Replies to Referee #2 on the manuscript 'Assessment of the error budget for stratospheric ozone profiles retrieved from OMPS limb-scatter measurements ' by C. Arosio et al.

We thank the reviewer for the time she/he spent reading the manuscript and constructively commenting on the paper. In the text below, we address the comments from the Referee #2. Referee's comments are shown in italic and authors' responses are highlighted in blue.

Dear authors, Congratulations on this comprehensive analysis and well-written paper. It was a pleasure to review this work. Error/uncertainty characterization of satellite retrievals is more and more gaining the attention it deserves, and this paper clearly demonstrates the way to go about this. It is (almost, see below) complete in terms of identified sources of errors, and the propagation using numerical techniques ensures sensitivity to those tricky situations where random errors turn into systematic contributions.
The only source of errors I did not see addressed is that related to horizontal smoothing: if the actual sensitivity is not centered on the tangent point (and the atmosphere is inhomogeneous) this will lead to additional errors/uncertainty. Addressing this is not trivial (you'd need horizontal AKs and a 3-D model of the atmosphere, e.g. from a reanalysis), but it has been done before, e.g. for MIPAS on ENVISAT(von Clarmann et al., AMT, 2009, and application in Cortesi et al., ACP, 2007). I think this deserves to be at least mentioned.
That said, my other comments are all minor. See below for the list.
Again, congratulations for this work!
A Referee.

Thank you for the appreciation and the positive review.
It is a good point about the horizontal smoothing and uncertainties coming from the inhomogeneity along the line of sight, we agree that it is also a relevant source of error especially in the presence of sharp gradients along the line of sight. Unfortunately, we are not able to investigate this issue properly because our radiative transfer model and retrieval approach are 1-D. To address your comment we added to the introduction some text about this topic, as follows:
'Another source of error which is not considered in this study but relevant for limb observations, is related to the horizontal smoothing, tackled for example in von Clarman et al. (2009) and Cortesi et al. (2007). Namely, limb observations smooth the atmospheric variability over a 250–400 km region around the tangent point (TP) along the line of sight (Zawada et al. 2008). In addition, a 1D retrieval cannot account for gradients along the line of sight, leading to an additional uncertainty component, which is not expected to be relevant on average, but might play a role for atmospheric scenes characterized by strong gradients along the line of sight, e.g. in the presence of the ozone hole (Zawada et al. 2008) or sharp reflectivity gradients.'

Abstract:
- Maybe add a few words on the impact of the cross section source?

We added a mention of this in the abstract, although we don't consider this contribution in the final error budget.

Intro:
- I'd reorder the arguments for the importance of the error characterization: foremost, correct uncertainty estimates are needed to ensure correct geophysical interpretation. Only in 2nd order are they important in the validation. Actually, the validation should validate also these

ex-ante/prognostic uncertainty estimates (as you do).

Thanks, we re-formulated this paragraph accordingly.

- Can you not create a short reference, e.g. Rault and Loughman (2012), hereafter RL12. And then change "The authors" into "RL12". To be checked against the AMTD style guidelines.

I have checked in the guidelines but could not find any information about such abbreviations, I will ask in the typesetting phase if it is possible.

Sect. 2:
- I guess there is also an official NASA retrieval that deserves to be mentioned before presenting the Bremen retrieval? Or is the Bremen retrieval the official product?

Thanks for pointing this out, we added a mention of the official NASA product at the beginning of the section.

- The SNR is in principle something to characterize random noise. But perhaps you

The reviewer is right, there are some systematic structures in the fit residuals, which are not explicitly taken into account. This however seems to have a negligible impact on the retrieval, as depicted in Fig.2 below (see discussion there).

- out of curiosity: how does your vertical resolution estimate compare to actually measuring the FWHM of the AK row? No need to invest much time in this, as you wish.

In Fig.1 below you can find a plot of the vertical resolution together with the FWHM of the AK, for an example case. As you can see they agree very well.

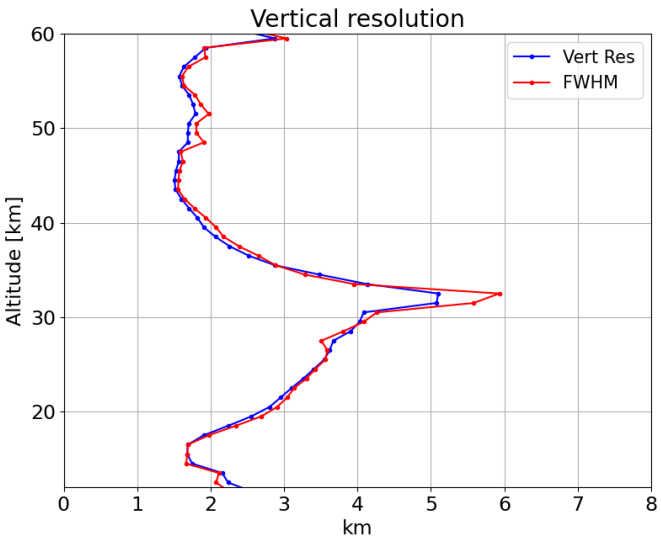

Figure 1: Comparison of vertical resolution with AK FWHM for an example case.

Sect. 3:
- if I understand correctly, you first derive a SNR from the residuals and then use this as starting point for the Gaussian generator, forcing the input errors to have a normal distribution.

Could you not take the actual residuals (of many retrievals) as a distribution to draw (randomly) from? That way, you'd also propagate potential systematic effects, no?

Thanks for this comment, which is directly related to the previous comment about systematic errors and SNR. Yes, we derive the SNR from the residuals and generate 50 normally distributed samples. We then compared the results of the usage of these 50 generated noise samples with the usage of the actual SNR sample from the measurement. In Fig. 2 below we show the results of this comparison, for a case of temperature perturbation (2K). The left panel shows the 50 retrieved profiles, using as inputs synthetic intensities with Gaussian SNR added, together with their mean in black, the average unperturbed profile in blue (same methodology) and a profile retrieved from synthetic intensity but using the actual SNR from the OMPS observation in green. On the right panel the relative differences 'black - blue' and 'green - blue' are shown. After looking at several cases, we found a very good agreement. This points out that possible systematic errors in the SNR are a secondary effect.

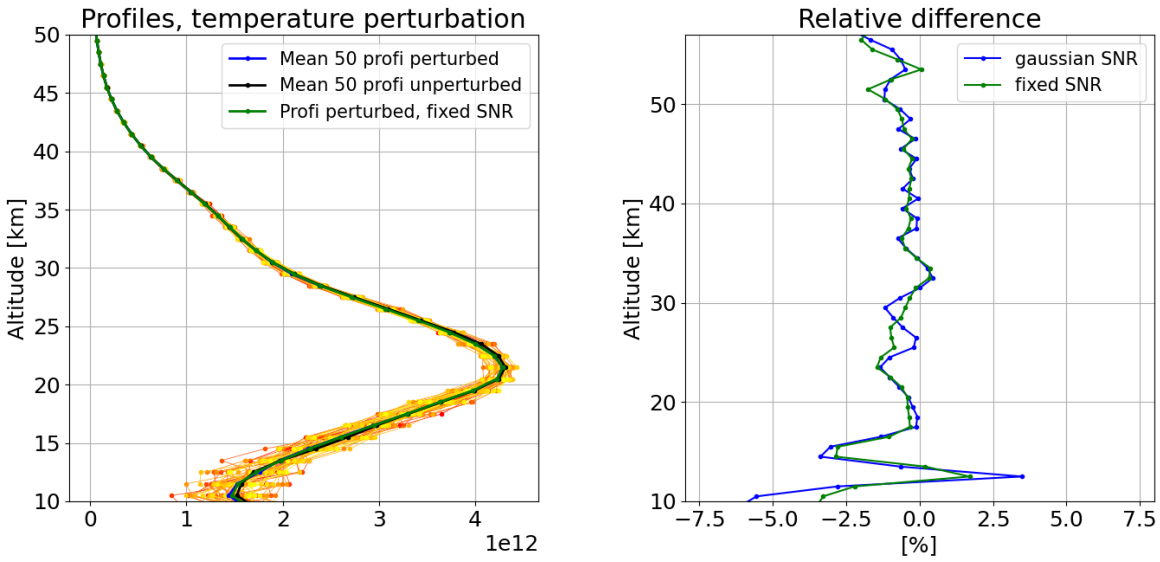

Figure 2: On the left panel, 50 synthetic perturbed retrievals with their mean (black) and the average unperturbed profiles (blue), when using Gaussian SNR samples. Over-plotted, a single perturbed profile retrieved using OMPS SNR. On the right, relative difference of the perturbed profiles (black and green) w.r.t. to the mean unperturbed, for the two cases.

Sect. 4:

- errors on these parameters can all be assumed to be random?

- Yes, to our knowledge the error on this parameters are essentially random. Only TH error could have systematic biases which are however not know (the known bias are accounted for in the pointing corrections from NASA)

- pointing errors: out of curiosity: is there any drift in the platform that would lead to a systematic but slowly changing pointing error that could cause a long-term drift in the retrieved profiles?

This is a very good question and the NASA team has been investigation pointing issues for quite some time and could not point out a clear drift in the pointing, although an introduced

jump-like correction may have enhanced the drift. An improved L1 data version is being processed by NASA and the possible the drift of the new time series will be checked.

Sect. 5:
- line 278: by "changing the used cross section", I guess you mean the source of the cross sections? For clarity, I'd rephrase to "cross section source" or "cross section database". This also needs to be changed in line 280.

Thanks, yes, in this way it is more clear.

Sect. 7:
- Eq. 11: you're entirely confident that they are all independent? Intuitively, I would agree. But maybe this assumption needs to be made explicit. EDIT: I guess Fig 13 demonstrates that they are not independent. What does that mean for Eq. 11?

Thanks for this comment, it is a point that can be improved. Fig. 13 of the manuscript shows that the parameter uncertainties are correlated, although it doesn't directly imply their interdependence. It is however sensible to assume some interdependence between them, e.g. increase in aerosol extinction impact also the reflectivity. However, since we have no relevant information regarding their dependence we have to assume them independent. A comment to Eq. 11 has been added.

Typos:
- abstract, line 3: complaint → compliant
- line 36 (Sheese et al. ref): \cite → \citep
- line 70: increases → decreases; up to → down to
- line 92: parameters → parameter

All the typos have been addressed.