GOSAT objectives can be pared down. In addition, these objectives should be related, at least in part, to the objectives of this research. i.e. How does this paper contribute to the objectives of the satellites and the scientific community? This question is briefly touched upon on l. 31, but needs to be developed.

We have broken the short paragraph at line 29 up, placing the argument mentioned at line 31 with the GOSAT objectives at lines 10-13.

2. p. 2 l. 24-26: Given the focus on zonal dependence, providing the latitude of Eureka in the text would be useful. Also, please include citations for PEARL and NDACC.

We provided the altitude of Eureka and added two citations.

3. p. 2 l. 29-31: As the main objective (and contribution) of this paper is to expand TANSO validation globally, more consideration of the issues of spatial coverage is needed, including a literature review of zonal biases (possibly does not exist for this particular TANSO product, but if that is the case this should be stated) and a description of mechanisms that make the Arctic non-representative (e.g. polar vortex changing vertical profiles of trace gases and reducing the accuracy of a priori information).

C6

This paragraph was removed as part of the correction for comment 1. Such a study for the TANSO-FTS data was performed by (Holl et al., 2016) and his results are discussed in Sect. 6.1.

4. p. 3 l. 5/Table 1: Please add year ranges to Table 1. It is not clear when mentioning the 2009-2013 time frame in the abstract whether all of the instrument measure consistently over that time period.

This is not so easy to address, all the instruments have downtime, and those operated at high latitudes are run in seasonal campaigns. Only two of the instruments came online during our study, Altzomoni and La Reunion Maito, and those dates have been added as footnotes to Table 1.

5. p. 3 l. 7-14: If this is an outline, put section numbers after each sentence. If this is an overview, this paragraph might fit better in the methods section rather than the introduction.

Appropriate section numbers are given on lines 6-7 and 15-19.

6. p. 3 l. 27: What type of coverage? Spatial? Spectral?

This is spatial coverage, added the qualifier to the text.

7. p. 3 l. 28: Is this paper the first time methodology for CO₂ is applied to CH₄? If so, the retrieval (or the at least the aspects that differentiate it from other TANSO retrievals) should be described more.

We added more details about the algorithm used, which is more in-line with our presentation of the MIPAS and NDACC data sets.

8. p. 3 l. 33: Do the 2011 updates have a reference (e.g. on the HITRAN website or used in a validation paper)? If so, please include a citation.

C7

There is not a citation for this. The HITRAN 2008 paper refers users to the HITRAN website.

9. p. 4 l. 4-6: This sentence, with the important conclusions of the referenced paper added, should be moved to the Introduction, at the end of the second paragraph.

This is a summary of what is already in the introduction. Page 2, lines 22–28 discuss this in greater detail and provide the actual, quantitative conclusions.

10. p. 4 l. 9-11: If information about MAESTRO is relevant, include a reference for the instrument; otherwise this sentence can be removed.

Technical details about MAESTRO have been removed.

11. p. 4 l. 16: Is 5km the lowest altitude for ACE-FTS measurements filtered using the recommended flags (e.g. not a priori values)? Listing the lowest altitude of the data used in this paper would be more relevant, particularly for the discussion on vertical range in subsequent sections.

5 km is approximate, some occultations may extend lower. As can be seen in Figs. 5 and 6, our data extend to 6 km.

12. p. 4 l. 33 and 36: Do the percentages listed apply to all trace gases or just methane? Please make this more clear.

Added specific references to CH₄. The validation paper de Maziere et al., (2008) focuses only on CH₄.

13. p. 5 l. 4-5: Because this is the data version used, the results from Waymark et al. (2013) should be summarized, as is done with the above papers.

C8

Waymark et al., (2013) contains a table summarizing the results for each gas. The result for CH₄ was added to the manuscript.

14. p. 5 l. 25: By "Initial guesses" do you mean a priori profiles? Please clarify.

The ESA MIPAS team makes a distinction between "initial guess" and a priori. The initial guess is a generic profile used to begin the least squares fitting iterations. The term a priori means that we have prior knowledge of the state of the atmosphere and that the initial profile has a level of accuracy that the "initial guess" may not.

15. p. 6 p. 5l. 6-7: What are the reasons for the outliers/discontinuities? Would these impacts the results of this study?

Outliers would hopefully have little effect on this study due the large sample size we use. We rely on the ESA MIPAS data quality flags to remove spurious results.

16. p. 6 l. 30: Using "some information" is vague and does not tell the reader the relevance of the included information.

We agree, "some information" is vague. That may have been written before we decided on what information to include in Table 1. The specific columns are now in place of the vague terminology.

17. p. 6 l. 32-35: Is there a reference for an inter-comparison of these instruments? How are inter-site differences due to different instrumentation accounted for in this study?

There is not a publication detailing exactly what the reviewer asks yet, but the reader is invited to see Table 1 for publications describing each site.

18. p. 6 l. 36: Are those references for the most recent versions of the retrieval software? The spectroscopy has presumably changed since

C9

1995/2004.

These references are not for the most recent versions of the software, but are the most recent publications about the software.

19. p. 6 l. 37: Does "harmonized" mean consistent between sites? Please use a more clear term.

Yes, harmonized means to "make consistent or compatible," so in our context we mean that the site operators have attempted to make the retrievals consistent between sites. This is the terminology used by the NDACC InfraRed Working Group.

20. p. 7 l. 10: Given the differences delineated in this section, have you done a covariance analysis or sensitivity test to assess whether the results of the comparison depend on the retrieval software used, instrument, or any other difference across NDACC sites?

Please see line 11, the start of Sect. 3. This comment is addressed as part of the reviewers Scientific Evaluation.

21. p. 7 l. 20: Are these measurements representative of year, season, location, tropopause height, etc.?

These data were randomly selected from our sample to avoid diurnal, seasonal, or zonal effects. This information has been added to the text.

22. p. 7 l. 25: Please define "sunset/sunrise measurement" in this context.

Noted that sunset and sunrise refer to the occultation direction. In solar occultation geometry, the satellite instrument makes a series of observations by directly viewing the sun while the limb of atmosphere intersects the line of sight. In each orbit there are two occultation opportunities: when the satellite enters the shadow

of the Earth, and when the satellite exits the shadow of the Earth. The instrument observes sunsets and sunrises, respectively.

23. p. 8 l. 5-6: Please summarize the results that the bias is consistent with.

This is in Sect. 2.3.2 where Laeng et al. (2015) is summarized.

24. p. 8 l. 24: Why was reducing coincident measurements an objective, if you could average them and thereby reduce potential bias (c.f. Kulawik et al. 2016)? Was this a data processing issue from the large number of coincidences? If the coincidence criteria varies by instrument, some sort of bootstrap or sensitivity test with a subset of data should be run to see if the VMRs are different with the more lax coincident criteria.

For consistency, we wanted to have sample sizes for the MIPAS comparison to be of the same order of magnitude as those for NDACC and ACE-FTS.

25. p. 9 l. 8-9: Including references for the approach mentioned would be useful.

Added a reference to Holl et al. (2016) for example.

26. p. 9 l. 19: Please define a z score and/or include a reference.

z-score is a standard parameter in statistics found in most undergraduate textbooks. It is perhaps more appropriately referred to as the "standard score," and it has been changed.

27. p. 9 l. 26-27/Figure 2: Does using only the first 200 observations of the year capture any time-varying spatial coverage of the satellite data? Would it be more appropriate to use the first 20 observations of each month of 2012, for instance?

C11

The orbit actually repeats its flyover-locations every couple of weeks. Using the first twenty occultations from each month would introduce a bias. However, the figure is purely qualitative. It shows the NDACC sites, that ACE-FTS favours high latitudes, and that there are no ESA MIPAS coincidences over the tropics.

28. p. 10 l. 5-6/Figure 3: Please provide a more clear description of what the pressure levels in the legends correspond to (i.e. the pressure widths of the averaging kernel rows vs. the pressure on the y-axis), as relates to the findings of the paper. (You could perhaps include the simple averaging kernel equation if useful, but if the text becomes too detailed I would suggest moving this description to an appendix/supplement.) Also, if these pressure levels are meant to be compared across instruments, using a single colour scale for panels a-d would be helpful, e.g. binned into ranges or following a colour gradient.

Since pressure is a retrieved parameter for some of the instruments, each retrieval is done on a unique pressure grid. To compute means of the averaging kernels, each kernel has to have the same vector length corresponding to a common pressure grid. Since Fig. 3 is illustrative, the pressure grid for each panel is not important, but that the means are taken at the same pressure level is. The pressure levels of the averaging kernels are given in each data product. A sentence was added to clarify this. The colours indicating each instrument are consistent throughout the manuscript.

29. p. 10 l. 8-12: Please add the implications for these comparisons. Related to the previous comment, the remarks on "full-width at the half-maximum values" would be more understandable by rewording the phrase "values when considering the location of the appropriate pressure level" and adding a more clear description of the averaging kernel widths (p. 9 l. 33-34).

C12

The implication of these findings is given on line 8, that TANSO-FTS is the instrument with coarser vertical resolution, and on line 25 after further description of the figure, that TANSO-FTS retrievals are highly dependent on their a priori. We've re-organized this paragraph to be more clear, introduced FWHM notation, and moved the definition of vertical resolution to the discussion of kernel widths.

30. p. 10 l. 25-26: How do you determine the influence of this dependence on the results? (This paragraph might need to be moved closer to the discussion of TANSO priors later in the manuscript.)

This paragraph is a description of Fig. 3 and follows the preceding discussion of MIPAS and ACE-FTS sensitivity naturally. The implication is explored further later in the manuscript, especially in the discussion, Sect. 8, which was greatly expanded to address comments from reviewer #1.

31. p. 10 l. 32: What is the implication of the flat trends over mid-latitudes and tropics, as relates to the objective of this paper to determine zonal dependence of the retrievals?

The trends are related to seasons, their southern hemisphere reversal is actually the phase difference of the seasons. This has been re-written to reflect that.

32. p. 11 l. 28: Please include a reference for the claim that the NDACC a priori/ measured pressure profiles are accurate. (This might fit better in Section 2.)

Added a reference to the NDACC retrievals paper (Sepúlveda et al., 2014), which quantifies the contributions of sources of error, including NCEP a priori.

33. p. 11 l. 27-29: I'm not sure I follow the logic here; does this just affirm that interpolating to a common pressure grid does not introduce additional bias or uncertainty?

C13

What we are defending here is not the choice of pressure grid, but the choice of doing our analysis in terms of pressure levels instead of altitude. We argue that pressure is more well known than altitude. We have added a statement specifying this.

34. p. 11 l. 30-32: Do these extrapolated values actually become part of the profile comparison? If not (as indicated in step 5 on p. 12 l. 10-11), why extrapolate these values at all? If so, profiles with extrapolated values included in the comparison would be problematic: the minimum altitudes for these instruments are so high that they tend to be in the region of the atmosphere where CH₄ varies significantly with altitude, and the extrapolation would have large uncertainties.

A common vertical grid is vital since we are finding the means of profiles. Extrapolation is needed because the vector lengths need to match those of TANSO-FTS averaging kernels. Extrapolated pressure levels are not used. A statement was added that explains the need for extrapolation and refers to Eq. 1.

35. p. 11 l. 32-33: Is this sentence a way of saying that the averaging kernel equals zero where no measurements exist? Wouldn't this zero out the extrapolation referred to in the preceding sentence?

In the non-overlapping region, we have set values of \hat{x} equal to x_a . When \mathbf{A} is multiplied by $(\hat{x} - x_a)$, $(\hat{x} - x_a)$ nullifies the contributions from the averaging kernel matrix at pressure levels where $(\hat{x} - x_a)$ is zero.

36. p. 12 l. 16-17: Do you look at the longitudinal variability for each zonal band? If so, does it vary between instruments?

The longitudes of coincidences are fairly uniformly distributed in each zone for ACE-FTS and MIPAS, but are not for NDACC.

C14

37. p. 12 l. 17-20: If my understanding is correct that you apply these filters only for individual points, as opposed to the entire profile, how do you account for heterogeneity in the underlying profile? For example, are all seasons represented at most altitudes ranges? If representation bias is not accounted for, the differences between the underlying measurements might account for some of the features illustrated in Figure 5, e.g. the strong agreement at high vs. low altitudes.

We use the data products as instructed by the product documentation, in some cases, specific points, usually at altitude extrema, are discarded, in others, the entire profile is discarded (e.g., when we say the ACE-FTS *quality_flag* cannot be equal to 4 at any altitude). Seasonal effects are smaller than zonal effects, the strongest being the increased cloud and humidity over the tropics, which strongly reduces the number of overlapping coincidences from MIPAS. Seasonal effects are shown in Fig. 4, and the strongest effect is for IMK MIPAS. The decrease in IMK MIPAS DOFs in Arctic winter corresponds to fewer measurements in the observation (ESA MIPAS, TANSO-FTS, and ACE-FTS do not have such a seasonal dependence). The features the reviewer mentions in Fig. 5 are attributed to the TANSO-FTS averaging kernels tending to zero at higher altitudes. Where the comparison is in very good agreement, the TANSO-FTS is approximately equal to its a priori, and the smoothed target profile has been transformed into approximately the a priori.

38. p. 12 l. 21-22: I find it surprising that measurements within the polar vortex did not impact the results, unless the problematic profiles were discarded through other filters or a priori values were used. Do Holl et al. (2016) apply the same data flags listed in this paper?

Holl et al. (2016) did use the same data quality flags, those for ACE-FTS are referred to in their manuscript. The primary reason the effect is negligible here is that the temporal duration of the polar vortex is small in our data set.

C15

39. p. 12 l. 28-29: Following on comments made on the manuscript's introduction, an explanation of why zonal biases may exist should be included. l. 29-31/Figure 5: Reiterating the last general comment, units of ppbv in the left-most panel would remove some of the extra text on the axes associated with the decimals and might be more intuitive for CH4.

Statements explaining this were added to the introduction, where Holl et al. (2016) is first discussed.

40. p. 12 l. 31-32: Do the sizes of the bins alter the comparisons? Are these zonal ranges narrow enough in the Northern Hemisphere? i.e. Are the profile differences at 50-60N comparable to 30-40N?

The *x*-axes of the left-most panels of Fig. 5 have been changed from ppmv to ppbv.

41. p. 12 l. 37: Please give a brief explanation of what this statistic tells us (or justification for using it) as opposed to the general correlation coefficient and/or include a citation.

Yes, the bins are narrow enough. Shrinking the zonal bin width results in a paucity of data over the tropics, and we tested different bin sizes and found no significant biases in any zone.

42. p. 13 l. 18-22: The pressure level at which CH4 decreases is the tropopause height. Unless I am misunderstanding this paragraph, the implication is that the tropopause heights of the instruments are different, which would very likely account for at least some of the profile differences observed. How do the calculated the tropopause heights compare among the various instruments? If they differ, it would indicate that some of the assumptions underlying the pressure interpolation (outlined in the paragraph on p. 11

C16

The Pearson correlation coefficient is the most common form of correlation coefficient and most appropriate for this type of study. The proper name distinguishes it from rank, distance, weighted, adjusted, and other forms of correlation coefficient. What the reviewer refers to as a "general correlation coefficient" is most likely the Pearson correlation coefficient.

43. p. 13 l. 25-29) might need to be reconsidered. Measurements for which the a priori values have a significant influence could be especially susceptible to tropopause height biases.

We believe the observed phenomenon does not represent the physics of the atmosphere. Since the TANSO-FTS averaging kernels fall off at around the tropopause height, its retrievals have low sensitivity at that altitude. The TANSO-FTS a priori may not accurately reflect the tropopause height, and therefore, neither will the retrievals. Conversely, the other instruments are capable of measuring the height of the tropopause, more accurately defined by the temperature minimum. We avoided discussing the physics of the atmosphere in this section, because we believe the artifact is due more to the fall off of the TANSO-FTS sensitivity and that the vertical resolution of all instruments is not good enough to perfectly reflect the shape of the CH₄ VMR vertical profile at the tropopause. The tropopause heights of the instruments tend to agree within uncertainty, but when, for example, ACE-FTS only has observations every 3 or 5 km, that uncertainty is necessarily large. We are not willing to say that the tropopause heights are different because of this feature. We agree that differences in tropopause height pose significant challenges, but the most relevant argument for using pressure instead of tangent altitude is that the tangent altitudes for TANSO-FTS are not known and computing them introduces additional errors. We were careful to say "this feature indicates that the altitude at which this VMR decrease differs" and not that the tropopause heights differ.

44. p. 13 l. 24/28: By "below 90 hPa" do you mean less than 90 hPa or
C17

at lower altitudes? Similarly, does "Above 100 hPa" mean greater than 100 hPa, or at higher altitudes? Looking at the figure, the reader can deduce the appropriate answer but would benefit from less confusing wording.

Referring to pressure is correct in this context since we are discussing a figure with pressure on the axis. We have reworded the statements here to refer to levels.

45. p. 13 l. 24-27: This result seems to me as one of the most important in the manuscript and deserves elaboration. Does the variability have any notable features? Does it depend on sensitivity (s) or a priori influence? Did you find covariance with latitude (i.e. within the 30 degree bins), tropopause height, or season?

This is explored in Sect. 8. We specifically measure the relationship on mean VMR difference with latitude.

46. p. 13 l. 33-34/Figure 6: Do zonal differences exist in the unsmoothed data? How do the unsmoothed data fit into the goal mentioned in the introduction for assessing the applicability of Holl et al. (2016) to lower latitudes?

The unsmoothed data didn't exhibit significant zonal differences, which is why we chose to show fewer figures. An explicit mention of the lack of zonal differences has been added to the text.

47. p. 13 l. 35-36: This sentence is confusing due to the the vague phrasing and verbosity (e.g. "actual differences one would expect"). Please reword.

The first half of this sentence has been re-written to read: "Fig. 6 shows the

mean differences between the TANSO-FTS data product and those of other instruments.”

48. p. 13 l. 37: More consistent across instruments? Or across altitudes for each instrument?

The sentence is re-worded to explicitly name differences as those profiles shown in Fig. 6.

49. p. 14 l. 9-11: Why would the differences between ACE-FTS and ESA MIPAS be smaller? Does this shed light on differences between each of these instruments and TANSO?

The difference between ACE-FTS and ESA MIPAS is smaller, and expectedly so, because both instruments' data products are more accurate throughout the altitude range shown due their increased sensitivity and decade-long retrieval development.

50. Section 7.1: Does any of the methodology apply to Section 6 (e.g. criteria for minimizing the dominance of the prior) and vice versa?

No, only the selection of coincident measurements and smoothing is common. In Sect. 6, we compare measurements at each pressure level and use whatever pressure levels are available. In Sect. 7, we are integrating the vertical profiles and set a requirement of a minimum overlap of retrieval levels.

51. p. 14 l. 21-22/Figure 8: The way this is described, it seems contradictory with the caption on Fig. 8, “The vertical range of partial column integration varies for each pair of coincident profiles.” If you mean that for each coincident measurement pair you match the vertical range of TANSO and each of the other vertical profiles, but that the vertical ranges across all coincident measurements vary, please describe this more explicitly somewhere in this section. Also, if that interpretation is

C19

correct, does the vertical range impact the distributions or correlations of the data? (This same question applies to Figure 9.)

There is some ambiguity here, that we have tried to correct. Each pair compared must have the same integration range. We didn't require every pair to have one range. That the magnitude of the partial columns varies with altitude is discussed on lines 19–24 and on page 15, lines 31–33, and statistics about the variability of altitude ranges are given in Table 3 and discussed on page 16.

52. p. 14 l. 34: Why a sensitivity threshold of 0.2? This seems a little low. The minimum of three pressure levels also seems low unless they are contiguous (i.e. don't skip filtered out data in the profile). If the data points do not adjoin each other, did you apply criteria on how far apart the levels can be?

This is partly because the sensitivity of TANSO-FTS is this low. The problem isn't just the sample size, however, with a higher threshold, the pairs that remain in the comparison tend to have short vertical ranges of integration. We are comparing smoothed data which are necessarily interpolated, so the altitudes of overlap are contiguous.

53. p. 14 l. 5-7: How different are the results when these 23% for the outliers in Figures 8 and 9?

There are no outliers, if they are shown, they lie beneath the “good” data. They all have small partial column values because the integration interval is small. This comment is a little confusing because, since we selected a sensitivity threshold, the 23% of the data that does not meet the criteria are excluded by the criteria. A comment about the behaviour of the excluded data has been added to the text.

54. p. 15 l. 10-20: Given that the values are all on a pressure grid, what is the advantage of integrating in altitude/z versus pressure/P? Do you account for water vapour (i.e. use “dry” P/T)?

C20

Reformulating the problem as an integral of pressure introduces a dependence on the vertical profile of the mass density ($dz = dP/(-\rho g)$). We did not convert to “dry” P - T because the coincident measurements are co-located so the humidity should be close to the same, but also, introducing a measurement of water vapour introduces the error in that measurement.

55. p. 15 l. 27: If partial columns with large gaps in the vertical are included (ref. my comment on p. 14 l.34), an uncertainty related with the interpolation should also be propagated through the calculation.

The data are smoothed and necessarily interpolated. The uncertainty of the interpolated data accounts for the interpolation process.

56. p. 15 l. 29/Figures 8 and 9: Given the emphasis of the paper on zonal dependencies, please add a colour scale to each plot associated with latitude (bin)?

Unfortunately, colourizing Figs. 8 and 9 with latitudes would not be helpful for our paper. You wouldn't see a latitudinal dependence on the correlation, but, you still see clustering of the data. Since the partial column depends on integration range and the lower limits of the ACE-FTS and MIPAS observations depend on latitude (water vapour, opacity and clouds increasing towards the tropics), then the latitudinal dependence in the partial columns is due to the observation geometry of those two satellites.

57. p. 15 l. 35-37: What would cause a bias in the intercept? Altitude range? Spectroscopy?

A non-zero, positive y -intercept means that the y -axis data, TANSO-FTS partial columns, are generally larger, by this amount, than those from other instruments. This is consistent with Fig. 5 which shows that the TANSO-FTS VMRs are generally greater than the other data products.

C21

58. p. 16 l. 8/Table 3: Why is the minimum altitude for the NDACC measurements so high? Figure 3 indicates that TANSO is at least somewhat sensitive closer to the surface than 3km.

3.3 km is actually very low for a remote sensing instrument. This value not for the NDACC stations, but for the overlap between TANSO-FTS and NDACC obeying our sensitivity and flagging criteria. The pressure maximum in the TANSO-FTS retrievals is always less than that of NDACC, and the sensitivity of TANSO-FTS falls off at the lower altitudes, as shown in Fig. 3e.

59. p. 16 l. 26: When combining the results, are all data weighed equally, or do you take into account the uncertainties of measurements? Is this the average across all latitudes, or is it a bias that is consistent for all latitudes? Have you also assessed whether altitude related biases exist in the combined data?

The least-squares regression is weighted, since each measurement has a unique uncertainty estimate. Each instrument is treated equally, so that with the largest sample size (MIPAS) has the highest contribution. Notes about the regression have been added to the text. Altitude-related biases exist due to the averaging kernels of TANSO-FTS. These can be inferred from Fig. 5.

60. p. 16 l. 28-29: Did you find a bias in the sub-tropics or mid-latitudes?

The bias is latitude dependent; it is the result of weighted least squares regression of all the data, which returned a slope and intercept that are statistically significant. The values discussed are approximations based on the regression results. The magnitude of the bias varies between the tropics and the poles. It is not clear enough in the manuscript that these results are from the least-squares regression and we have attempted to correct that.

C22

61. p. 16 l. 31-32: It is not clear what type of comparison was done? Regression? ANCOVA?

We compared each parameter with our results, as in Fig. 10, and looked at least squares regression results and correlation statistics. This has been specified in the test.

62. p. 17 l. 13: The mismatch in vertical extent you point out seems to indicate that these other satellites are not appropriate/useful for validation. If this is not your argument, please rephrase this sentence to make the argument more clear.

The vertical mismatch is unfortunate, and MIPAS and ACE-FTS are not the perfect data sets to compare to, but finding a comparable, satellite-based instrument with the same data product and an overlapping operation timeline is not so easy. This problem is also a result of our TANSO-FTS analysis and comparison. The sixteen ground-based observatories do have fully overlapping altitude ranges. We agree that this sentence and the preceding one are poorly written and do not convey our intent, so they have been rephrased.

63. p. 17 l. 15: Have you tried smoothing the TANSO profiles to NDACC to see if the agreement is robust?

Because NDACC has a finer vertical resolution than TANSO-FTS TIR, and is sensitive to more vertical structure in the profiles, this operation may not result in "smoothing." Furthermore, since the TANSO-FTS sensitivity is low and its retrieved profiles are close to the a priori over most of its altitude range, the result would be mostly mixing the TANSO-FTS a priori and the NDACC a priori together. Holl et al., (2016) measured the vertical resolution for each pair of measurements to determine which was the lower resolution instrument and in a small number of cases smoothed the TANSO-FTS profiles with the NDACC averaging kernels

C23

and a priori. We felt it was more important to maintain a consistent set of steps to create the data sets presented in Figs. 5, 6, 8, 9 and 10.

64. p. 17 l. 18-19: Given these biases, would you recommend "calibrating" the TANSO retrievals?

This would be case dependant, and because of the small magnitude of the bias, up to the researcher's discretion. Due to the spread of the data shown in Fig. 10 and the uncertainty of each instrument's retrieval, we would not recommend calibrating individual profiles, but would recommend that the bias is considered when a large number of profiles are combined.

65. p. 17 l. 19-21: Please include how this altitude feature varies (or doesn't) with altitude.

We are not certain of what the reviewer is asking us here. The feature is at a specific altitude, it does not vary with altitude, it is not present at other altitudes. We have re-written this sentence to ensure that which profiles are discussed is clear, and that we are discussing Fig. 5 is also clear.

66. p. 17 l. 25: It is not clear what "taken over altitude and latitude" means; please reword.

This refers to Fig. 10, but is a typo and should read "dependence of the mean differences, taken over altitude, on latitude." This is cumbersome and has been revised.

67. p. 17 l. 26-27: What improvements are expected in future versions of the retrieval (e.g. priors, spectroscopy)? Based on your results, what would you recommend needs the most/least attention to produce a more accurate data product? Given the limitations of this TANSO product, what applications would it be suited to?

C24

We have asked our co-authors who work on the TANSO retrievals what is next and included some comments in the conclusion about this. The data product is very important because of the altitude range that it covers, which covers the middle and upper troposphere – not covered by the other data products. This differs and compliments ground based observatories in its global spatial coverage. This is mentioned earlier in the conclusion, but maybe not strongly enough. We have added a statement at the beginning of the conclusion pointing out the importance of the data product.

1.4 Technical suggestions

Comments referring to the addition or removal of punctuation were included where I thought they might improve readability and are thus suggestions rather than corrections, except in cases where the serial comma in a list needs to be added.

1. p. 1 l. 3: Change "CO₂ and CH₄" to "carbon dioxide (CO₂) and methane (CH₄)"

We added definitions for CH₄ and CO₂.

2. p. 1 l. 15: Change "examine" to "examining"

The subject in this list is "we," so examine is correct here. We have added an "and" to better join clauses.

3. p. 2 l. 3-6: Add VMR (parts per notation) in parentheses after listed percentages.

Added ± 40 ppbv in parentheses.

4. p. 2 l. 5: Change "investigated" to "investigate"

Made the correction, thank you.

C25

5. p. 2 l. 7: Unless "over the equator" is specifically what is meant, change to "in the tropics"

Made the change.

6. p. 2 l. 29: Remove the comma after "local"

Made the change.

7. p. 2 l. 33: Change "made in coincidence" to "coincident"

Made the change.

8. p. 3 l. 30: Add a comma after "surface temperature"

Made the change.

9. p. 3 l. 32: Add a comma after "(Maksyutov et al., 2008; Saeki et al., 2013)"

Made the change.

10. p. 4 l. 9: Remove dash after "ACE"

This is the ACE-MAESTRO instrument.

11. p. 4 l. 19: Add comma after "Boone et al. (2005)"

Made the change.

12. p. 4 l. 24: Remove comma after "(Picone et al., 2002)"

Made the change.

13. p. 4 l. 25: The use of "and" does not seem correct. Could be replaced with "assimilated into" or "from" depending on the relationship between the met data and the model.

Changed "and" to "with."

C26

14. p. 4 l. 30: Add comma after "profiles"
Made the change.
15. p. 4 l. 35: Add comma after "Odin"
Left as is.
16. p. 5 l. 7: Awkward placement of "(inclination of 98)"
This has been removed.
17. p. 5 l. 9: Add comma after "cloud parameters"
Made the change.
18. p. 5 l. 11: Add comma after "2004"
Made the change.
19. p. 5 l. 12: Reword end of this sentence, e.g. change ", but" to "with"
Changed to "but with."
20. p. 5 l. 24: Add comma after "limb scan"
Made the change.
21. p. 5 l. 26: Add comma after "temperature"
Made the change.
22. p. 6 l. 20: Change the commas around "below 25km" to parentheses.
Made the change.
23. p. 7 l. 7: Rephrase "dynamical nature" to a more precise term.
Changed it to "characteristic atmospheric dynamics over Antarctica."
C27

24. p. 7 l. 15-16: The use of and phrasing after the semi-colon is awkward and makes the sentence unclear.
This is the correct use of a semi-colon, separating two list items which contain commas within them.
25. p. 8 l. 11-12: The parenthetical is awkwardly worded; please revise for clarity.
This sentence has been revised for clarity.
26. l.12: Change "differences is also" to "differences are also"
Correction made.
27. p. 8 l. 13: Please reword "When examining dates with several measurements" to make more clear.
Revised to "several measurements made on the same day."
28. p. 8 l. 16-17: The grammatical structure of this sentence is difficult to follow. Please reword.
Revised this sentence.
29. p. 9 l. 32: Add comma after "differ"
Made the change.
30. p. 10 l. 10: Change comma after "kernel" to semi-colon.
Replaced with a period.
31. p. 10 l. 13: Add comma after "role"
Made the change.

32. p. 10 l. 20: Add comma after "altitudes"
Left as is.
33. p. 10 l. 22: Change "70km. This is shown in Fig. 3e." to "70km (Fig. 3e)."
Moved this clause earlier in the text.
34. p. 10 l. 22: Add comma after "development"
Made the change.
35. p. 10 l. 27-28: Please reword. The structure of this sentence is difficult to follow, e.g. the verb ("are shown") appears twice and each has its own modifier.
This sentence had remnants of earlier changes, typos fixed, and further revisions made, thank you.
36. p. 11 l. 22-23: Remove parentheses, and add a period or semi-colon after "retrieval"
Made the change, using a period.
37. p. 11 l. 24: Please state which retrieval you're referring to (seems like the higher-resolution profile, but not self-evident).
Revised this clause.
38. p. 11 l. 27: Remove comma after "equilibrium"
Left as is.
39. p. 12 l. 21-22: Change "looking for" to "filtering" and remove ", and then filtering these events," to make the sentence more clear.
Made the change.

C29

40. p. 13 l. 20: I think a verb is missing after "VMR decrease" (e.g. "occurs").
Yes, there was a missing word.
41. p. 13 l. 33: Add a comma after "zonally"
Made the change.
42. p. 16 l. 29: Change "or" to "and" ("0.014 ppmv or 0.020 ppmv"), and "Pole" should be plural.
Made the change.
43. p. 17 l. 10: Remove commas after "sensitivity" and "product"
Made the change.
44. p. 17 l. 12: Remove comma after "altitudes"
Left as is.
45. p. 17 l. 12-13: Please reword "and that there is a limitation on the useful upper altitude of its data product of below 15 or 20km" to follow the clarity and structure of the beginning of the sentence.
This has been revised.
46. p. 17 l. 14: Add "upper" before "troposphere" (Without the addition, this sentence is misleading.)
Made the change.
47. p. 17 l. 20: Phrase starting with "and in a consistent manner" needs rewording for clarity.
This has been revised.

C30

48. Figures 2 and 10: The legend has two icons for every instrument, which adds extra visual clutter.

Left as is.

49. Figure 3: Several line colours do not appear in the legends of a-d. Instead of using a legend, you might consider labelling each line with the pressure using the same colour for the text.

There is not enough room to label every profile, so we only labelled 5 kernels for each instrument.

50. Figure 7: A heat map or similarly sequential colour scheme could be more helpful for this type of plot.

The bin widths are important here, smoothing the data would not really be helpful.

51. Figure 8: The “R” is missing on the R2 line of each sub-figure, and it looks like some other letters and numbers might also be missing.

These, and other characters in other figures were stripped from the pdf files when the AMT header was added. We will have to work with editors to ensure these characters remain in a revised manuscript.

52. Figure 10: The degree symbol is missing between parentheses on the x-axis label. Also, please add additional tick marks on the x-axis. You might consider including light gray grid lines behind the data.

Left background as is, will talk to editors about missing symbols.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-6, 2017.