Replies to Reviewer 2

We thank both reviewers for their knowledgeable and valuable comments. Our efforts in addressing them, together with the reviewers' suggestions, led to a revised manuscript that represents a great improvement with respect to the original version. In what follows, reviewer's comments are in black and authors' replies in red.

In general, the introduction in section 3 to 5 repeats the basics of radiative transfer, balancing instruments and optimal estimation, which are already described in detail in the cited literature. In my opinion, these sections could be easily shortened without loss of information.

We shortened the sections as suggested by the reviewer, also eliminating 13 basic equations.

The abstract (p.1 line 13) and section 3 describe the VESPA-22 back-end as an FFTS with 500 MHz bandwidth and 31kHz resolution. However, the schematic in Fig. 1(b) shows a system with 1GHz bandwidth. Is this the actual schematic of the instrument?

No, there was a mistake in the schematics. It has been corrected.

Also P4 line 1 states a 2GS/s sampling rate, which results in a 1 GHz Nyquist bandwidth.

The effective sampling rate is 1 GS/s. It was corrected in the revised manuscript.

P. 3 line 13: The sentence "the waveguide used by VESPA-22 polarizes the incoming radiation with a gain difference between the two polarization modes of 45 dB" is both incorrect and irrelevant. The rectangular waveguide supports only a single polarization by definition. The stated 45dB may refer to the feed horn, but the actual cross-polar with the offset reflector will be much worse. Bertagnolio 2012 states for the same instrument 35dB and 24dB, respectively. This should not affect the observations, so the sentence could be easily removed.

We agree with the reviewer and removed the sentence.

P. 6 line 28 states that the opacity in Fig 3 was calculated for water vapor profiles measured by AURA/MLS. However, the dominant effect in this figure is due to tropospheric water vapor which is not measured by MLS.

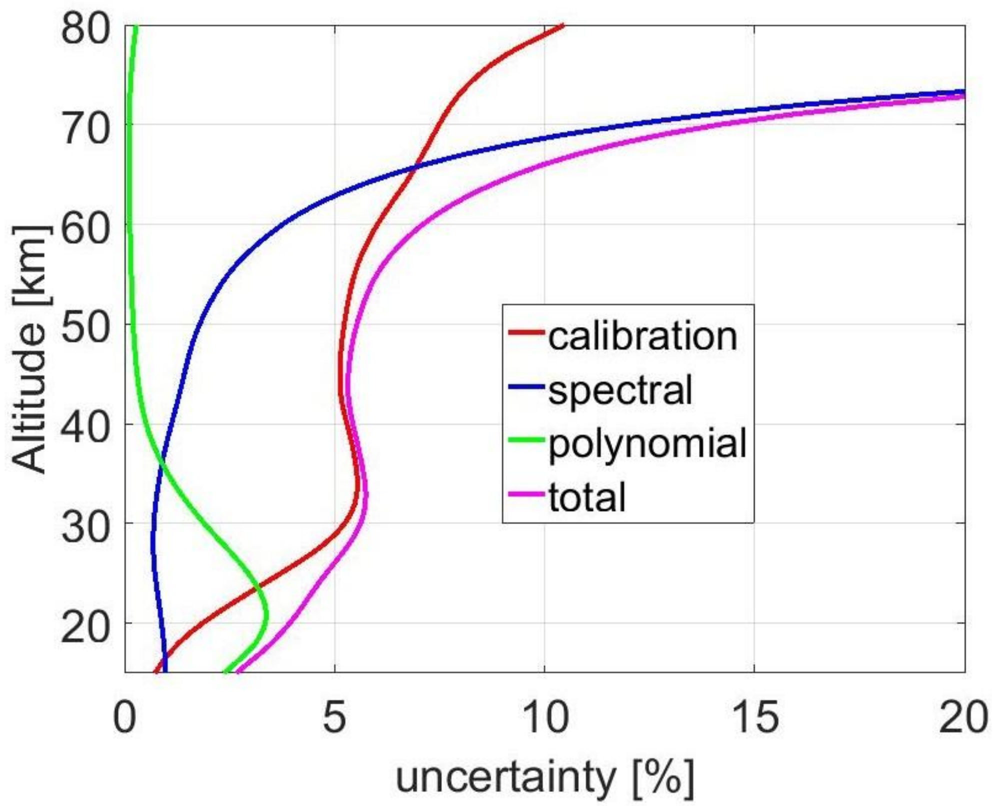
We clarified this statement in the revised manuscript. The profiles used were obtained merging tropospheric and stratospheric measurements collected by radiosondes from Eureka (80.0°N -85.9°W), Canada, and Aura/MLS, respectively.

P. 9 line 24 states that a 6 MHz wide interval around the line center is kept at 31kHz resolution, while all other channels are binned with 50 channels. In Fig 6(b) the interval without binning looks much smaller than 6MHz.

The interval without binning is indeed 6 MHz and was correctly showed in Figure 6b of the original manuscript. However, the x-axis unit in figure 6b was MHz and there was a 10^4 multiplicative factor on the bottom right of the figure. We realized that this indication was misleading and we changed the frequency unit for all the figures of the revised manuscript from MHz to GHz.

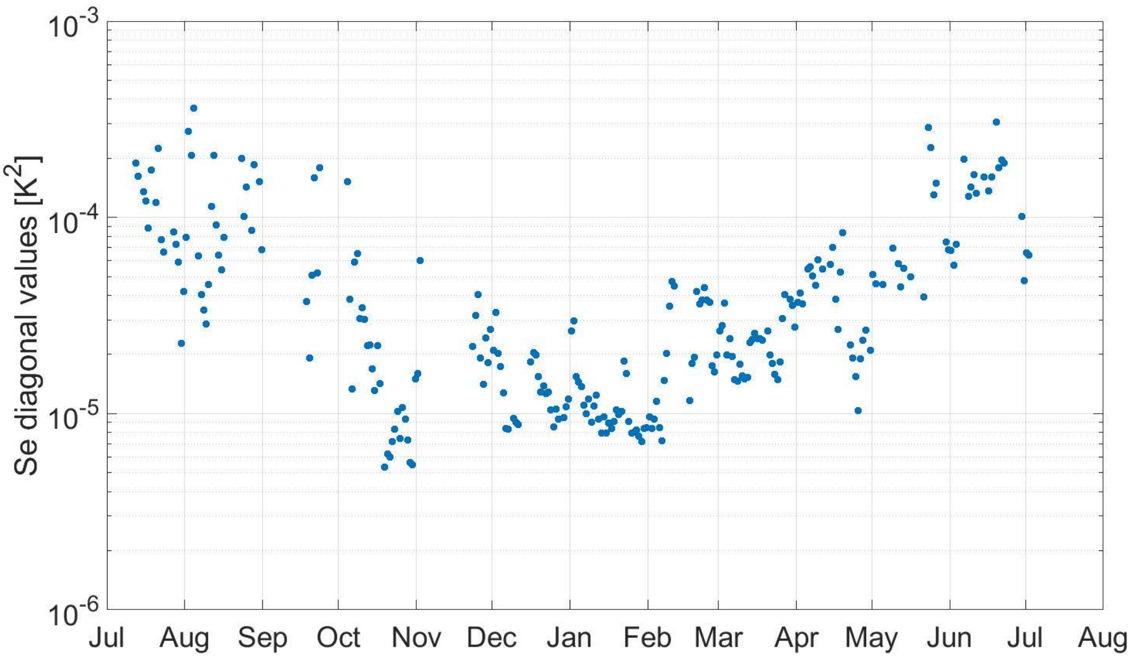
According to p. 14 line 1 the matrix with the measurement error Se was assumed with constant diagonal elements which are calculated from the residuals after an initial run of the retrieval. However, the central channels without binning will have a higher measurement error as the ones without binning. Please provide also the range of measurement errors which were used in Se for the daily retrievals. The retrieval errors in Fig 11 seem to be very small. Could this be an artifact of an underestimation of the measurement errors in Se due to the binning at the line wings?

The reviewer is correct and in the original manuscript Se was underestimated. In the new analysis inserted in the revised manuscript Se is now calculated before the smoothing process of the spectrum and this brought an increase of the retrieval uncertainty (now indicated as "spectral uncertainty") of approximately 70% over the entire vertical profile. The figure below (also inserted in the revised manuscript) shows the new error analysis where we also added the uncertainty due to the use of the second order polynomial (see later comment of this reviewer and a similar comment of reviewer 1).



The retrieval uncertainty (now “spectral uncertainty”) is still small, especially in the lower stratosphere, but we underline that it represents only the uncertainty due to spectral noise and potential spectral artifacts.

The figure below shows the time series of the Se values. The range of Se values has been indicated in the revised manuscript.



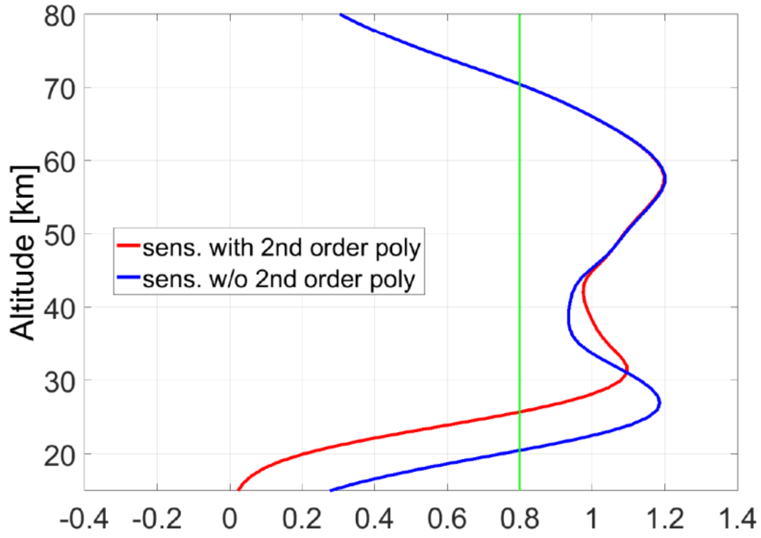
Where do the values for the apriori covariance in Fig. 5 come from?

The Sa matrix values were chosen empirically in order to optimize the characteristics of the retrieval (i.e., maximize sensitivity range and vertical resolution without introducing unphysical oscillations in the retrieved vertical profile). We added this information in the revised manuscript. In the new analysis, we also adopted a slightly rescaled σ profile with respect to the one used in in the original manuscript (and of course updated former figure 5, now figure 6).

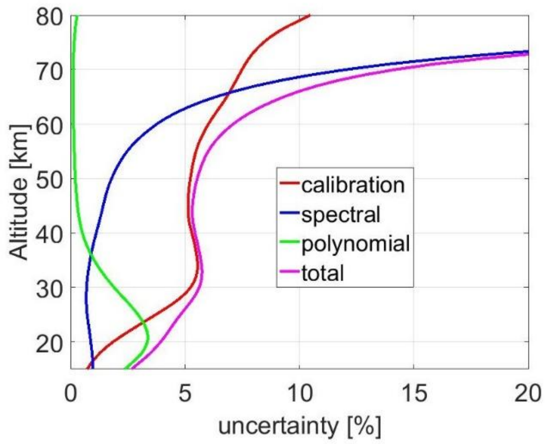
P. 14 line 7: "A second-order polynomial is also added to the retrieval . . .". At which stage is it taken into account in the retrieval, and how does it affect the measurement response at the lower altitudes? Is it really a second order polynomial, or just a straight line?

This issue was raised by both reviewer so we report here the same detailed answer we provided to reviewer 1.

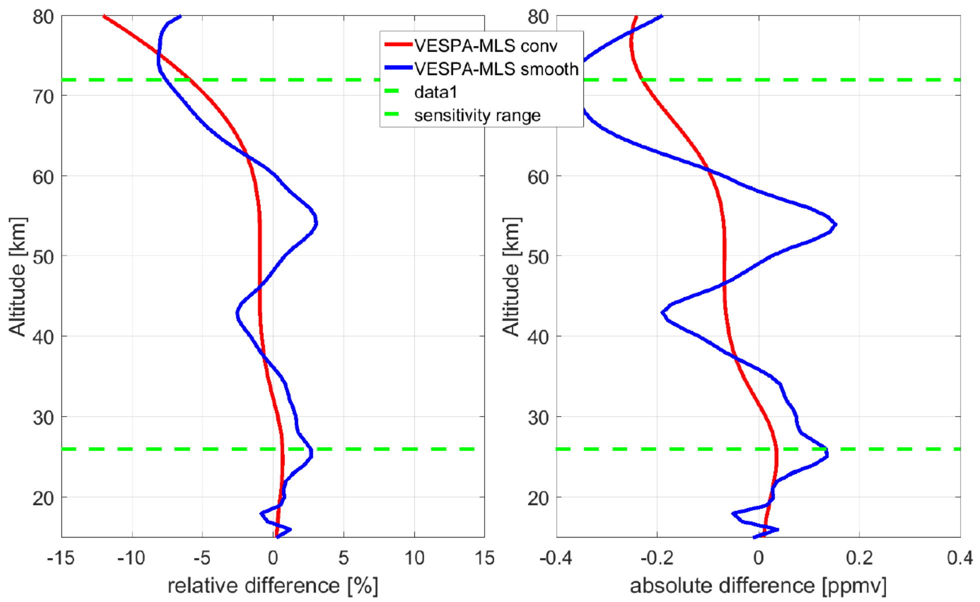
We added to the manuscript a discussion on how the retrieval process includes a second order polynomial. This is done by fitting it in each retrieval, individually. It is not a systematic term. It does have a large effect on the sensitivity below 30 km but this is already taken into account in the sensitivity profile of figure 8 (original manuscript, now figure 9). In the figure below, we show the difference in sensitivity between employing the second order polynomial or only a first order one.



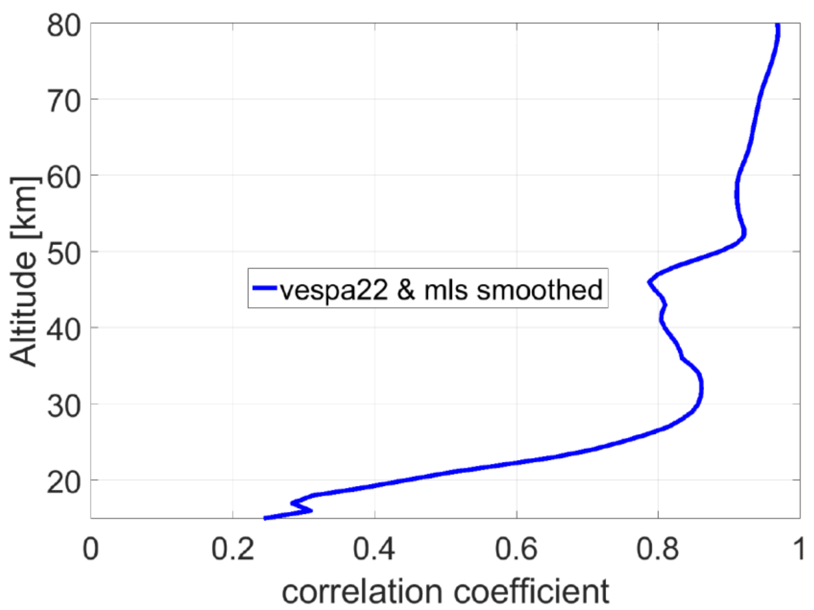
In order to address this issue raised by the reviewer, we estimated the uncertainty in the retrieved profile due to the use of the second order polynomial. The most rigorous way of doing it, in our opinion, is to take the uncertainty associated to the second order coefficient (Δa_2) in the retrieval process (say for example 20%, e.g., $a_2 = (-5 \pm 1) \times 10^{-3}$) and then perform two retrievals for the same spectrum with fixed values of the second order coefficient equal to $a_2 \pm \Delta a_2$ (in our example they would be $a_2 = -4 \times 10^{-3}$ and $a_2 = -6 \times 10^{-3}$). The resulting two vertical profiles would then provide an estimate for the uncertainty of the regular retrieved profile associated with the uncertainty in the second order coefficient. We found that the average uncertainty of the second order coefficient calculated by the optimal estimation routine over the entire dataset is 6%, with few retrievals showing more than 20%. We therefore decided to employ a fixed maximum uncertainty on the coefficient of 20% for the whole data set, rejecting from the data set those few retrievals (less than 5% of the total) that had an uncertainty larger than 20% in the determination of a_2. Displayed below is the updated figure related to the error analysis of the spectrum observed on 23 Dec, 2016, where we added the polynomial uncertainty to the other 2 sources. As expected the polynomial uncertainty has an impact mostly in the lower part of the retrieval.



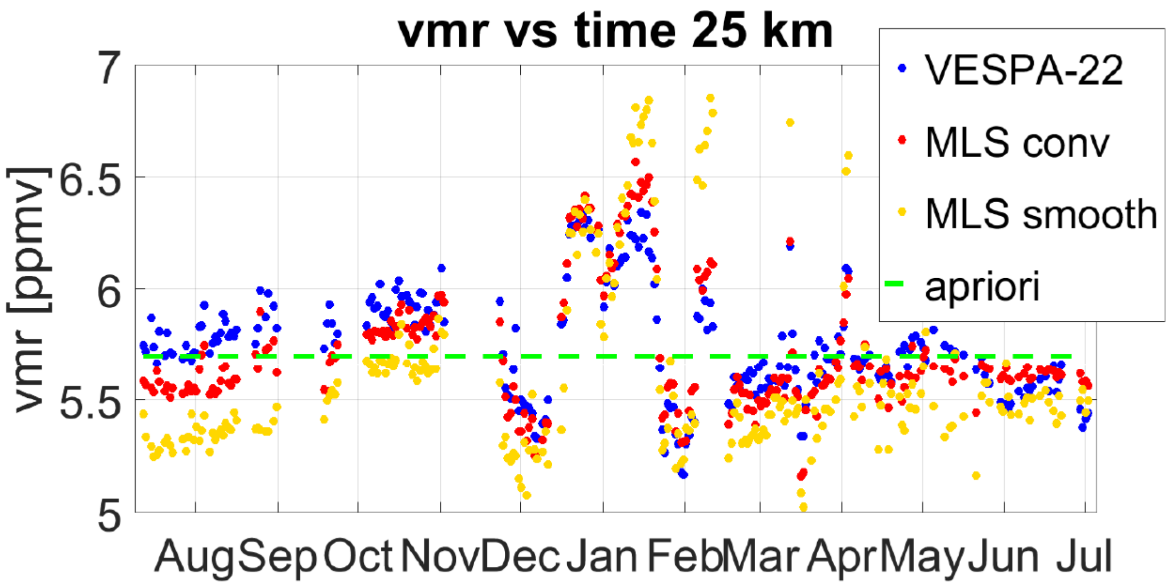
On top of it, in order to demonstrate the good quality of VESPA retrievals down to 25 km altitude, we changed the apriori used for the retrievals, which is now a fixed climatological profile up to about 48 km and a seasonal profile in the mesosphere (see also our reply to this specific issue raised by the reviewer in the following), and we introduced in the manuscript the correlation between VESPA profiles and MLS smoothed profiles. These are the original high resolution MLS profiles smoothed in the vertical with a running average of 10 km. This was done in order to decouple MLS profiles from the apriori and the averaging kernels used in VESPA retrievals (which was a short-coming of the correlation between VESPA and MLS convolved profiles shown in the originally submitted manuscript) and yet make the MLS vertical resolution somewhat similar to that characterizing VESPA profiles. Figures down below show the relative and absolute average differences between VESPA and MLS smoothed (blue), and between VESPA and MLS convolved (red),



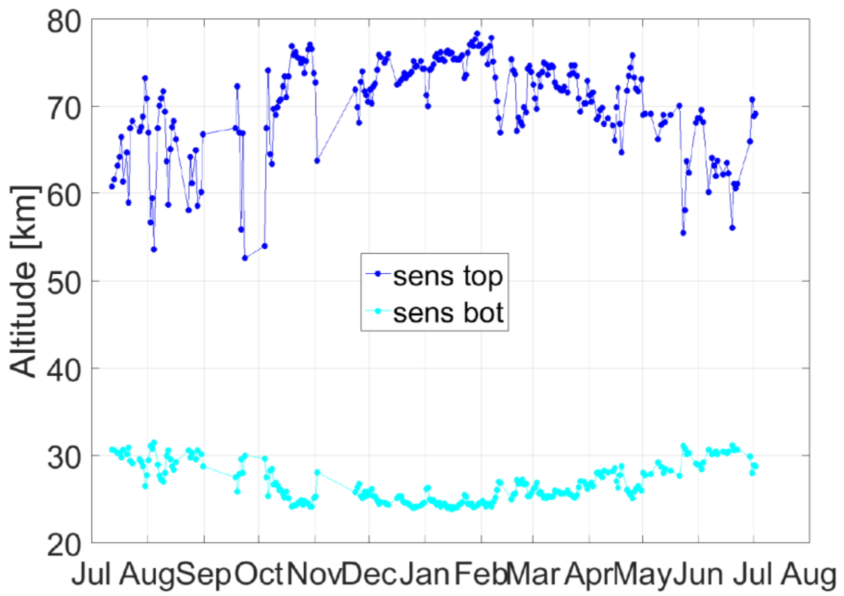
the correlation coefficient between VESPA and MLS smoothed profiles (blue),



and the time series at 25 km of VESPA22, MLS convolved and MLS smoothed.



All the figures above demonstrate in our opinion that VESPA22 retrievals are scientifically valuable in the sensitivity range indicated in the manuscript. Please note, however, that we also specified in the revised manuscript that the sensitivity range changes with seasons (see fig. 18 in the revised manuscript) and in summer it is approximately between 30 and 65 km. See figure below.



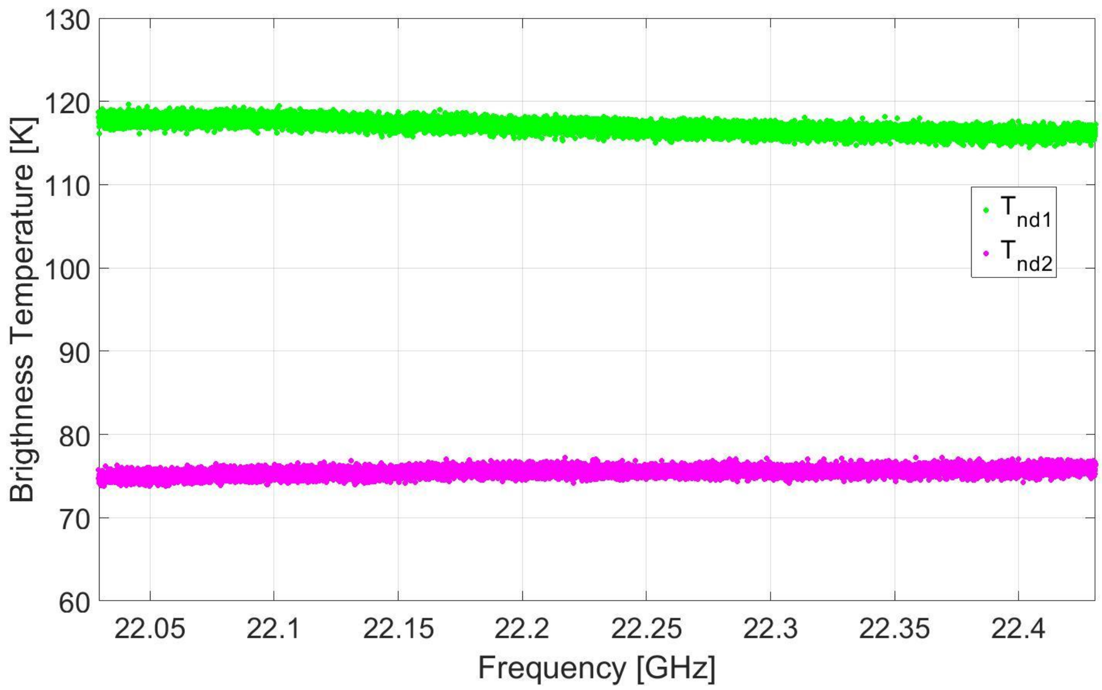
What was changed in the revised manuscript:

- Added discussion in Section 4 on how the second order polynomial is treated in the retrieval process;

- Modified figure 12 with the updated error analysis which includes the polynomial uncertainty;
- Modified figure 9 according to the new error analysis;
- Inserted the MLS smoothed data set, with its correlation with Vespa-22 profiles;
- Introduced the time series of the sensitivity interval, i.e., how the interval of accepted sensitivity changes through time. Added this time series in figure 18;
- Added two altitude levels in former figure 15 (figure 17 in the revised manuscript) to prove the good quality of VESPA-22 retrievals at the lower (25 km) and upper (75 km) limits of the sensitivity interval.

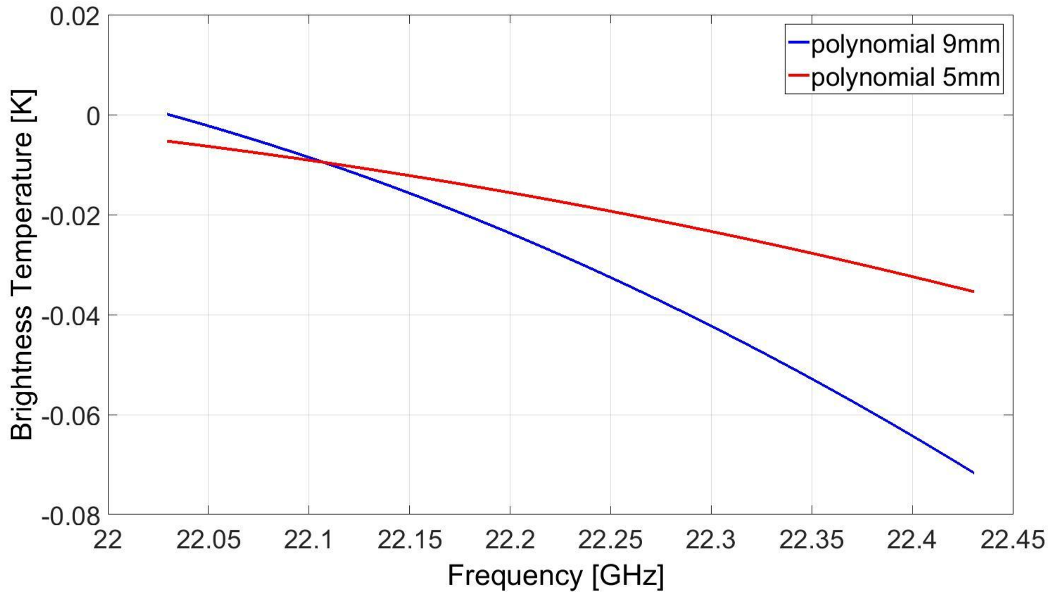
P. 10 line 7 "The noise diode produces a signal that can be considered constant in frequency within 1.5%". Where does this number come from, and can it be demonstrated in the VESPA22 measurements? Over which bandwidth is it valid? Presumably only 40MHz are used for the retrieval, but according to Eq. 20 the mean T_ND is calculated for a much wider frequency range.

The indicated 1.5% is half of the maximum difference between brightness temperature values over the spectral passband (see figure below). The noise diodes emission spectra are measured only during the LN_2 calibration. In the revised manuscript we rephrased the "constant in frequency" sentence with "The noise diode produces a signal that is measured to be quite stable in frequency. In fact, single-channel T_{nd} values are always within 1.5% of the spectral mean of the diode temperature brightness T_{nd}." This is valid over the full 400 MHz bandwidth used in the spectral inversion process.



Also the opacity of the Delrin sheet is assumed to be frequency independent. Is this really the case, or does e.g. the polynomial baseline fit change after changing the thickness of the Delrin sheet?

The reviewer is probably correct in suggesting that the delrin sheets have an opacity that varies over the 400 MHz of spectral measurement. We can infer it from the change of the polynomial baseline that we find when we change delrin sheet (in the figure below we indicate the average polynomial baselines when using a 5 mm or 9 mm sheet), as suggested by the reviewer.



However, we take into account the potential dependency of delrin opacity on frequency in the uncertainty associated to the measurement of the delrin opacity. Delrin opacity measurements have a maximum error (max-min/2) of approximately 2% but we conservatively estimated such an uncertainty to be ±10%.

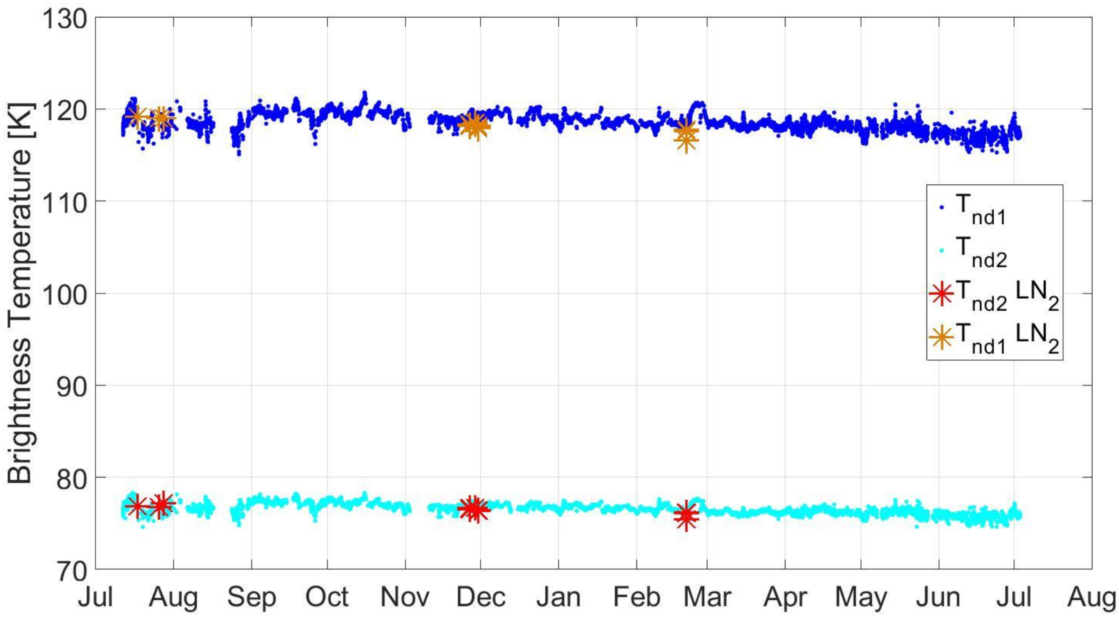
We added a phrase clarifying that the indicated τ_d is the delrin "mean opacity value over the spectral passband".

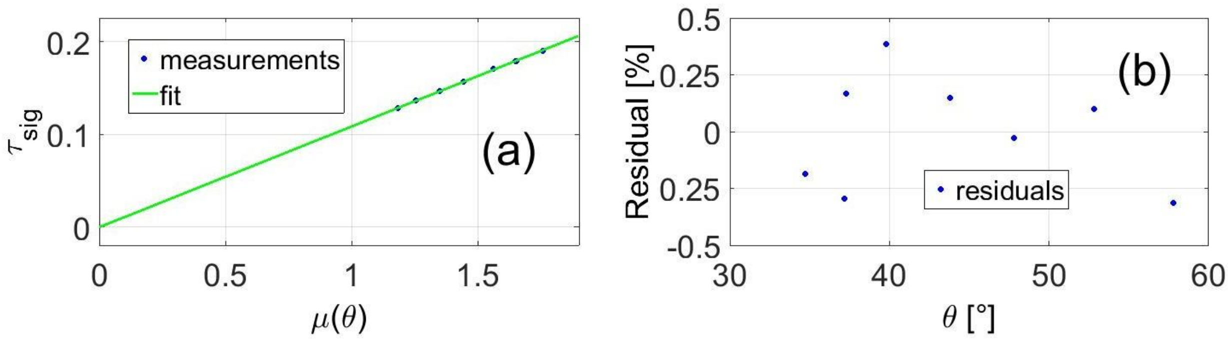
P. 12 line 6 claims that the noise diode temperatures measured with LN2 and tipping curve agree within 0.4%. However, the fluctuations of the tipping curve results in Fig. 4 seem to be in the order of several Kelvin, which should result in a higher discrepancy.

Are these fluctuations measurement errors caused e.g. by an inhomogeneous atmosphere, or do they represent a real fluctuation of the noise diode ENR caused e.g. by changes of the laboratory temperature? Which value was used in the actual retrievals?

The 0.4% agreement stated in the manuscript is calculated as the mean difference between the noise diodes temperature measured during the LN_2 calibration and the diodes temperature measured by means of the tipping curve immediately before or after the LN_2 calibration. We confirm that the number is correct and it is approximately 0.4%. We clarified this in the revised manuscript with the sentence "The mean relative difference between T_{nd} values calculated with LN_2 and with tipping curves carried out immediately before or after LN_2 calibrations is (0.4±0.4)% and (0.2±0.3)% for the calibration and the backup diodes, respectively."

Most of the fluctuations in the Fig. 4 of the original manuscript are caused by an inhomogeneous atmosphere. However, some are caused by changes in the ambient temperature inside the observatory (such as that visible on the end of February). We checked this against measurements of the temperature inside the observatory. As a result of this reviewer's comment, we decided to set a stricter rule for accepting tipping curve measurements and now use a threshold of 0.4 instead of 0.8 (see P12 L2 of original manuscript). The figure below shows the time series of T_{nd} values obtained by tipping curves that comply to this new criterion. In the revised manuscript this figure substitutes former Figure 4. In order to better discuss the tipping curve procedure we also added a new figure in the revised manuscript, also attached below.





P 11 line 3: "T_atm is the average temperature obtained from radiosonde data by weighting the tropospheric temperature vertical profile with the water vapor concentration profile." Is this weighting done with the water vapor vmr or with vmr*p ?

Yes, the weighting is done with the water vapor vmr*p.

Several occasions: replace "depending from" with "depending on"

We modified the manuscript as suggested.

The abstract (p.1 line 3) mentions an integration time in the order of hours, but the paper discusses only the results of daily mean values.

We calrified this in the abstract, mentioning first the potential capabilities of the instrument ("The integration time for a measurement ranges from 6 to 24 hours, depending on season and weather conditions"), and then specifying twice that for this work we use only 24-hour averaged spectra ("The VESPA-22 water vapor mixing ratio vertical profiles discussed in this work are obtained from 24-hour averaged spectra and are compared with version 4.2 of concurrent Aura/Microwave Limb Sounder (MLS) water vapor vertical profiles").

P. 2 line 17: The statement "Positive trends were observed during the last two decades" is followed by citations from 1999-2001. This sentence should be reworded or backed by more recent citations.

We modified the manuscript as suggested.

P. 3 line 10: instead of "full beam at half power" the term "full width at half maximum" (FWHM) would be more common and is also used in the cited Bertagnoli 2012 paper

We modified the manuscript as suggested.

P. 3 line 19: "Two noise diodes are inserted in the IF chain" should read "RF chain". Preferably "IF" and "RF" should be explained at first instance.

We modified the manuscript as suggested.

P. 4 line 19 mentions the brand name "eccofoam", but earlier the window was identified as LD-15 which is a different material.

We modified the manuscript as suggested.

P. 9 line 3: The statement "The ’zero’ signal is measured and subtracted to every acquired spectrum..." is misleading and could be understood that V0 is measured with every spectrum. Please clarify when and how often V0 is measured and whether it has a significant effect. Presumably it does not contribute at all to the result since the calibration with Eq. 13 and 16 uses always differences between two raw spectra.

The V_0 signal is measured every 15 minutes, at the beginning of each spectral integration which is the time resolution of the VESPA-22 spectral dataset. Although the measurement of V_0 is not required before the regular spectral measurement (as correctly argued by the reviewer), we do need to provide the FFTS with a "zero" signal at the input for its calibration, which VESPA performs every 15 minutes. We therefore measure and save such "zero" signal as a check on the performed FFTS calibration.

V_0 is also measured at the beginning of each tipping curve procedure (which is run every 30 minutes). Here we do not use S-R but S only and if we did not subtract the zero signal from S we could potentially introduce a small error in tipping curve measurements. However, V_0 is only about 0.5% of S and this error is possibly negligible.

We tried to clarify this aspect with the sentence "The "zero" signal, which amounts to approximately 0.5% of the incoming signal to the FFTS, is measured approximately every 15 minutes and it is subtracted to each S and R 15-minute integration spectra which are eventually saved on the control & acquisition PC (see figure 1b)." and by adding V_0 in the equation of Mu*Tau.

P. 9 line 5: "..and subtracts the counts number from these two sources" should probably read "...numbers...".

We modified the manuscript as suggested.

P. 13 line 16 states that the central 400MHz are used in the retrieval. Since the Fig 6a) shows only 40 MHz this is most likely a typo.

As mentioned above the frequency range showed in the figure is indeed 400MHz. This misunderstanding is caused by the factor 10^4 that was not clearly visible in the bottom right of the panel. We fixed this by using GHz instead of MHz.

P. 17 line 6+8: There seems to be a copy and paste error in the sentences "... integrated for second-degree polynomial 24 hours . . ." and "The cyan line is the retrieved by the inversion..."

We modified the manuscript as suggested.

P. 22 line 20 and following: These paragraphs repeat the principles of the Delrin balancing, Eq 42 is very similar to Eq.8. Also the details of the sheets and how their opacity is determined are presented. In my opinion this would fit better to section 4.1 where the instrument and calibration are discussed, than into this section 6 "retrieval uncertainty".

We modified the manuscript as suggested, removing equation 42 and moving the discussion on the measurement of the delrin sheets opacity to former Section 4.1, now Section 3.2.

P23 line 30: "...of the datasets obtained with the different models [. . .] respect to the reference model dataset..." Apparently a missing "with" in [. . .], otherwise the sentence does not make sense to me.

We modified this sentence in the revised manuscript and this comment no longer applies.

Fig. 10: The labels (c) and (d) are missing in the lower two subplots. Fig. 17: The labels (a) and (b) are missing in the two subplots.

These figures have been removed and this comment no longer applies.