Atmos. Meas. Tech. Discuss., 8, C3791–C3833, 2015
www.atmos-meas-tech-discuss.net/8/C3791/2015/

Atmospheric Measurement Techniques Discussions

Interactive comment on "MIPAS IMK/IAA CFC-11 (CCl_3F) and CFC-12 (CCl_2F_2) measurements: accuracy, precision and long-term stability" *by* E. Eckert et al.

E. Eckert et al.

ellen.eckert@kit.edu

Received and published: 11 November 2015

The authors thank the Reviewers for their thorough and insightful comments. In the following the original comments are inserted in *italic face* while our replies are printed in normal face.

Anonymous Referee #1
Comment: *SUMMARY*
The paper contains a number of comparisons of profiles of CFC-11 and CFC-12

retrieved from the Envisat MIPAS instrument by the IMK/IAA L2 processor, with collocated profiles from other instruments as well as a ground-based network. From these results the authors conclude that their CFC-11 values are 5-20% too high below 15km, but have a smaller (although unspecified) error above this. For the CFC-12 the accuracy is probably similar or better. As regards precision, the conclusion is that the noise SD is probably an underestimate of the actual precision. The CFC-11 data appear to have a negative drift of a few percent per decade, which is an order of magnitude smaller than the overall rate of decrease of atmospheric CFC-11. For CFC-12 the instrument drift is much larger (in the stratosphere) and of variable sign, and dominates the observed changes in MIPAS CFC-12 data.

MAJOR CONCERNS
1) STATED GOALS: Despite the title of the paper, the authors fail to provide any clear statement on either the accuracy or precision of their data (although I acknowledge that the long-term stability is explicitly, and convincingly, demonstrated). Instead, the authors provide is a list of comparisons with other datasets, from which one can discern relative biases; and measurements of SD of these differences, which contain a combination of precision of the two measurements and atmospheric variability.?

Reply 1: In the revised version the comparisons will be redone on the basis of the full error budget (not only noise) of MIPAS data, and, where available, also for the data of the comparison instruments. This will give much better access to accuracy and precision estimates.

Comment: *Accuracy: I appreciate that absolute accuracy is hard to establish but there is little attempt to rate the correlative datasets according to probable absolute accuracy eg I would guess that the cryo-sampler would be top of the list, and ILAS bottom. If the aim is to give an overall picture of the level of agreement of different*

AMTD
8, C3791–C3833, 2015

datasets, that's not reflected in the title of the paper. If the aim is to establish the absolute accuracy, only datasets known to be of comparable of better quality than MIPAS should be used.

Reply 2: Although we do not succeed in assigning a final number to the accuracy of MIPAS measurements, we extensively discuss the MIPAS accuracy in the paper. Thus we do not find it too misleading to mention the term 'accuracy' in the title. With respect to the superior accuracy of the cryosampler measurements the reviewer is certainly right, but given the low number of comparisons and the possible effect by less than perfect collocation it is questionable if the lessons learned from these comparisons are indeed more robust than those learned from comparisons with instruments which provide more numerous but less accurate measurements. Nevertheless we will give more weight on the accuracy of the comparison instruments in the revised version.

Comment: *Either way it would have been useful to have a single summary plot of the biases (difference v. altitude) from the different comparison datasets. And there should be at least a clearly stated numerical value, in the abstract and the summary, for the upper limit of the estimated accuracy for the full profiles of both CFC-11 and CFC-12.*

Reply 3: The Summary plots are an excellent idea. However, this means four additional plots (CFC-11 FR and RR; CFC-12 FR and RR). Since the biases are a function of altitude, we find it misleading to represent the accuracy by one scalar number.

Comment: *Precision: the only message seems to be that the nominal precision, ie propagation of noise into the profile retrieval, is probably too small, which (except in*

AMTD
8, C3791–C3833, 2015

the case of particularly noisy measurements) is hardly surprising. There are a number of obvious self-consistency tests that could have been used to quantify the precision of a satellite dataset, eg RMS difference between successive profiles, orbit intersections, SD of the mean in dynamically quiescent atmospheric conditions, but none is used here.

Reply 4: An upper bound of MIPAS noise will be determined by assessment of the CFC standard deviations for midlatitude summer conditions. Further, inclusion of the total MIPAS random error in the comparison will provide further information if the error estimates are realistic.

Comment: *Similarly, 'atmospheric variability' is often cited as a probably contributing factor to the observed scatter, but with no attempt to quantify this (my guess is that this is negligible for CFCs, apart from disturbed vortex conditions).*

Reply 5: We use natural variability as an explanation of the scatter only in a few cases. These are when the comparison is not based on collocated measurements (MIPAS-STR), for comparisons at high latitudes (e.g. ACE-FTS), and/or when a larger coincidence radius had to be used (ACE-FTS). The fact that fits are better when a smaller coincidence radius is used (ACE vs HIRDLS) supports our hypothesis that natural variability plays a role. Further (particularly in the case of the cryosampler comparisons) we see similar effects in other gases, too. This supports the natural variability hypothesis. This argument will be reported in the revised version.

Comment: *With so many of the other instruments measuring both CFC-11 and CFC-12, plotting the correlations between these two products might also have highlighted the different relative precisions of these instruments.*

AMTD

8, C3791–C3833, 2015

Reply 6: We are somewhat reluctant to further increase the already large number of figures.

Comment: *Again at the minimum there should be a citable number for the estimated precision of the CFC-11 and CFC-12 products.*

Reply 7: As stated above, the analysis will be repeated using the full MIPAS error estimates instead of noise only. This will give much better access to upper and lower bounds of MIPAS precision.

Comment: *Ultimately the question a user might be wondering is: is a single profile from MIPAS better than using climatology for a particular time/location? What about a monthly zonal mean? Or an annual global mean? The paper should make some reference to the expected atmospheric variability which ultimately define whether or not the product is likely to be useful at various resolutions.*

Reply 8: The scope of this paper is the assessment of the quality of single MIPAS vertical profiles. The randomization of errors by averaging depends on natural variability and its covariance (c.f. M. Toohey, and T. von Clarmann: Climatologies from satellite measurements: the impact of orbital sampling on the standard error of the mean mean. Atmos. Meas. Tech., 6(4), 937-948, 2013.). Such analyses are clearly beyond the scope of this paper. This paper aims at the characterization of single profiles in terms of precision, accuracy and stability. A paper on climatologies (with our data included) has recently been published by Tegtmeier et al., Earth Syst. Sci. Data Discuss., 8, 759-808, 2015, www.earth-syst-sci-data-discuss.net/8/759/2015/doi:10.5194/essdd-8-759-2015

AMTD
8, C3791–C3833, 2015

Comment: *A related concern is that there is widespread use of terms such as 'excellent agreement' and 'very good agreement' but with little objective basis for these statements, which might lead a casual reader to conclude, for example, that since the overall agreement between MIPAS and HIRDLS is described as 'excellent' (P7610, L14) while ACEFTS is merely 'good' (P7612, L9), the HIRDLS-MIPAS data must be a better match (ie more interchangeable) than ACE-MIPAS, while a comparison of the scatter in Figs 21 and 24 shows quite the opposite.*

Reply 9: We agree and will reword the respective paragraphs.

Comment: *2 STRUCTURE Firstly, given the similarity of the FR/OR results (certainly more similarity between the two modes than between MIPAS and any of the other instruments) the paper could be shortened by combining the two datasets for comparisons.*

Reply 10: We disagree. Retrievals in full resolution and reduced resolution are made with a different retrieval setup. If the resulting profiles are finally well consistent, this can be a result of the study (indeed we should mention this important result!) but should not be used as a premise.

Comment: *Comparisons are not helped by different types of plots and analyses being used for each different instrument rather than, say, one set of plots for limited profiles (cryosampler, MIPAS-B, MIPAS-STR, Mark IV). At least for the HIRDLS and ACE-FTS comparisons similar plots are used.*

Reply 11: The systematic behind our choice of types of plots is roughly this. For all instruments where large numbers of collocated measurement are available (HIRDLS,

AMTD
8, C3791–C3833, 2015

ACE-FTS), we use scatter plots, histograms, and the three-panel difference plots. The latter are also used for MIPAS-STR and ILAS. For measurements which do not provide profiles but point measurements (cryosampler, HATS) different types of plots are adequate, where single data points are shown. And for comparisons which involve a limited number of profiles (Mark IV, MIPAS-B) even another representation is adequate, where profiles are shown, but where the statistics panels are left out. Different data sets carry a different type of information. To force all comparisons into the same plot format would result in a loss of information – or a lot of empty panels.

Comment: *Section 2.2 discusses at length the various other instruments used in the comparisons, however it does seem like a collection of separate contributions (possibly from the various co-authors) without much attempt to harmonise or compare the information presented, or even to avoid repetition of previously-defined acronyms (MIPAS, KOPRA). In any case much of the information is more concisely (and clearly) summarised in Table 1, so this section could be considerably shortened by avoiding repetition of the same information and restriction to relevant details (eg I don't really need to know that the BONBON inlet has a gold pipe, when MIPAS-B made its maiden flight, the details of the cut-off criteria for ACE-FTS profiles, the location of the launch site of ADEOS).*

Reply 12: For the most part we agree. The only exception might be the ACE-FTS cutoff criterion, because this actually leads to a sampling bias if not correctly considered in the comparison. In the revised version, this part will be considerably shortened.

Comment: *On the other hand it would have been useful to include (in Table 1) the spectral resolution or sampling of the different interferometers, and flight altitude of the non-satellite measurements.*

AMTD
8, C3791–C3833, 2015

Reply 13: This is a good point indeed. This information will be included.

Comment: *Perhaps I missed it, but I couldn't find any reference to Table 1 in the main text either.*

Reply 14: A reference to this table is indeed missing and will be included in the revised version.

Comment: *MINOR COMMENTS a) The MIPAS-QWG 'recommends' that the term 'Optimised Resolution' (OR) is used instead of 'Reduced Resolution' (RR). Just mentioning it, that's all.*

Reply 15: We will add a note that the respectice ESA product is labelled OR.

Comment: *b) Given the repeated reference to 'MIPAS Envisat' in this paper, I would suggest shortening to just 'MIPAS', with MIPAS-B and MIPAS-STR being used to distinguish the balloon and airborne versions.*

Reply 16: Agreed.

Comment: *c) The authors frequently use 'comp.' followed by some reference. I don't know what this means. 'Compared with' suggests some contradictory opinion, but its use seems to be to cite supporting opinion.*

Reply 17: Thanks! We were not aware of this meaning and will reword the text

AMTD

8, C3791–C3833, 2015

accordingly.

Comment: *d) Eg '685 and 2410cm^{-1} (4.1 and 14.6μm)' (and several other places) - to me it seems better to keep the wavelength values in the same order as the preceding wavenumber values, so '(14.6 and 4.1μm)'. But that's just my opinion.*

Reply 18: We agree and will change this.

Comment: *e) It would be useful to have a plot showing the CFC-11 emission spectrum (or at least the 850cm^{-1} band) with, superimposed, the various spectral regions used by the other instruments. Certainly much clearer than trying to describe these in the body of the text.*

Reply 19: Excellent idea indeed! Such a plot will be provided.

Comment: *f) There is little reference to the source of spectroscopic data used by the different instrument teams. Could different spectroscopic data account for some of the observed differences?*

Reply 20: This information will be added and discussed.

Comment: *g) Comparing mean profiles (or mean differences) can sometimes be misleading since one dataset or the other may not be symmetrically distributed about its mean. It would be better also to include the median differences on such plots.*

AMTD
8, C3791–C3833, 2015

Reply 21: This is good idea. We will also provide the median.

Comment: *h) Mark IV comparisons: Could these just be due to a 1-2km altitude offset between the profiles? Are opposite biases seen for profiles of quantities which increase with height through the stratosphere, such as temperature or HNO_3?*

Reply 22: Laeng et al., AMT 7, 3971-3978, 2014 contains a MkIV-MIPAS ozone comparison, using the same geolocations (their Fig. 12). These results do not support the altitude offset hypothesis.

Comment: *i) For the figures, it would be helpful to have a clear title showing which instrument is being compared, or else add more detail in the caption, not just 'same as fig ... but for CFC-12'. Also, where relevant, list the latitude for the comparisons as well as the date.*

Reply 23: Here we might have over-exaggerated our anticipatory obedience. In former papers we were forced by the journal editors to apply the "same as" style, so we have applied it here as well. However, we agree that this makes reading difficult and will try to include explicit information in the figure captions.

Comment: *j) Caution may need to be applied when taking the mean of HIRDLS data. Some time ago - I don't know if this is still true - part of the radiometric calibration involved correcting the radiances to match the expected values from climatological profiles so naturally the mean values would agree with climatology but would not constitute an independent bias validation.*

AMTD
8, C3791–C3833, 2015

Reply 24: Many thanks for raising this point. Yes, a one-time normalization of the HIRDLS radiances relative to the Whole Atmosphere Community Climate Model (WACCM) was completed, and applied to all the HIRDLS CFC radiances. This has now been spelled out in Section 2.2.5. This affects the absolute CFC values but the morphologies and relative values are unchanged.

Comment: *k) With the HATS comparisons it wasn't clear whether the HATS data have been averaged to produce a true global mean, or if there is some latitude-weighting representing the locations of the stations. It seems that the oscillations in MIPAS CFC-11 are well correlated with those of CFC-12 and exhibit an annual cycle (min around March, max around September), which is worth mentioning. Also, I would have liked to have seen the FR time period included in these plots rather than as separate plots.*

Reply 25: The latitude-weighting is described in the last paragraph of Sec. 2.2.8 "HATS data": "Their measurements were weighted with the cosine of latitude and an average was calculated for each day." Day should be month, however, so this will be corrected in the revised version.
The correlation of the oscillations will be mentioned in the revised version and the FR and RR periods will be shown together in one plot.
Thanks for these ideas.

Comment: *l) I don't understand why the ESA CFC-11 and CFC-12 MIPAS products have been omitted from the intercomparison - they would have no colocation error. Indeed there is not even any mention of the existence of these products.*

Reply 26: We agree that the existence of these data should be mentioned and we will

AMTD
8, C3791–C3833, 2015

correct this in the revised version. We have not considered these data for comparison because they rely on the same measurements and are thus not independent. Such intercomparison would be an interesting study in its own right but is beyond the scope of this paper.

Comment: *SPECIFIC COMMENTS P7576, L9: for clarity, I suggest adding that these CFC-11 drifts are superimposed on an overall, larger, decreasing trend in atmospheric CFC-11.*

Reply 27: After trying out different wordings we finally found it confusing to merge drift and trend in one sentence. To make this clearer we will change "drift" to "instrument drift".

Comment: *P7579, L8: 0.035 and 0.0625 cm-1. The instrument spectral sampling for the FR/OR modes is 0.025 and 0.0625cm-1, so I would expect the actual unapodized spectral resolution to be larger that 0.0625cm-1 for the OR mode.*

Reply 28: Good point. Will be clarified.

Comment: *P7580, L5-20: This information on altitude range, vertical resolution, sampling, precision would be much clearer if presented on a diagram.*

Reply 29: We will try to include this information in a table.

Comment: *P7580, L8,L12: Altitude coverage down to 5km is claimed, but on P7579 L17 the lowest observation altitude is 7km.*

AMTD
8, C3791–C3833, 2015

Reply 30: We will change this to be consistent throughout the manuscript.

Comment: *P7582, L8: 'equivalent to the satellite version'. Presumably in FR mode?*

Reply 31: This will better be specified in the revised version.

Comment: *P7590, L19: The 1km grid is not actually the 'MIPAS Envisat grid', which implies some fundamental link to the instrument itself, but simply the grid chosen for the IMK/IAA profile retrievals.*

Reply 32: Agreed. This will be reworded.

Comment: *P7590, Eqs 4 and 5: would be simpler, and clearer, if Eq 5 came first and the the standard error could simply be defined as SD/sqrt(n).*

Reply 33: Agreed and changed.

Comment: *P7592, L14: It seems a bit too strong a claim to say that such deviations can be 'reliably' detected on the basis of a single example. To be convinced I would have liked to have seen the comparisons with the cryosampler shown as deviations about, say, the MIPAS monthly zonal mean to verify whether or not individual MIPAS profiles really do agree with the cryosampler in picking out deviations.*

Reply 34: Agreed; a monthly zonal mean will be included in the figure.

AMTD
8, C3791–C3833, 2015

Comment: *P7595, L28: the 'slight bump' in the HIRDLS data is not evident to me. Nevertheless, the MIPAS 150pptv peak at 23 km is curious - is this associated with MIPAS data from a particular latitude?*

Reply 35: The term 'bump' may be too strong. The text will be reworded.

Comment: *P7602, comparisons with ILAS: I think the only point worth making here is that the ILAS data is clearly wrong, so no need to dwell on the details.*

Reply 36: Agreed. This Section will be shortened.

Comment: *P7604, long-term stability: a couple of points worth making here are that the MIPAS detector non-linearity, translated as a radiometric gain error, would be partly compensated by the temperature retrieval so the residual effect on the CFC-11 retrieval is even smaller. Secondly, any such drift with time would not be expected to be latitude dependent and, as a percentage error, probably also altitude-independent, which is consistent with the uniformity shown on the left panel in Fig 16.*

Reply 37: We do not quite agree. Sensitivity studies were made involving the entire retrieval chain from end to end including the temperature retrieval.

Comment: *P7606, L2: I'm not convinced that these deviations represent 'proof' of large natural variability outside the winter polar vortex - as with the CFC-11 plots I would like to have seen comparisons of deviations from the zonal mean before being convinced that these aren't just random variations in the MIPAS profiles.*

AMTD
8, C3791–C3833, 2015

Reply 38: This conclusion particularly refers to the last panel, where the orange line is a multi-annual seasonal zonal mean.

Comment: *P7606, & Fig 18: here the MIPAS CFC-12 profiles extend to 50km but Table 1 lists the maximum altitude as 40km. In any case, the MIPAS oscillations are probably comparable with the vertical resolution of the retrieval at these altitudes so should not be interpreted as representating real atmospheric features.*

Reply 39: The altitude range will be restricted to below 40 km.

Comment: *P7608, L18: The suggestion here seems to be that the consistency between the CFC-11 and CFC-12 profiles measured by MIPAS, and those measured by MIPAS-STR, is evidence that both are retrieving the correct profiles but targeting different airmasses. Perhaps you could include some neighbouring MIPAS profiles as in the cryosampler comparisons to show that such local variability is indeed present?*

Reply 40: We will examine whether this is a systematic effect in the MIPAS Envisat measurements.

Comment: *P7648, L21: Again, I stuggle to see these multiple peaks in the HIRDLS data.*

Reply 41: Agreed, thanks!

Comment: *P7612, L29: so is most of the trend visible in the right plot of Fig 32 due to the change in non-linearity correction shown in the left hand plot? (see also comment*

AMTD
8, C3791–C3833, 2015

below on the Fig 16/32). If so, why not show what happens if the non-linearity trend is subtracted?

Reply 42: No! Drift correction does not nullify the trend but the difference to the HATS trend. Note that the panels have different color scales. Drifts are an order of magnitude smaller than the trends. Especially between 20 and 30 km - the region from which Kellmann et al. drew their conclusions - still look very similar despite the instrument drift.
No correction for the drift will be shown, since providing drift-corrected (real) trends goes beyond the scope of this paper.

Comment: *Figs 6, 9, 22 & 25: a minor point, but these might be clearer if the outline of the comparison histogram was superimposed on the MIPAS data rather than as separate plots in the bottom row. Perhaps also mark mean and median.*

Reply 43: Superimposing the panels seems to be a good idea, but the plot looks very busy so we decided not to do this. We will mark the mean and median.

Comment: *Figs 16 abd 32: it is not clear whether the plots on the right are the absolute trend in MIPAS data or if the effect of the changing non-linearity correction has been included.*

Reply 44: The trend is derived directly from the MIPAS data and no correction for the instrument drift has been applied in this plot. We will make this clearer in the revised version.

AMTD
8, C3791–C3833, 2015

Comment: *GRAMMATICAL/TYPOGRAPHICAL*
P7575, L3: 'abord' should be 'aboard'
P7578, L4: extra space after 'Organisation'
P7578, L12: 'deteriorated' suggested some inevitable and progressive worsening in instrument resolution, whereas in fact it was a deliberate, and instantaneous, change in the way the instrument was operated.
P7578, L17: 'A concluding summary is closing the paper' - clumsy wording - suggest 'The paper concludes with a summary'.
P7578, L23: 'Guyana', with a 'y' is a (different) country, formerly known as 'British Guiana'. Here it should be 'Guiana Space Centre' and 'French Guiana'.
P7579, L1: Suggest '10:00' and '22:00' local time as more conventional time notation.
P7579, L4: 'emission AT the limb' - suggest 'from the limb', the measurement was AT an altitude of 790km.
P7579, L10: 'In correspondence' - suggest: 'In consequence'
P7579, L26: 'was retrieved' should, pedantically, be 'were retrived'.
P7580, L9: 'is 3 up to' and 'about 7 at' - I assume the units are km?
Reply 45: Yes. Missing units will be inserted.
P7587, L2: 'edition' - How does one 'edit' a satellite?
Reply 46: Good question :-) . Will be reworded.
P7587, L4: I assume 'Nakajima et al' is the appropriate reference rather than part of the name of the satellite.
Reply 47: Yes. Will be corrected.
P7592, L18: 'excessive' - perhaps 'excess' is better.
P7593, L19: 'good' should be 'well' (adverb rather than adjective required)
P7597, L24: 'similar as for' - suggest 'similarly to'
P7599, L18: 'spectroscopical' - should be 'spectroscopic'
P7600, L5: should this be 'Sect 4.3' for details? (also P7612 L25).
P7604, L26: presumably 'right panel' for the MIPAS trends.
P7608, L18: 'ue to the fact, that' - remove comma after 'fact'.

AMTD
8, C3791–C3833, 2015

P7613, L1: 'extend' should be 'extent'.

Reply 48: Thanks a lot for spotting! These will all be corrected.

AMTD

8, C3791–C3833, 2015

Anonymous Referee #2

Comment: *The paper describes the validation of MIPAS IMK /IAA CFC-11 and CFC-12 measurements with respect to other satellite measurements (ACE, HIRDLS and ILAS) and measurements from balloon borne instruments (Cryosampler, Mark IV, MIPAS-balloon), stratospheric aircraft (MIPAS-STR) and ground based instruments (Halocarbon and other atmospheric trace species (HATS)). Measurements of the two species are validated distinguishing between the two phases of MIPAS mission, the so-called Full Resolution and Optimised Resolution phases. The paper is very long, since it describes, in the case of some balloon borne measurements, separately the comparison of the MIPAS collocated measurements for the different flights of these instruments....*

Reply 49: The major cause for the length of the paper is that we validate four data products (2 CFCs, each in two different measurement modes.

Comment: *...But then in general it lacks of a global discussion of all results of comparison aimed to quantify MIPAS bias with respect to all correlative measurements in the two different phases of MIPAS mission.*

Reply 50: After redoing the comparisons under consideration of the total MIPAS random error (not only noise) much more will be said on the overall comparison.

Comment: *Furthermore, some comparisons are only qualitative, while it would be very useful to quantify the differences and to compare them with the ones relative to other instruments.*

AMTD

8, C3791–C3833, 2015

Reply 51: Numbers will be given.

Comment: *In particular for the Cryosampler instrument, that is known to be characterized by an accuracy significantly better than MIPAS, no plot of the mean difference is reported.*

Reply 52: With only five comparisons, of which one seems to be an outlier, we consider mean differences as misleading. We think that statistical estimators should be used only if the sample is large enough.

Comment: *An attempt of characterizing precision of the measurements is performed, but not all contributing errors to the random error have been included in the error budget and comparison of the retrieval error with the standard deviation of the differences does not provide determined results since atmospheric variability may contribute and it is difficult to quantify.*

Reply 53: In the revised version, for MIPAS and for the comparison instruments where theses additional error estimates are available, the full random error will be considered. We did not do this in the discussion version because this leads to a inconsistent treatment of the various instruments, because these error estimates are not available for all instruments.

Comment: *The study of the temporal drift on the measurements induced by time dependent nonlinearities has been performed. Given the importance of the study of CFCs for climate studies and studies on ozone hole recovery, the validation of these measurements is needed before using them for climate studies, but major revisions addressing a better characterization of the bias and of the precision in the two phases*

AMTD
8, C3791–C3833, 2015

are needed before publishing the paper.

Reply 54: As said above, inclusion of the total MIPAS errors will give better access to the precision and accuracy.

Comment: *General comments*
1)The title mentions the characterization of accuracy, precision and long term stability of CFC-11 and CFC-12 measurements. Actually, the characterization of the precision is not adequate (see below) and also the bias is not always quantified. It would be probably better just to mention the validation of CFC-11 and CFC-12 measurements and its long term stability.

Reply 55: We consider the title as appropriate since we do discuss precision and bias. Mentioning terms in the title does not imply that one fixed number is presented for each of these terms. We will, however, present the results in a more quantitative way.

Comment: *2) Concerning the characterization of the bias between MIPAS and correlative measurements, for some instruments only the qualitative agreement between MIPAS and correlative measurements is evaluated, and in general no effort has been done for summarizing the results coming from the comparison of the different instruments.*

Reply 56: Results will be summarized in the revised version.

Comment: *Furthermore no mention is done of the estimated systematic errors of both MIPAS and correlative measurements that would help to understand if the*

AMTD
8, C3791–C3833, 2015

detected bias can be explained by their estimation.

Reply 57: The total error budget of MIPAS will be considered. However, total error estimates are not available for all comparison instruments.

Comment: *3) Concerning the characterization of precision, only the retrieval error is considered, but this error, as written in the paper, is not the only contribution to the random error.*

Reply 58: See above.

Comment: *The authors generally come to the conclusion that the retrieval error is an underestimation of the random error, if compared with the standard deviation of the differences, and this is due also to the fact that atmospheric variability, that is difficult to quantify, contributes to the standard deviation of the differences. More accurate methods are available in the literature that allow to estimate the precision (see for example Piccolo and Dudhia, ACP(2007)).*

Reply 59: As a cross-check for the MIPAS precision, the standard deviation of MIPAS data under quiescent atmospheric conditions will be assessed to get an upper bound of the precision.

Comment: *4) In Kelmann et al., ACP (2012) it is said that a discontinuity is present for both CFC-11 and CFC-12 at same altitudes between the two phases of MIPAS mission, as expected due to the different spectral and vertical resolution. Here this discussion is completely missing, while it would be very important to quantify possible*

AMTD
8, C3791–C3833, 2015

biases between MIPAS measurements in the two phases. Without this discussion, the need of distinguishing between the two phases is hard to be justified.

Reply 60: We will discuss this, but we do not think that this is of very large importance: The major differences encountered by Kellmann et al. is due to typically different averaging kernels of the two data versions. This does NOT imply that - within their diagnostic data - any of these data sets is necessarily incorrect. What Kellmann et al. have mentioned is not a bias in the narrower sense of a systematic error.

Comment: *5) In general, it is not sufficient to see if the agreement between MIPAS on ENVISAT and the correlative measurements is good, excellent or bad, it is needed to know how small is the bias at the different altitudes.*

Reply 61: This information is contained in the figures. While we are willing to present some numbers in the text, we feel that this type of information is much better communicated by figures.

Comment: *6) Sect. 4.1.1 The Cryosampler instrument is known to be characterized by an accuracy significantly better than the MIPAS one. The different flights are discussed separately and no plots of the differences are reported. Instead, it would be useful to provide plots containing the mean of the differences of all collocated measurements of the different flights of this instrument. This would allow to quantify the bias of MIPAS measurements with respect to measurements of this instrument at the different altitudes in both the phases.*

Reply 62: Difference plots would mean to subtract isolated point measurements from continuous profile measurements of finite resolution. Note that we have no evidence

AMTD
8, C3791–C3833, 2015

from the cryosampler measurements how the atmosphere behaves between two measurement points, which in the case of MIPAS, the contiguous profile is the result of the retrieval. We are reluctant to subtract quantities which represent so different entities, since we consider this misleading. Instead we extend the discussion of these plots in a more quantitative manner. Cryosampler measurements are point measurements which in no way represent their neighbourhood.

Comment: *Specific comments*
Introduction: This Section is very long, but it is mainly focused on the importance of measuring CFCs and only hints to the problems of validation, availability of other data, previous validation. Probably it should be mentioned the paper by Engel et al., Long term validation of ESA operational retrieval (version 6.0), AMTD, validating four species including CFC-11 and CFC-12 from MIPAS ESA processor with Cryosampler measurements.

Reply 63: Of course we will mention the new Engel et al paper in the revised version. The only reason why we have not mentioned it in our Discussion Paper is that it was not published when we submitted our manuscript.

Comment: *Line 13, pag. 7578 I would add that the fact that 'more scans in the vertical are performed per profile in the RR period' leads to an improved vertical resolution.*

Reply 64: Agreed.

Comment: *Sect.2.1 Line 8, pag. 7579: It is said that MIPAS spectral resolution in the first phase of measurements (until March 2004) is 0.035 cm-1, while it is 0.0625 cm-1*

AMTD
8, C3791–C3833, 2015

in the second phase. Actually, 0.035 cm-1 is the FWHM of the Instrument Line shape (ILS) in the full resolution phase, while 0.0625 cm-1 is the distance between two consecutive nodes of the ILS (or the spectral sampling) in the second phase of the mission(1/(2 MPD), with MPD being the Maximum Path Difference of the interferometer, equal to 20 cm in the FR phase and to 8 cm in the second phase). I suggest to use the same approach to define the spectral resolution in the two cases, i.e. the spectral sampling: 0.025 cm-1 for the FR phase and 0.0625 cm-1 for the Optimized resolution one.

Reply 65: This will be fixed in the revised version.

Comment: *Line 11, pag. 7579 The second phase of MIPAS measurements starting in January 2005 is mentioned in the paper as 'Reduced Resolution' (RR) phase, while MIPAS Quality Working Group agreed to call it 'Optimized Resolution' phase. Indeed, the discontinuity in MIPAS measurements that occurred in January 2005 does not involve only a change in the spectral resolution, but also a change in the vertical and horizontal resolution. The reduction in the measurement time due to the lower spectral resolution was exploited to perform more measurements both in the vertical and in the horizontal domain. This change allowed to obtain, at least for the original target species, an optimized compromise between spectral and spatial resolution. It is recommended to change Reduced Resolution (RR) to Optimised Resolution (OR) in all sections of the paper.*

Reply 66: We will mention that the related ESA product is called 'optimized resolution'. Since we have introduced our RR data product already before the ESA recommendation was made we are reluctant to change the name of our data product after so many years.

AMTD

8, C3791–C3833, 2015

Comment: *Sect. 2.2.3 L.11: It is said that the spectral resolution of MIPAS balloon is equivalent to the satellite version: considering that the spectral resolution is different in the two phases of MIPAS on ENVISAT mission, I would recommend to write explicitly the spectral resolution of MIPAS balloon instrument.*

Reply 67: Agreed.

Comment: *Line 18 You should also cite Arnone, E., Castelli, E., Papandrea, E., Carlotti, M., and Dinelli, B. M.: Extreme ozone depletion in the 2010âĂŽÄì2011 Arctic winter stratosphere as observed by MIPAS/ENVISAT using a 2-D tomographic approach, Atmos. Chem. Phys., 12, 9149-9165, doi:10.5194/acp-12-9149-2012, 2012.*

Reply 68: We do not quite see why this paper is relevant in this context.

Comment: *Sects. 4.2.1, 4.5.1 It would be useful to have the plots of the mean differences, including all collocations of the two flights, together with the standard deviation of the mean.*

Reply 69: Tukey's rule of 5 says: calculate the nth moment only if at least 5^n observations are available; i.e. the mean should be calculated only when at least 5 observations are available; the standard deviation only if there are more than 25 observations etc. We have only 3 matches with Mk IV, thus statistical analysis does not seem adequate to us.

Comment: *Sect. 4.1.3 Why are the results of the two flights not combined in order to*

AMTD
8, C3791–C3833, 2015

reduce the effect of random errors and to have a clearer determination of the bias? Probably you want to distinguish between different latitudes, but since results are very similar, the combination of all collocated differences would allow to reduce the random error.

Reply 70: The left panels are already quite busy, and the profile shapes below 15 km are quite different. This seems to us to counterbalance the modest advantage of lower noise of the mean.

Comment: *Why are systematic errors of the two instruments not shown? They could be compared with the detected bias between the two instruments!*

Reply 71: For the revised version we will include the total error for all instruments when available. For this version we have included noise only because for some instruments the total error estimate is not available and we wanted to be self-consistent.

Comment: *Sect. 4.1.6 Line 5: what does it mean that 'the ACE-FTS errors were estimated directly from the fit residuals'?*

Reply 72: The ACE-FTS errors are the random errors from the least-squares fitting process, the square root of the diagonal elements of the co-variance matrix. Also, the error budget of the version 3.5 ACE-FTS data contains an additional term in the reported error. This term is derived from the difference between a retrieved CO_2 volume mixing ratio (VMR) profile and the assumed CO_2 VMR profile employed in the pressure/temperature retrieval and is a measure of the ability of the retrieval system to reproduce the fixed input profile for the given occultation. (Boone et al. 2013 -

AMTD

8, C3791–C3833, 2015

http://www.ace.uwaterloo.ca/publications/2013/Version3.5retreivals2013.pdf)

Comment: *Line 8, pag. 7600. 'The drift estimate due to the detector aging is only based on drifts between 35° N and 35◦N DUE TO THE LACK OF DATA'?. It is not clear what 'lack of data' means. Actually, the full mission L1 dataset was released on the 22nd of May 2015 (https://earth.esa.int/web/guest/missions/esa-operational-eomissions/envisat/news/-/article/envisat-mipas-level-1b-dataset-processed-with-ipfversion-7-11) is available), so the full mission L1 data with Non-linearity correction were available for the comparison with L1 V5 without this correction. Or do the authors refer to the particular orbits provided for the study of the drifts?*

Reply 73: Yes, for the drift analysis particular orbits were provided, where the handling of the non-linearity correction was the only change compared to the bulk level 1 data.

Comment: *Last sentence: 'The agreement between MkIV and MIPAS is similarly good for FR and OR': I think we can learn more from these comparisons. For OR measurements MkIV is larger than MIPAS (and it is said that MkIV has a positive bias), for FR measurements MkIV is generally smaller than MIPAS. If MkIV can be considered a stable instrument, I think that this is a clear indication of a discontinuity between FR and OR CFC-11 (as also asserted by Kelmann et al., ACP, 2012)*

Reply 74: The discontinuity stated by Kellmann et al. refers simply to the effect of different averaging kernels, which will always cause a discontinuity when single altitudes are assessed. Furthermore, a comparison including three profiles does not seem adequate for us to infer a global discontinuity in the data.

AMTD

8, C3791–C3833, 2015

Comment: *Sect. 4.2.4 Line 23: 'Even though all values of the HATS measurements lie within the standard deviation of the MIPAS ENVISAT measurements': actually most of the HATS measurements are outside the standard deviation of MIPAS measurements. Please clarify.*

Reply 75: Agreed. Will be rephrased.

Comment: *Sect 4.3 I partly agree with the conclusions reported in this section. First of all it is not easy to extract numbers from the two maps, and a map where the ratio between drift and trend is reported would help. Surely the estimation of the instrumental drift is done only for altitude regions where the CFC-11 relative trend is the lowest, less than 10%, while the drift estimation is not smaller than 1%. Furthermore, the error in the trend estimation is surely larger than 10% in these regions. For the regions where the trend is larger, there is no estimation of the drift. And the instrumental drift is expected to be altitude dependent. In summary, the correction in the trend due to the drift is not negligible where the drift has been measured, nothing can be said in the regions where the drift has not been measured.*

Reply 76: Since the color bars in the left and the right panel of Fig. 16 are not the same, the trend is still larger, even at lower altitudes. Regarding the trends above 20 km, these are still significantly larger than the drift - where they were actually estimated. Even though no drifts could be estimated above 25 km there is no indication that drifts ten times larger than below this altitude are to be expected, so that the trends of down to -50% and up to more than 20% are still about an order of magnitude larger than the drift.

Comment: *Sect 4.4.1 See comment done for Sect. 4.1.1*

AMTD
8, C3791–C3833, 2015

Reply 77: See reply above.

Comment: *Sect 4.4.3 See comment done for Sect. 4.1.3*

Reply 78: See reply above.

Comment: *Sect. 4.4.6 Should the bump at the upper end of the profile be reduced if the mean of the profiles was computed weighting each profile with cosine of latitude?*

Reply 79: Since the bump is explained, we see no advantage in such a cosmetic operation. The bump simply results from the fact that the number of valid coincidences depends on altitude and that the altitude region where the bump is is dominated by lower latitude profiles.

Comment: *Sect. 4.4.7 Line 2, pag. 7613 : due to lack of data: see comment relative to Sect. 4.1.6*

Reply 80: A comment will be added.

Comment: *Sect. 4.5.1 I would highlight the common features in the results with ACE and HIRLDS, positive bias of MIPAS below 18 km, negative bias above.*

Reply 81: The common features between the comparisons will be highlighted in the conclusion.

AMTD
8, C3791–C3833, 2015

Comment: *Sect. 5 Summary and conclusions A discussions on the MIPAS bias with respect to the different instruments at the different altitudes, trying to find the common features and comparison between FR and OR measurements is missing. In Kelmann et al., ACP, 2012 it is said that a discontinuity is present for both species at same altitudes. Here this discussion is completely missing, even if this would justify the need of distinguishing between the two phases.*

Reply 82: We already have replied to the discontinuity issue above. The discussion of the common features of the various comparisons will be extended.

Comment: *Technical corrections*
Sect. 2.2.4 L.3 of pag. 7584: 'The shown 1-sigma error of the MIPAS-STR'. Where is it shown?

Reply 83: In Figure 4. Reference to figure will be added.

Comment: *Sect. 4.1.5 Line 14, pag. 7596: please write that the combined error is reported in the left panel.*

Reply 84: It is reported in the right panel.

Comment: *Sect. 4.2.1: Line 3, pag. 7601: MkIV measurements is represented by the 'black' curve.*

Reply 85: Will be corrected.

AMTD
8, C3791–C3833, 2015

Comment: *Sect. 4.2.2. Lines 25-26: I guess that you refer to the standard error of the mean, not to the standard error of the differences. The sentence starting with 'Even though ...' is very confusing, please make it clearer.*

Reply 86: No, we mean the standard error of the differences. For clarity we change to "standard error of the mean difference".

Comment: *Sect. 4.1.4 line 2, pag. 7595: 'showing minimum values around 16-17 km ..'. Please specify: minimum values in the differences.*

Reply 87: Agreed.

Comment: *Sect 4.4.3 Line 26-27: probably you should write 'standard deviation of the mean' instead of standard deviation of the differences.*

Reply 88: It is the standard deviations of the differences.

Comment: *See Sect. 4.1.2 The reference to Fig. 28 is missing.*

Reply 89: Thanks for spotting!

Comment: *Sect. 4.5.4 Line 12, pag. 7615: please replace: 'at an order' with 'of the order'.*

Reply 90: Agreed.

AMTD
8, C3791–C3833, 2015

Comment: *Sect. 4.6 Line 1, pag. 7615 Please remove the comma.*

Reply 91: Agreed.

AMTD
8, C3791–C3833, 2015

Anonymous Referee #3

Comment: *This is a validation paper for MIPAS CFC-11 and -12 by way of comparison to coincident measurements by other instruments. While fairly thorough, the paper is overly long, contains too much extraneous detail, and sometimes makes comparisons that add little to the result. The wording of several passages used colloquialisms that may obscure the meaning to a reader not fluent in English. I do not recommend publication in its current form. The paper needs to be rewritten to be more succinct, and clearer in its conclusions.*

Reply 92: Inclusion of the full MIPAS error budget will allow clearer conclusions. For the other points, see replies to the specific comments.

Comment: *Abstract: The software heritage of the processing system doesn't need to be in the abstract.*

Reply 93: Not clear what is meant. The level 1 data versions are critical for traceability of the data, and the processor used needs to be specified because there exist other processors with their own data products. Not specifying the processor used would lead to a huge confusion.

Comment: *Introduction: This includes the "story" of CFC-11 and -12 and its effect on ozone. This is just too much information that is likely already well known by anyone interested in this paper. The introduction could easily be reduced to one paragraph as to why you're comparing measurements of CFC-11 and -12 rather than why you're measuring these in the first place.*

AMTD

8, C3791–C3833, 2015

Reply 94: We think the introduction is the right place to put our work into a more general context and to motivate the study. Some of the analyses (particularly stability) is directly motivated by the issues mentioned in the introduction (expected change in trend.) Thus we have decided not to remove this information entirely.

Comment: *2.1: That Envisat was launched from Guyana is just more extraneous information (just as SCISAT was launched from Vandenberg, etc.) Most of what is the paragraph from pg 7579 line 19 through page 7580 line 19 can be put into a table.*

Reply 95: Agreed.

Comment *2.2.4, 2.2.5, 2.2.6, 2.2.7: While the pertinent details of other instruments should be reported, a lot of this information can go in tables.*

Reply 96: Agreed.

Comment *3.: While near-coincident observation against altitude is very useful, it does not always adequately remove meteorological variability (as noted in the paper âĂŽÄì particularly comparisons with HIRDLS). Reading the paper, I was wondering why the authors did not compare the ratio of results to a chemical clock like N2O (or even CH4), where these measurements were available. If the N2O results are reliable, such comparisons help clarify when a difference is truly because of meteorological variability or because of retrieval bias. I would have liked to see something like this addressed, and such comparisons, where possible, would make this a stronger paper.*

Reply 97: In principle this is an excellent idea. The N_2O MIAPS data product,

AMTD
8, C3791–C3833, 2015

however, may have its own problems and has not yet been validated. Thus the suggested method may raise more questions than it answers and we have not chosen this approach in order to avoid further uncertainties.

Comment: *Pg 7388, line 22: Change "artefacts" to "artifacts."*

Reply 98: There seem to be differences between British and American English. These issues will be homogenized by the language editor of the journal.

Comment: *Pg 7590, line 9: Add comma after "e.g." so it reads "...too small, e.g., if not all..."*

Reply 99: Agreed, thanks!

Comment: *Pg 7592, line 11: Suggest changing colloquial phrase "decently represented" to "well represented."*

Reply 100: Agreed.

Comment: *Pg. 7592, line 1-2: The wording of the sentence could apply to distance as well as mixing ratio. This is confusing to read. You could say "the closest co-coincident profiles have the highest bias" or something similar.*

Reply 101: Agreed.

AMTD
8, C3791–C3833, 2015

Comment: *Pg. 7594, line 7: Suggest changing "supposedly" to "presumably."*

Reply 102: Agreed.

Comment: *Pg. 7594, line 27: I don't mean to be facetious, but "steel blue" is a color of a car, not a line produced on paper.*

Reply 103: Agreed.

Comment: *Section 4.1.5 : This is a long, long paragraph âĂŽÄì more than 40 lines. Can this be made shorter, and better broken up into sections?*

Reply 104: Agreed; will be subdivided.

Comment: *A "global" comparison of MIPAS and HIRDLS on an altitude grid, and over the time scale of coincident measurements is not very useful. Pole-to-pole changes in the tropopause height plus secular change in CFC-11 make comparisons on an altitude grid less than definitive. If anything, show zonal means on seasonal timescales as in the appendix. This would avoid issues of (and paragraphs written about) artificial bumps.*

Reply 105: We will discuss Fig. A1 in order to highlight latitudinal differences. However, this does not make the global three panel figure obsolete, because it is needed for the discussion of the precision.

AMTD

8, C3791–C3833, 2015

Comment: *Section 4.1.6: Another long, long paragraph. Same issue with the global comparison on altitude grid.*

Reply 106: See above.

Comment: *Pg 7599, line 16: A profile may be curved but still "straight" within the error bar. That doesn't make it "suspicious."*

Reply 107: The error bars shown are the typical single profile errors. The increase is also seen in the mean profile. The increase of the mean is significant w.r.t. the standard error of the mean.

Comment: *Pg 7601: line 6: If a retrieval is negative, it's likely because they're retrieving the linear amount and not the logarithm, close to the noise limit. If the negative MkIV results have an error bar that crosses zero, then there's no problem.*

Reply 108: The error bar does not cross zero at all altitudes.

Comment: *4.2.3 : Looking the ILAS error bars in Figure 14, the results of ILAS are so uncertain as to make this comparison useless. Suggest cutting out comparisons to ILAS.*

Reply 109: Again: The error bars in the left panel are the single profile error bars. Since we have more than 5000 matches, the analysis can still become significant because the standard error of the mean difference is close to zero.

AMTD
8, C3791–C3833, 2015

Comment: *Pg 7602, line 18-19: The meaning of "general turning points of the profile" is not clear.*

Reply 110: Agreed.

Comment: *Pg 7606, line 3: "which provides proof of large natural variability." This is not proved beyond any metrological variability. What does "natural variability" mean here? How is this "natural variability" distinguished from the natural variability of, say, a chemically active species like ozone, PAN or chlorine nitrate?*

Reply 111: We understand that the reviewer means 'random errors' with 'metrological variability'. The standard deviation of the 1st April collocated data is much smaller than the difference between the collocated profiles and the seasonal mean profile. This rules out the 'metrological variability and proves the natural variability. Or was this a typo and it is meant 'meteorological variability'? The contrasting mention of chemical variability in the next sentence seems to suggest this interpretation. But we do not see that it is important in this context that the natural variation is caused by meteorological rather than chemical variation. We admit that it might not have been clear that our statement refers only to the last panel. This will be mentioned.

Comment: *Pg 7606, line 18-19: Unnecessary comment about MkIV not showing negative values.*

Reply 112: Agreed.

Comment: *Pg 7607 line 27- pg 7608, line 6: If you're trajectory calculations are not*

AMTD

8, C3791–C3833, 2015

good, then don't do the comparison. See comment about using N2O for accounting for metrological variability.

Reply 113: The trajectory calculations are good, except for one particular situation: Only on 31 March they cannot explain differences below 15 km.

Comment: *4.4.5: Paragraph too long.*

Reply 114: Agreed.

Comment: *Pg 7610, line 2: "seems to see" is an unnecessary colloquialism.*

Reply 115: Agreed.

Comment: *Pg 7610, line 10. "even though this effect is supposedly minor." Why should it be minor?*

Reply 116: Because the coincidences contain many measurements with quiescent atmospheric conditions and because the large number of available coincidences allow a narrow collocation criterion.

Comment: *Pg 7610, line 19: "seems to see" is an unnecessary colloquialism.*

Reply 117: Agreed.

AMTD
8, C3791–C3833, 2015

Comment: *4.4.6: Paragraph too long.*

Reply 118: Agreed.

Comment: *Pg 7611, line 25: If the "bump" is misleading, then don't do analyses that show it.*

Reply 119: We think that suspicious features in the data should be explained. Otherwise the bump could be harmful for the confidence in both ACE-FTS and MIPAS data. Thus we consider it as important to show that there is an explanation which does not need a data artefact as explanans.

Comment: *4.5.1: Refer to the profiles by dates, not "first case" or "second case."*

Reply 120: Agreed.

Comment: *Pg 4614, line 4: "vice-versa above that altitude" is not clear phrasing.*

Reply 121: Agreed.

Comment: *4.5.3: As is the case with CFC-11, the ILAS results and comparison add little to the paper.*

Reply 122: We agree that from this comparison we do not learn a lot about MIPAS, but we learn about ILAS. We will shorten the related paragraph.

AMTD
8, C3791–C3833, 2015

Comment: *Sec 5. Suggest reiterating what "FR" and "RR" mean for the reader who jumps to the conclusions section.*

Reply 123: Another reviewer has blamed us for re-introducing abbreviations, thus we are reluctant to do this. Instead, we simply avoid the use of the abbreviations in this section and use the full words instead.

Comment: *Table 2: Include standard deviations in average miss time and distance?*

Reply 124: Since we do not use this information in the analysis, we think that additional numbers might rather be confusing than clarifying.

Comment: *Figure 4 and others: It may be consistent to have the vertical axis of each graph go to 30 km, but it reduces the utility of those profiles that reach only, say, 20km. It's also a little confusing to have two plots of each comparison profile, e.g., MIPAS-STR on its grid and on a new grid. Best just to keep plots on the same grid as MIPAS.*

Reply 125: Agreed for the altitude axes. However, we would like to keep the original profile in the plot as well in order to allow comparison to other publications. To guide the eye of the reader we will use a thinner line though.

Comment: *Figure 18: To ask the reader to skip back to Figure 2 for the caption is too many figures back. Just repeat the caption using CFC-12.*

Reply 126: In previous papers of this publisher the 'same as Fig xy' style was

AMTD
8, C3791–C3833, 2015

requested by the journal editors and thus we applied it here in anticipatory obedience. But we agree that this is confusing and will try again to push the explicit self-containing figure captions.

Comment: *Figure 26-32: Same comment as above*

Reply 127: See above.

AMTD

8, C3791–C3833, 2015