# Reply to comments of reviewer #2

Gabriele Stiller et al.

Reviewer comments are in black, while our replies are in blue.

**General comments:**

The manuscript of Stiller et al. (2023) describes the new Version 8 (V8) of MIPAS CFC-11, CFC-12 and HCFC-22 measurements. The changes performed in the retrieval processes between the previous version (V5) and V8 are very well detailed and justified, and the improvements observed in the resulting data sets are substantial. The data are well characterized by their averaging kernels and uncertainties, and furthermore the authors introduce a new representation of their products on a coarser grid which will be very useful for the users.

Therefore, providing a description of these new V8 data sets is important for the scientific community and well in the scope of AMT. I recommend the publication in AMT, and I list some questions and suggestions below that might help clarifying a few points.

We thank the reviewer for their positive and encouraging assessment. In the following we reply to each of their comments.

**Specific comments:**

- p.3 Section 2 Description of MIPAS modes: It was not clear to me why while the NOM modes have two versions (FR and RR before and after Jan 2005 respectively), the MA has only one mode. I had to find some MIPAS documentation to see that the MA only started in Jan 2005 (right ?), so maybe put this information in the text here, for MIPAS non-expert.

Agreed, we will do this.

- p.4 CFC-12, l.96: “v5 CFC-12 might have a slight low bias”; and “Zhou et al. found good agreement”: Can you be more specific and give numbers? To check that v8 is an improvement, ideally one would like to see some validation of v8 in this manuscript. If not possible but the authors (unless you could consider it?), providing numbers (and sign of the bias) could help to check more quantitatively that there will be an improved agreement with the new version.

We think that additional validation would make this paper too long. We plan to provide the comparison with CFC observations from other instruments as a forthcoming paper. Nevertheless, we will be more specific and provide numbers on the differences of the v5 versus v8 data versions, in context with biases found for the v5 data versions.

- p. 5, figure1: I find it hard to follow. Maybe you could use different colors for the arrows concerning the 3 different species, or consider making 3 different schemes?

The problem with different colors is that we use color coding for the different categories of data (i.e. database, V5, V8) already. Beyond this, some arrows issue into other arrows, that is to say, some arrows represent two or more species. We thus prefer to leave the figure as is.

- p.10-14: Fig. 2 to 7: I would merge these 6 figures into 2 figures (Fig. 2 to 4; and Fig. 5 to 7). In Figs. 5-7, the legend describing what is the black lines/points is missing. Of course, the same remarks for the other species.

We will combine the figures as suggested (for the other species as well). For Figs. 5 to 7 we will explain that we have coloured only every 5th averaging kernel row, and the black lines are all averaging kernel rows for the retrieval grid altitudes in between.

- Sect 3.4: Table 4: Maybe a Figure as in von Clarmann 2009a would be nicer than numbers in a Table? Unless the authors think the exact numbers would be used by the community? They can decide. Same remark for Table 8 and 12.

We prefer to stay with the table and the numbers. Reviewer#1 made the point that model comparison could be done at the exact geolocation of the values along the vertical profile, which requires exact knowledge of the numbers.

- p. 13 l.182-183: "The horizontal smearing…is always narrower than the… horizontal distance between the nominal geolocations of two subsequent limb scans": can you give this horizontal distance in the text?

Sure, this will be done. For the FR phase the horizontal distance was about 500 km, while for the (NOM/MA/UTLS1) mode of the RR phase it was (410/430/290) km.

- p. 13, Sect. 3.5:

1) The authors describe the effect of new version vs v5 in term of profiles (biases, avoided negative vmr). How the uncertainty and the vertical and horizontal smoothings change from v5 to v8? Are they also improved or similar? Same question for the other species.

We will discuss changes in the vertical resolution and the spectral noise-induced uncertainty in the revised version in detail. In short, major changes occur for CFC-11 only. Here we have reduced the regularisation above 20 km in order to improve vertical resolution. This is now better than 3.5 km up to about 30 km altitude, but increases strongly from there on. As a result, the ESD (estimated standard deviation = retrieval uncertainty due to noise in the spectra) increases from about 6 pptv below 16 km up to 10 pptv at 30 km and larger above (see Figs. 11 and 12 in the manuscript, red line with crosses). For CFC-12 the vertical resolution and ESD are similar to the v5 data version. For HCFC-22, vertical resolution and ESD could both be slightly improved (however, by less than 0.5 km and less than 1 pptv, respectively) by using more spectral grid points in the retrieval as described in Section 5.1.

2) The authors give in the text some numbers for systematic differences between the 2 versions. Would it be possible to give some visualization of these changes? Some maps v5 vs v8 or a map of differences? It could be in a Supplement if the authors do not want to increase too much the size of the paper. Same remark for the other species.

We will provide difference plots (maps or monthly zonal mean cross sections) for all three species for illustration.

3) Is V8 closer to other reference measurements (does the bias given in this section go in the good direction with similar values than observed bias in the literature?) What is the impact of the new spectroscopy for CFC-12? Since it is the dominant systematic source of uncertainty, does it affect the bias here? Or is it only the regularization that matters more for this species?

Reviewer #1 had a comment pointing in the same direction. We will provide numbers on the differences of the v5 versus v8 data versions, and put them in context with biases found for the v5 data versions. And we will discuss the impact of the new spectroscopy in more detail.

4) I find it hard to see an improvement in the step between FR and RR by simply looking at Fig 9 and at Fig 12 of Kellmann et al. (2012). Because this is an important improvement for users and a strong motivation to change from v5 to v8, would it be possible to show this result more clearly? E.g. some examples of time-series with an automatic step detection and some numbers given for the bias before / after Jan 2005 for the v5 and v8? Same remark for CFC-11 (p.25,l. 185).

We will show time series as line plots for some example latitude/altitude bins of CFC-11 and CFC-12, comparing v5 and v8.

- p.17, Fig. 10: could be merged with Figs. 17 and 24.

We prefer to leave the Figures separated. By this, the figures are closer to the place in the manuscript where they are called. We reduce the number of figures already by combining the error budgets and averaging kernels for each species.

- p. 18, l. 222-227: Similar remark as above for CFC-12 p.4: the authors should give numbers of the observed bias in v5. And/or perform new validation here using reference measurements.

We think that additional validation would make this paper too long. We plan to provide the comparison with CFC observations from other instruments as a forthcoming paper. Nevertheless, we will be more specific and provide numbers on the differences of the v5 versus v8 data versions, and put them in context with biases found for the v5 data versions.

- p. 24, l.268-269: use FR, HR and MA as elsewhere (instead of "high resolution measurements,. . . ) ?

This will be corrected.

- p. 25, Sect.4.5:

1) Same remark as for CFC-12: hard to visualize how much the data set is improved without additional maps (e.g. the reference of Fig.8 is given, but we don't have the same fig. for v5 in order to compare).

We will provide difference plots (maps or monthly zonal mean cross sections as examples) for all three species for illustration.

2) Same remark for the step reduction between FR and RR as above (hard to see only by looking at Kellmann et al., 2012).

We will show time series as line plots for some example latitude/altitude bins of CFC-11 and CFC-12.

3) What is the part of the improvement that is due to the change of spectroscopy? What is the main driver of the improvement in the bias between FR and RR?

We have tested the retrieval by just exchanging the spectroscopic data, but leaving everything else, including the version of the measured spectra, as in the previous V5 retrieval. The change of spectroscopic data alone accounts for a reduction of about 4% over the altitude range of the profile up to $\sim$25 km. Additional changes of parameters (new level-1b data, the 2D temperature field, better information on pre-fitted constituents or a priori information on interferents) contribute another 2 to 5% of difference between V5 and V8 CFC-11. We will add this information to the revised version of the manuscript. Regarding the step between FR and RR data, we think that the improved temperature information (including the 2D fields) is a major contributor to the improvements of the retrievals. Further, any improvement in the retrieval chain before CFC-11, that affects the CFC-11 spectral region will contribute to an improved CFC-11 retrieval.

-p. 36, Sect. 5.5:

1) The maximum in the upper troposphere due to the old spectroscopy is also present in ACE-FTS and ground-based data using the same spectroscopy ?

We have checked ACE-FTS v3.5 versus 4.1 data. ACE-FTS retrievals have used the new spectroscopic data provided by Harrison (2016) from v4.0 on. There is a difference between the two data versions, with v4.1 being lower in general than v3.5. However, the data are too noisy to detect or exclude a maximum in the tropical upper troposphere. We need to keep in mind that the coverage of the Asian monsoon by ACE-FTS data is sparse. Ground-based remote sensing data have typically a too coarse vertical resolution for such a small-scale feature in the profiles. We have compared to tropical in situ data, however, we could not compare to data obtained in the Asian monsoons (presumably, there are none, or at least, we are not aware of any).

2) Again, hard to see the improvement without appropriate v5 map/figure.

We will provide difference plots (maps or monthly zonal mean cross sections as examples) for all three species for illustration.

- p. 37, Conclusions, l. 396-397: “Several implausible features in the V5 data.... Have disappeared or... reduced”. As said above for the 3 individual results sections, I think that giving concrete examples of these features present in v5 in figures or maps, then with the same examples “resolved” using v8, would help the visualization of the improvements.

We will provide figures on the comparison of version 5 and version 8 for illustration. However, these can be some examples only; otherwise the number of figures would be immense.