Dear Reviewer,

Thank you very much for your comments on our paper. We took your comments into account in the revised version of the manuscript. Please find below our detailed replies (black font) on your comments (blue font).

Reviewer#1 comments:

The paper discusses various methods for assessing the actual random error of satellite-derived profiles of atmospheric data, comparing these with the reported random error, and showing results of various methods of assessment.

I have no objection to the content of the paper, or the analysis, but I offer some suggestions for improvements and clarifications which the authors may wish to consider.

1) Firstly, There ought to be an initial statement of the meaning of 'random uncertainty' as used in the title of this paper. I think it may be close to the lab definition at the start of section 3.2 but it should be at the start, along with some discussion of various alternative interpretations that have also been used (e.g. those briefly listed in Table 1).

In the revised version, we added a short definition of random uncertainty.

2) I can think of two further methods which could also at least be mentioned, if not applied.
a) (in addition to methods considered in 3.3) For continuous limb-scanning instruments retrieving profiles every few hundred km along the orbit, one could compare each profile with a profile interpolated from the profiles immediately before and afterwards along the orbit. While this still has some component of natural variability, that would be reduced by the linear interpolation. This has an advantage over the method of using orbital intersections since the time gap is smaller and all three profiles are likely to be measured with the same day or night illumination. However, this would not work for tomographic retrievals.

Thank you for the interesting idea. However, the interpolated profile cannot be considered as “measurement”: it can be different from reality, and its uncertainty will be correlated with uncertainties of data used for interpolation. In terms of validation of uncertainties, interpolation will not bring additional information.

For the limb instruments with dense sampling, one can consider comparison of pairs of consecutive measurements along the orbit. However, the spatial separation of a few hundred km is not sufficiently small, so that the variance of differences contain the typical patterns of small-scale variability.

b) (in addition to methods considered in 3.4) Measurement of variation about the zonal mean. This seems to be partly covered in 3.4.2 but I was thinking of much narrower latitude bands. This particularly suits solar-occultation instruments which typically make 14 measurements in each of two very tightly-constrained latitude bands every 24 hours, and also tomographic retrievals since any along-orbit correlation is likely to be negligible over half an orbit. One could then dispense

with any time (apart from within the same day) or longitudinal constraint on matching - hence more comparisons with polar-orbiting instruments - and the only additional information required is \sigma_nat on c.15deg longitude scale which is, obviously, the same for all instruments and, in the summer stratosphere, quite possibly negligible

The latitude band can be narrower than that used for illustration in Section 3.4.2. However, evaluation of sample variance requires large number of measurements. For the particular case of solar occultation, it requires combining the data from several days /latitudes.

For measurements along the orbit, the natural variability is so large that any estimates of random uncertainties are not reliable (see requirements in Sect 3.4.2)

3) Although the various methods that are discussed are applied to different instruments, there is no summary table or plot comparing the results from the different methods applied, eg, to just one instrument, so that the methods can be directly compared.

In the revised version, we included a figure comparing ex-post uncertainties for MIPAS at 20°S-20°N in 2007 estimated by different methods and the corresponding discussion.

Minor points/typographical corrections:

Section 2 - it would be helpful in each subsection to have just an initial sentence describing the type of instrument/observation.

In the revised version, we added a short description of type of instruments and also the main principle of ozone profile retrievals. The revised Table 1 also contains information about the retrieval method.

Generally, use 'en' dashes ($--$) to indicate a range of numbers rather than hyphens (eg Figure captions, P11 L20, P13 L4-5 L18, P17 L15).

Corrected

P5 L9 (&L17): A large chi-squared value seems more likely to indicate the presence of residual spectral features, eg systematic errors in the forward model, than correctness of the assumed random error.

This sentence begins with “If the theoretical model describes the experimental data correctly,…” i.e., this is the statement for the case when there are no systematic (and big) misfit of spectra.

P5 L20: 'em' dashes are required here ($---$ in LaTeX),

P7 L19/20: 'which represent ... is different': sigma^2_0,nat is treated as both plural and singular in this sentence.

Corrected

P8 L9: Note that such collocated measurements necessarily involve comparing ascending and descending nodes of the orbit, so likely to involve different day/night conditions.

In the revised version, we added this note.

P8 Eq (5): presumably D(\rho) depends differently on the magnitude of each coordinate of \rho (and in any case some scaling is required to convert between the time and space coordinates).

Yes, $D(\boldsymbol{\rho})$ is usually anisotropic (illustrations can be found in Sofieva et al., 2021).

P9 L11: S_12 (upper case here, lower case elsewhere)

P9 Fig 15 caption: '20011' should presumably be '2011'.

P10 L16: "true" - initial pair of double-quote marks show as ",,"

Corrected

P10 L22: "not dense" - I suggest "sparse"

Changed as suggested.

P10 L32: Here it seems that "a-posteriori" and "ex-post" mean the same thing but elsewhere both are used individually so it is less clear that their meanings are the same. Also "a posteriori" is sometimes hyphenated, sometimes not (P16 L21)

In the revised version, “a posteriori” is changed to “ex-post” everywhere except its first definitions and introduction of von Clarmann terminology.

P11 L2: Since it is a direct part of the sentence, I would suggest "von Clarmann et al (2020)" rather than "(von Clarmann et al., 2020)" (also P17 L14)

Corrected

P13 Fig 25: I was initially impressed with the consistency of the \sigma_nat values shown in the lower plots, but then I realised that these are very similar to the sample SDs shown in the upper plot, somewhat contradicting condition (b) mentioned on P14 L2.

We agree. The smallest natural variability is in the tropics, but the random uncertainty estimates for limb-instruments are usually smaller. In the revised version, we added this note.

P14 L23: pedantically it should perhaps be noted that \epsilon_y,z refer to random errors scaled to x rather than associated with the original measurements (to me it seems more natural to have eg y = c_y t + e_y)

We follow the original formulation of Stoffelen (1998).

P12 L11: I may have missed it, but what is σ^2_0,var ?

It should be $\sigma^2_{0,nat}$. We corrected this.