

Author's response

Ozone trend profiles in the stratosphere: combining ground-based data over Central Europe to consider uncertainties

Bernet et al., Atmos. Chem. Phys. Discuss., 2018, <https://doi.org/10.5194/acp-2018-1213>

Dear Referees, dear Editor Jens-Uwe Grooß,

we would like to thank the referees for the detailed investigation of the manuscript and for the constructive and elaborated comments and suggestions. We have taken the remarks into account and present a detailed answer in the following.

According to the Editor's comment, we repeated our analysis without excluding negative values, to avoid the appearance of an artificial positive bias. The text and all figures have been adapted accordingly.

We attach a revised version of the manuscript with marked changes and hope that we have responded satisfactorily to the suggestions and remarks. The referee's comments are given in italic typeface, our responses are given in blue, and the corresponding changes in the manuscript in grey.

Kind regards,

Leonie Bernet (on behalf of all co-authors)

Contents

1 Author's response to Referee #1	2
1.1 General comment	2
1.2 Specific comments	2
2 Author's response to Referee #2	13
2.1 General comment	13
2.2 Major comments	13
2.3 Minor comments	18
2.4 Technical corrections	34
3 Bibliography	40
4 Marked-up manuscript	42

1 Author's response to Referee #1

1.1 General comment

This is the first review of the paper titled "Ozone trend profiles in the stratosphere: combining ground-based data over Central Europe to consider uncertainties" by Leonie Bernet et al. The paper describes effect of the sampling and instrumental artifacts on derived trends. Paper focuses on the Microwave time series from Bern, Switzerland and investigates step changes in the record due to modification to the instrument in 2009. The impact of step changes in the record on the trend and uncertainties is tested with the synthetic data record. The temporal sampling and time period selected for trend analyses are also tested by authors to determine uncertainties in the derived stratospheric ozone trends. Authors compare other ozone records from several local and regional NDACC stations with Bern Microwave ozone record and find positive trends in the upper stratosphere, consistent between ground-based instruments and satellite records (Aura MLS record). Derived trends in the upper stratosphere are similar to previously published analysis (i.e. WMO Ozone assessment, 2014; Steinbrecht et al, 2017; Petropavlovskikh et al , 2018). Larger differences in derived trends are found in the middle and lower stratosphere. Authors conclude that their proposed anomaly tool will help to improve trend estimates for ground-based stations.

This is a very well written paper. Analyses of ground-based data records for Europe are done with great detail to instrumental artifacts, sampling biases and trend model assessment to assess impact of anomalies on the trends and uncertainties. Comparisons with satellite overpass data help to identify differences in the considered ground-based records and suggest steps to improve homogenization of the GB records. The detailed comments are provided below.

We thank the referee for the positive feedback and give detailed answers to the comments in the following.

1.2 Specific comments

p.1 abstract, line 9 . "We compare our improved GROMOS trend estimates with results DERIVED from other ground-based station OZONE RECORDS () in Central Europe." Add " derived and "ozone records". You might also indicate that these are NDACC records.

Corrected.

p.1 line 20 "Antarctic ozone started" -> "stratosphere ozone over Antarctica started "

Corrected.

p.2 line1 "Outside of the polar vortex, however, increasing ozone is more difficult to detect" - why? I always thought that due to large dynamical variability in the vortex position over Antarctic (or Arctic) it is more difficult to detect a trend as compared to less variability observed over other locations, especially in the upper stratosphere. Are you discussing low stratosphere trends here?

The text was adapted to follow your comment.

Outside of the polar regions, however, differences in ozone recovery are observed depending on altitude

and latitude.

p.2 line 28 "All of these studies" - I does not feel right to use "these studies" at the beginning of the paragraph. May be you can change it to the "As discussed above, several published studies found similar in magnitude trends in the corrected satellite records . . ."

The above presented studies agree on positive ozone trends in the upper stratosphere with some differences in magnitude, [...]

p.3 line 23, "altitude information of the ozone molecules" -> "information about vertical distribution of ozone molecules"?

Corrected.

p.4 line 2, "subtracting mean bias" - was it annual mean or monthly mean bias?

It was the mean over the whole overlapping period. We adapted the sentence to clarify.

[...] by subtracting the mean bias profile averaged over the whole overlap period ($FB_{mean} - FFTS_{mean}$) from all FB profiles (Moreira et al., 2015).

p.4, line 16 (or line 18) "3-days moving window" (or 30-days moving window) - should it be "3-days moving average" ("30-days moving average")?

It has been corrected to "moving median window", which is more precise than "average", because the median has been used as statistical average in the smoothing window (see also comment from Ref. #2, p.4, l.16).

p.4, line 27 At the end of the paragraph, please provide a brief description of main differences in the processing of these two station records (i.e different spectroscopic datasets, averaging time period of the measurements, different error estimates, vertical resolution, smoothing, external information used in retrieval such as temperature profiles), some information that will let reader know if there are differences between the two retrieval methods such that they would create differences in the processed data originating from the same measurement.

There are indeed differences in the retrievals of GROMOS and SOMORA instruments, mainly because the retrievals are optimised for each instrument. We agree that it would be a good idea to compare in more detail the processing differences of both instruments. However, this would also require a deep analysis of how the different parameters affect the ozone data to draw conclusions about the differences in ozone profiles, which would be beyond the scope of the present study. We therefore refrain from adding more detailed information about the processing in the manuscript and think that the description in Sect. 2.1. fulfils the requirements for the further analysis in our study. For more information regarding the retrieval, please refer to Moreira et al. (2015) for GROMOS and Maillard Barras et al. (2009) and Maillard Barras et al. (2015) for SOMORA.

p.5, line 1 "influence" -> interference?

Corrected.

The lidars can only retrieve ozone profiles during clear-sky nights due to scattering on cloud particles and the interference with sunlight.

p.5 line 2 "averaged measurement error" - is it derived for one night of averaging or for the entire month? Is it standard deviation of averaged profiles or error propagation analyses?

The "averaged measurement error" mentioned here is the mean value of all uncertainty profiles given in the study period. For each individual lidar ozone profile averaged over one night, the uncertainty is given by the statistical error of the signal. We adapted the text to clarify this.

In this study we limit the data to the altitude range where the measurement error averaged over the whole study period is below 10 % [...].

p.5 line 5 "host" sounds like lidar is temporary located in France. Would "operated" be a better way to describe it?

Corrected.

p.5 line 18. Instead of "those" use Dobson . Also, at the end sentence add "for normalization".

Corrected.

p.6 line 11 . "in the time period FROM Jan, 1995. . ." - add "from"

Corrected.

p.6, line 11."The different GB instruments. . ." - add GB (ground-based). May be you can abbreviate ground based earlier in the text.

We adapted the sentence after changing the structure of the methodology parts (see major comment 3 from Referee #2).

The stratospheric ozone profile data used in the present study comes from different ground-based instruments that measure in Central Europe (France, Germany, and Switzerland, Table 1).

p.6, line 27. "We therefore interpolate all lidar and ozonesonde profiles . . ." Interpolation from altitude to pressure level requires knowledge of temperature . Temperature is typically available for ozonesondes from radiosonde measurement. But it is not always the case for the lidar (unless lidar also measures temperature). What temperature profiles is used to interpolate OHP and MOH lidar data from altitude to pressure level? Is it taken from reanalyses? ECMWF?

The pressure and temperature data in the OHP lidar files are profiles of radiosondes launched in the nearby Nîmes station located at about 200 km, completed by NCEP operational data above the burst altitude of the balloon. At Hohenpeissenberg, temperature and pressure profiles from the co-located ozonesonde profiles (from MOH) or radiosonde profiles from Munich Oberschleissheim are used. Above ~27 km, the data is completed by co-located lidar temperature measurements, which are also used to extrapolate the radiosonde pressure data. We added a sentence to the manuscript to include this information.

[...] In case that the data were given in number density ($molecules\,cm^{-3}$), the VMR was calculated with the air pressure and temperature provided by the same instrument for ozonesondes or co-located ozone- or radiosonde data for lidars. For lidar measurements, these sonde temperature and pressure profiles are completed above the balloon burst by operational model data from the National Center for Environmental Prediction (NCEP) at OHP and by lidar temperature measurements and extrapolated radiosonde pressure data at MOH.

p.7, lines 26-27. "...was corrected by its mean relative difference to GROMOS". Was the mean calculated over the entire period of the GROMOS/GB instrument time series?

Yes, the data were corrected by the mean value of the relative difference between GROMOS and the respective other instrument, by averaging the relative differences of the whole time period. This correction helped to better identify periods with anomalies, because constant offsets between GROMOS and another dataset were thus removed, and the same anomaly criteria could then be applied for all instruments. According to the comment of Referee #2 (p. 7, l. 25-27), this paragraph was adapted and an equation was inserted to clarify the matter.

To identify these anomalies we used a debiased relative difference (RD_{debiased}), given by

$$RD_{\text{debiased},i,X} = RD_{i,X} - \overline{RD}_X, \quad (1)$$

where \overline{RD}_X is the mean relative difference of GROMOS to the data set X over the whole period.

p.7 line 29 "the respective month was identified as anomaly" Was anomaly identified in GROMOS data or in any of the compared instrumental records?

The idea of the anomaly detection was to find out if GROMOS data show anomalies compared to the other instruments. We therefore only identify anomalies in the GROMOS records. The sentence was adapted to be more precise.

When this debiased relative difference was larger than 10% for at least three instruments, the respective month was identified as anomaly in the GROMOS data.

p.9 line 7 "the degrees of freedom of the trend fit" - can you please discuss this here - how you derive it.

The degrees of freedom of the trend fit are the number of data points minus the number of fit variables, including relevant correlation terms. We adapted the text to specify this as followed.

The degrees of freedom are the number of data points minus the number of fitted variables. The latter are the number of the coefficients of the trend model plus the number of relevant correlation terms inferred by

the procedure described above.

p.9 line 21. "added to the corresponding part of covariance matrix" - can you provide more details? May be an example can be provided at this point xto demonstrate the process?

A sentence with an example was added. A more detailed example with an example covariance matrix is given in von Clarmann et al. (2010).

For example, to account for a bias in the summer months of 2014, a fixed bias uncertainty is added to all variances and covariances of these months in \mathbf{S}_{bias} .

p.9 line 33-34. "value of the corresponding dataset at each altitude" - did you already remove bias earlier in the process? And thus, this will be the same value for all datasets, yes?

For each trend profile, the mean profile of the corresponding instrument is used to determine the trend in percent per decade, so it is not the same value for all datasets. For example, to determine the trend in percent for the lidar at MOH, the model trend output for the MOH lidar (in ppm per decade) is divided by the mean ozone profile from the MOH lidar (in ppm). The sentence has been adapted as followed.

$$\text{Lidartrend}_{\text{MOH}}(\%) = \frac{\text{Lidartrend}_{\text{MOH}}(\text{ppm})}{\text{Lidar}_{\text{MOH}} \text{ ozone mean value}} \cdot 100 \quad (2)$$

The trends are always given in % per decade, which is determined at each altitude level from the regression model output in ppm per decade by dividing it for each data set by its ozone mean value of the whole period.

p.11, line 15-17 "this can be explained by the low number of OHP profiles..." - Why, if the two measurements sense the same ozone, there should be no large bias. Otherwise, what you are implying is that some large biases average out if one has a large number of coincident profiles. . . . I wonder if you should use coincident lidar measurements from OHP to evaluate if large biases between GROMOS and OHP sondes are indeed result of the spatial variability in ozone. Perhaps, you can plot PDF of subset of data (full or coincident) for both instruments to evaluate how they sample ozone variability over the same time-period. It will indicate if one instrument is not capable of obtaining low/high ozone, perhaps due to AK smoothing...

The OHP ozonesonde seems to have some outlier profiles. For example, there seems to be a problem with the sonde temperature between Oct. 2002 and July 2007. The OHP ozonesonde data will be harmonized, but the reprocessed data are not available yet for the present study. We have now replaced a few profiles with a corrected version, but it does almost not affect the comparison. The ozonesondes and the lidar at OHP are usually not measuring at the same time. Therefore we can not compare coincident profiles directly. We adapted the text to clarify the influence of the small sample size of the OHP ozonesondes (see the comment p.11, l.15-17 of Referee # 2).

p.11, line 30-33. You indicated that positive anomalies in the middle and lower stratosphere in 2000 were confirmed through anomaly detection method. You also found positive bias in the upper stratosphere in 2000, although it was not confirmed by comparisons with other instruments (MLS and SOMORA). Would the middle and lower stratosphere biases suggests the instrumental anomaly in upper stratosphere in GROMOS as it seems to affect the entire profile?

You are right, an anomaly at a specific altitude might indeed affect also other altitudes in the GROMOS profiles. That's why we considered in the trend programme a bias at all altitudes as soon as an anomaly was detected at a specific altitude.

p.12, line 21-22, why amplitude A is negative? Are coefficients in Eq (3) derived from observations? Why 0.1 ppm was selected for monthly uncertainty? Earlier you indicated that trend is 0.1 ppm per decade. What confidence level of the trend detection is expected in these time series?

The amplitude A was now corrected to be positive. The coefficients g and h have been adapted to be negative, in order to obtain a realistic seasonal cycle with minimal ozone values in winter. The artificial time series was chosen in such a way that it is similar to a real ozone time series at around 10 hPa, where the mean GROMOS ozone value is approximately 7 ppm. The uncertainty of each monthly ozone value was set to 0.1 ppm, which was assumed to be a realistic uncertainty. This is a bit smaller than GROMOS observational error due to thermal noise at 10 hPa. The confidence level for the resulting trend is 99%. If a larger uncertainty had been chosen, e.g. 0.4 ppm, the confidence level of the resulting trend would have been 95%.

p.12, line 21-22 (continued), Also, please indicate the length of the artificial time series (yes, it can be deduced from Figure 5, but should be written in the first paragraph of section 4.1).

We added the time period of the artificial time series in the text.

The artificial time series used for this purpose consists of monthly ozone values from Jan. 1997 to Dec. 2017 and is given by [...]

p.12, line 29. "(~0.4 ppm) " - should it be 0.35 or 0.36 based on $a=7$ ppm in Eq (3) and the rate of change in time series is very small ($0.1/120 \times 12$ (month per year) $\times 20$ (years) = 0.2ppm)

Five percent of the average ozone value of 7 ppm would indeed be 0.35 ppm. However, the anomalies in the artificial time series were only added to summer months in the years 2014, 2015 and 2017. In the summer months the ozone values in the artificial time series for the corresponding years are approximately 7.9 ppm in June and 8.2 ppm in July and August. 5% thus correspond to 0.395 ppm in June and 0.41 ppm in July and August, or ~ 0.4 ppm on average for each summer. We slightly changed the sentence to clarify that we only increase the ozone values for the summer months.

[...] We increased the monthly ozone values in the summer months (June, July, Aug.) of 2014, 2015 and 2017 by 5% (≈ 0.4 ppm).

p.13, line 4, you provide uncertainty now as 5 %? What is it in ppm? Is it still 0.35 ppm? Then 0.1 ppm uncertainty for series A was 1.4 %? It is a bit confusing to switch between % and ppm. Also, it is not clear what "corresponding months" means. Do you mean that uncertainties for JJA months in 2014, 2017 and 2017 were increased to 5 % while the rest of month in the entire time series has 0.1 ppm uncertainty?

Yes, your description is correct, uncertainties were increased in the summer months of 2014, 2015 and 2017 to 5% of the mean ozone value (0.36ppm) and remained 0.1 ppm for the rest of the time series. We adapted the text and Table 2 to be clearer. The uncertainties are now always given in ppm, and only the added anomaly is given in percent. "Corresponding months" were replaced by "affected summer months".

[...] and set the uncertainties of the affected summer months to 0.36 ppm (case C in Fig. 6 and Table 2). This uncertainty value corresponds to 5 % of the overall mean ozone value. The uncertainty of the months without anomalies remained 0.1 ppm.

p.13, line 12-13 you state that trend estimate improves " the higher the uncertainties are chosen". Did you test model with larger than 5 % uncertainties? Why 5 % was selected in the first place?

Yes, we tested the trend model with uncertainties for anomalies larger than 5% and the bias uncertainties larger than 5 ppm. We decided to use 5% for the anomalies because it is a value close to the "real" added anomaly. The same was done later on for the GROMOS trend estimates, where the uncertainty for the anomalies was chosen based on the estimated "real" anomaly (compared to the other instruments), as described in Sect. 4.1.

Concerning the bias uncertainty (5 ppm), we made a sensitivity test and found out that 5 ppm is the smallest value where the trend remains constant. For larger bias uncertainties, the trend estimates remain the same. For smaller bias uncertainties, the a priori bias (which would be a bias of zero) gets more confidence which leads to different trend values. We therefore defined 5 ppm as the smallest bias uncertainty for which the bias is fully estimated from the data instead of being influenced by an a priori null bias.

p.14 line 11-15, "GROMOS trend peak is observed at slightly lower altitude". " We observe a slight shift in the peak" of AK. - what do you mean? Can you explain the shift and when it occurred in time series? Or you mean that the maximum weight in AK profile for layer 35-40 km is higher by 2-4 km than the assumed peak of the AK at the middle of that layer?

Since the new retrieval in 2009, the averaging kernel (AVK) peak does not always correspond exactly to the expected altitude. This means that at some pressure levels, the maximal amount of information is retrieved from slightly higher or lower altitudes than indicated. At pressure levels below ~ 10 hPa, the AVK peak is one or two pressure levels lower ($\sim 2-4$ km), above ~ 3 hPa the AVK peak is one or two pressure levels higher up ($\sim 2-4$ km) than expected (see Fig. 1 (c) in this response). We are aware of this problem and will try to solve it in a new retrieval version. For the present study, however, it is not problematic to use the data. The direct comparison with GROMOS is not affected by this problem because the data is convolved with GROMOS AVK. Only for the trend comparison, where no convolution is applied, it is important to mention this effect. We adapted the text to stress that it is a shift in altitude and not in time.

(Sect. 4.3:) We observe that after the instrument upgrade in 2009, the averaging kernels peak at higher altitudes than expected, [...].

(Sect. 4.4:) This difference might be related to the averaging kernels of GROMOS, which indicate that

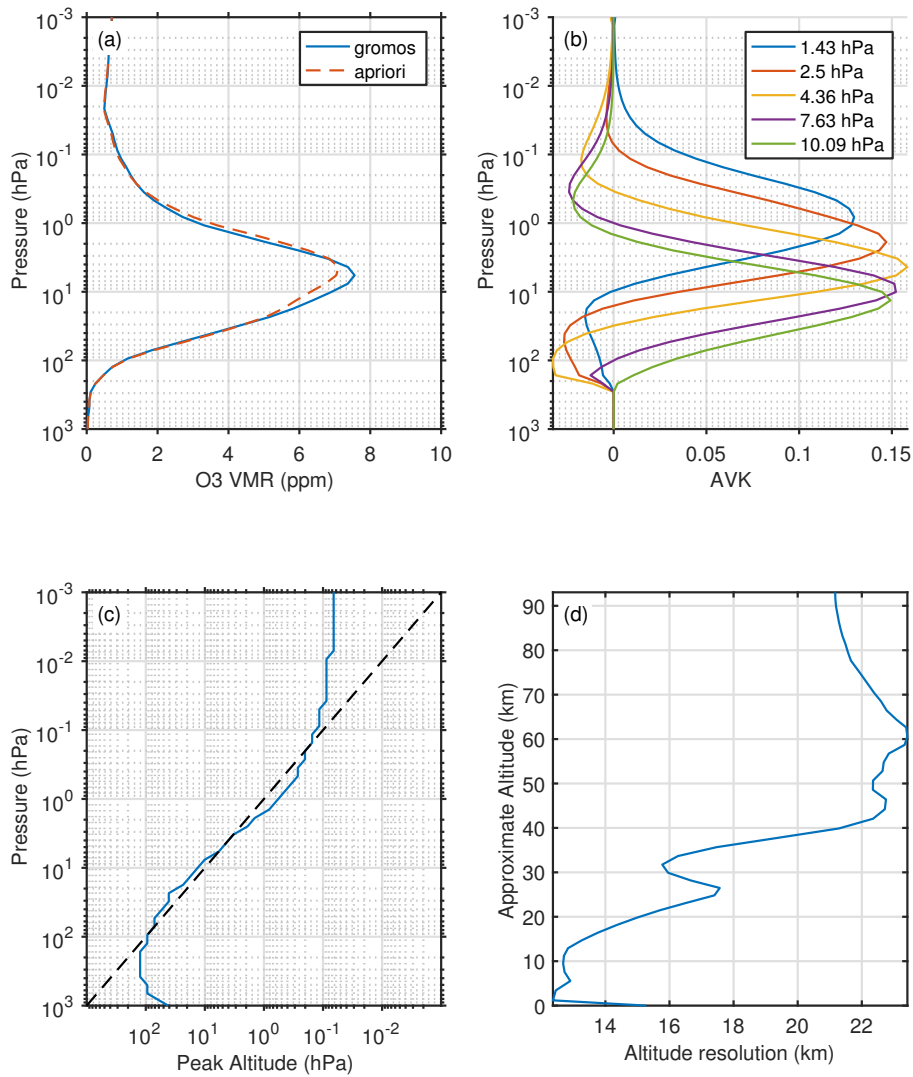
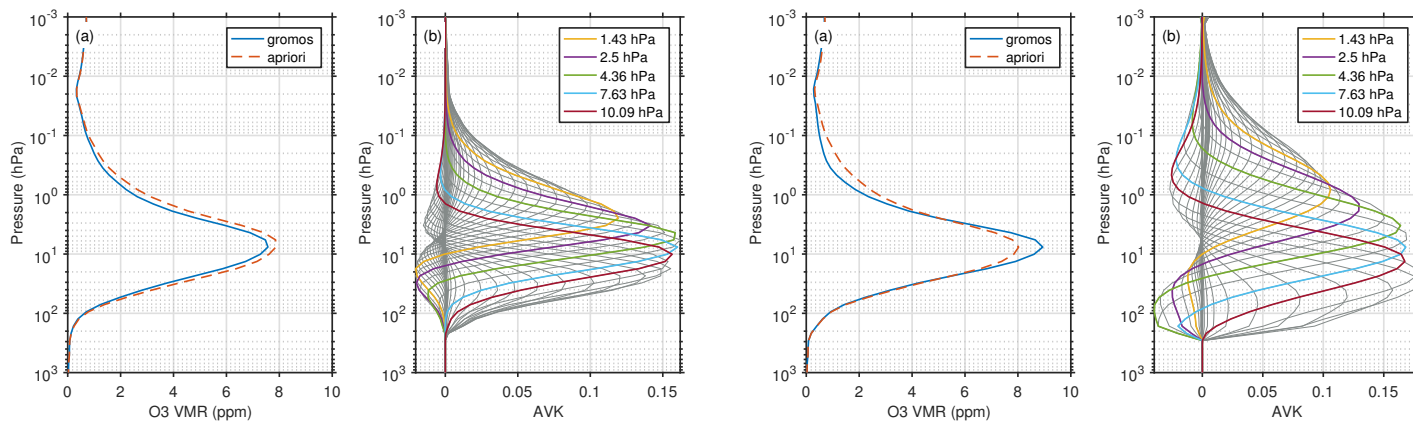


Figure 1: Retrieval figures for a GROMOS retrieval from 02.02.2017. (a) Retrieved and a priori profile, (b) Averaging kernels (AVK) for specific pressure levels, (c) Peak of the AVKs, (d) Altitude resolution (full width at half maximum of the AVKs).

GROMOS retrieves information from higher altitudes than expected (~ 2 km difference).

p.14 line 11-15 (continued), Then, you also suggest that the full width of the AK is between 20 and 25 km. If it is centered at ~ 36.5 km, then you have to assume that the averaged (AK weighted) trend is estimated between 26.5 - 46.5 km altitude range.

It's true that due to the altitude resolution, GROMOS trends at a given altitude are also affected by trends at other altitudes. However, this smoothing due to the averaging kernels (AVK) is not directly related to the altitude of the AVK peak. We therefore decided to remove the corresponding sentence ("Furthermore, one has to keep in mind the coarse altitude resolution of GROMOS, which is at 15 this altitude between 20 to 25 km.") and hope that it clarifies this paragraph.



(a) Example for a filter bench (FB) retrieval

(b) Example for a FFTS retrieval

Figure 2: GROMOS retrieved profile and corresponding AVKs for two sample dates before (a) and after (b) the instrument upgrade in 2009. In both situations, the atmospheric transmittance was 0.4.

p.14 line 11-15 (continued), Can you please provide AK example for the layer?

Below we present an example for the GROMOS averaging kernels during the filter bench period (Fig. 2a in this response) and after the update to the FFTS (Fig. 2b in this response). The differences between the AVKs after the instrument upgrade have been analysed and are accounted for in the harmonisation (see comment below, p. 14, line 16-17).

p.14 line 11-15 (continued), If you apply the AK smoothing to the lidar profile and calculate trends. Would these trends be different from trends derived by integrating ozone profile between 35 and 40 km?

Estimating the trends of the convolved lidar data and integrating it over the mentioned altitude range leads roughly to similar results. However, we did not consider trends from convolved data in the study, because trends are better represented when using the unconvolved data. The convolution might introduce artificial trends, especially towards the edges of the range where lidar data is available. Also, the above described change in the Gromos AVK after the instrument upgrade can introduce some artificial trends. This shift has been corrected in the GROMOS data by the harmonization, but it might introduce an artefact when estimating trends from convolved profiles from other instruments.

p.14, line 16-17. The harmonization of time series before/after 2009 (application of bias and seasonal cycle correction) does not necessarily means that the entire bias in the retrieved profiles will be removed. Say, the change in the instrument optical characterization would produce a shift in the vertical distribution of the AK weights, the upgraded instrument would start sampling at higher/lower altitudes. And since ozone trends can be different at higher/lower altitudes, it would affect the homogenized time series, even after you remove the step-change bias. Can you show the AK before and after 2009 (for approximately the same atmospheric conditions)?

You are right that the instrument upgrade in 2009 led to changes in the AVK (see example cases in Fig. 2 in this response). These differences can also affect the time series and trends. To analyse this effect, we applied the mean AVK before and after the upgrade to a same ozone profile. The difference between the

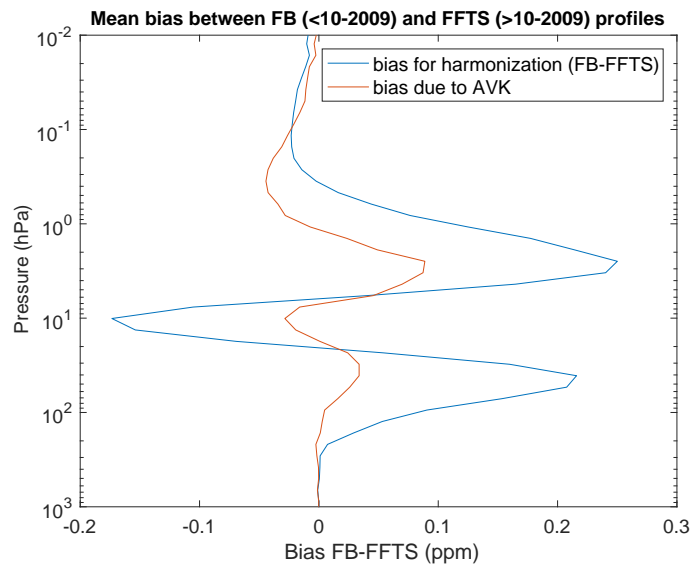


Figure 3: Difference between the mean of filter bench (FB) profiles before 10-2009 and Fast Fourier Transform Spectrometer (FFTS) profiles after 10-2009 during the overlapping measurement period, used for the harmonization of FB data. The red line indicates the difference that is only caused by the difference in averaging kernels.

resulting convolved profiles shows the effect of the different AVKs on the profile (Fig. 3 in this response). We observed that this difference profile (bias due to AVK) looks similar to the difference profile that is applied to the data for the harmonization. The profile used for the harmonization has larger amplitudes in the peaks, indicating that additional differences are also corrected for. We conclude that the harmonization that we applied corrects for the effect that is due to differences in the AVK after the instrumental upgrade.

p.15, line 1. Can the temporal sampling test be done on the artificial time series (change in sampling from 1 hour to 3 minutes)?

The temporal sampling should not affect the artificial time series as long as the diurnal and seasonal cycles are well represented.

p.15, line 15-16. It is confusing to show "ozonesonde" results at altitude above 30 km. Maybe you should stop test at 30km.

It might be indeed confusing that a trend profile based on ozonesonde measurements shows data above 30 km, even though only the ozonesonde measurement time and not the data itself is used. Keeping this information, however, gives information about possible effects in the upper stratosphere when an instrument measures only several times a week at noon. We therefore agree with Referee #2 (comment p.15, l.16) and propose to keep the figure as it is.

p.16,line17-is it 45km or 42km?

The trend peak of SOMORA is indeed at ~ 1.8 hPa (~ 43 km), so we adapted the text.

We observe that all instruments that measure in the upper stratosphere show a trend maximum between ~ 4 and ~ 1.8 hPa (between ~ 37 km and 43 km), [...]

p.16, line 22 "derive" - this sentence seems to have some missing text...

Corrected, the word "derive" was misused and replaced by "deviate".

p.16, line 30 - change "our" to "derived" or "estimated". Also "discrepant results" means "differences in derived trends"?

The sentence was reformulated.

In the middle and lower stratosphere, at altitudes below 5 hPa (≈ 36 km), the estimated trends differ from each other.

Figures 6 b - add uncertainty bars to the results. You can also add circle slightly shifted horizontally to show where they fall.

We adopted the figure as suggested.

7 and 8 - please consider to change either green or blue colors - they are too close to be easily distinguished in the plot. Also, please consider to change uncertainty envelopes from solid color to shaded or not even filled envelopes - it is hard to distinguish between different instruments.

We agree that the choice of color is difficult to discern, and we changed the colors to blue, red, and yellow. An important point in the trend profile figures is to discern if different trend profiles overlap within their uncertainties, and if they are significantly different from zero. It is less important, however, to discern the exact amplitude of the uncertainty. Therefore we decide to keep the uncertainty shading as it is, but added solid lines to the uncertainty envelopes, to be better distinguishable.

10. - the same comment as above - selection of distinct colors and reduction in envelopes. May be the plot can be separate in two plots to reduce overlapping information - one for MOH and one for OHP.

We adapted the colors and separated the figure in two subfigures as suggested.

2 Author's response to Referee #2

2.1 General comment

Short resume: *Outlying measurements are a nuisance to many types of analysis. Time series analysis in particular can be very sensitive to outliers, impacting both the estimated regression coefficients and their uncertainty. This topic is especially relevant for ozone profile trend assessments as the actual ozone trend is expected to be quite small (not more than $\sim 2\%$ per decade) which makes robust detection tricky and susceptible to a handful of outliers. This paper starts off by presenting a comprehensive analysis of periods of outlying measurements in the GROMOS time series. Where most other studies stop here and simply reject the outliers from further analysis, the authors apply a regression method proposed by von Clarmann et al (2010) which uses the timing of the detected/suspected outlying data and infers their magnitude along with other components in the regression model, such as the linear trend. Bernet et al. illustrate the value of the method using synthetic data, before applying it to the actual GROMOS time series. The methodology is explained well and the application is very relevant and adequately elaborated. The last part of the paper moves slightly away from the core subject of the paper, and briefly discusses the impact of sampling and length of the time series on the GROMOS trend result. These are also compared to trends obtained from neighbouring ground-based data records and Aura MLS, however, the adopted outlier detection and incorporation method are not for these data sets.*

Recommendation: *This paper is very well written and the material presented supports the claims made by the authors. The subject is scientifically very relevant and fits the scope of ACP. I would therefore highly recommend to publish this paper as soon as my comments below have been incorporated.*

Thank you very much for your positive feedback and the recommendation. The detailed answers to your comments are given in the following

2.2 Major comments

2.2.1 Order of vertical regridding and smoothing

p.7, l.7-9: The order of regridding and smoothing is unintuitive. First you downgrade highly resolved vertical profiles to a coarser grid, then convolve these with the MWR averaging kernel and a priori. In reality, the MWR instrument does not vertically smooth a coarse atmosphere. Shouldn't the procedure be the other way around? First smooth at best available vertical resolution, then regrid to the coarse grid. Hence you should oversample the averaging kernel in one direction to match the sampling resolution of the x_h , as in Eqs. 10-11 of Keppens et al. (2015, AMT).

We changed the AVK smoothing to the approach proposed by Keppens et al. (2015) (Eq. (15)) by interpolating the rows of the AVK to the highly resolved grid from the respective instrument. The changes compared to the approach used so far are small, but we recompiled all figures with this new approach and adapted the text in Sect. 3.1.2.

The profiles having higher vertical resolution than GROMOS were convolved by the averaging kernel matrix according to Connor et al. (1991), with

$$x_{conv} = x_a + \mathbf{AVK}(x_h - x_{a,h}) \quad (3)$$

where x_{conv} is the resulting convolved profile, x_a is the a priori profile used in the GROMOS retrieval, $x_{a,h}$ is the same a priori profile, but interpolated to the grid of the highly resolved measurement, x_h is the profile of the highly resolved instrument, and **AVK** is the corresponding averaging kernel matrix from GROMOS. The rows of the **AVK** have been interpolated to the grid of the highly resolved instrument and scaled to conserve the measurement response (Keppens et al., 2015).

2.2.2 Unconventional choice of definition relative difference

p.7, l.24: The relative difference is defined as $\Delta = 100(GRX)/GR$. Why did you not choose $\Delta' = 100(GRX)/X$, which is the more widely adopted definition? The ratio between both definitions is $\Delta/\Delta' = X/GR \geq 0$ (you removed all negative values; p.6, l.24-25). Your expression will be more sensitive to detect $GR < X$, the conventional expression is more sensitive to detect $GR > X$. Did you test the sensitivity of the detected periods to this definition?

We agree that the sensitivity for these two definitions is not the same. We therefore tested the anomaly detection also with the conventional definition and found that the two approaches lead only to small differences in the detected anomalies. We had chosen $\Delta = 100(GRX)/GR$ as definition for relative differences because the aim is to compare all the different instruments with GROMOS. We think that having GROMOS as a constant reference when calculating the relative differences is therefore more appropriate than having different references for each instrument comparison. Because of this advantage, and because the difference between both approaches is small, we decided to keep the definition as it is in the manuscript.

2.2.3 Description far away from application

The description of the methodology (Sects. 2.2 and 2.3) is quite far away from the actual application. This makes it a slightly more difficult. You could consider explaining the outlier detection and regression methodologies right before their application in Sects. 3 and 4, rather than in Sect. 2.

We followed your suggestion and changed the manuscript accordingly. We moved the comparison methodology in the beginning of Sect. 3 (Time series comparison), and the explanation of the trend model to Sect. 4 (Ozone trend estimations). Some section titles and the introductions of the affected sections have been adapted as followed.

2 Ozone data sets

The stratospheric ozone profile data used in the present study comes from different ground-based instruments that measure in Central Europe (Table 1). They are all part of the Network for the Detection of Atmospheric Composition Change (NDACC, 2018). In addition, we used data from the Microwave Limb Sounder (MLS) on board the Aura satellite (Aura/MLS). All data from the different stations are compared to data from the GROMOS radiometer located in Bern, Switzerland (46.95° N, 7.44° E, 574 m above sea level (a.s.l.)). The aerological station (MeteoSwiss) in Payerne (Switzerland, 46.8° N, 7.0° E, 491 m a.s.l.) is located 40 km south-west of Bern, which ensures comparable stratospheric measurements. The Meteorological Observatory Hohenpeissenberg (MOH, Germany, 47.8° N, 11.0° E, 980 m a.s.l.) is located 290 km north-east of Bern, and the Observatory of Haute Provence (OHP, France, 43.9° N, 5.7° E, 650 m a.s.l.) lies 360 km south-west of Bern. Even if stratospheric trace gases generally show small horizontal variability, the distance between the different stations, especially between MOH and OHP, may lead to some differences in measured ozone.

3 Time series comparison

To identify potential anomalies in the GROMOS data, we compared the data with the other described data sets in the time period from Jan. 1995 to Dec. 2017, except for some instruments that cover a shorter time period (Table 1).

3.1 Comparison methodology

To compare GROMOS with the other instruments, the different data sets have been processed to compare consistent quantities and smoothed to the GROMOS grid. Taking relative differences between the data sets made it possible to identify anomalous periods in the GROMOS time series.

4 Ozone trend estimations

Ozone trends are estimated in the present study by using a multilinear parametric trend model (von Clarmann et al., 2010). By comparing the GROMOS data to the other data sets as described above (Sect. 3.3), we have confirmed some anomalous periods in the GROMOS time series. To improve the GROMOS trend estimates, we use now these detected anomalies and consider them in the regression fit. In the following, the trend model (Sect. 4.1) will first be applied to an artificial time series to test and illustrate the approach of considering anomalies in the regression (Sect. 4.2). It will then be applied on GROMOS data (Sect. 4.3), before comparing the resulting GROMOS trends to trends from the other instruments (Sect. 4.4).

2.2.4 Outliers are only detected & considered for GROMOS

You should state clearly that the proposed outlier detection and outlier incorporation methods are only applied to the analysis of the GROMOS time series. You could have done that for the other instruments as well, this would have resulted in a more fair comparison and improved the discussion in Sect. 4.5.

We understand your remark that including the trend correction also for the other data sets would have improved the discussion. Unfortunately, applying the method to all data sets was beyond the scope of the current study. Further, GROMOS data show larger anomalies than most of the other data sets (see e.g. their better agreement in Fig. 3), which makes the correction more urgent for GROMOS than for most of the other data sets. We added a sentence to clarify that the method was only applied to GROMOS data (see comment p.9, l.29).

2.2.5 Show trend results when outliers are removed

Many, if not most, researchers remove outlying data points prior to regression. I am interested in seeing how this simple approach performs in the particular case of real GROMOS data (Sect. 4.2, Fig. 7) and of synthetic data (Sect. 4.1, Fig. 6). I acknowledge that this will not allow to draw conclusions that apply generally, since the result depends on the number and character of the outliers. At least this shows whether it was worth to invest extra coding time and analysis effort.

We consider it a brute method to simply discard data of an anomalous data subset. To use the data in a conventional scheme or to discard them is sort of black and white view to the problem. Any additional additive component in the covariance matrix, regardless if only in the diagonal of \mathbf{S}_y , or a block encompassing multiple data points, will reduce the weight of the anomalous data, and thus we have the entire grey-scale between the "black" and "white" option available. This gives us greater flexibility.

Enhanced diagonal entries (case C in the artificial test time series) simply gives less weight to the anomalous data. Test cases D and E, via the correlated block in the covariance matrix, keep information on the relative differences between the anomalous data but largely ignore differences between the anomalous and the regular data. The latter approach performs better than simple downweighting. Since downweighting is a step towards discarding anomalous data, the poorer performance of setup C in comparison with D and E indicates that the covariance approach might indeed be better.

Rejection of anomalous data might be an option if the fraction of anomalous data is very small, but if a larger fraction of the data is subject to anomalies, we think that it is worthwhile to have a method which can save these data. Also, the trend uncertainties increase the more data is omitted in the time series. We concede that the method depends on two assumptions: (1) We need a priori knowledge on which data points have this anomaly, and (2) the anomaly must indeed be bias-like. This method (cases D or E) will not work for independent outliers, but we state these caveats in the paper.

Concerning the extra effort: The software tool has not been developed for this particular purpose. The motivation for the development was to be able to cope with inhomogeneous data sets biased against each other. But, having this tool available anyway, why not using it, since it has been shown to be useful also for the treatment of anomalies?

We also tested omitting anomalies in the artificial time series. This approach leads in this test data to similar results as the advanced uncertainty approach and is close to the true trend. This can be explained by the fact, that the artificial time series only consists of an ideal seasonal cycle without any noise, that is perfectly reproduced by the trend model, also if some data are missing. Nevertheless, even in this case omitting the anomalies leads to increased trend uncertainties. In a real time series, however, the data is less predictable. The GROMOS trend estimate when omitting anomalies is close to the trend estimate that simply weights the anomalies differently, but different from the corrected trend estimate where biases are fitted for anomalies.

We conclude that omitting data in the time series increases trend uncertainties and leads to a loss of information about data in the anomalous periods. We added the following paragraph to Section 4.1 to mention the disadvantage of omitting data.

A simple way to handle such anomalies would be to omit anomalous data in the time series. This, however, would increase trend uncertainties and lead to a loss of important information. Therefore, we use the presented approach to handle anomalous observations in the time series when estimating the trend. To account for these anomalies in the trend estimation, we make use of the fact that the user of the trend programme has several options to manipulate the error covariance matrix \mathbf{S}_y . [...]

2.2.6 Expand discussion to trend uncertainty

The discussion in Sect. 4.1 is very insightful but does not treat another important factor in trend analysis: trend uncertainty. Can you elaborate by how much the outliers affect the trend uncertainty and whether the more advanced method is also able to provide more robust estimates of trend uncertainty? One element that determines trend uncertainty is the noise in the time series. The example shown in Sect. 4.1 is hence an idealised case without random uncertainties. The few added outliers will therefore have a larger impact on the trend estimate than in the more realistic case where noise would have been present. Have you

studied this effect? And please add a comment that the influence of outliers may be smaller in the case random errors would have been simulated as well.

The trend uncertainty is estimated by the algorithm by propagating the S_y matrix onto the resulting trend. By default, noise in the data as well as autocorrelated error terms (which can be understood as atmospheric variability not covered by the trend model) are included. To account for anomalies in the data, we do not restrict ourselves to diagnosing their impact by a sensitivity study (case B) but we immunize the trend tool against anomalies, which is possible under the assumption that we know which data points are affected by the anomaly. For this purpose, we use the special feature of the trend program, which is that we can assign correlated uncertainties to subsets of the data. Just assigning a larger uncertainty to the anomaly-affected data points already reduces the impact of the anomaly from 33% down to 3%. Using the special feature to assign a fully correlated uncertainty (bias of unknown magnitude) to the data makes the trend estimate even more robust and the impact of the anomaly is almost fully removed. After this "immunization", the anomalies increase the trend uncertainty from ± 0.010399 ppm/dec to ± 0.010815 ppm/dec (last column in Table 2, but the precise digits are not shown in the table). These trend estimates include noise, the auto-correlated terms, and, for cases B to E, also the anomaly-induced uncertainties. This strategy might not have become clear from the original text. We have thus tried to rewrite parts of Section 4.1 (see Sect. 4.2 in the new manuscript). We do not agree that presence of noise in the data makes the effect of anomalies smaller. We can safely consider noise and anomalies as independent sources of error. Thus, their effects will add up quadratically.

2.2.7 Positive trend at lower altitudes?

Please tone down the statement on p.14, l.33-35: "This result implies that ozone might also recover at lower altitudes, but the uncertainties, the dependency on anomalies, and the insignificance of the corrected trend show that this result is less robust." I read it as (1) there is a recovery (2) but the result is less robust. So, do you claim a recovery in this altitude region, or not? In fact, it is the region where GROMOS trends differ most from those obtained by other satellite- and ground-based techniques. Hence, more caution is needed in how this is phrased.

We agree that a recovery can only be claimed if the result is robust. We therefore rephrased the sentence as followed.

The larger uncertainties in the lower stratosphere, the dependency on anomalies, and the insignificance of the corrected trend show that this positive trend in the lower stratosphere is less robust than the trend at higher altitudes.

2.3 Minor comments

p.1, l.0: The title does not convey the main message of the paper. It should be improved so readers will find your work more easily. First, the manuscript is all about detecting unusual observations and taking these into account in the regression. Hence the "to consider uncertainties" is too vague, I believe you should be more specific. Uncertainties can mean anything, really. Second, the presence of "combining ground-based data" suggests that you merged data sets prior to regression, which is not the case. This is misleading, which is definitely undesired for a title of a paper. What about something along the lines of "Combined study of ground-based ozone profile data over Central Europe to detect and incorporate anomalous observations in the analysis of trends in the stratosphere"

Thank you for your suggestions, the title has been adapted to better convey the main message.

(New title:) Ground-based ozone profiles over Central Europe: incorporate anomalous observations in the analysis of stratospheric ozone trends

p.1, l.1-15: The abstract should be improved, as it is not clear exactly what the paper adds to the existing body of literature. For instance, "[...] an approach for handling suspicious anomalies [...]" does not say how you handle anomalies. This can be done in many ways. In my view, you should state clearly (1) that you analysed multiple ground-based data records to detect anomalous periods in the Bern MWR data record, and (2) that you then used this information to regress the Bern time series for long-term trends and the temporary bias for the detected anomalous data taking periods.

We adapted the abstract to clarify the anomaly approach.

We present an approach for handling suspicious anomalies in trend estimations. For this, we analysed multiple ground-based stratospheric ozone records in Central Europe to identify anomalous periods in data from the Ground-based Millimetre-wave Ozone Spectrometer (GROMOS) located in Bern, Switzerland. The detected anomalies were then used to estimate ozone trends from the GROMOS time series by considering the anomalous observations in the regression. We compare our improved GROMOS trend estimate with results derived from the other ground-based ozone records (lidars, ozonesondes, and microwave radiometers), that are all part of the Network for the Detection of Atmospheric Composition Change (NDACC).

p.1, l.18: "[...] in recent years [...]". Add LOTUS Report and replace (WMO, 2014) by (WMO, 2018) to the citation. Also, in the submitted version you sometimes refer to (WMO, 2014) and sometimes to (Pawson et al, 2014). Choose just one of these where relevant.

We added the LOTUS report here, and changed Pawson et al. (2014) to WMO (2014) in the whole manuscript.

p.2, l.6-8: Incorrect language. The linear trend fitted represents the linear component in the time series after all other natural variations are accounted for (solar, QBO, ENSO, aerosol, ...). This linear term is expected to receive contributions from decreasing ODS and increasing GHG, and possibly other unidentified processes. It is not possible to fit a linear term due to ODS and another linear term due to GHG, since these terms

are fully dependent. Hence, it is incorrect to state that "a trend might be masked by natural variability and factors such as GHG increases and changing dynamics". So this needs to be rephrased.

The purpose of this sentence was not to state that a trend can be masked by factors such as GHG, but by natural variability. In addition we wanted to illustrate the complexity of ozone trend detection by mentioning some of the many processes that influence stratospheric ozone. We rephrased the sentence.

[...] detecting a small trend is a difficult task. Many factors influence stratospheric ozone such as variations in atmospheric dynamics, solar irradiance or volcanic aerosols and the increase of greenhouse gases (WMO, 2014). Further, ozone trends might be masked by natural variability. Other important sources for trend uncertainties are [...]

p.2, l.10: Different in what sense? If you combine data records from different instruments, ok. But here such a merging is not considered. So a "changing resolution" or a "changing sampling pattern" is more relevant for a time series composed of measurements by one instrument.

It's true that the first three aspects (instrumental drift, abrupt changes, biases) are reasons for trend uncertainties within a single time series, whereas the last point (sampling issues) is mostly relevant when comparing trends from different instruments. However, all of them can lead to differences between trend estimates from different data sets, which is the topic of this paragraph. We adapted the sentence as followed to clarify this aspect.

Other important sources for trend uncertainties are instrumental drifts, abrupt changes, biases, or sampling issues e.g. due to instrumental differences in sampling patterns or in vertical or temporal resolution.

p.2, l.11: Add Frith et al. (2017) to the citation. The paper is more recent but also deals with ozone profiles which is closer to what you report in the manuscript.

Corrected.

p.2, l.12-14: The influence of outliers at start and end points on the trend estimates is well known and not discovered by Bai et al as suggested by the current phrasing. I propose to move the citation to the end of the phrase such as (e.g., Bai et al., 2017).

Corrected as suggested.

Biases in ozone data sets can lead to important differences in trend estimates, especially when they are occurring in the beginning or the end of the considered trend period (e.g. Bai et al., 2017).

p.2, l.20: I am not sure there is evidence that all data sets used by Steinbrecht et al. (2017) have been successfully corrected for drift. Hubert et al. reported drift in parts of the HALOE, SCIAMACHY and GOMOS data sets. These have not been corrected. And so far no one has proven that the OSIRIS correction is removing the entire drift. Please rephrase this sentence.

Compared to older studies (Harris et al., 2015; WMO, 2014), Steinbrecht et al. used newer data sets with reduced drifts (e.g. SAGE-OSIRIS) and omitted older data sets with large drifts (SAGE-GOMOS), see

Table 3 in Steinbrecht et al. (2017). We rephrased the sentence indicating that they used "data sets with small drifts", instead of "drift corrected data sets".

The study of Steinbrecht et al. (2017) summarizes recent trend estimates using only updated satellite data sets with small drifts.

p.2, l.34: Add a reference to the LOTUS report as it contains the most recent and complete set of trend estimates.

We added the LOTUS report as reference.

p.3, l.1: Confusing language "handle [...] anomalies by considering anomalies [...]". It would help to define up front what you mean with "anomaly". Is it due to the instrument, due to natural processes, ...? A great deal of readers will read anomaly as something that is not problematic, as it is caused by natural processes. Others will interpret it as an issue to solve.

We adapted the sentence and defined "anomalies" in the introduction (see comment below, p.3, l.12).

The present study proposes an approach to handle the problem of anomalous observations in time series by considering the anomalies when estimating trends.

p.3, l.5: "[...] resulting in a corrected trend estimate [...]". Corrected with respect to what? Perhaps "improved" is more in place here?

We agree that in this part of the manuscript, an "improved trend" is more appropriate to stress the fact of the improvement. We changed the term from "corrected trend" to "improved trend" in this paragraph. Later on, however, the correction of the anomalies in the trend estimation is described (e.g. Sect. 4.2), which illustrates that the trends are corrected with respect to the trends where anomalies were not considered. We therefore decided to keep the term "corrected trend" in the part of the manuscript that follows the description of how this "correction" is performed.

p.3, l.12: The word "anomaly" is often also used for anomalous geophysical events. This class of anomalies is not the subject of this manuscript. In caption of Fig. 1 you use "anomaly" in the sense of a "deviation from a climatology", which agrees with how many readers will interpret it. To avoid any confusion I suggest to define "anomaly" unambiguously at the beginning of the manuscript (and perhaps repeated here and there in the text). And are inhomogeneity (see comment below) and anomaly the same for you?

We now defined anomalies in the introduction and repeated it later on. We also removed the word "anomaly" in the caption of Fig. 1 and in the corresponding text (Sect. 3.2) to avoid confusion. Concerning the inhomogeneities, see comment below (p. 8, l.15).

(Introduction:) We define anomalies as periods where the data deviates from the other data sets.

(Sect. 3.2:) This ozone deviation is calculated by the ratio of the deseasonalised monthly means (difference between each individual monthly mean and the corresponding climatology of this month) and the monthly climatology.

p.3, l.11: Section 2 is a tad long. If you consider trimming the manuscript (which is not what I propose), then this section would be a good place.

We agree that the data and methods section is a bit long, but we think that it is important to give information about the used instruments. Also, it is essential for the study to explain details of the trend programme and how it is used to account for anomalies in the trend estimates. However, following your suggestion, we reordered the method part by moving the time series and trend description to the respective section (see major comment 3).

p.4, l.16: Add "mean" or "median" to the moving window to clarify what statistical measure you used for the location of the smoothed data.

"Moving window" has been replaced by "moving median window" (see also comment from Referee #1, p.4, line 16 (or line 18)).

p.5, l.17: Replace by "[...] using concurrent total column ozone data from the Dobson [...]". This to clarify that the corrections are done using data taken at the same/similar time.

Corrected.

The profiles are normalised using concurrent total column ozone from the Dobson spectrometer at Arosa [...]

p.6, l.4: You mentioned the spectral line for the ground-based MWR, so it would be nice to have the frequency of the line for Aura MLS as well.

We agree and adapted the manuscript.

It provides profiles of different trace gases in the atmosphere, with a vertical resolution of ~ 3 km. Stratospheric ozone is retrieved by using the spectral band centred at 240 GHz.

p.6, l.7: Improve language "stable with insignificant drifts", this is twice the same information. Either use "stable", or more precisely "stable within X% per decade", or just "because there are no drifts between".

Corrected.

We chose the Aura/MLS data for our study because there are no drifts between 20 and 40 km (Hubert et al., 2016).

p.6, l.9: Clarify what data is compared. E.g. "Comparison of time series". (Further on in the paper there is also a comparison of the trends.)

After moving this subsection to Sect. 3 (Time series comparison) according to your suggestion in major comment 3, we changed the title of this subsection to "Comparison methodology". Because this is a subsection of the section "Time series comparison", it is not necessary to include "time series" again.

p.6, l.17: Why is the GROMOS retrieval accurate between 31-0.8 hPa? The retrieval is not accurate simply because the measurement response is high. The instrument can still be biased (large systematic error) and even imprecise (large random uncertainty). The use of "accurate" suggests a small uncertainty. Is that really the case?

"Accurate" is indeed not the appropriate wording. We adapted the sentence.

In this study we concentrate on the altitude range between 31 and 0.8 hPa, where the a priori contribution to GROMOS profiles is low (see Sect. 2.1).

p.6, l.22-24: The native measurement coordinate system for lidar is number density versus altitude. You need to explain how the lidar data are converted as well. For ozonesonde this is trivial due to the temperature data from the attached radiosonde.

See the comment of Referee #1, p.6, l. 27.

p.6, l.26-29: Please clarify the details of the regridding procedure. The description is confusing and I am not sure I understand the order of operations and on what observed quantities.

We reformulated the sentence and hope that it is clearer now. To have similar grids for the different instruments, we interpolated the altitude-based data first to a regular altitude grid, and averaged then the provided pressure data to obtain a pressure grid with the same grid points.

The lidar and ozonesonde data were therefore linearly interpolated to a regular spaced altitude grid of 100 m for the ozonesonde at OHP and Payerne, and 300 m for the lidars and the ozonesonde at MOH. The mean profile of the interpolated pressure data then built the new pressure grid for the ozone data.

p.6, l.26-29: Also, is the interpolation a simple linear interpolation or is it a more advanced method such as described in Calisesi et al. (2005)?

We used a simple linear interpolation and added this in the manuscript.

The lidar and ozonesonde data were therefore linearly interpolated to [...]

p.7, l.3 and p.7, l.21: Remove "direct" from "direct comparison", since there are no indirect comparisons in this manuscript.

We changed the subsection title to "GROMOS comparison and anomalies".

p.7, l.25-27: The use of "corrected" relative difference is misleading. Why not e.g. "debiased" relative difference or "variations of the" relative difference? You simply remove the mean level to look at deviations of the relative difference w.r.t. its mean level. Please abstain from using "corrected relative difference" elsewhere in the text as well.

We changed it to "debiased relative difference" and added a formula to the text.

Based on the relative differences we identified periods where GROMOS differs from the other instruments. To identify these anomalies we used a debiased relative difference (RD_{debiased}), given by

$$RD_{\text{debiased},i,X} = RD_{i,X} - \overline{RD}_X, \quad (4)$$

where \overline{RD}_X is the mean relative difference of GROMOS to the data set X over the whole period.

p.7, l.25-27: Add formula to clarify that you remove a constant offset in RD. In other words, assuming RD_i is the monthly mean of month i and the mean of RD over all months is μRD , then you debiased the RD time series using $RD_i - \mu RD$.

We added a formula and reformulated the text.

For comparison with GROMOS we computed relative differences between the monthly mean values of the different data sets and the monthly mean of the coincident GROMOS profiles. The relative difference (RD) for a specific month i has been calculated by subtracting the monthly ozone value of the data set (X_i) from the corresponding GROMOS monthly mean (GR_i), using the GROMOS monthly mean as a reference:

$$RD_{i,X} = (GR_i - X_i) / GR_i \cdot 100. \quad (5)$$

p.8, l.19-20: Von Clarmann present an approach to consider "known", "identified" or "suspected" inhomogeneities in the trend analysis. Their method does not identify the time of an inhomogeneity, but co-fits the magnitude of the temporal offset. Second remark, state right here that you apply the von Clarmann method and code.

We added "known" and "suspected" to the sentence. We added a sentence to clarify that we use the method from von Clarmann et al. (2010) in our analyses.

Von Clarmann et al. present an approach to consider known or suspected inhomogeneities in the trend analysis. [...] We use the method and code from von Clarmann et al. in our trend analyses [...].

p.8, l.15-...: Now you start to use "inhomogeneities" rather than "anomalies". Is there a difference between these terms? Please define these clearly, if they are. If not, then state that as well.

"Inhomogeneities" is the general term that we use for the method by von Clarmann et al. (2010). As you mentioned in your comment below (p.8, l.21-22), this can include known causes such as changes in measurement systems or sampling irregularities, but also unknown instrumental issues. The unknown instrumental issues in our data lead to inhomogeneities in some months. We call these inhomogeneities in some months *anomalies*. We added the following sentence to clarify this.

We use the method and code from von Clarmann et al. in our trend analyses to account for inhomogeneities. The inhomogeneities are in our case anomalies in some months that we identified as described in Sect. 3.1.2.

p.8., l.15: Strengthen the link with previous Section 2.2. State asap that you will now use the times of the anomalies detected by the previously described method (Sect. 2.2) in the regression scheme (Sect. 2.3).

We adapted the manuscript and reformulated parts of the paragraph.

The inhomogeneities are in our case anomalies in some months that we identified as described in Sect. 3.1.2. [...]

To account for the anomalies in the time series when estimating the trends we adapted \mathbf{S}_y in two steps. First, we increased the uncertainties for months and altitudes where anomalies were identified (using the method described in Sect. 3.1.2).

p.8, l.15-16 and/or l.23-24: Clarify that the full error covariance matrix describes the correlation between measurements, not between pressure levels.

Corrected.

The total uncertainty of the data set is represented by a full error covariance matrix \mathbf{S}_y that describes for each pressure level covariances between the measurements in time.

p.8, l.21-22: Remove the entire phrase "Another reason for [...]". Instead add this possible cause (unknown origin) to the end of the phrase on l.16-18 "Inhomogeneous data can originate from [...] spatial or temporal sampling." (which currently lists known causes). The method you described in Sect 2.2 identifies periods of anomalies, but does not identify the cause of the anomaly. It can be anything in the list of l.16-18, but also unknown instrumental issues.

Corrected.

Inhomogeneous data can originate from changes in the measurement system (e.g. changes in calibration standards or merging of data sets with different instrumental modes), from irregularities in spatial or temporal sampling, or from unknown instrumental issues.

p.9, l.5: "no error correlation", but in what domain? Clarify that you deal with the correlation between measurements, not between pressure levels.

Corrected.

The off-diagonal elements of \mathbf{S}_{instr} are set to zero, assuming no error correlation between the measurements in time.

p.9, l.5: You introduce a second VCM (variance-covariance matrix). It would help to have a separate symbol for the different VCM's (S_1 , S_2 , S_{instr} , $S_{autocorr}$ or whatever) and use in-line formulas to explain when you add these and when not. This allows you to get away from unclear language such as "additional covariance matrix" and "enhanced covariance matrix". You could just refer to it as $S' = S_1 + S_2$. This also

We introduce as suggested different symbols for the different covariance matrices and include the corresponding equation in the manuscript.

The covariance matrix is for each instrument given by

$$S_y = S_{\text{instr}} + S_{\text{autocorr}} + S_{\text{bias}} \quad (6)$$

where S_{instr} gives the monthly uncertainty estimates for the instrument, S_{autocorr} accounts for residuals autocorrelated in time which are caused by atmospheric variation not captured by the trend model, and S_{bias} describes the bias uncertainties when a bias is considered.

p.9, l.12: Clarify "which means the time correlated residuals" are not considered (S2 or Sautocorr).

[...] which means that the time correlated residuals (S_{autocorr}) are not considered for the GROMOS trend.

p.9, l.15: Now you switch back to using "anomalies", not "inhomogeneities". Be consistent or mention that both terms have identical meaning.

See comment p.8, l.15-...

p.9, l.16: It is important to link with previous section. "[...] for months where anomalies were identified (using the method described in Sect. 2.2)".

First, we increased the uncertainties for months where anomalies were identified (using the method described in Sect. 3.1.2).

p.9, l.18: Are you using the so-called "corrected relative differences" or not?

Yes, we used the mean of all debiased relative differences, with the aim to obtain a representative value for each anomaly. We adapted the text to be clear.

For this purpose, we set the diagonal elements of S_{instr} for the respective month i to a value obtained from the mean difference to all instruments ($RD_{\text{debiased},i}$) and the mean GROMOS ozone value.

p.9, l.24: Replace "low" values by "small" values. E.g. -25% is a very low value, but not a small value when compared to -5%. I know the example is not applicable here, but I would generally suggest to refrain from using "low/high" and use "small/large" instead when referring to the magnitude of values (not the sign).

Corrected.

p.9, l.29: Clearly state that above procedure is only applied to GROMOS data. You did not consider outliers in any other time series than GROMOS.

We added the following sentence to the beginning of the paragraph. Further, we clarified this again in the beginning of the section about trend comparison (Sect. 4.4).

(Sect. 4.1:) The described procedure for anomaly consideration is only applied to the GROMOS time

series, whereas the other trend estimates were not corrected for anomalies.

4.4 Trend comparison

The GROMOS trend profile (corrected as described above) and trend profiles for all instruments at the other measurement stations (uncorrected) are shown in Fig. 10.

p.9, l.32-34: The percentage trend is obtained by dividing the ppmv trend by the mean level. If there is a trend the mean value will depend on the period over which the mean is computed. Did you use exactly same period for all trends? Since Aura/MLS has the shortest record, I would guess the ppmv trend should be scaled to 2005-2017 mean values. Has this been done? If not, please state whether this effect is negligible.

We did indeed not use the same time period to calculate the percentage trend for the different data sets. However, we tested the difference and found that the effect is negligible. For GROMOS, for example, the difference between the percentage trend calculated over the whole period and over the period 2005-2017 (period of Aura/MLS trends) is smaller than 0.05 throughout the whole profile, which corresponds to a difference between both trend profiles of $\sim 1\%$ in the middle stratosphere and 2% at the edge of the upper stratosphere. This is a very small differences for a trend of $\sim 2\%$, and we therefore decide to use the available time series period for each data set, as it was done so far.

p.10, l.1: You could remove ", as described by Tiao et al. (1990)", as this is just introductory-level Gaussian statistics.

Corrected.

p.11, l.1-2: You refer to "another temporal sampling", another than the other instruments? Another than the usual GROMOS sampling?

Shown in Fig. 2 is the complete GROMOS time series with its full sampling (30min to 1h), and not only coincident data with one of the other instruments. We adapted the sentence.

[...] whereas the GROMOS data shown here represent the complete GROMOS time series with its high temporal sampling.

p.11, l.9-10: Remove ambiguity. You deal with relative differences of monthly mean values of time coincident pairs, not of the monthly mean value of the relative difference of time coincident pairs. Figure 3 suggests the latter however.

We adapted the figure captions (Fig. 3 and Fig. 4) to "Relative differences of monthly means" and refer to the new Eq. (2) in the text.

Shown are the monthly relative differences of time coincident pairs of GROMOS (GR) and the convolved data set X, with GROMOS data as reference value (Eq. (2)).

p.11, l.15-17: The large outliers in the OHP comparison time series (Fig. 3) are all negative. The explanation is not satisfying, if not incorrect. If outliers are due to the small sample, then large positive peaks should be observed as well. Unless there is a systematic component in both the sampling pattern and the ozone fields. The latter seems highly unlikely to me. Can't it, instead, be due to the presence of large positive outliers in the OHP ozonesonde record?

The small sample is indeed not the cause for the differences, but it makes them more visible when comparing monthly means. Having for example one ozonesonde profile with large differences to Gromos within a month and several profiles with very good agreement, the monthly mean would not be affected much by the one profile with bad agreement. However, when we have in some cases only one or two profiles per month, the monthly mean is highly affected in case that the difference to GROMOS is large for these profiles. We reformulated the text to clarify this aspect. See also the comment of Referee #1, p. 11, l. 15-17.

We further observe that the relative difference between GROMOS and the OHP ozonesonde shows some important peaks in the last decade, indicating that the ozonesonde often measures more ozone than GROMOS. The ozonesonde data seems to have some outlier profiles. When comparing the monthly means of coincident pairs, these outliers are even more visible, because only a small number of OHP ozonesonde profiles are available per month (only four profiles on average).

p.11, l.9: Does Fig. 3 show the "corrected relative difference"? I guess not, since the mean level seems to differ from zero. This should be clarified.

Fig. 3 shows indeed the "normal" relative difference (RD). The debiased relative difference (RD_{debiased}) was only used to detect the anomalies, but is not shown in the figures. This makes it possible to see the "true" differences between the instruments in the figures. The debiased relative difference is just a correction of RD to make the anomaly detection easier, by removing a potential constant offset. This has been clarified by using the acronyms RD and RD_{debiased} in the figure captions of Fig. 3 and Fig. 4.

p.11, l.21: Doesn't Fig. 4 show the "corrected relative difference" (or debiased relative difference, see earlier comment above)? Mention this clearly in the text, and abstain from using "corrected relative difference", I like "debiased relative difference" better.

See comment above. To clarify we modified the caption of Fig. 4.

The black areas in the lowest panel show periods where at least three data sets (or two data sets above 2 hPa) have a debiased relative difference (RD_{debiased}) larger than 10%.

p.11, l.34: Replace by "[...] comparing GROMOS to SOMORA [...]". There is only one reference instrument in the US during the period before Aura/MLS.

Corrected.

[...] were not confirmed by comparing GROMOS to SOMORA, and might thus be real ozone variations.

p.12, l.1: Add "[...] at these altitudes, especially prior to the start of the MLS measurements in 2004." And you could then e.g. also mention the step change around 2005 visible in panel (a) around 2005 which may be related to the change of SOMORA front end (see also comment below; p.28, Fig.3).

We added the proposed sentence. Concerning the step change, see comment below (p. 28, Fig. 3).

[...] which makes the anomaly detection less robust at these altitudes, especially prior to the start of Aura/MLS measurements in 2004.

p.12, l.7-8: It is not entirely clear whether you are saying here that you disprove the claim by Steinbrecht (2009)? If yes, then end the phrase by removing the ambiguity. For instance: "[...] due to instrumental issues of GROMOS and not due to sampling."

We adapted the text as proposed.

[...] which are probably due to instrumental issues of GROMOS and not only due to sampling differences.

p.12, l.7-8: Are there any leads on what could be the cause of the instrumental issues? If yes, it may be of interest to the reader.

We checked several parameters such as instrumental problems, noise temperature, or atmospheric opacity but could not find any clear explanation for the anomalies so far.

p.12, l.21: A negative amplitude has no meaning. Is this a typo?

The amplitude should indeed be positive, it was adapted so that g and h are negative, in order to obtain the desired phase of the seasonal cycle (see also comment of Referee #1, p.12, l. 21-22).

p.12, l.26: Vague language "This might be due to some error propagation". This clause does not hold additional information and may as well be removed.

We removed the sentence as proposed.

p.13, l.4: You arbitrarily chose 5%, which is $3.5 = 5/(0.1/7)$ times larger than $1.4\% = 0.1/7$. What if you increase to even higher uncertainties, e.g. in the limit of infinite uncertainty? This case would be equivalent to removing the anomaly from the time series prior to regression. (See also my major comment)

Indeed, increasing the uncertainty would improve the trend estimate (as mentioned in the manuscript), especially in the perfect, noise-free artificial time series. However, we decided to use a realistic value for this uncertainty, since the trend is only affected marginally when using an even higher uncertainty (see your comment below, p.13, l.34). See also the comment of Referee #1 (p.13, line 12-13) concerning the choice of the uncertainty and your major comment concerning the option to remove the anomalies.

p.13, l.6: Text is inconsistent with caption of Fig. 6. Shouldn't it be "[...] once without (case D) and once with (case E) the increased [...]"?

You are right, we corrected the text.

p.13, l.7: Add to the last phrase that a 5 ppm correlated block essentially leads to a free fit of the magnitude bias.

Adding the correlated block to \mathbf{S}_y corresponds to an unknown bias of the data subset that is affected by anomalies, and leads to a free fit of the bias magnitude. This bias is represented in \mathbf{S}_y as a fully correlated block of (5 ppm)². It has an expectation value of zero and an uncertainty of 5 ppm.

p.13, l.27: Repeat by how much you increased the uncertainty, so a reader does not have to go back all the way to Sect 2.3.

[...] we increased the uncertainties (diagonal elements of \mathbf{S}_y) for the months and altitudes that were detected as anomalies. The uncertainties for these anomalous data have been set to a value obtained from the difference to the other datasets ($RD_{debiased}$), as described in Sect. 4.1.

p.13, l.34: "[...] decreases the trend slightly in the lower stratosphere, but the differences are small." Doesn't this point to a too small scaling of the uncertainties?

The uncertainties for the anomalies has been set to a realistic value based on the relative difference detected by comparing GROMOS data to data from the other instruments. Setting them to higher values affects the trend only slightly. Therefore we propose the more advanced approach, where a bias is fitted by considering a correlated block in the covariance matrix.

p.14, l.4: I still don't like the phrasing "corrected" trend profile. Corrected w.r.t. what? See also my comment p.3, l.5.

See the answer to your comment for p.3, l.5.

p.14, l.4 and l.30-31: "The corrected trend profile agrees well with recent satellite-based ozone trends [...]". I am not convinced that the case III results agree well with other studies. They are consistent with Steinbrecht (2017; Fig.5, Table 6) between 10-2 hPa (what you call "middle" stratosphere). However, there is tension between the results at 30-10 hPa ("lower" stratosphere). The other studies find smaller trend values, more closely to 0-1% per decade.

Adapted as followed.

[...] The corrected trend profile is consistent with recent satellite-based ozone trends (e.g. Steinbrecht et al., 2017) in the middle stratosphere.

Compared to other studies, the GROMOS trends in the lower stratosphere are slightly larger than trends of most merged satellite data sets.

[...] This is consistent with satellite-based trends from recent studies in the middle stratosphere, but is still larger than most satellite trends in the lower stratosphere.

p.14, l.13-14: Somewhat confusing phrasing "a slight shift in the peak height". A shift with respect to what? Is the peak shifting in time, which is what I first think of in the context of time series analysis? More likely you mean that the peak sensitivity of the GROMOS retrievals is not at the true altitude?

Text adapted, see also comment of Referee #1 (p.14 line 11-15).

(Sect. 4.3:) We observe that after the instrument upgrade in 2009, the averaging kernels peak at higher altitudes than expected, [...].

(Sect. 4.4:) This difference might be related to the averaging kernels of GROMOS, which indicate that GROMOS retrieves information from higher altitudes than expected (~ 2 km difference).

p.14, l.30-31: See earlier comment, I would refrain from using "agree". The results are consistent, but do not agree well especially below the 10 hPa level. The bias corrected trend does lead to a better agreement

In summary, correcting the GROMOS trend with our anomaly approach leads to a trend profile of ~ 2 to 2.5 % per decade in the lower and middle stratosphere. This is consistent with satellite-based trends from recent studies in the middle stratosphere, but is still larger than most satellite trends in the lower stratosphere.

p.15, l.16: Unlike Reviewer 1, I like the fact that the full altitude range is shown. This section is not about a comparison to actual ozonesonde or lidar trends, but about the impact of sampling on GROMOS trends. I would keep the altitude range in Fig. 8 as is.

Thank you for this statement, we agree and keep it as it is (and expanded it to other instruments, see next comment).

p.15, l.17-19: This is a very nice piece of information. Out of curiosity, have you tested to randomly sub-sample the GROMOS time series to MOH and NPay, then regress? Such an analysis would add information about the random nature of the impact of sampling. The more positive sonde trend at 3 hPa may as well turn into a less positive trend if another temporal GROMOS sampling was picked but with same sample size as the ozonesonde time series. Would be nice, but not necessary.

Thank you for this input. To check how the trends look like with other sampling rates, we added also trends with sampling rates of the other lidar (OHP) and ozonesondes (MOH and OHP). The larger trends at 40 km when using only data at ozonesonde times is confirmed by the additional sampling rates of the other two ozonesondes. We adapted the figure in the manuscript and the text as followed.

(Sect. 4.3.1:) The GROMOS time series is used for these trend estimates, once using only measurements at the time of lidar measurements, and once only at the time of ozonesonde launches. The differences to the trend that uses the complete GROMOS sampling are not significant, but still important, especially between 35 and 40 km and in the lower stratosphere. Using the sampling of the MOH lidar leads to larger trends (~ 3 % per decade) than using the normal sampling (~ 2 % per decade) below 5 hPa. The OHP lidar

sampling, however, leads to smaller trends than the normal sampling above 10 hPa. This suggests that selecting different night measurements within a month can lead to trend differences.

All three ozonesonde samplings result in a larger trend than normal or lidar sampling above 5 hPa. [...]

(Sect. 4.4:) [...] distance of ~ 600 km between the MOH and OHP stations might also explain some differences between the lidar trends. Furthermore, our sampling results show that the lidar sampling at MOH leads to a larger trend in the lower stratosphere than using a continuous sampling, whereas OHP lidar sampling leads to a lower trend in the middle stratosphere. The large lidar trend at MOH and the comparable low OHP lidar trend might therefore also be partly explained by the sampling rate of the lidars.

p.15, l.26: Replace "estimations" by "regression".

Corrected.

p.15, l.26-27: Did you average the GROMOS trends or the GROMOS time series? If the former, how did you do that and does the trend of vertically averaged data agree with the vertical average of the trends?

We averaged the GROMOS trends over the three different pressure ranges. We think that this should be quasi the same as the trend of vertically averaged data.

p.16, l.5-6: Vague language "The trend dependency [...] is controversial to [...]". Please rephrase, I do not understand what you mean.

The trend magnitudes depend on the starting year of the regression, which is controversial to the definition of a linear trend that does not change with time.

p.16, l.6-7: "This suggests that the true trend might not be linear, or that some interannual variations or anomalies are not captured by the trend model" This is known for a long time. Please rephrase to "This illustrates that [...]".

Corrected.

p.17, l.2-3: "A thorough harmonization would be necessary to correct the trend for this change." In fact, your manuscript describes a method that does not require harmonisation. I understand the extra work effort, but it is a pity for Section 4.5 that you did not treat outlying data in the non-GROMOS regression analyses. This would have made an even more interesting comparison.

We agree that it would be a good idea to improve the trends also for the other data sets. Nevertheless, it was more urgent to consider anomalies in the GROMOS data because there are some clear anomalies compared to the other data. See also major comment 4.

p.17, l.4: It is unclear whether you have used these newer SOMORA retrievals or not. Please clarify in the revised manuscript.

The SOMORA trend shows a positive peak at 30 km, which is probably due to homogenization problems that are corrected in the new retrieval version of SOMORA, which is, however, not yet used in our analyses (Maillard Barras et al., 2019).

p.26, Fig.1: "[...] indicate months without data." Missing data does not imply there are no data, and you mean there are no (screened) data during these months.

The white lines indicate months where no measurements were available due to instrumental issues.

p.26, Fig.1: Mention the period over which the climatology is computed. So, "Deviation from GROMOS monthly mean 1997-2017 climatology" would make a better description.

Corrected.

Deviation from GROMOS monthly mean climatology (1997 to 2017), smoothed by a moving median window of 3 months.

p.28, Fig.3: "Monthly means of relative differences" implies you compute relative difference $100 \times (GR-X)/GR$ for each coincident pair, then take the monthly average. This is not how you described it in Sect. 2.2.

We adapted this wrong formulation, see comment above (comment p.11, l.9-10).

p.28, Fig.3: Do these time series represent the "relative difference" or the "corrected relative difference"? If these are corrected relative differences, then how can it be that most curves seem to have a negative mean value where zero would be expected?

They represent the "uncorrected" relative difference, see comment above (comment p.11, l.9).

p.28, Fig.3: Panel (a) shows a step around 2004-2006. Is this due to the change of front end in SOMORA in 2005 (p.4, l.21)? If yes, then it would be nice to at least mention this in the last paragraph of p.11.

This step change might indeed be related to the front-end change of SOMORA in 2005. We added the following sentence to the manuscript.

A step change between SOMORA and GROMOS is visible in 2005, which might be related to the SOMORA front-end change in 2005.

p.29, Fig.4: Do all panels represent the "relative difference" or the "corrected relative difference"? Clarify in the caption in the latter case.

All panels show the relative differences – and not the debiased (corrected) relative differences – to show the true difference to GROMOS. The debiased relative difference has only been used to detect the anomalies (see also the answer to comment p.11, l.9).

p.34, Fig.9: See my earlier comment. Did you average the trend, or average the time series then derive the trend? In the first case, how exactly did you compute the error on averaged trend?

We averaged the trends and the uncertainties after estimating them from the time series (see also comment p.15, l.26-27). Since the different trend profile points that are averaged for a specific pressure range are not independent, we decided to use the arithmetic mean of the uncertainties, which leads to slightly larger uncertainties than using a standard error (standard deviation divided by the square root of the sample size).

p.34, Fig.9: Mention significance level of the error bars. I assume that these represent 1σ since some bars that do not cross the zero level are greyed out.

We adapted the figure to show the 2σ uncertainty and mentioned it in the figure caption.

p.37, Table.2: Add time unit in first column. It should be "ppm/decade".

Corrected.

p.37, Table.2: Move third column ("Monthly uncert.") to first column in section "Parameters in the trend programme". No uncertainty has been added in the time series.

Corrected.

p.37, Table.2: Use consistent number of significant digits. I suggest three digits for all numbers quoted.

We agree and adapted the number of digits to be consistent for the estimated trends in Table 2. For the other numbers, we kept it as it is because we do not think that it is necessary to add zeros to all of them.

2.4 Technical corrections

We adapted the manuscript to the suggested technical corrections and show the changes in the following.

p.1, l.4-5: Replace by "[...] anomalies that may all mask a [...]".

Trends and their uncertainties are influenced by factors such as instrumental drifts, sampling patterns, discontinuities, biases, or short-term anomalies that may all mask a potential ozone recovery.

p.1, l.7: Remove "[...] to improve the derived trend profiles [...]". This is evident.

We present an approach for handling suspicious anomalies in trend estimations.

p.1, l.11: Replace by "[...] in agreement with satellite [...]".

[...] providing a confirmation of ozone recovery in the upper stratosphere in agreement with satellite observations.

p.1, l.17: What is a "serious" decrease? Perhaps "fast" or "large" is more appropriate language?

After the large stratospheric ozone decrease due to ozone depleting substances (ODS) [...]

p.1, l.20: Replace by "[...] Antarctic ozone concentrations started [...]".

Corrected, also in accordance with suggestion of Referee #1.

As a consequence, stratospheric ozone concentrations over Antarctica started to increase again [...]

p.2, l.1: Replace by "[...] increasing ozone levels are more difficult [...]".

Changed according to the comment of Ref. #1.

Outside of the polar regions, however, differences in ozone recovery are observed depending on altitude and latitude.

p.2, l.1: Replace by "[...] to detect and these depend on altitude [...]".

See previous comment.

p.2, l.7: Add "s" to "[...] greenhouse gas increases [...]".

Adapted to "the increase of greenhouse gases".

p.2, l.19-20: Replace by "[...] with no or small drifts [...]".

[...] Sofieva et al. (2017) used only stable satellite products with no or small drifts.

p.2, l.25: Remove "data" from "[...] data steps [...]".

Corrected.

p.2, l.27: Move the "yet" to "[...] has not yet been [...]".

[...] but it has not yet been applied to ground-based data.

p.3, l.21: Replace by "[...] They measure the 142 GHz line where ozone molecules [...]". Currently you have "[...] atmospheric emission [...] molecules emit [...]".

They measure the 142 GHz line where ozone molecules emit microwave radiation due to rotational transitions.

p.3, l.29: Replace by "[...] is the main focus [...]".

The **G**round-based **M**illimetre-wave **O**zone **S**pectrometer (GROMOS) located in Bern, Switzerland, is the main focus of this study (Kämpfer, 1995; Peter, 1997).

p.4, l.10: Replace "instead" by "rather than from".

Corrected.

p.4, l.12: "data corrections".

Corrected.

p.4, l.13: "troposphere" is possibly more precise terminology here, rather than "atmosphere"? This then also links better with next phrase which contains "tropospheric humidity".

Because the stratospheric signal is weak in an opaque and humid troposphere, [...]

p.4, l.16: "[...] 3-day moving [...]". Drop "s".

[...] 3-day moving median window [...]

p.4, l.18: Replace by "[...] outliers exceeding four times [...]"

[...] where the instrumental system temperature showed outliers exceeding four times the standard deviation within a 30-day moving median window.

p.4, l.18: "[...] 30-day moving [...]". Drop "s".

[...] 30-day moving median window [...]

p.4, l.34: Move the last phrase "The lidar can only [...] the influence of sunlight." to the first paragraph of Sect. 2.1.2. This to avoid confusion that this information holds for all lidars, not just the one installed at Hohenpeissenberg which is the subject of this paragraph.

Corrected.

p.5, l.9: Remove newline, and ensure the phrase "More detailed information [...]" is next to the one ending in "[...] with the GROMOS limits." Since the phrase only contains information for OHP.

Corrected.

p.5, l.28: Replace by "[...] 30 km, above which the balloon [...]". The balloon does not usually burst at "30 km", but at "30 km or above".

Ozonesonde data are limited to altitudes up to ~ 30 km, above which the balloon usually bursts.

p.6, l.2: Replace by "[...] Aura satellite, launched in mid 2004, [...]"

The microwave limb sounder (MLS) on the Aura satellite, launched in mid 2004 [...]

p.7, l.4: Replace by "The vertical resolution of GROMOS and SOMORA are usually coarser than for the other instruments. [...]"

The vertical resolution of GROMOS and SOMORA is usually coarser than for the other instruments. When comparing profiles directly with GROMOS profiles, the different vertical resolution of the instruments has to be considered.

p.7, l.17: Replace by "[...] GROMOS and SOMORA have a higher temporal resolution than the other instruments. [...]"

Corrected.

p.8, l.18: "temporally correlated".

Such inhomogeneities lead to groups of temporally correlated data [...]

p.8, l.22: Remove "that lead to anomalies". Otherwise "[...] reason for inhomogeneities might be [...] issues that lead to anomalies [...]".

Several sentences have been rephrased.

Inhomogeneous data can originate from changes in the measurement system (e.g. changes in calibration standards or merging of data sets with different instrumental modes), from irregularities in spatial or temporal sampling, or from unknown instrumental issues. [...] We use their method and code in our trend analyses to account for inhomogeneities. The inhomogeneities are in our case anomalies in some months that we identified as described in Sect. 3.1.2.

p.9, l.7: Replace "adjusted" by "scaled", because that is what you do.

Corrected.

p.9, l.8: Replace by "[...] the trend fit becomes unity."

Corrected.

p.9, l.20: Some irrelevant information. Replace by "[...] In a second step, we account for biases [...]".

In a second step, we account for biases in the data subsets where anomalies were detected.

p.9, l.27: Drop "have" from "we have found".

Corrected.

p.9, l.31: Remove "and the Aura/MLS trend starts in Jan. 2005", as this is duplicated in the next phrase.

Corrected.

p.9, l.32: Replace by "Aura/MLS cover the shortest trend period, starting in [...]".

Corrected.

Aura/MLS covers the shortest trend period, starting only in Jan. 2005.

p.9, l.33: Add "regression" to "regression model output", to avoid any confusion with e.g. CCM's.

Corrected.

p.10, l.28: Replace altitude "classes" by "ranges".

Corrected.

p.11, l.1: "[...] whereas the GROMOS data shown here [...]".

Corrected.

p.12, l.20: Replace by "[...] per decade, i.e. $b = 0.1$ [...]".

Corrected.

p.12, l.25-26: Replace by "The residuals are of order 10^{-6} and increase towards the start and end of the time series (Fig. 5(b))."

Corrected.

p.13, l.12-13: Rephrase, it is hard to read. "We further found that the trend estimate is closer to its true value when higher uncertainties are chosen [...]".

Corrected.

p.14, l.11: Replace by "[...] even though the GROMOS [...]".

Corrected.

p.16, l.22: How do lidar data "derive" from satellite data? Did you mean "deviate", perhaps?

Yes, it was a mistake, it was corrected to "deviate".

p.17, l.29: Replace by "[...] in the trend analysis [...]".

Corrected.

p.26, Fig. 1: Not "stripes" but "lines".

Corrected.

p.32, Fig.7: Remove the space before "case II" in the parenthesis on the second line of the caption.

Corrected.

3 Bibliography

- Bai, K., Chang, N.-B., Shi, R., Yu, H., and Gao, W. (2017). An inter-comparison of multi-decadal observational and reanalysis data sets for global total ozone trends and variability analysis. *J. Geophys. Res. Atmos.*, 122(13):1–21.
- Connor, B. J., Parrish, A., and Tsou, J.-J. (1991). Detection of stratospheric ozone trends by ground-based microwave observations. In James L. McElroy and McNeal, R. J., editors, *Proc. SPIE 1491, Remote Sens. Atmos. Chem.*, volume 1491, pages 218–230, Orlando, FL, United States. International Society for Optics and Photonics.
- Harris, N. R. P., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C. D., Bourassa, A., Davis, S. M., Degenstein, D., Delcloo, A., Frith, S. M., Froidevaux, L., Godin-Beekmann, S., Jones, N., Kurylo, M. J., Kyrölä, E., Laine, M., Leblanc, S. T., Lambert, J.-C., Liley, B., Mahieu, E., Maycock, A., de Mazière, M., Parrish, A. D., Querel, R., Rosenlof, K. H., Roth, C., Sioris, C., Staehelin, J., Stolarski, R. S., Stübi, R., Tamminen, J., Vigouroux, C., Walker, K. A., Wang, H. J., Wild, J., and Zawodny, J. M. (2015). Past changes in the vertical distribution of ozone - Part 3: Analysis and interpretation of trends. *Atmos. Chem. Phys.*, 15(17):9965–9982.
- Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J. L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., Van Gijssel, J. A. E., Van Malderen, R., Von Der Gathen, P., Walker, K. A., Wolfram, E., and Zawodny, J. M. (2016). Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records. *Atmos. Meas. Tech.*, 9(6):2497–2534.
- Kämpfer, N. (1995). Microwave remote sensing of the atmosphere in Switzerland. *Opt. Eng.*, 34:2413–2424.
- Keppens, A., Lambert, J. C., Granville, J., Miles, G., Siddans, R., Van Peet, J. C., Van Der A, R. J., Hubert, D., Verhoelst, T., Delcloo, A., Godin-Beekmann, S., Kivi, R., Stübi, R., and Zehner, C. (2015). Round-robin evaluation of nadir ozone profile retrievals: Methodology and application to MetOp-A GOME-2. *Atmos. Meas. Tech.*, 8(5):2093–2120.
- Maillard Barras, E., Haefele, A., Stübi, R., and Ruffieux, D. (2015). A method to derive the Site Atmospheric State Best Estimate (SASBE) of ozone profiles from radiosonde and passive microwave data. *Atmos. Meas. Tech. Discuss.*, pages 3399–3422.
- Maillard Barras, E., Ruffieux, D., and Hocke, K. (2009). Stratospheric ozone profiles over Switzerland measured by SOMORA, ozonesonde and MLS/AURA satellite. *Int. J. Remote Sens.*, 30(15-16):4033–4041.
- Maillard Barras et al. (2019). Trends of stratospheric ozone daily cycle, in preparation.
- Moreira, L., Hocke, K., Eckert, E., Von Clarmann, T., and Kämpfer, N. (2015). Trend analysis of the 20-year time series of stratospheric ozone profiles observed by the GROMOS microwave radiometer at Bern. *Atmos. Chem. Phys.*, 15(19):10999–11009.
- NDACC (2018). Network for the Detection of Atmospheric Composition Change.
- Peter, R. (1997). *The Ground-based Millimeter-wave Ozone Spectrometer - GROMOS*. Research report no. 97-13, Institute of Applied Physics, University of Bern, Switzerland.
- Sofieva, V. F., Kyrölä, E., Laine, M., Tamminen, J., Degenstein, D., Bourassa, A., Roth, C., Zawada, D., Weber, M., Rozanov, A., Rahpoe, N., Stiller, G. P., Laeng, A., von Clarmann, T., Walker, K. A., Sheese, P., Hubert, D., van Roozendaal, M., Zehner, C., Damadeo, R. P., Zawodny, J., Kramarova, N., and Bhartia, P. K. (2017). Merged SAGE II, Ozone_cci and OMPS ozone profiles dataset and evaluation of ozone trends in the stratosphere. *Atmos. Chem. Phys.*, 17:12533–12552.
- Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., Frith, S. M., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke, K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., García, O., Jones, N., Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D., and Tummon, F. (2017). An update on ozone profile trends for the period 2000 to 2016. *Atmos. Chem. Phys.*, 17:10675–10690.
- von Clarmann, T., Stiller, G. P., Grabowski, U., Eckert, E., and Orphal, J. (2010). Technical note: Trend estimation from irregularly sampled, correlated data. *Atmos. Chem. Phys.*, 10(14):6737–6747.

~~Ozone trend~~ Ground-based ozone profiles in the stratosphere: ~~combining ground-based data over Central Europe to consider~~ ~~uncertainties: incorporate anomalous observations in the analysis of~~ stratospheric ozone trends

Leonie Bernet^{1,2}, Thomas von Clarmann³, Sophie Godin-Beekmann⁴, Gérard Ancellet⁴, Eliane Maillard Barras⁵, René Stübi⁵, Wolfgang Steinbrecht⁶, Niklaus Kämpfer^{1,2}, and Klemens Hocke^{1,2}

¹Institute of Applied Physics, University of Bern, Bern, Switzerland

²Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland

³Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research, Karlsruhe, Germany

⁴Centre National de la Recherche Scientifique, Université de Versailles Saint-Quentin-en-Yvelines, Guyancourt, France

⁵MeteoSwiss, Payerne, Switzerland

⁶Deutscher Wetterdienst, Hohenpeissenberg, Germany

Correspondence: Leonie Bernet (leonie.bernet@iap.unibe.ch)

Abstract. Observing stratospheric ozone is essential to assess if the Montreal Protocol has succeeded to save the ozone layer by banning ozone depleting substances. Recent studies have reported positive trends indicating that ozone is recovering in the upper stratosphere at mid-latitudes, but the trend magnitudes differ and uncertainties are still high. Trends and their uncertainties are influenced by factors such as instrumental drifts, sampling patterns, discontinuities, biases, or short-term anomalies that ~~all might may all~~ mask a potential ozone recovery. The present study investigates how anomalies, temporal measurement sampling rates and trend period lengths influence resulting trends. We present an approach for handling suspicious anomalies in trend estimations ~~to improve the derived trend profiles. The approach was applied to data from a~~. For this, we analysed multiple ground-based stratospheric ozone records in Central Europe to identify anomalous periods in data from the Ground-based Millimetre-wave Ozone Spectrometer (GROMOS) located in Bern, Switzerland. The detected anomalies were then used
10 to estimate ozone trends from the GROMOS time series by considering the anomalous observations in the regression. We compare our improved GROMOS trend estimate with results ~~from other ground stations derived from the other ground-based ozone records~~ (lidars, ozonesondes, and microwave radiometers) ~~in Central Europe~~, that are all part of the Network for the Detection of Atmospheric Composition Change (NDACC). The data indicate positive trends of 1 to 3 % per decade at an altitude of about ~~40~~39 km (3 hPa), providing a confirmation of ozone recovery in the upper stratosphere in agreement ~~to~~ with satellite observations.
15 At lower altitudes, the ground station data show inconsistent trend results, which emphasize the importance of ongoing research on ~~lower stratospheric ozone trends~~ ozone trends in the lower stratosphere. Our presented method of a combined analysis of ground station data provides a useful approach to recognize and to reduce uncertainties in stratospheric ozone trends by considering anomalies in the trend estimation. We conclude that stratospheric trend estimations still need improvement and that our approach provides a tool that can also be useful for other data sets.

1 Introduction

After the ~~serious-large~~ stratospheric ozone decrease due to ozone depleting substances (ODS) [Molina and Rowland, 1974; Chubachi, 1984; Farman et al., 1985], signs of an ozone recovery have been reported in recent years [e.g. ~~WMO, 2014~~WMO, 2018; SPARC/IO3C/GAW, 2019]. Implementing the Montreal Protocol (1987) ~~have-has~~ succeeded to reduce ODS emissions so that the total chlorine concentration has been decreasing since 1997 [Jones et al., 2011]. As a consequence, ~~Antarctic-ozone~~ stratospheric ozone concentrations over Antarctica started to increase again, as shown by recent studies [Solomon et al., 2016; Kuttippurath and Nair, 2017; Pazmiñ, 2018; Strahan and Douglass, 2018]. Outside of the polar regions, however, ~~increasing ozone is more difficult to detect and depends~~ differences in ozone recovery are observed depending on altitude and latitude. It is still controversial if ozone is recovering in the lower stratosphere [Ball et al., 2018; Chipperfield et al., 2018; Stone et al., 2018; Wargan et al., 2018], whereas broad consensus exists that stratospheric ozone has stopped to decline in the upper stratosphere since the end of the 1990s [Newchurch et al., 2003; Reinsel et al., 2005; Steinbrecht et al., 2006; Stolarski and Frith, 2006; Zanis et al., 2006; Steinbrecht et al., 2009a; Shepherd et al., 2014; ~~WMO, 2014~~WMO, 2014, 2018; SPARC/IO3C/GAW, 2019]. Recently estimated trends for upper stratospheric ozone are positive, but they are still different in magnitude and significance, because detecting a small trend is a difficult task. ~~A trend might be masked by natural variability and many~~ Many factors influence stratospheric ozone such as ~~greenhouse gas increase and~~ variations in atmospheric dynamics, solar irradiance or volcanic aerosols and the increase of greenhouse gases [Pawson et al., 2014~~WMO, 2014~~]. Further, ozone trends might be masked by natural variability.

Other important sources for trend uncertainties are instrumental drifts, abrupt changes, biases, or sampling issues e.g. due to ~~different instrumental differences in sampling patterns or in~~ vertical or temporal resolution ~~or different sampling patterns~~ [Pawson et al., 2014; Damadeo et al., 2018]. Satellite drifts have been included in trend uncertainties in several studies [e.g. Stolarski and Frith, 2006; Frith et al., 2017]. Possible statistical methods to consider abrupt changes in a time series are for example presented by Bates et al. [2012]. ~~Bai et al. [2017] found that biases~~ Biases in ozone data sets can lead to important differences in trend estimates, especially when they are occurring in the beginning or the end of the considered trend period [e.g. Bai et al., 2017]. The influence of non-uniform sampling patterns on trends was illustrated by Millá [2016]. Also Damadeo et al. [2018] showed that accounting for temporal and spatial sampling biases and diurnal variability changes satellite based trends.

To account for several of the mentioned factors that influence trend estimates, different approaches were published following the Scientific assessment of Ozone Depletion of the World Meteorological Organization (WMO) in 2014 [WMO, 2014], with the aim to reduce uncertainties in trend estimates. Drifts in single satellite data sets were, for example, considered in the studies by Eckert et al. [2014] or Bourassa et al. [2018], whereas Sofieva and Kyrö [2017] used only stable satellite products with no or small drifts. The study of Steinbrecht et al. [2017] summarizes recent trend estimates using only updated satellite data sets ~~that are drift-corrected, with small drifts.~~ The drifts were mainly identified by Hubert et al. [2016] and ~~not yet were not~~ considered in the trend estimates by Harris et al. [2015] or WMO [2014]. Steinbrecht et al. also used data from a large range of ground stations, but possible biases or anomalies in these ground-based data were not considered. The resulting ground-based trends

consequently show some important differences, and were not used in their final merged trend profile. *Ball et al.* [2017] used an advanced trend estimation method that considers ~~data~~ steps in satellite time series or biases due to measurement artefacts. Their Bayesian method uses a priori information about the different satellite data sets and results in an optimal merged ozone composite, but it has not ~~been applied yet yet been applied~~ to ground-based datasets.

5 ~~All of these studies using corrected satellite data~~ The above presented studies agree on positive ozone trends in the upper stratosphere with some differences in magnitude, and show varying trends in the middle and lower stratosphere. This agreement is more difficult to observe in ground-based data sets, where the data variability is larger due to strong regional variability [*Steinbrecht et al.*, 2017; ~~*Pawson et al.*, 2014~~ WMO, 2014]. Because of this larger variability, considering instrumental biases or regional anomalies is of special importance for trend estimations derived from ground-based data. In addition to *Steinbrecht et al.* [2017] and *WMO* [2014], several other studies presented ground-based trends of stratospheric ozone profiles [e.g. *Steinbrecht et al.*, 2009a; *Nair et al.*, 2013, 2015; *Harris et al.*, 2015; SPARC/IO3C/GAW, 2019], but biases in the data sets that might influence the resulting trends have not been considered yet.

The present study proposes an approach to handle the problem of ~~anomalies in trend data sets by considering anomalies~~ anomalous observations in time series by considering the anomalies when estimating trends ~~from ground-based data sets~~. For this purpose, we present the updated data set of the ground-based microwave radiometer GROMOS (**G**round-based **M**illimetre-wave **O**zone **S**pectrometer) located in Bern, Switzerland. We determine its trends with a multilinear parametric trend model [*von Clarmann et al.*, 2010] by considering anomalies and uncertainties in the time series, resulting in ~~a corrected~~ an improved trend estimate. To identify ~~biased periods~~ such anomalies in the GROMOS data set, we compare the GROMOS data with other ground-based data sets (lidars, ozonesondes, and microwave radiometers) in Central Europe (Sect. 3). We define anomalies as periods where the data deviates from the other data sets. Before applying our trend approach to the GROMOS time series (Sect. 4.3), we tested it with an artificial time series (Sect. 4.2). Not only anomalies in a time series influence resulting trends, but also sampling patterns and the choice of the trend period. We therefore present a short analysis of temporal sampling rate and trend period length based on the GROMOS data set (Sect. 4.3.1 and 4.3.1). Finally, we ~~compared the corrected~~ compare the improved GROMOS trend with the trends from the other data sets used (Sect. 4.4).

25 2 ~~Data and methods~~ Ozone data sets

~~Our study presents how to consider anomalies in a time series to improve resulting trend estimates. We apply this approach to data from the GROMOS microwave radiometer and compare the time series and the trend with different ground-based ozone data sets.~~

2.1 **Ozone data sets**

30 The stratospheric ozone profile data used in the present study comes from different ground-based instruments that measure in Central Europe (~~France, Germany, and Switzerland~~ Table 1). They are all part of the Network for the Detection of Atmospheric Composition Change [*NDACC, 2018* NDACC, 2019]. In addition, we used data from the Microwave Limb Sounder (MLS) on

board the Aura satellite (Aura/MLS). ~~We consider data from Jan. 1995 to Dec. 2017, except for some instruments that cover a shorter time period (Table 1).~~ All data from the different stations are compared to data from the GROMOS radiometer located in Bern, Switzerland (46.95° N, 7.44° E, 574 m above sea level (a.s.l.)). The aerological station (MeteoSwiss) in Payerne (Switzerland, 46.8° N, 7.0° E, 491 m a.s.l.) is located 40 km south-west of Bern, which ensures comparable stratospheric measurements. The Meteorological Observatory Hohenpeissenberg (MOH, Germany, 47.8° N, 11.0° E, 980 m a.s.l.) is located 290 km north-east of Bern, and the Observatory of Haute Provence (OHP, France, 43.9° N, 5.7° E, 650 m a.s.l.) lies 360 km south-west of Bern. Even if stratospheric trace gases generally show small horizontal variability, the distance between the different stations, especially between MOH and OHP, may lead to some differences in measured ozone.

2.0.1 ~~Microwave radiometers~~

2.1 Microwave radiometers

We use data from two microwave radiometers, both located in Switzerland. They measure ~~atmospheric emission at the~~ 142 GHz ~~, line~~ where ozone molecules emit microwave radiation due to rotational transitions. The spectral line measured is pressure broadened and thus contains ~~altitude information of the~~ information about the vertical distribution of ozone molecules. To obtain a vertical ozone profile, the received radiative intensity is compared to the spectrum simulated by the Atmospheric Radiative Transfer Simulator 2 (ARTS2, *Eriksson et al.* [2011]). By using an optimal estimation method according to *Rodgers* [2000], the best estimate of the vertical profile of ozone volume mixing ratio is then retrieved from the measured spectrum. This is done by the software tool Qpack2, which provides together with ARTS2 an entire retrieval environment [*Eriksson and Jimé*, 2005].

The **Ground-based Millimetre-wave Ozone Spectrometer** (GROMOS) located in Bern ~~, Switzerland (46.95 N, 7.44 E) at an altitude of 574 above sea level (a.s.l.) is is the~~ main focus of this study [*Kä*, 1995; *Peter*, 1997]. GROMOS has been measuring continuously ozone spectra since November 1994. Before October 2009, the measurements were performed by means of a filter bench (FB) with an integration time of 1 h. Since October 2009, a Fast Fourier Transform Spectrometer (FFTS) with an integration time of 30 min has been used. An overlap measurement period of almost 2 years (Oct. 2009 to July 2011) was used to homogenize the FB data, by subtracting the mean bias profile ~~of the averaged over the whole~~ overlap period ($FB_{mean} - FFTS_{mean}$) from all FB profiles [*Moreira et al.*, 2015]. These homogenized ozone data are available on the NDACC web page (<ftp.cpc.ncep.noaa.gov/ndacc/station/bern/hdf/mwave/>). The FFTS retrieval used in the present study (version 2021) uses variable errors in the a priori covariance matrix of around 30 % in the stratosphere and 70 % in the mesosphere, and a constant measurement error of 0.8 K (*Moreira*, 2017). The retrieved profiles have a vertical resolution of ~ 15 to 25 km in the stratosphere. We concentrate in this study on the middle and upper stratosphere between 31 hPa (≈ 24 km) and 0.8 hPa (≈ 49 km), where the retrieved ozone is quasi independent of the a priori information. This is assured by limiting the altitude range to the altitudes where the area of the averaging kernels (measurement response) is larger than 0.8, which means that more than 80 % of the information comes from the observation ~~instead rather than from~~ the a priori data (*Rodgers*, 2000). More information about the homogenization as well as the parameters used in the retrieval can be found in *Moreira et al.*

[2015]. Besides the described data harmonization to account for the instrument upgrade, we performed some additional data ~~correction~~corrections. Because the stratospheric signal is weak in an opaque and humid ~~atmosphere~~troposphere, we discarded measurements when the atmospheric transmittance was smaller than 0.3. Excluding measurements in such a way should not result in a sampling bias because tropospheric humidity is uncorrelated to stratospheric ozone. Also, the data have been corrected for outliers at each pressure level by removing values that exceed four times the standard deviation within a ~~3-days-moving~~3-day moving median window. Profiles were excluded completely when more than 50 % of their values were missing (e.g. due to outlier detection). Furthermore, we omitted profiles where the instrumental system temperature showed outliers ~~-,identified~~by an exceed-of exceeding four times the standard deviation within a ~~30-days-moving~~30-day moving median window.

The second microwave radiometer used in this study is the Stratospheric Ozone Monitoring Radiometer (SOMORA). It was built in 2000 as an update of the GROMOS radiometer and is located ~~at the aerological station (MeteoSwiss) in Payerne (46.8 N, 7.0 E) at 491 a.s.l. in Payerne~~ since 2002. Some instrumental changes were performed in the beginning of 2005 (front-end change) and in October 2010, when the acousto-optical spectrometer of SOMORA has been upgraded to an FFTS [Maillard Barras *et al.*, 2015]. The data have been harmonized to account for the spectrometer change. The instrument covers an altitude range from 25 to 60 km with a temporal resolution of 30 min to 1 h. In this study, we consider SOMORA data at an altitude range between 18 hPa (≈ 27 km) and 0.8 hPa (≈ 49 km). For more information about SOMORA, refer to Calisesi [2003] concerning the instrumental setup and Maillard Barras *et al.* [2009, 2015] concerning the operational version of the ozone retrievals used in the present study.

2.1.1 Lidars

2.2 Lidars

We use data from two differential absorption lidar (DIAL, Schotland [1974]) instruments in Germany and France. The instruments emit laser pulses at two different wavelengths, where one is absorbed by ozone molecules and the other is not. Comparing the backscattered signal at these two wavelengths provides information on the vertical ozone distribution in the atmosphere. The lidars can only retrieve ozone profiles during clear-sky nights due to scattering on cloud particles and the interference with sunlight.

The lidar at the Meteorological Observatory Hohenpeissenberg (MOH, ~~47.8 N, 11.0 E, 980 a.s.l.) in Germany~~) has been operating since 1987, emitting laser pulses at 308 and 353 nm [Werner *et al.*, 1983; Steinbrecht *et al.*, 2009b]. ~~The lidar can only retrieve ozone profiles during clear-sky nights due to scattering on cloud particles and the influence of sunlight.~~ On average, it retrieves eight night-profiles in a month. In this study we limit the data to the altitude range where the ~~averaged measurement error~~measurement error averaged over the whole study period is below 10 % (below ~~43~~42 km or 2 hPa). The lower altitude limit was set to the chosen limit of GROMOS at 31 hPa (≈ 24 km).

The Observatory of Haute Provence (OHP, ~~43.9 N, 5.7 E, 650 a.s.l.) in France hosts~~) operates a lidar that ~~measures has been measuring~~ in its current setup since the end of 1993 [Godin-Beekmann *et al.*, 2003]. The lidar emits laser pulses at 308 nm and 355 nm, as first described by Godin and Mé [1989]. The instrument measures on average 11 profiles per month. We use OHP

lidar profiles below 40 km (≈ 32.7 hPa), where the averaged measurement error remains below 10 %. As lower altitude limit we use 31 hPa (≈ 24 km) to be consistent with the GROMOS limits.

More detailed information about the lidars and ozonesonde used can be found for example in *Godin et al.* [1999] and *Nair et al.* [2011, 2012].

5 2.2.1 Ozonesondes

2.3 Ozonesondes

The three mentioned observatories at Payerne, MOH and OHP also provide weekly ozonesonde measurements. The ozonesonde measurements in Payerne are usually performed three times a week at 11:00 UTC [*Jeannet and Stii*, 2007], resulting in 13 profiles per month on average. The meteorological balloon carried a Brewer Mast sonde (BM, *Brewer and Milford* [1960]) until 10 September 2002, which was then replaced by an electrochemical concentration cell (ECC, *Komhyr* [1969]). ~~The profiles are normalised using~~ The profiles are normalized using concurrent total column ozone from the Dobson spectrometer at Arosa (46.77° N, 9.7° E, 1850 m a.s.l., *Favaro et al.* [2002]). ~~If those~~ If the Dobson data are not available, forecast ozone column estimates based on GOME-2 (Global Ozone Monitoring Experiment-2) data are used (<http://www.temis.nl/uvradiation/nrt/uvindex.php>).

15 Ozone soundings at MOH are performed two to three times per week with a BM sonde (on average 10 profiles per month). Three different radiosonde types have been used since 1995, all carrying a BM ozonesonde, without major changes in its performance since 1974 [*Steinbrecht and Kö*, 2016]. The profiles are normalized by on-site Dobson or Brewer spectrophotometers and, if not available, by satellite data [*Steinbrecht and Kö*, 2016].

20 At OHP, ECC ozonesondes have been used since 1991 with several instrumental changes [*Gaudel and Ancellet*, 2015]. The data are normalized with total column ozone measured by a Dobson spectrophotometer until 2007 and an ultraviolet-visible SAOZ (Systè~~Systè~~me d'Analyse par Observation Zé~~Zé~~énithale) spectrometer afterwards [*Guirlet et al.*, 2000; *Nair et al.*, 2011]. In our analysed period, four profiles are on average available per month.

Ozonesonde data are limited to altitudes up to ~ 30 km, ~~where,~~ above which the balloon usually bursts. Therefore, we used ozonesonde profiles only below 30 km, which is a threshold value for Brewer Mast ozonesondes with precision and accuracy 25 below ± 5 % (Smit and Kley, 1996). For normalization, the correction factor (CF), which is the ratio of total column ozone from the reference instrument to the total ozone from the sonde, has been applied to all ozonesonde profiles. At all measurement stations, we discarded profiles when their CF was larger than 1.2 or smaller than 0.8 (Harris et al., 1998; Smit and ASOPOS Panel, 2013). We further excluded ~~negative values and~~ profiles with extreme jumps or constant ozone values, as well as profiles with constant or decreasing altitude values.

2.3.1 ~~Aura/MLS~~

2.4 Aura/MLS

The microwave limb sounder (MLS) on the Aura satellite, ~~which was~~ launched in mid 2004, measures microwave emission from the Earth in five broad spectral bands [Guirlet *et al.*, 2006]. It provides profiles of different ~~molecules~~ trace gases in the atmosphere, ~~including stratospheric ozone~~, with a vertical resolution of ~ 3 km. Stratospheric ozone is retrieved by using the spectral band centred at 240 GHz. We used ozone data from Aura/MLS version 4.2 above Bern with a spatial coincidence of $\pm 1^\circ$ latitude and $\pm 8^\circ$ longitude, where the satellite passes twice a day (around 02:00 and 13:00 UTC). More information about the MLS instrument and the data product can be found, e.g., in Waters *et al.* [2006]. We chose the Aura/MLS data for our study because ~~it was identified to be stable with insignificant~~ there are no drifts between 20 and 40 km [Hubert *et al.*, 2016].

10 2.5 ~~Data comparison~~

~~To verify-~~

3 Time series comparison

To identify potential anomalies in the GROMOS data ~~and detect possible anomalies~~, we compared ~~all described data sets with the GROMOS data from Bern~~ the data with the other described data sets in the time period ~~from~~ Jan. 1995 to Dec. ~~2017~~. ~~The different instrument sites are all located in Central Europe. Payerne is located 40 south-west of Bern, which ensures comparable stratospheric measurements. Hohenpeissenberg is located 290 north-east of Bern, and the OHP lies 360 south-west of Bern. Even if stratospheric trace gases generally show small horizontal variability, the distance between the different stations, especially between MOH and OHP, may lead to some differences in measured ozone.~~ 2017, except for some instruments that cover a shorter time period (Table 1).

20 3.1 Comparison methodology

To compare GROMOS with the other instruments, the different data sets have been processed to compare consistent quantities and smoothed to the GROMOS grid. Taking relative differences between the data sets made it possible to identify anomalous periods in the GROMOS time series.

3.1.1 ~~Data processing~~ for comparison

25 ~~The GROMOS retrieval is accurate~~ In this study we concentrate on the altitude range between 31 and 0.8 hPa, ~~where the a priori contribution to GROMOS profiles is low (see Sect. 2.1).~~ We therefore limit all instrument data ~~in this study~~ to this altitude range, and divide it in three parts. For convenience, they will be referred to as lower stratosphere between 31 and 13 hPa (≈ 24 to 29 km), middle stratosphere between 13 and 3 hPa (≈ 29 to 39 km) and upper stratosphere between 3 and 0.8 hPa (≈ 39 to

49 km). The limits for the upper stratosphere agree with the common definition [e.g. *Ramaswamy et al.*, 2001], whereas the here defined lower and middle stratosphere are usually referred to as middle stratosphere in other studies.

Most of the instruments provide volume mixing ratios (VMR) of ozone in parts per million (ppm). In case that the data were given in number density (molecules cm^{-3}), ~~as for example for the OHP ozonesonde, the~~ VMR was calculated with the air pressure and temperature provided by the same instrument ~~or collocated radiosonde data. Data that can not be physically explained such as negative ozone values have been excluded for all instruments.~~ for ozonesondes or co-located ozone- or radiosonde data for lidars. For lidar measurements, these sonde temperature and pressure profiles are completed above the balloon burst by operational model data from the National Center for Environmental Prediction (NCEP) at OHP and by lidar temperature measurements and extrapolated radiosonde pressure data at MOH.

The GROMOS, SOMORA and Aura/MLS profiles have a constant pressure grid, which is not the case for the lidar and ozonesonde data. ~~We therefore interpolated all~~ The lidar and ozonesonde profiles ~~to a pressure grid based on a data were therefore linearly interpolated to a~~ regular spaced altitude grid of 100 m for the ozonesonde at OHP and Payerne, and 300 m for the lidars and the ozonesonde at MOH. The mean profile of the interpolated pressure data then built the new pressure grid for the ozone data. These interpolated lidar and ozonesonde data are used for the trend estimations. For the direct comparison with GROMOS, the data were adapted to the GROMOS grid, which is described in the next section (Sect. 3.1.2). Our figures generally show both pressure and geometric altitude. The geometric altitude is approximated by the mean altitude grid from GROMOS, which is for each retrieved profile determined from operational model data of the European Centre for Medium-Range Weather Forecasts (ECMWF).

3.1.2 ~~Direct comparison with~~ GROMOS comparison and anomalies

The vertical resolution of GROMOS and SOMORA is usually coarser than for the other instruments, ~~except for SOMORA~~. When comparing ~~their~~ profiles directly with GROMOS profiles, the different vertical resolution of the instruments has to be considered. Smoothing the profiles of the different instruments by GROMOS' averaging kernels makes it possible to compare the profiles with GROMOS without biases due to resolution or a priori information [*Tsou et al.*, 1995]. The profiles having higher vertical resolution than GROMOS were ~~interpolated to the GROMOS pressure grid, and were then~~ convolved by the averaging kernel matrix according to *Connor et al.* [1991], with

$$\mathbf{x}_{conv} = \mathbf{x}_a + \mathbf{AVK}(\mathbf{x}_h - \mathbf{x}_{a,h}) \quad (1)$$

where \mathbf{x}_{conv} is the resulting convolved profile, \mathbf{x}_a is the a priori profile used in the GROMOS retrieval, ~~\mathbf{x}_h is the profile of the higher resolved instrument,~~ $\mathbf{x}_{a,h}$ is the same a priori profile, but interpolated to the ~~GROMOS grid,~~ grid of the highly resolved measurement, \mathbf{x}_h is the profile of the highly resolved instrument, and \mathbf{AVK} is the corresponding averaging kernel matrix from GROMOS. ~~When \mathbf{x}_h covered a smaller altitude range than GROMOS, the profile \mathbf{x}_h was extended by the a priori profile of GROMOS to cover the same altitude range. After the convolution, these additional values were removed again to obtain the original altitude range of \mathbf{x}_h .~~ The rows of the \mathbf{AVK} have been interpolated to the grid of the highly resolved instrument and scaled to conserve the vertical sensitivity [*Keppens et al.*, 2015]. The SOMORA profiles have a similar vertical resolution

as profiles from GROMOS and were thus not convolved, because this would require a more advanced comparison method as proposed by *Rodgers and Connor* [2003] or *Calisesi et al.* [2005]. GROMOS ~~has and SOMORA have~~ a higher temporal resolution than the other instruments ~~(except SOMORA)~~. For SOMORA, only profiles coincident in time with GROMOS have been selected. For the other instruments, a mean of GROMOS data at the time of the corresponding measurement was used, with a time coincidence of ± 30 min. Only for the lidars, GROMOS data were averaged over the whole lidar measurement time (usually one night).

For ~~direct~~ comparison with GROMOS we computed relative differences between the monthly mean values of the different data sets and the monthly mean of the coincident GROMOS profiles. The relative ~~differences have difference (RD) for a specific month i has~~ been calculated by subtracting ~~each the~~ monthly ozone value of the data set ~~X (X_i)~~ from the corresponding GROMOS monthly mean (GR_i), using the GROMOS monthly mean ~~of the corresponding month~~ as a reference ~~(($GR - X$)/ $GR - 100$).~~ ~~Based on these:~~

$$\underline{RD_{i,X}} = (GR_i - X_i)/GR_i \cdot 100. \quad (2)$$

~~Based on the~~ relative differences we identified periods where GROMOS ~~showed biased values compared to~~ differs from the other instruments. To identify these anomalies we used a ~~corrected relative difference to GROMOS. For this the relative difference time series for each instrument was corrected by its~~ debiased relative difference (RD_{debiased}), given by

$$\underline{RD_{\text{debiased},i,X}} = RD_{i,X} - \overline{RD_X}, \quad (3)$$

where $\overline{RD_X}$ is the mean relative difference ~~to GROMOS of GROMOS to the data set X over the whole period~~. This made it possible to ignore a potential constant offset of the instruments and to concentrate on periods with temporally large differences to GROMOS. When this ~~corrected debiased~~ relative difference was larger than 10 % for at least three instruments, the respective month was identified as anomaly ~~in the GROMOS data~~. Above 2 hPa, where only SOMORA and Aura/MLS data are available, both data sets need to have a ~~corrected relative difference debiased relative difference to GROMOS~~ larger than 10 % to be identified as anomaly.

3.2 Trend model

4 Time series comparison

~~To verify the GROMOS data and identify potential anomalies, we first have a look at the GROMOS time series (Sect. 3.1) and compare then the data with instrument data in Central Europe (Sect. 3.2).~~

3.1 GROMOS time series

The monthly means of the GROMOS time series (Fig. 1 (a)) clearly depict the maximum of ozone VMR between 10 hPa and 5 hPa and the seasonal ozone variation, with increased spring–summer ozone in the middle stratosphere and increased autumn–winter values in the upper stratosphere [*Moreira et al.*, 2016]. Figure 1 (b) shows GROMOS' relative deviations from

the monthly climatology (monthly means over the whole period 1995 to 2017). ~~This ozone anomaly~~ This ozone deviation is calculated by the ratio of the deseasonalised monthly means (difference between each individual monthly mean and the corresponding climatology of this month) and the monthly climatology. We observe some periods where GROMOS data deviates from its usual values, mostly distinguishable between the lower/middle stratosphere and the upper stratosphere. In the lower/middle stratosphere we observe negative anomalies (less ozone than usual) in 1995 to 1997 and 2005 to 2006 and positive anomalies (more ozone than usual) in 1998 to 2000, in 2014 to 2015 and in 2017. In the upper stratosphere the data show negative anomalies in 2016 and positive anomalies in 2000 and 2002 to 2003. Strong but short-term positive anomalies are visible in 1997 in the upper stratosphere and in the beginning of 2014. The positive anomaly in the upper stratosphere in 1997 is due to some missing data in November 1997 because of an instrumental upgrade, leading to a ~~higher~~ larger monthly mean value than usual. Besides this we did not detect any systematic instrumental issues in the GROMOS data that could explain the anomalies. Therefore, we compare the GROMOS data with the presented ground-based ~~instruments from Central Europe, data sets~~, as well as with Aura/MLS data, to check whether the observed anomalies are due to natural variability or due to unexplained instrumental issues.

3.2 Comparison of different data sets

We compared GROMOS with ground-based and Aura/MLS data and averaged it over three altitude ranges (Fig. 2). The different data sets have been smoothed with the averaging kernels of GROMOS to make a direct comparison possible, as described in Sect. 3.1.2. Due to the similar vertical resolution of GROMOS and SOMORA, the SOMORA profiles have not been smoothed by GROMOS' averaging kernels, despite differences between their a priori data and averaging kernels. This might lead to larger differences between GROMOS and SOMORA than between GROMOS and the other instruments. To avoid that an instrument does not cover the full range of one of the three altitude ~~classes~~ ranges, all ozonesonde data have been cut at 30 km (≈ 11.5 hPa), all lidar data at 3 hPa (≈ 40 ~~39~~ km), and all SOMORA data below 13 hPa (≈ 29 km) for this analysis. The different instrument time series shown in Fig. 2 contain only data that are coincident with GROMOS measurements as described in Sect. 3.1.2, whereas the ~~here shown GROMOS data~~ GROMOS data shown here represent the complete GROMOS time series with ~~another, usually finer~~ its high temporal sampling. This might lead to some sampling differences that are not considered in this figure.

The different data sets agree well, showing, however, periods where some instruments deviate more from GROMOS than others. In the upper stratosphere (Fig. 2 (a)), GROMOS and SOMORA agree well most of the time, but GROMOS reports slightly ~~lower~~ less ozone than SOMORA and also ~~lower~~ smaller values than Aura/MLS. A step change between SOMORA and GROMOS is visible in 2005, which might be related to the SOMORA front-end change in 2005. In the middle stratosphere (Fig. 2 (b)), both lidars exceed the other instrument data in the last years, starting in 2004 at OHP and in 2010 at MOH. Similar deviations of the MOH and OHP lidars have also been observed by *Steinbrecht et al.* [2017], as can be seen in the latitudinal lidar averages of their study.

Differences between the data sets in the lower stratosphere can be better seen in Fig. 3. Shown are the monthly relative differences of time coincident pairs of GROMOS (GR) and the convolved data set X, with GROMOS data as reference value

~~((GR - X)/GR - 100), Eq. (2)).~~ The mean relative difference of all instruments compared to GROMOS (black line in Fig. 3) generally lies within $\sim \pm 10\%$. However, there are some periods with larger deviations, where GROMOS measures ~~lower~~ less ozone than the other instruments (negative relative difference) in 1995 to 1997 and in 2006 in the lower stratosphere, and in 2016 in the middle and upper stratosphere. We further observe that the relative difference between GROMOS and the OHP ozonesonde shows some important peaks in the last decade, indicating that the sonde often measures ~~higher~~ more ozone than GROMOS. ~~This can be explained by the low~~ The ozonesonde data seems to have some outlier profiles. When comparing the monthly means of coincident pairs, these outliers are even more visible, because only a small number of OHP ozonesonde profiles are available per month (only four profiles on average), ~~which leads occasionally to large differences in the monthly mean when comparing only few profiles with time coincident GROMOS data.~~

For a broader picture, the same relative differences to GROMOS are shown in Fig. 4, but each panel represents an individual instrument and all altitude levels are shown. The anomalies where at least three data sets (or two above 2 hPa) deviate by more than 10% from GROMOS (as described in Sect. 3.1.2) are shown in the lowest panel in black. In addition to the negative anomalies observed already in the other figures (e.g. in 2006 in the lower stratosphere and in 2016 in the upper stratosphere), we also observe positive anomalies in the lower/middle stratosphere in 2000 ~~, 2014, 2015, and 2017.~~ and 2014. The negative anomalies in 1995 to 1997 in the lower stratosphere that we observed in Fig. 3 were only partly detected as anomalies with our anomaly criteria.

To summarize our comparison results, we observed some periods with anomalies compared to GROMOS' climatology (Fig. 1). Some of these anomalies were also observed when comparing GROMOS to the different data sets (Fig. 2, Fig. 3, and Fig. 4). This implies that the source of the anomalies are local variations in Bern or instrumental issues of GROMOS rather than broad atmospheric variability. We can thus conclude that the observed negative GROMOS anomalies in the lower stratosphere in 2006 and in the upper stratosphere in 2016 are biases in the GROMOS time series. The same is the case for the positive anomalies in 2000 ~~, and 2014, 2015 and 2017~~ in the lower and middle stratosphere, also for some summer months in 2015, 2016, and 2017. In contrast to these confirmed anomalies, the GROMOS anomalies in the lower stratosphere in 1995 to 1997 (negative) and 1998 and 1999 (positive) as observed in Fig. 1 are small when comparing to the other instruments and are thus only confirmed for a few months by our anomaly detection. The biased periods in the upper stratosphere in 2000 and 2002 to 2003 (positive anomalies) were not confirmed by comparing GROMOS to ~~the other data sets~~ SOMORA, and might thus be real ozone variations. However, we have to keep in mind that less instruments (only SOMORA and Aura/MLS) provide data for comparison above 2 hPa, which makes the anomaly detection less robust at these altitudes, especially prior to the start of Aura/MLS measurements in 2004. Our results are consistent with those reported by *Moreira et al.* [2017]. They compared GROMOS with Aura/MLS data and also observed positive deviation of GROMOS in the middle stratosphere in 2014 and 2015, as well as a negative deviation in the upper stratosphere in 2016. ~~at. [2018a] al. [2019a]~~ found similar anomalies in the GROMOS time series by comparing ground-based data sets to several satellite products (see also ~~at. [2018b]).~~ SPARC/IO3C/GAW [2019]). Some of our detected biased periods were also found by *Steinbrecht et al.* [2009a] who compared different ground-based instruments, for example the GROMOS anomaly in 2006. They attribute the observed biases to sampling

differences, but also to irregularities in some data sets. In fact, our results confirm irregularities in the GROMOS time series, which are probably due to instrumental issues of GROMOS and not only due to sampling differences.

4 Ozone trend estimations

Ozone trends are estimated in the present study by using a multilinear parametric trend model [von Clarmann et al., 2010].

5 By comparing the GROMOS data to the other instruments data sets as described above (Sect. 3.2), we have confirmed some biased-anomalous periods in the GROMOS time series. To improve the GROMOS trend estimates, we use now these detected anomalies by considering and consider them in the regression fit (Sect. 4.3). In the following section, this method, the trend model (Sect. 4.1) will first be applied to an artificial time series to test and illustrate the approach of considering anomalies in the regression (Sect. 4.2). It will then be applied on GROMOS data (Sect. 4.3), before comparing the resulting GROMOS trends to trends from the other instruments (Sect. 4.4).

4.1 Trend model

To estimate stratospheric ozone trends we applied the multilinear parametric trend model from *von Clarmann et al.* [2010] to the monthly means of all individual data sets. The model fits the following regression function:

$$\begin{aligned}
 y(t) = & a + b \cdot t \\
 & + c \cdot \text{QBO}_{30\text{hPa}}(t) + d \cdot \text{QBO}_{50\text{hPa}}(t) \\
 & + e \cdot \text{F10.7}(t) \\
 & + f \cdot \text{MEI}(t) \\
 & + \sum_{n=1}^4 \left(g_n \cdot \sin\left(\frac{2\pi t}{l_n}\right) + h_n \cdot \cos\left(\frac{2\pi t}{l_n}\right) \right)
 \end{aligned} \tag{4}$$

where ~~t is the monthly time vector~~, $y(t)$ represents the estimated ozone time series, t is the monthly time vector, and a to h are coefficients that are fitted in the trend model. $\text{QBO}_{30\text{hPa}}$ and $\text{QBO}_{50\text{hPa}}$ are the normalized Singapore winds at 30 and 50 hPa, used as indices of the quasi-biennial oscillation (QBO, available at <http://www.geo.fu-berlin.de/met/ag/strat/produkte/qbo/singapore.dat>). F10.7 is the solar flux at a wavelength of 10.7 cm used to represent the solar activity (measured in Ottawa and Penticton, Canada [~~Canada, 2018~~ Canada, 2019]). The El Niño Southern Oscillation (ENSO) is considered by the Multivariate ENSO Index (MEI), that combines six meteorological variables measured over the tropical Pacific [*Wolter and Timlin*, 1998]. The F10.7 data and the MEI data are available via <https://www.esrl.noaa.gov/psd/data/climateindices/list/>. In addition to the described natural oscillations, we used four periodic oscillations with different wavelengths l_n to account for annual ($l_1 = 12$ months) and semi-annual (~~$l_n = 6$~~ $l_2 = 6$ months) oscillation as well as two further overtones of the annual cycle (~~and~~ $l_3 = 4$ months and $l_4 = 3$ months).

The strength of the model used from von Clarmann et al. [2010] used in our study is that it can consider inhomogeneities in data sets, by considering a full error covariance matrix (\mathbf{S}_y) when reducing the cost function (χ^2). Inhomogeneous data

can originate from changes in the measurement system (e.g. changes in calibration standards or merging of data sets with different instrumental modes), ~~or~~ from irregularities in spatial or temporal sampling, or from unknown instrumental issues. Such inhomogeneities lead to groups of ~~temporal-correlated data~~ temporally correlated data errors, that can, if not considered, change significance and slope of the estimated trend [von Clarmann *et al.*, 2010]. Von Clarmann *et al.* present an approach to consider ~~such~~ known or suspected inhomogeneities in the trend analysis. They divide the data into multiple subsets which are assumed to be biased with respect to each other and which are characterized by diagonal blocks in the data covariance matrix. ~~Another reason for inhomogeneities might be unknown instrumental issues that lead to anomalies, which is treated in the present study.~~ We use their method and code in our trend analyses to account for inhomogeneities. The inhomogeneities are in our case anomalies in some months that we identified as described in Sect. 3.1.2.

~~To consider inhomogeneities, the~~ The total uncertainty of the data set is represented by a full error covariance matrix \mathbf{S}_y at that describes for each pressure level. ~~The covariances between the measurements in time. The covariance matrix is for each instrument given by~~

$$\mathbf{S}_y = \mathbf{S}_{instr} + \mathbf{S}_{autocorr} + \mathbf{S}_{bias} \quad (5)$$

where \mathbf{S}_{instr} gives the monthly uncertainty estimates for the instrument, $\mathbf{S}_{autocorr}$ accounts for residuals autocorrelated in time which are caused by atmospheric variation not captured by the trend model, and \mathbf{S}_{bias} describes the bias uncertainties when a bias is considered. The diagonal elements of $\mathbf{S}_{y_{instr}}$ are set to the monthly means of the measurement uncertainties for each instrument. For lidar data this is on average 4 % for the OHP lidar and 6 % at MOH between ~ 20 and 40 km, with ~~lowest~~ smallest uncertainties in the middle stratosphere. For the ozonesonde, the uncertainty is assumed to be 5 % for all ECC sondes and 10 % for BM sondes (Harris *et al.*, 1998; Smit and ASOPOS Panel, 2013). For SOMORA we use uncertainties composed of smoothing and observational error, ranging from 1 to 2 % in the middle stratosphere and 2 to 7 % in the upper stratosphere. The ~~used~~ Aura/MLS uncertainties used range between 2 and 5 % throughout the stratosphere. For ~~the GROMOS trends,~~ GROMOS we use uncertainty estimates as described by *Moreira et al.* [2015], composed of the standard error (σ/\sqrt{DOF} with standard deviation σ and degrees of freedom DOF) of the monthly means, an instrumental uncertainty (measurement noise), and an estimated systematic instrument uncertainty obtained from cross-comparison. The resulting uncertainty values are approximately 5 % in the middle stratosphere, 6 to 8 % in the lower stratosphere and 6 to 7 % in the upper stratosphere. The off-diagonal elements of $\mathbf{S}_{y_{instr}}$ are ~~first~~ set to zero, assuming no error correlation ~~between the measurements in time. The additional covariance matrix $\mathbf{S}_{autocorr}$, which is added to \mathbf{S}_{instr} , is first also set to zero.~~ In a second step iteration, autocorrelation coefficients are inferred from residuals of the initial trend fit. ~~These are used to build another covariance matrix which is added to the data error covariance matrix.~~ The mean variance of ~~this additional covariance matrix is adjusted~~ $\mathbf{S}_{autocorr}$ is scaled such that the χ^2 divided by the degrees of freedom of the trend fit with the ~~enhanced covariance matrix is unity.~~ new \mathbf{S}_y becomes unity. The degrees of freedom are the number of data points minus the number of fitted variables. The latter are the number of the coefficients of the trend model plus the number of relevant correlation terms inferred by the procedure described above. This additional covariance matrix $\mathbf{S}_{autocorr}$ represents contributions to the fit residuals which are not caused by data errors but by phenomena that are not represented by the trend model. ~~This second trend fit is only performed.~~ $\mathbf{S}_{autocorr}$ is only considered if

the initial normalized χ^2 is larger than unity, which is not the case if the assumed data errors are larger than the fit residuals [Stiller *et al.*, 2012]. The more sophisticated uncertainty estimates that we use for GROMOS are larger than the residuals in the first regression fit ($\chi^2 < 1$), which means that the ~~correlated residuals~~ time correlated residuals (S_{autocorr}) are not considered for the GROMOS trend. For all the other instruments, however, correlated residuals are considered, because we use simple
5 instrumental uncertainties that are usually smaller than the fitted residuals.

To account for the anomalies in the time series when estimating the trends we adapted S_y in two steps. First, we increased the uncertainties for months and altitudes where anomalies were identified ~~-(using the method described in Sect. 3.1.2)~~. For this purpose, we set the diagonal elements of $S_{y_{\text{instr}}}$ for the respective months-month i to a value obtained from the mean difference to all instruments ($RD_{\text{debiased},i}$) and the mean GROMOS ozone value at each altitude. Assuming for example that GROMOS
10 deviates from all instruments on average by 10% in August 2010-2014 at 10 hPa, and the overall mean August value at this altitude would be 7 ppm. In this case we would assume an uncertainty of 0.7 ppm for this biased month at this altitude level. In a second step, we ~~used the ability of the trend programme to~~ account for biases in ~~a data subset~~. the data subsets where anomalies were detected. For this, a fully correlated block composed of the squared estimated bias uncertainties is added to the ~~corresponding part of the covariance matrix~~ part of S_{bias} that is concerned by anomalies. For example, to account for a bias
15 in the summer months of 2014, a fixed bias uncertainty is added to all variances and covariances of these months in S_{bias} . Considering the bias in this way is mathematically equivalent to treating the bias as an additional fit variable that is fitted to the regression model with an optimal estimation method, as shown by von Clarmann *et al.* [2001]. The value chosen for the bias uncertainty determines how much the bias is estimated from the data itself. For low-small values, the bias will be close to the a priori value, which would be a bias of zero, for high-large values it will be estimated completely from the given data. The
20 bias can thus be estimated from the data itself, which makes the method more robust because it does not depend on an a priori choice of the bias [Stiller *et al.*, 2012]. In a sensitivity test we ~~have~~ found that assuming a-an uncertainty value of 5 ppm for the correlated block permits a reliable fit, where the bias is fitted independently of the a priori zero bias.

The described procedure for anomaly consideration is only applied to the GROMOS time series, whereas the other trend estimates were not corrected for anomalies. Our ozone trend estimates always start in Jan. 1997, which is the most likely turning
25 point for ozone recovery due to the decrease of ODS [Jones *et al.*, 2011], and all end in Dec. 2017. Exceptions are the trends from SOMORA and Aura/MLS. The SOMORA trend starts in Jan. 2000 when the instrument started to measure ozone ~~and the Aura/MLS trend starts in Jan. 2005~~. Aura/MLS trends build covers the shortest trend period, starting only in Jan. 2005. The trends are always given in % per decade, which is determined from the at each altitude level from the regression model output in ppm per decade by dividing it by the overall for each data set by its ozone mean value of the ~~corresponding data set at each~~
30 altitude level whole period. We declare a trend to be significantly different from zero at 95 % confidence interval, as soon as its absolute value exceeds twice its uncertainty, ~~as described by ?~~. This statistically inferred confidence interval is based on the assumed instrumental uncertainties. Unknown drifts of the data sets, however, are not considered in this claim.

4.2 Artificial time series analysis

The trend programme from *von Clarmann et al.* [2010] can handle uncertainties in a flexible way, which makes it possible to account for anomalies in a time series, as described above (Sect. 4.1). To investigate how anomalies can be best considered in the trend programme, we first test the programme with a simple artificial time series that consists of monthly ozone values with a seasonal cycle and a linear trend. This and then try to use the specific features of the trend programme to immunize it against anomalies. The artificial time series $y(t)$ used for this purpose consists of monthly ozone values from Jan. 1997 to Dec. 2017 and is given by

$$y(t) = a + b \cdot t + g \cdot \sin\left(\frac{2\pi t}{l_n}\right) + h \cdot \cos\left(\frac{2\pi t}{l_n}\right) \quad (6)$$

with the monthly time vector t , a constant ozone value $a = 7$ ppm and a trend of 0.1 ppm per decade, with i.e. $b = 0.1/120$ ppm per month. We consider a simple seasonal oscillation with an amplitude $A = 1$ ppm such that $A^2 = g^2 + h^2$ (e.g. $g = h = \sqrt{0.5}$ ppm) and a wavelength $l_n = 12$ months. The uncertainty was assumed to be 0.1 ppm for each monthly ozone value, which was considered in the diagonal elements of $S_y S_y$. No noise was superimposed to the data. This artificial time series (shown in Fig. 5 (a)) is later on referred to as case A. The estimated trend for this simple time series corresponds quasi perfectly to the assumed time series' trend (0.1 ppm per decade), which proves that the trend programme works well. The residuals (Fig. 5 (b)) are small (6 order of magnitudes), but are of order 10^{-6} and increase towards the time series edges. This might be due to some error propagation, but can be neglected for this study because of its small magnitude start and end of the time series (Fig. 5(b)).

To investigate how the trend programme reacts to anomalies in the time series, we performed the following sensitivity study. We increased the monthly ozone values by 5% (≈ 0.4) in the summer months (June, July, Aug.) of 2014, 2015 and 2017. When estimating the trend for 2017 by 5% (≈ 0.4 ppm). Since we are interested in the net effect of the anomalies, again no noise was superimposed on the test data. The same error covariance matrix S_y as in case A has been used. For this modified time series (case B), we observe a trend of ~ 0.13 ppm per decade instead of the expected ~ 0.1 ppm per decade (Fig. 6 and Table 2). Assuming that a real time series contains such suspicious anomalies due to for example instrumental issues, they would distort the true trend.

A simple way to handle such anomalies would be to omit anomalous data in the time series. This, however, would increase trend uncertainties and lead to a loss of important information. Therefore, we use the presented approach to handle anomalous observations in the time series when estimating the trend. To account for these anomalies in the trend estimation, we applied different possible corrections by modifying the error covariance matrix S_y in the trend model. The used parameters for the different corrections are summarized in Table 2, and Fig. 6 presents the resulting linear trends. In a first attempt we decreased the weight of the anomalies in the time series by increasing their uncertainties (diagonal elements of S_y) and set the uncertainties of the corresponding months to affected summer months to 0.36 ppm (case C in Fig. 6 and Table 2). This uncertainty value corresponds to 5% of the overall mean ozone value (case C). Afterwards, the uncertainty of the months without anomalies remained 0.1 ppm. No additional

error correlations between the anomaly-affected data points were considered. The impact of the anomaly is already reduced from about 33 % to about 3 %. The estimated error of the trend has slightly increased because with this modified covariance matrix which represents larger data errors the data set contains less information.

In a next step, we added a correlated block to the covariance matrix for the months containing anomalies, and applied the S_y once ~~with~~ without (case D) and once ~~without~~ with (case E) the increased diagonal elements of S_y . ~~The bias uncertainty in~~ Adding the correlated block ~~was set to 5~~.

The results show that all of the corrected trend estimations (cases C, D, and E) improve the trends and lead to almost the true trend of 0.1 (Fig. 6 (b)). They are thus able to correct for the artificially added anomalies. Considering a correlated bias block in to S_y corresponds to an unknown bias of the data subset that is affected by anomalies, and leads to a free fit of the bias magnitude. This bias is represented in S_y ~~(case~~ as a fully correlated block of $(5 \text{ ppm})^2$. It has an expectation value of zero and an uncertainty of 5 ppm. With this approach (cases D and E) ~~leads to even better results than only increasing the diagonal elements of S_y (case C). Consequently,~~ the trend obtained from the anomaly-affected data is almost identical with the trend obtained from the original data (case A). This implies that the trend estimation has successfully been immunized against the anomaly.

In summary, we found that the trend estimates based on anomaly-affected data are largely improved by consideration of the trends ~~are improved by correcting the~~ anomalies in the trend estimation, ~~whereas the uncorrected estimate distorts the trend unintentionally.~~ covariance matrix, while without this a largely erroneous trend is found. For this purpose it is not necessary to know the magnitude of the systematic anomaly. It is only necessary to know which of the data points are affected. We further found that the trend estimate is closer to ~~the true trend, the higher the~~ its true value when higher uncertainties are chosen (diagonal elements of S_y or correlated block in S_y) ~~; not shown~~. This can be explained by the fact that the trend analysis ~~allows additional uncertainties represented by S_y allow~~ the bias to vary as in an optimal estimation scheme ~~where the bias is a fit variable itself~~ [von Clarmann *et al.*, 2001]. When estimating the bias, the ~~higher larger~~ the bias uncertainty is, the less confidence is accounted to the a priori knowledge about the bias (that would be a bias of zero), and the bias is then determined directly from the data as if it would be an additionally fitted variable. Based on our test with the artificial time series we conclude that our method succeeds to handle suspicious anomalies in a time series, leading to an estimated trend close to the trend that would be expected without anomalies.

4.3 GROMOS trends

The GROMOS time series has been used for trend estimations in *Moreira et al.* [2015], who found a significant positive trend in the upper stratosphere. In recent years, GROMOS showed some anomalies leading to ~~higher larger~~ trends than expected in the middle atmosphere, as for example shown by *Steinbrecht et al.* [2017] or *al.* [2018b]. ~~These higher~~ *SPARC/IO3C/GAW* [2019]. ~~These larger~~ trends motivated us to improve GROMOS trend estimates by accounting for the observed anomalies. In the following, we present the trend profiles of the GROMOS time series by considering the detected anomalies in the trend programme with the different correction methods that were introduced in Sect. 4.24.1. In a first step, we estimated the trend without considering ~~inhomogeneities~~ anomalies in the data (case I in Fig. 7). The uncertainties of the data that are used as diagonal elements

in S_y range between 5 and 8 % and are composed as proposed by *Moreira et al.* [2015] (see Sect 4.1). In a second step (case II), we increased the uncertainties (diagonal elements of S_y) ~~as described in Sect. 4.1~~ for the months and altitudes that were detected as anomalies, ~~but still assume an uncorrelated error (case II)~~. ~~In a third step~~, The uncertainties for these anomalous data have been set to a value obtained from the difference to the other datasets ($RD_{debiased}$), as described in Sect. 4.1. Finally, we considered a full-fully correlated block for the periods where anomalies were detected to fit a bias (case III). The bias uncertainty was set to 5 ppm at all altitudes, which ensures that the bias is fully estimated from the data. The diagonal elements of S_y stayed the same as in case II.

The trend profiles in Fig. 7 report an uncorrected GROMOS ozone trend (case I) of about 3 % per decade in the middle stratosphere from 30 to 5 hPa, which decreases above to around -4 % per decade at 0.8 hPa. Correcting the trend by increasing the uncertainty for months with anomalies (case II) decreases the trend slightly in the lower stratosphere, but the differences are small. Using a correlated block in the covariance matrix to estimate the bias of each anomaly in an optimized way (case III) decreases the trend by around 1 % per decade in the lower stratosphere. The trend profile of this optimized trend estimation has a trend of 2.2 % per decade in the lower and middle stratosphere and peaks at approximately 4 hPa with a trend of 2.52.7 % per decade. The corrected trend profile agrees well is consistent with recent satellite-based ozone trends (~~e.g. Steinbrecht et al., 2017~~) ~~between ~ 20 and 40~~ . [e.g. Steinbrecht et al., 2017] in the middle stratosphere. In the upper stratosphere it decreases again, to -2.62.4 % per decade at 0.8 hPa.

All these different GROMOS trend estimates, considering anomalies in different ways, show a significant positive trend of ~ 2.5 % per decade at around 4 hPa (≈ 37 km) altitude, and a trend decrease above. This agreement indicates that the trend at these altitudes is only marginally affected by the identified anomalies and rather robust. We can thus conclude that the trend of around 2.5 % per decade is a sign for an ozone recovery in the altitude range of 35 to 40 km. This result is consistent with trends derived from merged satellite data sets as for example found in *Ball et al.* [2017], *Frith et al.* [2017], *Sofieva and Kyrö* [2017] or *Steinbrecht et al.* [2017], ~~even if, even though~~ the GROMOS trend peak is observed at slightly lower altitudes. A possible reason for this difference in the trend peak altitude might be related to the averaging kernels of the current GROMOS retrieval version. We observe ~~a slight shift in the peak height of the averaging kernels that after the instrument upgrade in~~ 2009, the averaging kernels peak at higher altitudes than expected, indicating that the information is retrieved from slightly higher altitudes (~ 2 to 4 km). ~~Furthermore, one has to keep in mind the coarse altitude resolution of GROMOS, which is at this altitude between 20 to 25.~~ It is therefore possible that the trend peak altitude does not exactly correspond to the true peak altitude. The instrumental upgrade in 2009 led to a change in the averaging kernels. This change, however, should not influence the trend estimates because the data have been harmonized (see Sect. 2.1) and thus corrected for such effects. The harmonization also corrects for possible effects due to changes in the temporal resolution (from 1 h to 30 min).

In the upper stratosphere (above 2 hPa), the GROMOS trend estimates are mostly insignificant but negative. This is probably influenced by the negative trend observed in the mesosphere. A negative ozone trend in the mesosphere is consistent with theory, because increased methane emissions lead to an enhanced ozone loss cycle above 45 km [*Pawson et al., 2014; Brasseur and Solomon, 2005; WMO, 2014*]. However, this is not further investigated in the present study because of the low-small mesospheric measurement response in the GROMOS filter bench data (before 2009).

In the middle and lower stratosphere (below 5 hPa) using different anomaly corrections results in largest trend differences, with trends ranging from 2 to 3 % per decade. This result suggests that the GROMOS anomalies mostly affect these altitudes between 30 and 5 hPa. The corrected GROMOS trend (case III) is not significantly different from zero below ~~1723~~ hPa, ~~and~~ ~~but~~ cases I (uncorrected) and II (corrected by S_y) show significant trends. Compared to other studies, the GROMOS trends in the lower stratosphere are slightly ~~higher~~ larger than trends of most merged satellite data sets, ~~but the corrected GROMOS trend lies within their uncertainty bars.~~

In summary, correcting the GROMOS trend with our anomaly approach leads to a trend profile of ~ 2 to 2.5 % per decade in the lower and middle stratosphere, ~~which agrees.~~ This is consistent with satellite-based trends from recent studies in the middle stratosphere, but is still larger than most satellite trends in the lower stratosphere. The GROMOS trends are almost not affected by anomalies at 4 hPa (≈ 37 km), suggesting a robust ozone recovery of 2.5 % per decade at this altitude. At lower altitudes, trends are more affected by the detected anomalies, and the corrected trend estimate shows a trend of 2 % per decade. ~~This result implies that ozone might also recover at lower altitudes, but the uncertainties, the~~ The larger uncertainties in the lower stratosphere, the dependency on anomalies, and the insignificance of the corrected trend show that this ~~result~~ positive trend in the lower stratosphere is less robust than the trend at higher altitudes.

15 4.4 Influence of temporal sampling on trends

4.3.1 Influence of temporal sampling on trends

Stratospheric ozone at northern mid-latitudes has a strong seasonal cycle of ~ 16 % [Moreira *et al.*, 2016] and a moderate diurnal cycle of 3 to 6 % (Schanz *et al.*, 2014; Studer *et al.*, 2014). The sampling rate of ozone data is thus important for trend estimates, because measurement dependencies on season or time of the day might influence the resulting trends. An important characteristic of microwave radiometers is their measurement continuity, being able to measure during day and night, as well as during almost all weather conditions (except for an opaque atmosphere) and thus during all seasons. Other ground-based instruments such as lidars are temporally more restricted because they typically acquire data during clear-sky nights only. Clear-sky situations are more frequent during high-pressure events which vary with season and location [Steinbrecht *et al.*, 2006; Hatzaki *et al.*, 2014]. The lidar measurements thus do not only depend on the daily cycle, but also on location and season. The seasonal dependency for ozonesondes might be smaller, but they are only launched during daytime.

Figure 8 gives an example of how the measurement sampling rate can influence resulting trends. The GROMOS time series is used for these trend estimates, once using only measurements at the time of ~~ozonesonde launches in Payern~~ lidar measurements, and once only at the time of ~~lidar measurements at MOH~~ ozonesonde launches. The differences to the trend that uses the complete GROMOS sampling are not significant, but still important, especially at \sim between 35 and 40 km and in the lower stratosphere. ~~The lidar sampling leads to higher~~ Using the sampling of the MOH lidar leads to larger trends (~ 3 % per decade) than using the normal sampling (~ 2 % per decade) below 5 hPa. ~~Above 5~~ The OHP lidar sampling, however, leads to smaller trends than the normal sampling above 10 hPa, using ozonesonde sampling results in a higher trend than using. This suggests that selecting different night measurements within a month can lead to trend differences.

All three ozonesonde samplings result in a larger trend than normal or lidar sampling above 5 hPa. Even if ozonesondes are not measuring at those altitudes, the result shows that measuring with an ozonesonde sampling (e.g. only at noon) might influence the trend at these altitudes. Our findings suggest that the time of the measurement (day or night) or the number and the timing of measurements within a month can influence the resulting trend estimates. Results concerning the time dependences of trends based on SOMORA data can be found in *at.* [2018e] *al.* [2019b]. We conclude that sampling differences have to be kept in mind when comparing trend estimates from instruments with different sampling rates or measurement times.

4.4 Influence of time period on trends

4.3.1 Influence of time period on trends

An important factor that influences trend estimates is the length and starting year of the trend period. Several studies have shown that the choice of starting or end point affects substantially the resulting trend [Harris *et al.*, 2015; Weber *et al.*, 2018]. Further, the number of years in the trend period are crucial for the trend estimate [Vyushin *et al.*, 2007; Millá, 2016]. We investigate how the GROMOS trends change when the ~~estimations start~~ regression starts in different years. For this, we average the GROMOS trends over three altitude ranges and determine the trend for periods of different lengths, all ending in Dec. 2017 but starting in different years (Fig. 9). As expected, the uncertainties increase with decreasing period length, and trends starting after 2010 are thus not even shown. Consequently, trends become insignificant for short trend periods. In the lower and middle stratosphere, more than 11 years are needed to detect a significant positive trend (at 95 % confidence level) in the GROMOS data, whereas the 23 years considered are not enough to detect a significant trend in the upper stratosphere (above 3 hPa). Weatherhead *et al.* [2000] and Vyushin *et al.* [2007] state that at least 20 to 30 years are needed to detect a significant trend at mid-latitudes, but their results apply to total column ozone, which can not directly be compared with our ozone profiles. In general we observe that the magnitude of the trend estimates highly depends on the starting year. Furthermore, the trends start to increase in 1997 (middle stratosphere) or 1998 (lower stratosphere), probably due to the turning point in ODS. The later the trend period starts after this turning point, the ~~higher~~ larger the trend estimate is, which has also been observed by Harris *et al.* [2015]. Starting the trend for example in the year 2000, as it is done in other studies [e.g. Steinbrecht *et al.*, 2017; ~~at.~~, 2018b] SPARC/IO3C/GAW, 2019], increases the GROMOS trend by almost 2 % per decade compared to the trend that starts in 1997. ~~The trend dependency~~ The trend magnitudes depend on the starting year of the regression, which is controversial to the definition of a linear trend ~~, which that~~ does not change with time. This ~~suggests~~ illustrates that the true trend might not be linear, or that some interannual variations or anomalies are not captured by the trend model. Nevertheless, our findings demonstrate that it is important to consider the starting year and the trend period length when comparing trend estimates from different instruments or different studies.

30 4.4 Trend comparison

~~Trend~~ The GROMOS trend profile (corrected as described above) and trend profiles for all instruments at the ~~three measurement stations~~ other measurement stations (uncorrected) are shown in Fig. 10. The trend profiles agree on a positive trend in the upper

stratosphere, whereas they differ at lower altitudes. Due to the given uncertainties, most of the trend profiles are not significantly different from zero at 95 % confidence interval. Only GROMOS, the lidars and the ozonesondes at OHP show significant trends in some parts of the stratosphere (bold lines in Fig. 10).

We observe that all instruments that measure in the upper stratosphere show a trend maximum between ~ 4 and $\sim 1.51.8$ hPa (between ~ 37 km and 4543 km), which indicates that ozone is recovering at these altitudes. The trend maxima range from ~ 1 to 3 % per decade, which is comparable with recent, mainly satellite-based ozone trends for northern mid-latitudes e.g. [Ball et al., 2017](#); [Frith et al., 2017](#); [Sofieva and Kyrö, 2017](#); [Steinbrecht et al., 2017](#). (e.g. [Ball et al., 2017](#); [Frith et al., 2017](#); [Sofieva et al., 2017](#)). Only the lidar trend at MOH is larger throughout the whole stratosphere, with ~ 3 % per decade in the middle stratosphere and 4 to 5 % per decade between 5 and 2 hPa. *Nair et al.* [2015] observed similar trend results for the MOH lidar, even if they consider five less years with a trend period ranging from 1997 to 2012. *Steinbrecht et al.* [2006] found that lidar data at MOH and OHP ~~derive deviate~~ from reference satellite data above 35 km before 2003, with ~~lower-less~~ ozone at MOH and ~~higher more~~ ozone at OHP compared to SAGE satellite data (Stratospheric Aerosol and Gas Experiment, *McCormick et al.* [1989]). Opposite drifts are reported by *Eckert et al.* [2014] after 2002 compared to MIPAS satellite data (Michelson Interferometer for Passive Atmospheric Sounding, *Fischer et al.* [2008]). Combining those drifts might explain our ~~high-large~~ MOH trends and ~~lower-smaller~~ OHP trends. The distance of ~ 600 km between the MOH and OHP stations might also explain some differences between the lidar trends. ~~The Furthermore, our sampling results show that the lidar sampling at MOH leads to a larger trend in the lower stratosphere than using a continuous sampling, whereas OHP lidar sampling leads to a lower trend in the middle stratosphere. The large lidar trend at MOH and the comparable low OHP lidar trend might therefore also be partly explained by the sampling rate of the lidars.~~ The GROMOS trend peaks at slightly lower altitudes than the trends of the other instruments. This difference might be related to the averaging kernels of GROMOS, which indicate that GROMOS retrieves information from higher altitudes than expected (~ 2 km difference).

In the middle and lower stratosphere, at altitudes below 5 hPa (≈ 36 km), ~~our trends show discrepant results~~ ~~the estimated trends differ from each other~~. The microwave radiometers and the MOH lidar report trends of 0 to 3 % per decade, and also the ozonesonde at OHP confirms this positive trend. However, the ozonesondes at MOH and Payerne as well as Aura/MLS and the OHP lidar report a trend of around 0 to -2 % per decade below 5 hPa. Some of these observed trend differences can be explained by instrumental changes or differences in processing algorithms and instrument setup. The discrepancy between ozonesonde and lidar trends at OHP, for example, are possibly due to the change of the pressure-temperature radiosonde manufacturer in 2007, which resulted in a step change in bias between ozonesonde and lidar observations. A thorough harmonization would be necessary to correct the trend for this change. The SOMORA trend shows a positive peak at 30 km, which is probably due to homogenization problems that are corrected in the new retrieval version of SOMORA, ~~which is, however, not yet used in our analyses~~ [[al., 2018](#); [eal., 2019b](#)]. Furthermore, we have shown that sampling rates and starting years have an important effect on the resulting trend. ~~Our results show that the lidar sampling at MOH leads to a higher trend in the lower stratosphere than using a continuous sampling. The high lidar trend at MOH might therefore partly be explained by the sampling rate of the lidar.~~ ~~The trend~~ ~~The trend~~ period lengths differ between SOMORA, Aura/MLS and the other data sets, which might also partly explain differences in trend estimates. To explain the remaining trend differences in the lower stratosphere, further corrections

e.g. for anomalies, instrumental changes or sampling rates would be necessary. In brief, trends in the lower stratosphere are not yet clear. For some instruments, significant positive trends are reported, but for many other instruments trends are negative and mostly not significantly different from zero. This result reflects the currently ongoing discussion about lower stratospheric ozone trends [e.g. *Ball et al.*, 2018; *Chipperfield et al.*, 2018; *Stone et al.*, 2018].

5 In summary, our ground-based instrument data agree that ozone is recovering around 3 hPa ($40 \approx 39$ km) with trends ranging between 1 and 3 % per decade for most data sets. In the lower and middle stratosphere between 24 and 37 km (≈ 31 and 4 hPa), however, the trends disagree, suggesting that further research is needed to explain the differences between ground-based trends in the lower stratosphere.

5 Conclusions

10 Our study emphasizes that natural or instrumental anomalies in a time series affect ground-based stratospheric ozone trends. We found that the ozone time series from the GROMOS radiometer (Bern, Switzerland) shows some unexplained anomalies. Accounting for these anomalies in the trend estimation can substantially improve the resulting trends. We further compared different ground-based ozone trend profiles and found an agreement on ozone recovery at around 40 km over Central Europe. At the same time, we observed trend differences ranging between -2 and 3 % per decade at lower altitudes.

15 We compared the GROMOS time series with data from other ground-based instruments in Central Europe and found that they generally agree within ± 10 %. Periods with larger discrepancies have been identified and confirmed to be anomalies in the GROMOS time series. We did not find the origins of these anomalies and assume that they are due to instrumental issues of GROMOS. The identified anomalies have been considered in the GROMOS trend estimations because they can distort the trend. By testing this approach first on a theoretical time series and then with the real GROMOS data, we have shown
20 that identifying anomalies in a time series and considering them in the trend ~~programme~~-analysis makes the resulting trend estimates more accurate. With this method, we propose an approach of advanced trend analysis based on the work of *von Clarmann et al.* [2010] that may also be applied to other ground- or satellite-based data sets to obtain more consistent trend results.

Comparing the GROMOS trend with other ground-based trends in Central Europe suggests that ozone is recovering in the
25 upper stratosphere between around 4 and 1.5 hPa (≈ 37 km and 45 km). This result confirms recent, mainly satellite-based studies. At other altitudes, we have observed contrasting trend estimates. We have shown that the observed differences can partly be explained by different sampling rates and starting years. Other reasons might be instrumental changes or nonconformity in measurement techniques, instrumental systems or processing approaches. Further, the spatial distance between some stations might explain some trend differences, because different air masses can be measured, especially in winter when polar
30 air extends over parts of Europe. Accounting for anomalies in the different data sets as proposed in the present study might be a first step to improve trend estimates. Combined with further corrections, e.g. for sampling rates or instrumental differences, this approach may help to decrease discrepancies between trend estimates from different instruments. In many other studies, the observed trend differences are less apparent because the ground-based data are averaged over latitudinal bands [e.g. *WMO*,

2014; Harris *et al.*, 2015; Steinbrecht *et al.*, 2017]. Nevertheless, it is important to be aware that ground-based trend estimates differ considerably, especially in the lower stratosphere. Exploring the origin of the differences and improving the trend profiles in a similar way as we did for GROMOS may be an important further step on the way to monitor the development of the ozone layer. To summarize, we have shown that anomalies in time series need to be considered when estimating trends.

- 5 Our ground-based results confirm that ozone is recovering in the upper stratosphere above Central Europe, and emphasize the urgency to further investigate lower stratospheric ozone [changes](#). The presented approach to improve trend estimates can help in this endeavour.

Data availability. All ground-based data used in this study are available at <ftp://ftp.cpc.ncep.noaa.gov/ndacc>. The MLS data from the Aura satellite are available at <http://avdc.gsfc.nasa.gov>.

- 10 *Author contributions.* L. Bernet and K. Hocke designed the concept and the methodology. L. Bernet carried out the analysis and prepared the manuscript. T. von Clarmann provided the trend model. All co-authors contributed to the manuscript preparation and the interpretation of the results.

Competing interests. The authors declare that they have no competing interests. T. von Clarmann is co-editor of Atmospheric Chemistry and Physics but he is not involved in the evaluation of this paper.

- 15 *Acknowledgements.* [The authors thank the Aura/MLS team for providing the satellite data.](#) This study has been funded by the Swiss National Science Foundation, grant number 200021-165516.

References

- Bai, K., Chang, N.-B., Shi, R., Yu, H., and Gao, W.: An inter-comparison of multi-decadal observational and reanalysis data sets for global total ozone trends and variability analysis, *J. Geophys. Res. Atmos.*, 122, 1–21, <https://doi.org/10.1002/2016JD025835>, 2017.
- Ball, W. T., Alsing, J., Mortlock, D. J., Rozanov, E. V., Tummon, F., and Haigh, J. D.: Reconciling differences in stratospheric ozone composites, *Atmos. Chem. Phys.*, 17, 12 269–12 302, <https://doi.org/10.5194/acp-2017-142>, 2017.
- Ball, W. T., Alsing, J., Mortlock, D. J., Staehelin, J., Haigh, J. D., Peter, T., Tummon, F., Stübi, R., Stenke, A., Anderson, J., Bourassa, A., Davis, S. M., Degenstein, D., Frith, S. M., Froidevaux, L., Roth, C., Sofieva, V., Wang, R., Wild, J., Yu, P., Ziemke, J. R., and Rozanov, E. V.: Evidence for continuous decline in lower stratospheric ozone offsetting ozone layer recovery, *Atmos. Chem. Phys.*, 18, 1379–1394, <https://doi.org/10.5194/acp-18-1379-2018>, 2018.
- 10 Bates, B. C., Chandler, R. E., and Bowman, A. W.: Trend estimation and change point detection in individual climatic series using flexible regression methods, *J. Geophys. Res. Atmos.*, 117, 1–9, <https://doi.org/10.1029/2011JD017077>, 2012.
- Bourassa, A. E., Roth, C. Z., Zawada, D. J., Rieger, L. A., McLinden, C. A., and Degenstein, D. A.: Drift-corrected Odin-OSIRIS ozone product: Algorithm and updated stratospheric ozone trends, *Atmos. Meas. Tech.*, 11, 489–498, <https://doi.org/10.5194/amt-11-489-2018>, 2018.
- 15 Brasseur, G. P. and Solomon, S.: *Aeronomy of the Middle Atmosphere*, Springer, Dordrecht, The Netherlands, 3 edn., 2005.
- Brewer, A. W. and Milford, J. R.: The Oxford-Kew Ozone Sonde, in: *Proc. R. Soc. Lond. Ser. A, Mathematical and Physical Sciences*, vol. 256, pp. 470–495, Royal Society, <http://www.jstor.org/stable/2413928>, 1960.
- Calisesi, Y.: The stratospheric ozone monitoring radiometer SOMORA: NDSC application document, Research report no. 2003-11, Institute of Applied Physics, University of Bern, Switzerland, 2003.
- 20 Calisesi, Y., Soebijanta, V. T., and van Oss, R.: Regridding of remote soundings: Formulation and application to ozone profile comparison, *J. Geophys. Res.*, 110, D23 306, <https://doi.org/10.1029/2005JD006122>, 2005.
- Chipperfield, M. P., Dhomse, S., Hossaini, R., Feng, W., Santee, M. L., Weber, M., Burrows, J. P., Wild, J. D., Loyola, D., and Coldewey-Egbers, M.: On the Cause of Recent Variations in Lower Stratospheric Ozone, *Geophys. Res. Lett.*, pp. 5718–5726, <https://doi.org/10.1029/2018GL078071>, 2018.
- 25 Chubachi, S.: A special ozone observation at Syowa station, Antarctica from February 1982 to January 1983, in: *Atmos. Ozone, Proc. Quadrenn. Ozone Symp.*, edited by Zerefos, C. S. and Ghazi, A., pp. 285–286, Halkidiki, Greece, 1984.
- Connor, B. J., Parrish, A., and Tsou, J.-J.: Detection of stratospheric ozone trends by ground-based microwave observations, in: *Proc. SPIE 1491, Remote Sens. Atmos. Chem.*, edited by James L. McElroy and McNeal, R. J., vol. 1491, pp. 218–230, International Society for Optics and Photonics, Orlando, FL, United States, <https://doi.org/10.1117/12.46665>, 1991.
- 30 Damadeo, R. P., Zawodny, J. M., Remsberg, E. E., and Walker, K. A.: The Impact of Non-uniform Sampling on Stratospheric Ozone Trends Derived from Occultation Instruments, *Atmos. Chem. Phys.*, 18, 535–554, <https://doi.org/10.5194/acp-2017-575>, 2018.
- Eckert, E., von Clarmann, T., Kiefer, M., Stiller, G. P., Lossow, S., Glatthor, N., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Leblanc, T., McDermid, S., Pastel, M., Steinbrecht, W., Swart, D. P. J., Walker, K. A., and Bernath, P. F.: Drift-corrected trends and periodic variations in MIPAS IMK/IAA ozone measurements, *Atmos. Chem. Phys.*, 14, 2571–2589, <https://doi.org/10.5194/acp-14-2571-2014>, 2014.
- 35 Eriksson, P., Jiménez, C., and Buehler, S. A.: Qpack, a general tool for instrument simulation and retrieval work, *J. Quant. Spectrosc. Radiat. Transf.*, 91, 47–64, <https://doi.org/10.1016/j.jqsrt.2004.05.050>, 2005.

- Eriksson, P., Buehler, S., Davis, C., Emde, C., and Lemke, O.: ARTS, the atmospheric radiative transfer simulator, version 2, *J. Quant. Spectrosc. Radiat. Transf.*, 112, 1551–1558, <https://doi.org/10.1016/j.jqsrt.2011.03.001>, 2011.
- Farman, J. C., Gardiner, B. G., and Shanklin, J. D.: Large losses of total ozone in Antarctica reveal seasonal ClO_x/NO_x interaction, *Nature*, 315, 207–210, <https://doi.org/10.1038/315207a0>, 1985.
- 5 Favaro, G., Jeannot, P., and Stübi, R.: Re-evaluation and trend analysis of the Payerne ozone soundings, Scientific report no. 63, MeteoSwiss, Zürich, Switzerland, 2002.
- Fischer, H., Birk, M., Blom, C., Carli, B., Carlotti, M., Von Clarmann, T., Delbouille, L., Dudhia, A., Ehhalt, D., Endemann, M., Flaud, J. M., Gessner, R., Kleinert, A., Koopman, R., Langen, J., López-Puertas, M., Mosner, P., Nett, H., Oelhaf, H., Perron, G., Remedios, J., Ridolfi, M., Stiller, G., and Zander, R.: MIPAS: An instrument for atmospheric and climate research, *Atmos. Chem. Phys.*, 8, 2151–2188, <https://doi.org/10.5194/acp-8-2151-2008>, 2008.
- 10 Frith, S. M., Stolarski, R. S., Kramarova, N. A., and McPeters, R. D.: Estimating Uncertainties in the SBUV Version 8.6 Merged Profile Ozone Dataset, *Atmos. Chem. Phys.*, 17, 14 695–14 707, <https://doi.org/10.5194/acp-2017-412>, 2017.
- Gaudel, A., Ancellet, G., and Godin-Beekmann, S.: Analysis of 20 years of tropospheric ozone vertical profiles by lidar and ECC at Observatoire de Haute Provence (OHP) at 44°N, 6.7°E, *Atmos. Environ.*, 113, 78–89, <https://doi.org/10.1016/j.atmosenv.2015.04.028>, 2015.
- 15 Godin, S., Mégie, G., and Pelon, J.: Systematic lidar measurements of the stratospheric ozone vertical distribution, *Geophys. Res. Lett.*, 16, 547–550, 1989.
- Godin, S., Carswell, A. I., Donovan, D. P., Claude, H., Steinbrecht, W., McDermid, I. S., McGee, T. J., Gross, M. R., Nakane, H., Swart, D. P., Bergwerff, H. B., Uchino, O., von der Gathen, P., and Neuber, R.: Ozone differential absorption lidar algorithm intercomparison, *Appl. Opt.*, 38, 6225–6236, <https://doi.org/10.1364/AO.38.006225>, 1999.
- 20 Godin-Beekmann, S., Porteneuve, J., and Garnier, A.: Systematic DIAL lidar monitoring of the stratospheric ozone vertical distribution at Observatoire de Haute-Provence (43.92°N, 5.71°E), *J. Environ. Monit.*, 5, 57–67, <https://doi.org/10.1039/b205880d>, 2003.
- Guirlet, M., Keckhut, P., Godin, S., and Mégie, G.: Description of the long-term ozone data series obtained from different instrumental techniques at a single location: The Observatoire de Haute-Provence (43.9°N, 5.7°E), *Ann. Geophys.*, 18, 1325–1339, <https://doi.org/10.1007/s00585-000-1325-y>, 2000.
- 25 Harris, N. R. P., Hudson, R., and Phillips, C.: SPARC/IOC/GAW, Assessment of trends in the vertical distribution of ozone, SPARC Report No. 1, WMO Ozone Research and Monitoring Project Report No. 43, Geneva, Switzerland, 1998.
- Harris, N. R. P., Hassler, B., Tummon, F., Bodeker, G. E., Hubert, D., Petropavlovskikh, I., Steinbrecht, W., Anderson, J., Bhartia, P. K., Boone, C. D., Bourassa, A., Davis, S. M., Degenstein, D., Delcloo, A., Frith, S. M., Froidevaux, L., Godin-Beekmann, S., Jones, N., Kurylo, M. J., Kyrölä, E., Laine, M., Leblanc, S. T., Lambert, J.-C., Liley, B., Mahieu, E., Maycock, A., de Mazière, M., Parrish, A. D., Querel, R., Rosenlof, K. H., Roth, C., Sioris, C., Staehelin, J., Stolarski, R. S., Stübi, R., Tamminen, J., Vigouroux, C., Walker, K. A., Wang, H. J., Wild, J., and Zawodny, J. M.: Past changes in the vertical distribution of ozone - Part 3: Analysis and interpretation of trends, *Atmos. Chem. Phys.*, 15, 9965–9982, <https://doi.org/10.5194/acp-15-9965-2015>, 2015.
- Hatzaki, M., Flocas, H. A., Simmonds, I., Kouroutzoglou, J., Keay, K., and Rudeva, I.: Seasonal aspects of an objective climatology of anticyclones affecting the mediterranean, *J. Clim.*, 27, 9272–9289, <https://doi.org/10.1175/JCLI-D-14-00186.1>, 2014.
- 35 Hubert, D., Lambert, J.-C., Verhoelst, T., Granville, J., Keppens, A., Baray, J. L., Bourassa, A. E., Cortesi, U., Degenstein, D. A., Froidevaux, L., Godin-Beekmann, S., Hoppel, K. W., Johnson, B. J., Kyrölä, E., Leblanc, T., Lichtenberg, G., Marchand, M., McElroy, C. T., Murtagh, D., Nakane, H., Portafaix, T., Querel, R., Russell, J. M., Salvador, J., Smit, H. G. J., Stebel, K., Steinbrecht, W., Strawbridge, K. B., Stübi, R., Swart, D. P. J., Taha, G., Tarasick, D. W., Thompson, A. M., Urban, J., Van Gijsel, J. A. E., Van Malderen, R., Von Der Gathen, P.,

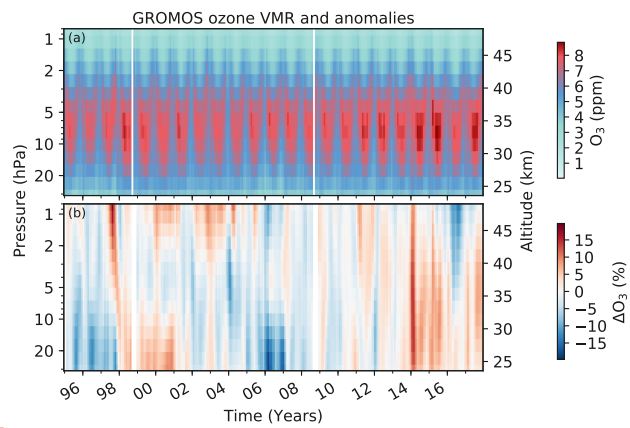
- Walker, K. A., Wolfram, E., and Zawodny, J. M.: Ground-based assessment of the bias and long-term stability of 14 limb and occultation ozone profile data records, *Atmos. Meas. Tech.*, 9, 2497–2534, <https://doi.org/10.5194/amt-9-2497-2016>, 2016.
- Hubert et al.: The temporal and spatial homogeneity of ozone profile data records obtained by ozonesonde, lidar and microwave radiometer networks, in preparation, 2019.
- 5 Jeannot, P., Stübi, R., Levrat, G., Viatte, P., and Staehelin, J.: Ozone balloon soundings at Payerne (Switzerland): Reevaluation of the time series 1967-2002 and trend analysis, *J. Geophys. Res. Atmos.*, 112, <https://doi.org/10.1029/2005JD006862>, 2007.
- Jones, A., Urban, J., Murtagh, D. P., Sanchez, C., Walker, K. A., Livesey, N. J., Froidevaux, L., and Santee, M. L.: Analysis of HCl and ClO time series in the upper stratosphere using satellite data sets, *Atmos. Chem. Phys.*, 11, 5321–5333, <https://doi.org/10.5194/acp-11-5321-2011>, 2011.
- 10 Kämpfer, N.: Microwave remote sensing of the atmosphere in Switzerland, *Opt. Eng.*, 34, 2413–2424, 1995.
- Keppens, A., Lambert, J. C., Granville, J., Miles, G., Siddans, R., Van Peet, J. C., Van Der A, R. J., Hubert, D., Verhoelst, T., Delcloo, A., Godin-Beekmann, S., Kivi, R., Stübi, R., and Zehner, C.: Round-robin evaluation of nadir ozone profile retrievals: Methodology and application to MetOp-A GOME-2, *Atmos. Meas. Tech.*, 8, 2093–2120, <https://doi.org/10.5194/amt-8-2093-2015>, 2015.
- Komhyr, W. D.: Electrochemical concentration cells for gas analysis, *Ann. Géophysique*, 25, 203–210, 1969.
- 15 Kuttippurath, J. and Nair, P. J.: The signs of Antarctic ozone hole recovery, *Sci. Rep.*, 7, <https://doi.org/10.1038/s41598-017-00722-7>, 2017.
- Maillard Barras, E., Ruffieux, D., and Hocke, K.: Stratospheric ozone profiles over Switzerland measured by SOMORA, ozonesonde and MLS/AURA satellite, *Int. J. Remote Sens.*, 30, 4033–4041, <https://doi.org/10.1080/01431160902821890>, 2009.
- Maillard Barras, E., Haeefe, A., Stübi, R., and Ruffieux, D.: A method to derive the Site Atmospheric State Best Estimate (SASBE) of ozone profiles from radiosonde and passive microwave data, *Atmos. Meas. Tech. Discuss.*, pp. 3399–3422, [https://doi.org/10.5194/amtd-8-3399-](https://doi.org/10.5194/amtd-8-3399-2015)
- 20 2015, 2015.
- Maillard Barras et al.: Trends of stratospheric ozone daily cycle, in preparation, 2019.
- McCormick, M., Zawodny, J., Veiga, R., Larsen, J., and Wang, P.: An overview of SAGE I and II ozone measurements, *Planet. Space Sci.*, 37, 1567–1586, [https://doi.org/10.1016/0032-0633\(89\)90146-3](https://doi.org/10.1016/0032-0633(89)90146-3), 1989.
- Millán, L. F., Livesey, N. J., Santee, M. L., Neu, J. L., Manney, G. L., and Fuller, R. A.: Case Studies of the Impact of Orbital Sampling on
- 25 Stratospheric Trend Detection and Derivation of Tropical Vertical Velocities: Solar Occultation versus Limb Emission Sounding, *Atmos. Chem. Phys.*, 16, 11 521–11 534, <https://doi.org/10.5194/acp-2016-356>, 2016.
- Molina, M. J. and Rowland, F. S.: Stratospheric sink for chlorofluoromethanes: chlorine atom-catalysed destruction of ozone, *Nature*, 249, 810–812, <https://doi.org/10.1038/249810a0>, 1974.
- Moreira, L.: Trends and oscillations in middle atmospheric ozone observed by a ground-based microwave radiometer, Ph.d. thesis, University
- 30 of Bern, Switzerland, 2017.
- Moreira, L., Hocke, K., Eckert, E., Von Clarmann, T., and Kämpfer, N.: Trend analysis of the 20-year time series of stratospheric ozone profiles observed by the GROMOS microwave radiometer at Bern, *Atmos. Chem. Phys.*, 15, 10 999–11 009, <https://doi.org/10.5194/acp-15-10999-2015>, 2015.
- Moreira, L., Hocke, K., Navas-Guzmán, F., Eckert, E., von Clarmann, T., and Kämpfer, N.: The natural oscillations in stratospheric
- 35 ozone observed by the GROMOS microwave radiometer at the NDACC station Bern, *Atmos. Chem. Phys.*, 16, 10 455–10 467, <https://doi.org/10.5194/acp-2016-66>, 2016.
- Moreira, L., Hocke, K., and Kämpfer, N.: Comparison of ozone profiles and influences from the tertiary ozone maximum in the night-to-day ratio above Switzerland, *Atmos. Chem. Phys.*, 17, 10 259–10 268, <https://doi.org/10.5194/acp-2017-274>, 2017.

- Nair, P. J., Godin-Beekmann, S., Pazmiño, A., Hauchecorne, A., Ancellet, G., Petropavlovskikh, I., Flynn, L. E., and Froidevaux, L.: Coherence of long-term stratospheric ozone vertical distribution time series used for the study of ozone recovery at a northern mid-latitude station, *Atmos. Chem. Phys.*, 11, 4957–4975, <https://doi.org/10.5194/acp-11-4957-2011>, 2011.
- 5 Nair, P. J., Godin-Beekmann, S., Froidevaux, L., Flynn, L. E., Zawodny, J. M., Russell, J. M., Pazmiño, A., Ancellet, G., Steinbrecht, W., Claude, H., Leblanc, T., McDermid, S., Van Gijssel, J. A. E., Johnson, B., Thomas, A., Hubert, D., Lambert, J.-C., Nakane, H., and Swart, D. P. J.: Relative drifts and stability of satellite and ground-based stratospheric ozone profiles at NDACC lidar stations, *Atmos. Meas. Tech.*, 5, 1301–1318, <https://doi.org/10.5194/amt-5-1301-2012>, 2012.
- 10 Nair, P. J., Godin-Beekmann, S., Kuttippurath, J., Ancellet, G., Goutail, F., Pazmiño, A., Froidevaux, L., Zawodny, J. M., Evans, R. D., Wang, H. J., Anderson, J., and Pastel, M.: Ozone trends derived from the total column and vertical profiles at a northern mid-latitude station, *Atmos. Chem. Phys.*, 13, 10373–10384, <https://doi.org/10.5194/acp-13-10373-2013>, 2013.
- Nair, P. J., Froidevaux, L., Kuttippurath, J., Zawodny, J. M., Russell, J. M., Steinbrecht, W., Claude, H., Leblanc, T., Van Gijssel, J. A., Johnson, B., Swart, D. P., Thomas, A., Querel, R., Wang, R., and Anderson, J.: Subtropical and midlatitude ozone trends in the stratosphere: Implications for recovery, *J. Geophys. Res. Atmos.*, 120, 7247–7257, <https://doi.org/10.1002/2014JD022371>, 2015.
- 15 National Research Council of Canada: Latest Solar Radio Flux Report from DRAO, Penticton, <http://www.spaceweather.ca/solarflux/sx-4-en.php>, last access: 12 March 2019, 2019.
- NDACC: Network for the Detection of Atmospheric Composition Change, <http://www.ndaccdemo.org>, last access: 12 March 2019, 2019.
- Newchurch, M. J., Yang, E.-S., Cunnold, D. M., Reinsel, G. C., Zawodny, J. M., and Russell, J. M.: Evidence for slowdown in stratospheric ozone loss: First stage of ozone recovery, *J. Geophys. Res. Atmos.*, 108, <https://doi.org/10.1029/2003JD003471>, 2003.
- 20 Parkinson, C. L., Ward, A., and King, M. D., eds.: *Earth Science Reference Handbook - A Guide to NASA's Earth Science Program and Earth Observing Satellite Missions*, National Aeronautics and Space Administration, Washington, D. C., 2006.
- Pazmiño, A., Godin-beekmann, S., Hauchecorne, A., Claud, C., Khaykin, S., Goutail, F., Wolfram, E., Salvador, J., and Quel, E.: Multiple symptoms of total ozone recovery inside the Antarctic vortex during austral spring, *Atmos. Chem. Phys.*, 18, 7557–7572, 2018.
- Peter, R.: *The Ground-based Millimeter-wave Ozone Spectrometer - GROMOS*, Research report no. 97-13, Institute of Applied Physics, University of Bern, Switzerland, 1997.
- 25 Ramaswamy, V., Chanin, M.-L., Angell, J., Barnett, J., Gaffen, D., Gelman, M., Keckhut, P., Koshelkov, Y., Labitzke, K., Lin, J.-J. R., O'Neill, A., Nash, J., Randel, W., Rood, R., Shine, K., Shiotani, M., and Swinbank, R.: Stratospheric temperature trends: Observations and model simulations, *Rev. Geophys.*, 39, 71–122, <https://doi.org/10.1029/1999RG000065>, 2001.
- Reinsel, G. C., Miller, A. J., Weatherhead, E. C., Flynn, L. E., Nagatani, R. M., Tiao, G. C., and Wuebbles, D. J.: Trend analysis of total ozone data for turnaround and dynamical contributions, *J. Geophys. Res. D Atmos.*, 110, <https://doi.org/10.1029/2004JD004662>, 2005.
- 30 Rodgers, C. D.: *Inverse Methods for Atmospheric Sounding: Theory and Practice*, World Scientific Publishing Co. Pte. Ltd, Singapore, 2000.
- Rodgers, C. D. and Connor, B. J.: Intercomparison of remote sounding instruments, *J. Geophys. Res.*, 108, 46–48, <https://doi.org/10.1029/2002JD002299>, 2003.
- Schanz, A., Hocke, K., and Kämpfer, N.: Daily ozone cycle in the stratosphere: Global, regional and seasonal behaviour modelled with the Whole Atmosphere Community Climate Model, *Atmos. Chem. Phys.*, 14, 7645–7663, <https://doi.org/10.5194/acp-14-7645-2014>, 2014.
- 35 Schotland, R. M.: Errors in the Lidar Measurement of Atmospheric Gases by Differential Absorption, *J. Appl. Meteorol.*, 13, 71–77, [https://doi.org/10.1175/1520-0450\(1974\)013<0071:EITLMO>2.0.CO;2](https://doi.org/10.1175/1520-0450(1974)013<0071:EITLMO>2.0.CO;2), 1974.

- Shepherd, T. G., Plummer, D. A., Scinocca, J. F., Hegglin, M. I., Fioletov, V. E., Reader, M. C., Remsberg, E., von Clarmann, T., and Wang, H. J.: Reconciliation of halogen-induced ozone loss with the total-column ozone record, *Nat. Geosci.*, 7, 443–449, <https://doi.org/10.1038/ngeo2155>, 2014.
- Smit, H. G. J. and ASOPOS Panel: Quality Assurance and Quality Control for Ozone Sonde Measurements in GAW, World Meteorological Organization, Global Atmosphere Watch, GAW Report No. 201, Geneva, Switzerland, 2013.
- 5 Smit, H. G. J. and Kley, D.: Jülich Ozone Sonde Intercomparison Experiment (JOSIE), World Meteorological Organization, Global Atmosphere Watch, Report No. 130, 1996.
- Sofieva, V. F., Kyrölä, E., Laine, M., Tamminen, J., Degenstein, D., Bourassa, A., Roth, C., Zawada, D., Weber, M., Rozanov, A., Rahpoe, N., Stiller, G. P., Laeng, A., von Clarmann, T., Walker, K. A., Sheese, P., Hubert, D., van Roozendaal, M., Zehner, C., Damadeo, R. P., Zawodny, J., Kramarova, N., and Bhartia, P. K.: Merged SAGE II, Ozone_cci and OMPS ozone profiles dataset and evaluation of ozone trends in the stratosphere, *Atmos. Chem. Phys.*, 17, 12 533–12 552, <https://doi.org/10.5194/acp-2017-598>, 2017.
- 10 Solomon, S., Ivy, D. J., Kinnison, D., Mills, M. J., Neely, R. R., and Schmidt, A.: Emergence of healing in the Antarctic ozone layer, *Science*, 353, 269–274, <https://doi.org/10.1126/science.aae0061>, 2016.
- SPARC/IO3C/GAW: SPARC/IO3C/GAW Report on Long-term Ozone Trends and Uncertainties in the Stratosphere, Petropavlovskikh, I., Godin-Beekmann, S., Hubert, D., Damadeo, R., Hassler, B., Sofieva, V. (eds.), SPARC Report No. 9, GAW Report No. 241, WCRP-17/2018, <https://doi.org/10.17874/f899e57a20b>, www.sparc-climate.org/publications/sparc-reports, 2019.
- 15 Steinbrecht, W., Claude, H., Schöenborn, F., McDermid, I. S., Leblanc, T., Godin, S., Song, T., Swart, D. P. J., Meijer, Y. J., Bodeker, G. E., Connor, B. J., Kämpfer, N., Hocke, K., Calisesi, Y., Schneider, N., de la Noë, J., Parrish, A. D., Boyd, I. S., Brühl, C., Steil, B., Giorgetta, M. A., Manzini, E., Thomason, L. W., Zawodny, J. M., McCormick, M. P., Russell, J. M., Bhartia, P. K., Stolarski, R. S., and Hollandsworth-Frith, S. M.: Long-term evolution of upper stratospheric ozone at selected stations of the Network for the Detection of Stratospheric Change (NDSC), *J. Geophys. Res. Atmos.*, 111, <https://doi.org/10.1029/2005JD006454>, 2006.
- 20 Steinbrecht, W., Claude, H., Schöenborn, F., McDermid, I. S., Leblanc, T., Godin-Beekmann, S., Keckhut, P., Hauchecorne, A., Van Gijssel, J. A. E., Swart, D. P. J., Bodeker, G. E., Parrish, A. D., Boyd, I. S., Kämpfer, N., Hocke, K., Stolarski, R. S., Frith, S. M., Thomason, L. W., Remsberg, E. E., von Savigny, C., Rozanov, A., and Burrows, J. P.: Ozone and temperature trends in the upper stratosphere at five stations of the Network for the Detection of Atmospheric Composition Change, *Int. J. Remote Sens.*, 30, 3875–3886, <https://doi.org/10.1080/01431160902821841>, 2009a.
- 25 Steinbrecht, W., McGee, T. J., Twigg, L. W., Claude, H., Schöenborn, F., Sumnicht, G. K., and Silbert, D.: Intercomparison of stratospheric ozone and temperature profiles during the October 2005 Hohenpeißenberg Ozone Profiling Experiment (HOPE), *Atmos. Meas. Tech.*, 2, 125–145, <https://doi.org/10.5194/amt-2-125-2009>, 2009b.
- 30 Steinbrecht, W., Köhler, U., Schöenborn, F., Adelwart, M., Petersen, D., Heinen, M., Munier, H., Claude, H., Steiner, S., and Attmannspacher, W.: 50 Years of Routine Ozone Soundings at Hohenpeissenberg, Quadrennial Ozone Symposium of the International Ozone Commission, Edinburgh, United Kingdom, 4–9 September 2016, QOS2016-28-1, 2016.
- Steinbrecht, W., Froidevaux, L., Fuller, R., Wang, R., Anderson, J., Roth, C., Bourassa, A., Degenstein, D., Damadeo, R., Zawodny, J., Frith, S. M., McPeters, R., Bhartia, P., Wild, J., Long, C., Davis, S., Rosenlof, K., Sofieva, V., Walker, K., Rahpoe, N., Rozanov, A., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Kramarova, N., Godin-Beekmann, S., Leblanc, T., Querel, R., Swart, D., Boyd, I., Hocke, K., Kämpfer, N., Maillard Barras, E., Moreira, L., Nedoluha, G., Vigouroux, C., Blumenstock, T., Schneider, M., García, O., Jones, N., Mahieu, E., Smale, D., Kotkamp, M., Robinson, J., Petropavlovskikh, I., Harris, N., Hassler, B., Hubert, D., and Tummon, F.: An update
- 35

- on ozone profile trends for the period 2000 to 2016, *Atmos. Chem. Phys.*, 17, 10 675–10 690, <https://doi.org/10.5194/acp-17-10675-2017>, 2017.
- 5 Stiller, G. P., von Clarmann, T., Haanel, F., Funke, B., Glatthor, N., Grabowski, U., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., and López-Puertas, M.: Observed temporal evolution of global mean age of stratospheric air for the 2002 to 2010 period, *Atmos. Chem. Phys.*, 12, 3311–3331, <https://doi.org/10.5194/acp-12-3311-2012>, 2012.
- Stolarski, R. S. and Frith, S. M.: Search for evidence of trend slow-down in the long-term TOMS/SBUV total ozone data record: the importance of instrument drift uncertainty, *Atmos. Chem. Phys.*, 6, 4057–4065, <https://doi.org/10.5194/acpd-6-3883-2006>, 2006.
- Stone, K. A., Solomon, S., and Kinnison, D. E.: On the Identification of Ozone Recovery, *Geophys. Res. Lett.*, 45, 5158–5165, <https://doi.org/10.1029/2018GL077955>, 2018.
- 10 Strahan, S. E. and Douglass, A. R.: Decline in Antarctic Ozone Depletion and Lower Stratospheric Chlorine Determined From Aura Microwave Limb Sounder Observations, *Geophys. Res. Lett.*, 45, 382–390, <https://doi.org/10.1002/2017GL074830>, 2018.
- Studer, S., Hocke, K., Schanz, A., Schmidt, H., and Kämpfer, N.: A climatology of the diurnal variations of stratospheric and mesospheric ozone over Bern, Switzerland, *Atmos. Chem. Phys.*, 14, 5905–5919, <https://doi.org/10.5194/acpd-13-22445-2013>, 2014.
- Tsou, J. J., Connor, B. J., Parrish, A. D., McDerimid, I. S., and Chu, W. P.: Ground-based microwave monitoring of middle atmosphere ozone: Comparison to lidar and Stratospheric and Gas Experiment II satellite observations, *J. Geo.*, 100, 3005–3016, 1995.
- 15 von Clarmann, T., Grabowski, U., and Kiefer, M.: On the role of non-random errors in inverse problems in radiative transfer and other applications, *J. Quant. Spectrosc. Radiat. Transf.*, 71, 39–46, [https://doi.org/10.1016/S0022-4073\(01\)00010-3](https://doi.org/10.1016/S0022-4073(01)00010-3), 2001.
- von Clarmann, T., Stiller, G. P., Grabowski, U., Eckert, E., and Orphal, J.: Technical note: Trend estimation from irregularly sampled, correlated data, *Atmos. Chem. Phys.*, 10, 6737–6747, <https://doi.org/10.5194/acp-10-6737-2010>, 2010.
- 20 Vyushin, D. I., Fioletov, V. E., and Shepherd, T. G.: Impact of long-range correlations on trend detection in total ozone, *J. Geophys. Res.*, 112, <https://doi.org/10.1029/2006JD008168>, 2007.
- Wargan, K., Orbe, C., Pawson, S., Ziemke, J. R., Oman, L. D., Olsen, M. A., Coy, L., and Emma Knowland, K.: Recent Decline in Extratropical Lower Stratospheric Ozone Attributed to Circulation Changes, *Geophys. Res. Lett.*, 45, 5166–5176, <https://doi.org/10.1029/2018GL077406>, 2018.
- 25 Waters, J. W., Froidevaux, L., Harwood, R. S., Jarnot, R. F., Pickett, H. M., Read, W. G., Siegel, P. H., Cofield, R. E., Filipiak, M. J., Flower, D. A., Holden, J. R., Lau, G. K., Livesey, N. J., Manney, G. L., Pumphrey, H. C., Santee, M. L., Wu, D. L., Cuddy, D. T., Lay, R. R., Loo, M. S., Perun, V. S., Schwartz, M. J., Stek, P. C., Thurstans, R. P., Boyles, M. A., Chandra, K. M., Chavez, M. C., Chen, G. S., Chudasama, B. V., Dodge, R., Fuller, R. A., Girard, M. A., Jiang, J. H., Jiang, Y., Knosp, B. W., Labelle, R. C., Lam, J. C., Lee, K. A., Miller, D., Oswald, J. E., Patel, N. C., Pukala, D. M., Quintero, O., Scaff, D. M., Van Snyder, W., Tope, M. C., Wagner, P. A., and Walch, M. J.: The Earth Observing System Microwave Limb Sounder (EOS MLS) on the aura satellite, *IEEE Trans. Geosci. Remote Sens.*, 44, 1075–1092, <https://doi.org/10.1109/TGRS.2006.873771>, 2006.
- 30 Weatherhead, E. C., Reinsel, G. C., Tiao, G. C., Jackman, C. H., Bishop, L., Hollandsworth, S. M., Deluisi, J., Keller, T., Oltmans, S. J., Fleming, E. L., Wuebbles, D. J., Kerr, J. B., Miller, A. J., Herman, J., McPeters, R., Nagatani, R. M., and Frederick, J. E.: Detecting the recovery of total column ozone, 105, 201–210, <https://doi.org/https://doi.org/10.1029/2000JD900063>, 2000.
- Weber, M., Coldewey-Egbers, M., Fioletov, V. E., Frith, S. M., Wild, J. D., Burrows, J. P., Long, C. S., and Loyola, D.: Total ozone trends from 1979 to 2016 derived from five merged observational datasets – the emergence into ozone recovery, *Atmos. Chem. Phys.*, 18, 2097–2117, <https://doi.org/10.5194/acp-2017-853>, 2018.

- Werner, J., Rothe, K. W., and Walther, H.: Monitoring of the stratospheric ozone layer by laser radar, *Appl. Phys. B Photophysics Laser Chem.*, 32, 113–118, <https://doi.org/10.1007/BF00688815>, 1983.
- WMO: Scientific Assessment of Ozone Depletion: 2014, p. 416, World Meteorological Organization, Global Ozone Research and Monitoring Project - Report No. 55, Geneva, Switzerland, 2014.
- 5 WMO: Scientific Assessment of Ozone Depletion: 2018, p. 588, World Meteorological Organization, Global Ozone Research and Monitoring Project - Report No. 58, Geneva, Switzerland, 2018.
- Wolter, K. and Timlin, M. S.: Measuring the strength of ENSO events: How does 1997/98 rank?, *Weather*, 53, 315–324, <https://doi.org/10.1002/j.1477-8696.1998.tb06408.x>, 1998.
- Zanis, P., Maillard, E., Stachelin, J., Zerefos, C., Kosmidis, E., Tourpali, K., and Wohltmann, I.: On the turnaround of stratospheric ozone trends deduced from the reevaluated Umkehr record of Arosa, Switzerland, *J. Geophys. Res. Atmos.*, 111, <https://doi.org/10.1029/2005JD006886>, 2006.
- 10



figures/fig01

Figure 1. (a) Monthly means of ozone volume mixing ratio (VMR) measured by GROMOS (Ground-based Millimetre-wave Ozone Spectrometer) at Bern from Jan. 1995 to Dec. 2017. The white stripes lines indicate months with missing data where no measurements were available due to instrumental issues. (b) Deviation from GROMOS multi-year-monthly means (ozone mean climatology). The anomalies in (b) 1997 to 2017 are smoothed by a moving average median window of 3 months.

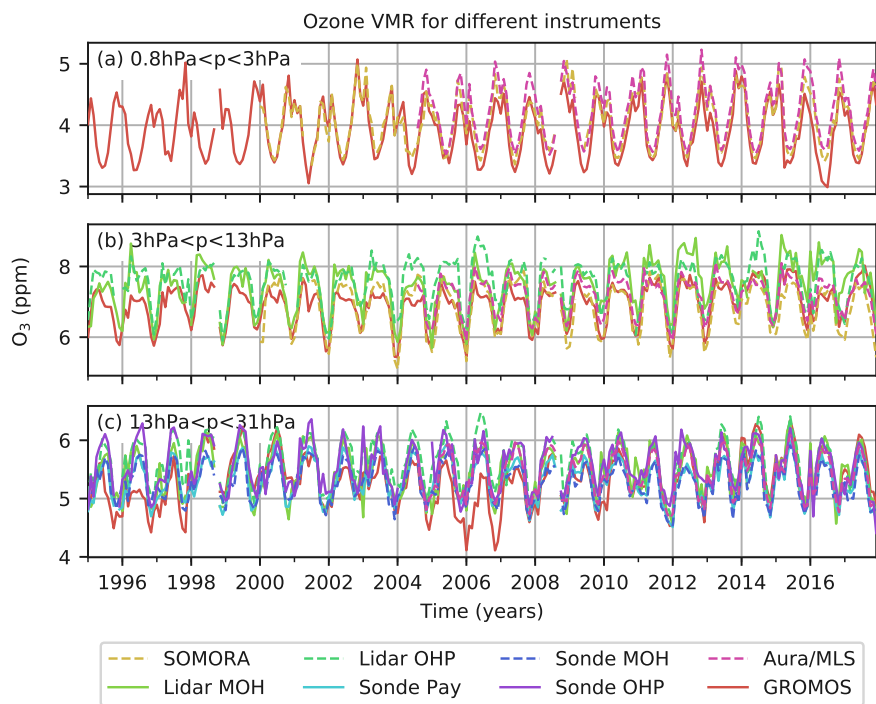


Figure 2. Monthly means of ozone VMR from the microwave radiometers GROMOS (Bern) and SOMORA (Payerne), the lidars at the observatories of Hohenpeissenberg (MOH) and Haute Provence (OHP), the ozonesonde measurements at MOH, OHP, and Payerne, as well as Aura/MLS data above Bern. The data have been averaged over three altitude ranges.

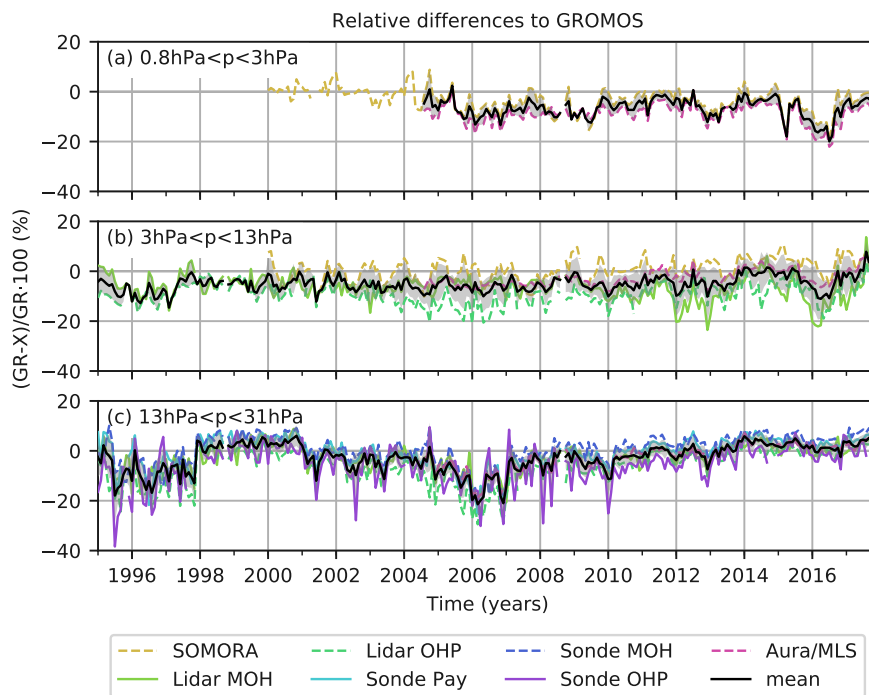


Figure 3. Monthly means of relative differences (RD) of monthly means between GROMOS and coincident pairs of SOMORA, lidars (MOH, OHP), ozonesondes (Payerne, MOH, OHP) and Aura/MLS, averaged over three altitude ranges. The relative difference RD is calculated given by $(\text{GROMOS} - X) / \text{GROMOS} \cdot 100$ ($(GR - X) / GR \cdot 100$), where GR represents GROMOS monthly means and X represents monthly means of the other data sets. The black lines show the mean values of all differences RDs and the grey shaded areas shows its standard deviation.

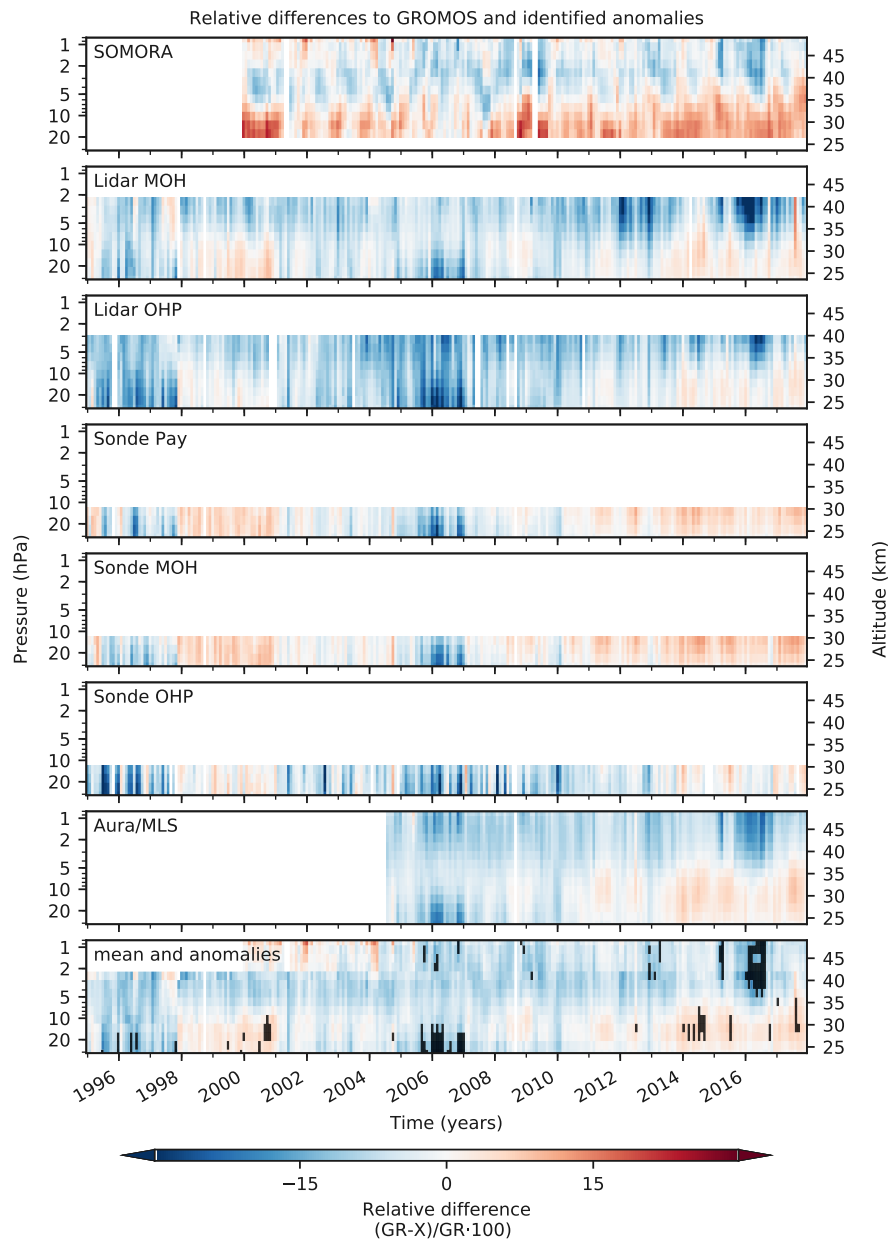


Figure 4. Monthly means of relative differences (RD) of monthly means between GROMOS and SOMORA, lidars (MOH and OHP), ozonesondes (Payerne, MOH, OHP) and Aura/MLS. The lowest panel shows the mean of all relative differences RDs. The black areas in the lowest panel show periods where at least three data sets (or two data sets above 2 hPa) have a debiased relative bias-of-difference ($RD_{debiased}$) larger than 10%.

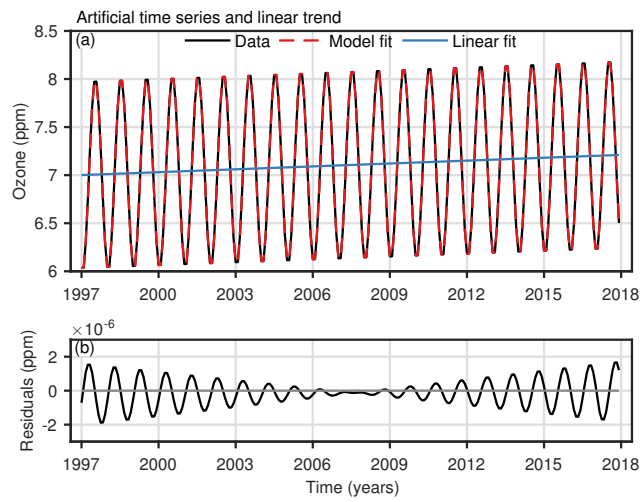


Figure 5. (a) Artificial ozone time series (composed of a simple seasonal cycle and a linear trend) and its model fit and linear trend estimation. (b) Trend model residuals (data – model fit).

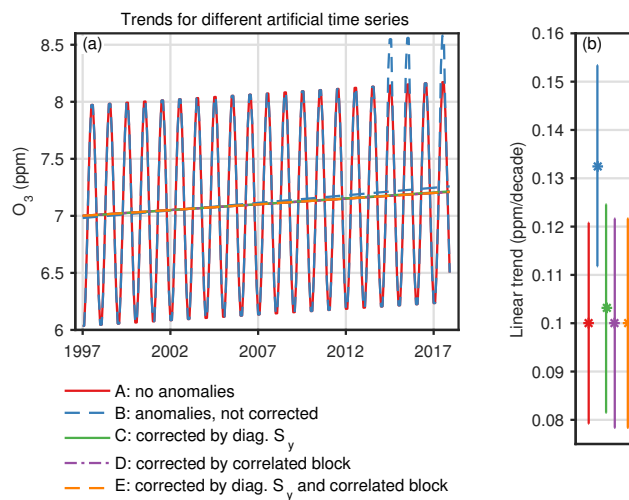


Figure 6. (a) Ozone time series and linear trends for the same artificial ozone time series as in Fig. 5 (case A). Anomalies were added to the time series in the summer months 2014, 2015 and 2017 in case B. Different corrections have been applied to account for those anomalies in the trend fit, represented by the cases C, D, and to E. (b) Linear trend estimates for the time series without anomalies (case A) and the time series with added anomalies with different trend estimates, considering where the anomalies were considered in different ways (cases C, D, E) or not considering considered at all (case B) the anomalies. For cases A, D, and E the trend is approximately the same (~ 0.1) and therefore almost not discernible in panel (b).

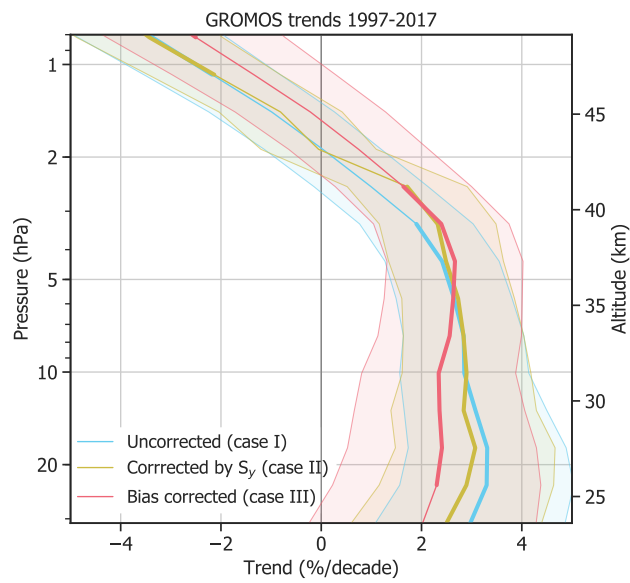


Figure 7. GROMOS ozone trends from Jan. 1997 to Dec. 2017, without considering anomalies (case I), considering anomalies in the diagonal elements of S_y (case II), and considering a correlated bias block for anomalies (case III). The shaded areas show two standard deviations (σ), whereas the bold lines represent altitudes where the trend profile is significantly different from zero at 95 % confidence interval ($|\text{trend}| > 2\sigma$).

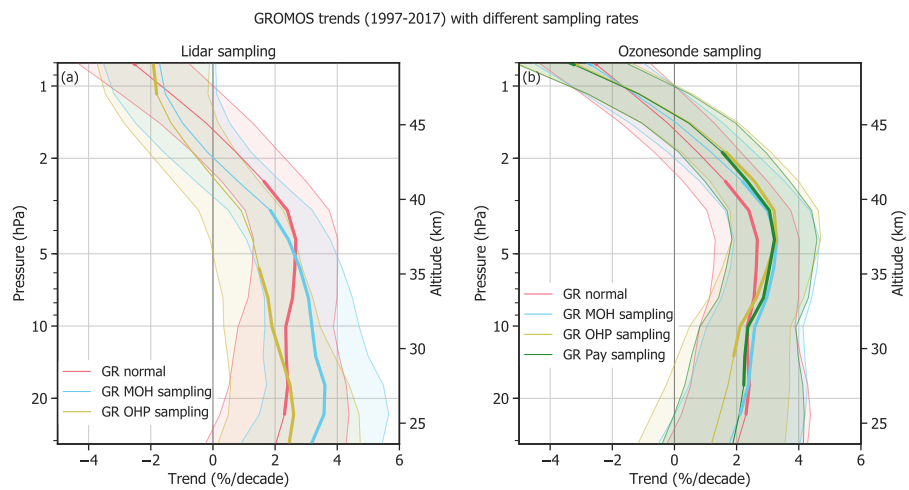


Figure 8. GROMOS (GR) trends from Jan. 1997 to Dec. 2017 using its high temporal resolution (normal data set with correction for anomalies), as well as the temporal sampling rate of the MOH lidar different lidars (a) and the Payerne ozonesonde ozonesondes (b). The shaded areas show the $2-2\sigma$ uncertainty and bold lines identify trends that are significantly different from zero.

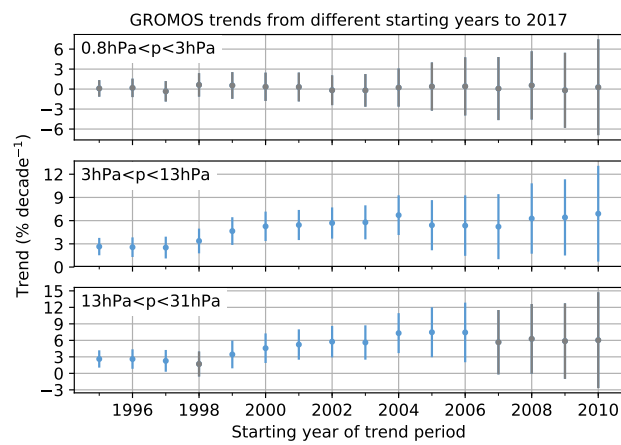


Figure 9. GROMOS trends averaged over three altitude ranges starting in different years, all ending in Dec. 2017. [The error bars show the \$2\sigma\$ uncertainties.](#) Trend estimates that are not significantly different from zero at 95 % confidence interval are shown in grey.

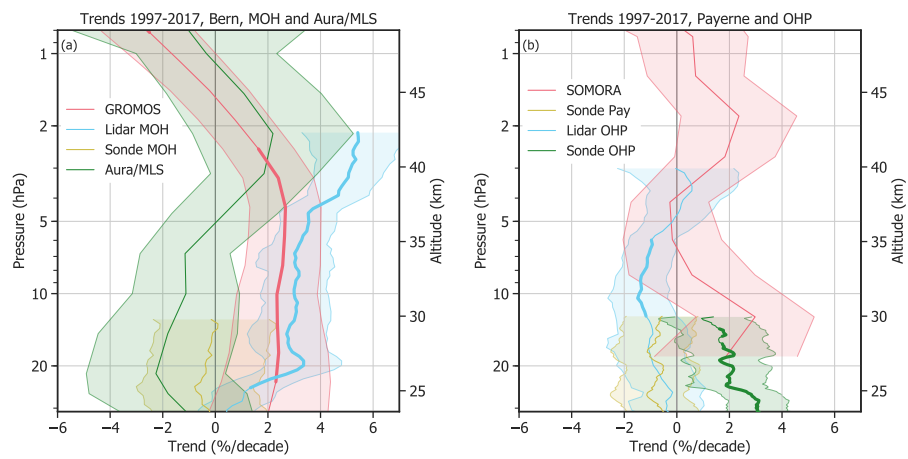


Figure 10. Ozone trends of different ground-based instruments in Central Europe and Aura/MLS (over Bern, Switzerland). The $2\text{-}\sigma$ uncertainty is shown by shaded areas. Bold lines indicate trends that are significantly different from zero (at 95 % confidence interval). The trends at Bern and MOH as well as the Aura/MLS trend are shown in (a), whereas (b) shows the trends from datasets at Payerne and OHP.

Table 1. Information about measurement stations, instruments, and data used in the present study.

Station	Instrument	Altitude range	Measurement rate	Analysis period
Bern, Switzerland 46.95 ° N, 7.44 ° E, 574 m	GROMOS	31–0.8 hPa	30 min to 1 h	01/1995–12/2017
Payerne, Switzerland 46.8° N, 7.0° E, 491 m	SOMORA	31–0.8 hPa	30 min to 1 h	01/2000–12/2017
	Ozonesonde	24–30 km	13 profiles/month ^a	01/1995–12/2017
Hohenpeissenberg, Germany 47.8° N, 11.0° E, 980 m	Lidar	24–42.3 km	8 profiles/month ^a	01/1995–12/ 2017
	Ozonesonde	24–30 km	10 profiles/month ^a	01/1995–12/2017
Haute Provence, France 43.9° N, 5.7° E, 650 m	Lidar	24–39.9 km	11 profiles/month ^a	01/1995–12/2017
	Ozonesonde	24–30 km	4 profiles/month ^a	01/1995–12/2017
Aura Satellite, above Bern 46.95±1° N, 7.44±8° E	MLS	31–0.8 hPa	2 overpasses/day	08/2004–12/2017 ^b

^a Averaged number of profiles per month in the analysed period.

^b For the trend calculations, data from Jan. 2005 to to Dec. 2017 are used.

Table 2. Summary of the artificial ozone time series and the different model parameters used to correct the trend estimation for artificially added anomalies. ~~The estimated trends are given in per decade (/dec).~~

<u>Case</u>	Parameters in the artificial time series		Parameters in the trend programme			
	True trend	<u>Added anomalies</u>	Monthly uncert. ^a	<u>Uncert. for anomalies^b</u>	Bias uncert. ^c	Estimated trend
A	0.1 ppm/decade	<u>–</u>	0.1 ppm	–	–	–0.100 ± 0.010 ppm/decade
B	0.1 ppm/decade	0.15 %	<u>0.1 ppm</u>	–	–	<u>0.133 ± 0.010 ppm/decade</u>
C	0.1 ppm/decade	0.15 %	50.1 % ppm	<u>0.36 ppm</u>	–	<u>0.103 ± 0.011 ppm/decade</u>
D	0.1 ppm/decade	0.15 %	<u>0.1 ppm</u>	–	5 ppm	<u>0.100 ± 0.011 ppm/decade</u>
E	0.1 ppm/decade	<u>5 %</u>	0.1 ppm	50.36 % ppm	5 % ppm	50.100 ± 0.011 ppm/decade

^aUncertainty considered in the diagonal elements of the covariance matrix S_y .

^bUncertainty considered in the diagonal elements of S_y for months with anomalies.

^cBias uncertainty considered in the off-diagonal elements of S_y for months with anomalies, set as a correlated block.

