amtd-6-6097-2013

Intercomparison of stratospheric ozone profiles for the assessment of the upgraded GROMOS radiometer at Bern

S. Studer, K. Hocke, M. Pastel, S. Godin-Beekmann, and N. Kämpfer

Final Response to the Reviewers and the Editor

Firstly we like to apologize for the time delay with the revision of the article. The reason is that the first author left the university after getting the PhD degree. Though she always wanted to make the revision by herself, she told me on 4 February 2014 that she cannot manage the revision because of other duties and tasks. Circumstances went different to what she thought, and time is running fast.

She told me (Klemens Hocke) that I should take care on the revision of her paper and I agreed. The order of the authors of the manuscript will remain unchanged. Of course it is a bit difficult for me to work with her text files, data files, and her programs. On the other hand I am optimistic to carry out the moderate revision in a few weeks.

Point-to-Point Response to Reviewer 1:

Thank you for your interest in our study. We agree that the GROMOS radiometer plays an important role in the NDACC network and that an assessment of its data quality is mandatory. As suggested by you, we will work on a better presentation. For example, the literature references are quite untidy and we will improve the explanations and the structure of the text.

Point 1 of your main comments: "1) The new FFTS and the original FB backends measured in parallel for over 2 years, for data harmonization/calibration. Keeping in mind the main objective for the study - to get a long-time ozone record - one would expect a detailed comparison of these perfectly collocated data. However, this part is missing (comparisons of FFTS and FB are scattered in sections and presented together

with other datasets). From my point of view, it would be very advantageous for the paper to present comparison from the old and new GROMOS detectors in a dedicated section/subsection. Focus should be on (i) overall agreement, (ii) temporal evolution of profile differences, seasonal cycles, (iii) representation of diurnal variations. This section should contain also discussion of differences, underlying reasons, and possible solutions for data homogenization."

We will clarify the purpose of our study in the revised manuscript. Our planing is to publish two articles. The main purpose of the present article is the quality assessment of the new FFT spectrometer. We know from past intercomparisons that the old filter bench (FB) of GROMOS made a good job and was within 10 percent compared to satellite and ozonesonde data. However it is better to perform intercomparisons of GROMOS FFTS with independent data such as from ozonesondes and Aura/MLS. The main purpose of the present article is to present and discuss these intercomparisons. So the data of GROMOS FB play a secondary role. However the present study already shows that a later harmonization of the ozone time series from FB and FFTS will be successful (strong adjustments are not necessary).

The harmonization of the data sets will be done in a future article and is the task of the new PhD student. The future article will derive uncertainties and will describe the harmonization process in all details. Finally trend analysis of the harmonized time series will be performed.

Please excuse us if we won't follow you in all of your advices, since it would be too much to perform the data harmonization already in the present article. In any case we sort and clarify the text so that the reader recognizes our purpose in a better way.

Point 2 and Point 3 of your main comments: 2) The choice of satellite instruments. Please explain your choice of satellite instruments. As follows from section 3.2, both MIPAS and SABER have rather significant biases with respect to ozonesonde and lidar observations. Why you prefer using these data (are they conclusive for validation?), and not using the data from the instruments, which have small biases with respect to ground-based observations, like OSIRIS and GOMOS? Related to ACE-FTS data, comparing monthly mean data with individual ozone profiles looks very strange. This cannot be conclusive vice versa, can be misleading. I would recommend avoiding such kind of comparisons, and simply not including this dataset due to a very small number of collocated data. No data - no show.

3) It is well known that the quality of ozone profiles in ERA-Interim is insufficient. The figures presented in the paper demonstrate this clearly. But again: the main objective is assessment of the new GROMOS instrument. Thus only reliable data should be used. (Validation of ERA-Interim is definitely outside the scope of the paper). I suggest removing all comparisons with ERA-Interim.

We agree that sampling of ACE/FTS ozone profiles is poor at mid-latitudes and we follow your advice so that we keep out ACE/FTS in the revised paper. We also follow your suggestion to keep out the ERA-Interim data. It is enough that we have the intercomparison with ACE and ERA Interim in the discussion paper. We are often asked by other scientists how is the quality of the ERA Interim ozone product? Is it usable for scientific studies? That was the reason why we included ERA Interim in the intercomparisons. However we see your point too.

Generally we concentrated on intercomparisons with satellite limb sounders. We agree that comparisons with nadir sounders such as GOME are useful too. Indeed there is a nice validation report of GOME-2 using the Bern (GROMOS) data http://o3msaf.fmi.fi/docs/vr/Validation_Report_NOP_NHP_OOP_OHP_Jun_2013.pdf

This validation report and your comment are a good motivation to work in future with GOME data which were not so much known to us before, - particularly the ozone profiles from GOME.

Since there are many scientific papers based on TIMED/Saber and MIPAS ozone data, we think that intercomparisons with these experiments are important. Knowing of systematic biases or at least getting an impression about the biases are the most important result which we can achieve from intercomparisons.

Point 4 of your main comments: 4) Averaging kernel smoothing (or its description) is incorrect. In order to get the profiles in the same vertical resolution, the 1st profile should be convolved with the av- eraging kernel of the 2nd profile AND the 2nd profile should be convolved with the averaging kernel of the 1st profile. Alternatively, if use Eq.(1) only, AVK should represent the DIFFERENTIAL smoothing (i.e., the difference in vertical resolution). For ozonesonde data this second smoothing (low-resolution profiles with the averaging kernel of high-resolution profile) can be ignored, the difference in vertical resolution is too large. However, for satellite data it should be taken into account.

The most studies which we know follow the approach of

Tsou, J. J., Connor, B. J., Parrish, A., McDermid, I. S., and Chu, W. P.: Ground-based microwave monitoring of middle atmosphere ozone: comparison to lidar and stratospheric 15 and Gas Experiment II satellite observations, J. Geophys. Res.-Atmos 100, 30053016, doi:10.1029/94JD02947, 1995. 6110

which takes the ozone profile from an ozonesonde, lidar or a satellite limb sounder as the "true profile with super high resolution and without a priori influence". We agree that this assumption is not fulfilled by 100%. On the other hand we are sure that the intercomparison results only would marginally change while the programming effort of the "two kernel" (or "regridding")-method is extremely high. We cannot fulfill your wish since it is beyond our manpower resources. It would be good if the science community can provide validated "open access programs" for advanced "two-kernel" methods so that not each single scientist has to invent the wheel again. Once I asked Yasmine Calisesi for such a routine but she did not give the routine to me since the routine was not user-friendly enough.

Calisesi, Y., V. T. Soebijanta, and R. van Oss (2005), Regridding of remote soundings: Formulation and application to ozone profile comparison, J. Geophys. Res., 110, D23306, doi:10.1029/2005JD006122.

In the revised version we mention the shortcoming of the Tsou et al. approach and we will tell our wish for a user-friendly program following the regridding method of Calisesi.

Your questions in the Detailed comments: 2) p. 6099, last line measure : : :ozone profiles with the same accuracy. Is there the instrument ageing? How does it affect the accuracy of ozone profiles?

We will add a sentence about the instrument ageing. A microwave radiometer requires a good technical staff. Fortunately things are getting easier now: the FFT spectrometer is much easier to maintain than a filter bench or an acousto optical spectrometer which require regular control checks/adjustments by an expert.

4) P.6109, l. 20-21. only coincident GROMOS profiles between 5 and 9 p.m. in winter and between 7 and 11 p.m. in summer have been taken into account - Why?

Good question, I guess that the lidar has the best quality before noon. I am going to ask the OHP lidar expert Maud Pastel for this.

Thank you for the other comments which we are going to include in the revised version!

Point-to-Point Response to Reviewer 2:

1) the discussion of the retrieval errors is too brief and the errors estimation is not used to interpret the results of the comparisons,

Thank you for this comment. We are going to provide more information about the uncertainties. In the previous version the discussion of the systematic biases were in the focus since the biases can only be found by intercomparisons of independent and coincident measurements. Our retrieval is capable to propagate some error sources (e.g. thermal noise of spectrum) from the ozone line spectrum to the ozone profile. So, we can address your wish in the revised article and that the article will be improved by this.

2) the comparison between the FB and FFTS retrievals should be discussed alone in a dedicated section to clearly show possible bias,

Your point is similar to point 1 of the other reviewer whom we answered:

We will clarify the purpose of our study in the revised manuscript. Our planing is to publish two articles. The main purpose of the present article is the quality assessment of the new FFT spectrometer. We know from past intercomparisons that the old filter bench (FB) of GROMOS made a good job and was within 10 percent compared to satellite and ozonesonde data. However it is better to perform intercomparisons of GROMOS FFTS with independent data such as from ozonesondes and Aura/MLS. The main purpose of the present article is to present and discuss these intercomparisons. So the data of GROMOS FB play a secondary role. However the present study already shows that a later harmonization of the ozone time series from FB and FFTS will be successful (strong adjustments are not necessary).

The harmonization of the data sets will be done in a future article and is the task of the new PhD student. The future article will derive uncertainties and will describe the harmonization process in all details. Finally trend analysis of the harmonized time series will be performed.

Please excuse us if we won't follow you in all of your advices, since it would be too much to perform the data harmonization already in the present article. In any case we sort and clarify the text so that the reader recognizes our purpose in a better way.

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.

Thank you, we are going to check it. The right panels in Figure 5 show for example the standard deviation. However it is more like the natural ozone variability and not the instrument precision. We will add informations about the instrument precision. It is interesting that FFTS has a much better precision than FB. This means that the time step of the ozone profiles of FFTS can be reduced by a factor of 2 or more compared to FB.

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 ?

Thank you, we will add these informations. If I remember right than the binning in the line center is over 3 channels and outside the binning is about much more channels. Actually 150 frequency points are enough for a spectrum inversion. I will look into the programs and provide these informations. The FB integrates over 3 min intervals while the FFTS integrates 20 sec spectra. More details will be in the revised article.

- 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.

I think one can roughly estimate the AVK center and the vertical resolution in Fig. 1. However we are going to provide a plot with the a priori contribution in addition, since we forgot it.

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

yes, the retrieval was coordinated parallel for FB and FFTS. However the FB required some tuning, e.g. because of 1-2 bad channels, which were of course not applied to the FFTS retrieval

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 (30different. 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).

You are right, the AVKs should not be cut at 0. Unfortunately the article was produced with a lack of time at the end of the 4-year PhD interval, and this explains some of the errors. There seems to be an unknown force that many PhD students shift many things into the last year.

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).

Measurementerrors 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.

this will be done parallel to the wish of the other reviewer to provide a detailed error discussion. We spent a lot of time for having the retrieval as it is now. One impression was that provision of flexible and channel-dependent error bars are disliked by the inversion method leading to unstable solutions and oscillations in the profile. For the derivation of a long-term ozone series it is better not to go to the edge of what is possible. For case studies it is possible to select small error bars and to retrieve with "high risk/high gain", e.g. ozone profiles with high vertical resolution resolving a double peak in the ozone layer.

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.

I am not aware of another ozone microwave radiometer beside GROMOS and SOMORA in central Europe. We compared GROMOS with SOMORA in Payerne, however the SOMORA retrieval was not optimized. Thus we could not use SOMORA for intercomparison in the present article. The Bordeaux ozone radiometer is not operating since a long time. We agree that there might be some possibilities e.g. Hohenpeissenberg lidar however it also was a question of time since the FFTS retrieval was lately finished, ca. an half year before end of the PhD time.

Your other comments to errors, standard deviations are certainly be included in the revised version.

Thank you for your detailed comments and your effort!