The manuscript “Intercomparison of stratospheric ozone profiles for the assessment of the upgraded GROMOS radiometer at Bern” by Studer et al. presents a comparison of the observations by the new version of the GROMOS radiometer with a ground-based lidar and various satellite observations, sondes and ECMWF re-analysis. The instrumental upgrade consists of using a new FFT spectrometer instead of a filter bench. A direct comparison of observations obtained with the two backends operating in parallel is also presented and reveals differences between the two versions of the instrument.
This is an important analysis to characterize the performances of the new instrumental version and most of all, to allow for a continuity with the observations performed since 1994.

The paper is very well written and show the good quality of the measurements though I believe that additional discussions and information are needed to fully achieve its objectives. My main comments are: 1) the discussion of the retrieval errors is too brief and the errors estimation is not used to interpret the results of the comparisons, 2) the comparison between the FB and FFTS retrievals should be discussed alone in a dedicated section to clearly show possible bias, 3) The standard deviation calculated for seasonal means analysis should be clarified and it is not fully used in the discussion. I provide here below the detailed comments.

In conclusion, I believe that the paper is well suited for the journal and should be published after few improvements.

Detailed comments:

P 6101, line 21: The 32768 “channels” are probably binned for the retrieval process. What is the spectral resolution of the inverted spectrum ?

P 6101, line 26:
More details about the measurements when the 2 backends are in parallel are needed. Are the 8-sec integrated spectra simultaneously observed by the FB and the FFTS? Is the value of the thermal noise correlation between FB and FFTS spectra available? It should be high? Is the spectral baseline the same on the FB and FFTS spectra? The time resolutions are 60 mn and 30 mn for the FB and the FFTS, respectively (Table 1). Why are they different ?

P 6102, line 20:

- Since the comparisons between GROMOS and satellite data are shown up to 0.1 hPa, it would be interesting to show the a priori contamination up to this altitude. At the highest altitudes, the averaging kernels center is not necessary at the retrieval altitudes, therefore it would also be interesting to show the relation AVK-center vs altitudes.

- What are the assumptions for the measurements and O3 a priori errors ? Is the retrieval setting the same for both FB and FFTS?

P 6102, line 26: Because of the different integration times, the signal-to-noise ratio with the FFTS should be larger than the one with the FB (~30%). Also the spectrometer bandwidths are slightly different. I then expect differences between the FB and FFTS AVKs and a priori contaminations, especially at the edges of the vertical retrieval range. How big are the differences? They should be quantitatively described in the paper.

P 6103, line 6: I believe the AVKs have negative values that are not seen because the x-absciss starts

from 0. If I am right, this should be mentioned in the figure caption (though I would prefer to see the negative part on the plot).

P 6103, line 10: How is the bias estimated? Are the spectroscopic corrections applied to both FFTS and FB retrievals?

P 6103, line 15: Are the spectroscopic parameters in the current version of the catalogs the same as in the 1998 catalogs? I think it would be good to provide the value of the parameters (pressure broadening parameter, line strength and central frequency) used in the inversions.

P 6103, line 26: The description of the retrieval errors should be improved. First the retrieval error from the FFTS spectrum should be different than the one from FB spectrum not only because the FFTS spectra noise is larger (smaller time resolution) but also because the a priori contamination may be different (see earlier comment). It is then important to provide errors for both integration times. Second, in order to help to understand the differences (bias and standard deviation) between the O3 profiles derived from the FFTS and FB, it is important to provide the systematic, random and smoothing errors over the full altitude range discussed in the paper (50 to 0.1 hPa). Measurement-errors correlation between FFTS and FB should also be discussed. Finally I would recommend to plot the results instead of only providing the values.

The error estimations are based on a old error analysis (before 1997) and I am wondering if the assumptions are still up to date. I would expect a better characterisation of measurements/retrieval errors based on almost 20 years of data. Especially I have in mind errors due to the FB which, to my understanding, was not taken into account by Peter (1997) and errors due the spectral baseline.

P 6103, line 29: the thermal noise decreases as 1/SQRT(tint) and not 1/tint.

P 6104, line 9: why the analysis include only a single ground-based instrument while several O3 profilers are available in nearby NDACC sites? I think other instruments are important for a quantitative assessment of GROMOS performances and can provide important additional information. For example, other microwave radiometers reach 70 km while the lidar is limited below 50 km and provide night/daytime obervations.

P 6108, line 17: errors are given up to 3 hPa while the upper limit of the sonde is 10 hPa (line 11). It doesn't look consistent.

P 6109, line 6: should the probability be 78 % and not 88 % ?

P 6110, line 14: I would prefer “GROMOS a priori profile” instead of “microwave a priori profile”

Sections 5 and 6: General comment:
The comparison between GROMOS FFTS and FB should be discussed in a separate section and not together with the satellite data. I think this comparison is an important result of the analysis and deserves a more detailed discussion. In particular, the mean differences and the standard deviations of the differences should be provided as well as their time variations during the two years period. An important point is to check if the retrieval errors estimation can explain the differences. It should be possible to explain most of the differences since the atmospheric conditions are the same for both observations as well as most of the instrumental/retrieval parameters.

P 6111, line 12: I think that the standard deviation of the differences should be provided and discussed since it gives information on the precision and correlation of the observations and on the atmospheric differences between both sites.

P 6112, line 16: "top panels show**s**" → "top panels show"

P 6112, line 19: I don't understand the profiles in the 3rd panel of each row. Is-it the standard deviation of the retrieved profiles or the standard deviation of the differences ? Since the values are small and the FFTS-STD is not shown, I guess it is the standard deviation of the differences with the FFTS. Am I right? It should be made clearer.

P 6113, line 7: The difference with the FB should be discussed. For instance between 50 and 20 hPa, the difference is relatively large (up to 10% at 50 hPa) and it is larger than the differences with the satellite data. What is the reason for such differences ? Is-it due to different a priori contamination or different baseline or ...? Does the use of the FFTS significantly improves the observations compared to the FB ?

P 6113, line 10: the information about single retrieval precision for each instrument is needed to correctly interpret the standard-deviation results. To my understanding only accuracy is given in the section presenting the instruments. it is worth noting that the vertical oscillation is also on the FFTS-FB standard-deviation. This should gives some indications about its origin.
As already mentioned, I think the analysis of the differences between the FFTS and FB deserves a separate section.

P 6113, line 11: "ozone profiles show**s** " → "ozone profiles show "

P6114, line10: It has been mentioned in the paper that most of the sondes drift toward Bern. Is-it true for all seasons ?

P6114: the standard deviations FB-FFTS in Fig. 6a and 6b are similar or larger than the ones obtained with other satellite data. For instance in summer, a STD of 2.5% is found between 50 and 2 hPa for the FB-FFTS against 1-2 % for the satellite data. I would expect the opposite. Also I am surprised by the low value of the standard deviations obtained from the comparisons with other satellites. It is smaller than the single retrieval precision of the satellite data. May-be I misunderstand the definition of the standard deviation shown in the paper (see my comment about P 6112, line 19).

P6114, line 24: Is-it the standard-deviation of the FFTS O3 profile? It should also be clarified in the caption of Fig 8.

P 6116, line 4: "... and shown **are** the ..." → "... and shown at the ..."

P 6116, line 5: is it the standard deviation of the differences ?

P6116, Section 6.4
I think that this section should be placed before the comparison sections since it contains information relevant for the comparisons. For instance this section could be Section 4.3.

P6116, line 27: Are the satellite data smoothed with the GROMOS AVK ?

P6117, line 25: I would mention that between 50 and 20 hPa, the difference can reach 10% which is, for me, a large value since the observations are performed by the same instrument and are simultaneous.

Figures 7,8,9 and 10: it should be indicated in the captions that the time series is made of monthly mean data.

Figures 2-9: In order to make the figures clearer I would recommend to show the comparisons between FFTS and lidar, FB and ozone-sondes in the same figures (Fig. 2, 3, 4, 7,8). Comparisons with satellite and ECMWF data should be shown in figures 5, 6, 9.