Response to Reviewers

ACPD Manuscript doi:10.5194/acpd-10-11401-2010
Title: Multi sensor reanalysis of total ozone

We would like to thank both reviewers for their effort. The paper is both lengthy and complicated, and we realize the reviewing it is not a trivial matter. Virtually all the comments and suggestions of both reviewers have led to changes and -in our eyes- improvements to the paper.

Based on the suggestions of the reviewers, we have changed the emphasis of section 5. Its scope has been widened to give more information about the quality of the ozone reanalysis. The heading of this section has been changed accordingly from "OmF and OmA analysis" to "Evaluation of the MSR data". Also the discussion about significance of the OmF and OmA has been extended.

Another mayor change has been to the figures. Figure 2 (the correction to the 14 satellite datasets over the Netherlands) has been replaced by two figures, showing the 14 satellite dataset minus the ground data before and after the corrections have been applied. We have looked at the regional variations, but insufficient additional insight can be gained from these. We decided not to include them in the paper. An additional figure has been included that shows the monthly averaged MSR level 4 minus ground in a similar way as we show two satellite datasets in figure 1. This is to illustrate the importance of our corrections. We decided to keep figure 4. This figure is the only illustration of the horizontal resolution of the MSR level 4 dataset, which we think is important. The discussion of this figure in the text has been extended. We have decided to delete figure 3, as we think it adds little to the paper.

In the evaluation of the MSR data we added a Figure about the geographical distribution of the offset between MSR level 4 data and individual ground stations. This illustrates that individual ground stations still show an offset, but there are no regional dependencies.

Response to Reviewer #1 (original review text in italic)

The paper is technical and the thorough description of the satellite retrieval correction process and the findings about the properties of the satellite data are the main strength of the paper. The time series showing the biases of the satellite retrievals against observations for De Bilt (Figure 2) is an interesting piece of information which could be made more general by showing the bias time series also for other areas.

We have made Figure 2 more general by showing the global averages of the observations against all observations of the ground stations. Although we have done this also for several latitude zones with similar results, we have only included the Figure with a global average. On request of reviewer 2 we have made this plot for both the uncorrected as the corrected satellite observations. The new Figure 2a and 2b are added at the end of this file.

A shortcoming of the paper is that the examination of the MSR re-analysis data set is not sufficient in its present form. The paper does not include an evaluation of the MSR data set with independent data. The inter-annual variability and trends of the data sets, arguably the motivation for the whole effort, are hardly discussed, nor is the data set compared with alternative long-term ozone analyses.

There are many sources of independent data, e.g. model data, SOAZ network, lidar or microwave observation, etc., however, considering the needed time period, spatial coverage and availability, we consider the Brewer/Dobson data as the most suitable data for the evaluation.
To evaluate the quality of the MSR data we have compared the level 4 data with the Brewer/Dobson ground data to show that the offset, inter-annual variability and trends have disappeared from the data. We included an extra Figure to show the global distribution of the offset of the MSR compared to ground data (see Figure 6 at the end of this file). We also added Figure 5 to show an example how the data has been improved compared to the original satellite-ground differences as shown in Figure 1 (as suggested later by the reviewer). The following text has been added to discuss the MSR data quality:

"Fig. 5 gives an example of the MSR level 4 data set compared to ground data: it shows the same time period and location as for Fig. 1. Where Fig. 1 clearly showed systematic deviations, most notable the seasonal cycles, in Fig. 5 no seasonal cycle or trends are visible. Still a small offset remains between MSR and the ground observations on this location, which can be caused by either the ground observations or the MSR. Therefore, the MSR data has been compared with all available ground observations from the WOUDC database. One might argue that these ground observations are already used in generating the MSR data, but at least it is a consistency check and it still shows the quality of a level 4 data set compared to the "ground truth". When fitting the MSR data to the ground observations the fitted offset was found to be smaller than 0.2 DU, and both the trend (0.02 DU/year) and seasonal variation (effective ozone temperature dependence of -0.006 DU/ºK) were negligible. The geographical distribution of the offset between MSR level 4 data and individual ground stations is shown in Fig. 6. In Europe some stations can no longer be distinguished due to its high concentrations of ground stations. On average no offset is present, and only a few outliers are visible. No systematic structures are obvious in the geographical distribution."

Analyzing the variability and trends in total ozone is not trivial, regarding all phenomena one have to consider: QBO, ENSO, solar cycle, effects from stratospheric volcanic emissions, and emissions of ozone depleting gasses. We have started to work on this, but we feel this is the subject of a new paper.

Only a few figures - although as such quite interesting as the one showing the Antarctic ozone deficit - are included but not sufficiently discussed.

Figure 3 has been removed on request of reviewer 2 to reduce the number of Figures. To discuss more about the Figures 4 and 5, we have the added the following text:

"One example drawn from the MSR ozone analysis data set is shown in Fig. 3, which shows the MSR ozone field derived for 15 April 1992 and 24 September 2002 at 12.00 UTC. These examples illustrate the resolution of the data set, which allows monitoring events like the split of the ozone hole in 2002 with 6-hourly time steps. No discontinuity across the date line is present in the images, which is often seen in gridded level 2 data. The 6-hourly instantaneous and monthly mean ozone fields are available on the TEMIS web site, http://www.temis.nl/. For UV radiation studies the daily ozone fields at local noon are also made available on this web site. In Fig. 4 the average ozone mass deficit over Antarctica in the period 21-30 September is shown for the period 1978-2008. The ozone mass deficit is defined as the total amount of ozone needed to fill the ozone columns below 60º South to a level of 220 DU. The period covered by the MSR data set shows the beginning of the ozone hole in the eighties when CFCs entered the stratosphere and the more or less stable period afterwards when the Montreal protocol was endorsed. Exceptional year was 2002 when the ozone hole broke up (see Fig. 3), where after one of the two fragments was quickly dissolved. Record holder was the year 2003 with a mass deficit of $42.2 \cdot 10^9$ kg on 26 September."

It would be good to get a better balance between the description of the constructions of the MSR data set and its scientific value. For instance section 2.2 and 2.3 could be made much shorter because they contain too much details which could be omitted or referred to in the literature.

We have shortened section 2.2, but we feel that the information in Section 2.3 is necessary for a good understanding of our method. The information given there is special for our own method and therefore referring to the literature is not possible. Furthermore, we have changed the paper substantially, as a result of the review, with more focus on the scientific discussion of the MSR data.
The data from ground station “Paramaribo” is now available from the WOUDC. So the remark about this data in section 2.2 is no longer valid, and we have taken the liberty to remove it.

The conclusion section remains a bit vague. It should contain more about the results (satellite biases, correction model , assimilation performance) and should possibly be extended to the evaluation, trends and variability in the MSR. The more general statements of the potential usefulness of the data set would be better omitted or placed in the introduction.

To reflect the additions in the text and to add the issues mentioned by the reviewer, the following text has been added to the conclusions about the evaluation of the MSR:

“It has been shown that the MSR level 2 data presented no drift and an insignificant SZA and effective ozone temperature dependence as compare to the ground observations (see Table 2). The final MSR level 4 data is on average comparable with the ground data: the fitted offset, trend and seasonality in the comparison between MSR level 4 data and all observations of the WOUDC data base were negligible. The maximum fitted offset is 0.2 DU. All the systematic effects found in the satellite data are removed by the simple corrections (using a few basic parameters) applied to the satellite observations in the period 1978-2008.”

As said earlier, we think that analyzing the variability and trends in total ozone is beyond the scope of this paper considering all issues involved.

Finally, I would like to encourage the Authors to draw a more optimistic conclusion as "As the true amount of ozone in the atmosphere is not known, it is not possible to draw conclusions from this work about the quality of an individual dataset." I think the paper has shown that with a well chosen reference data set one can learn more about the properties of the individual data sets.

We had no intention to sound pessimistic in our conclusion, and changed the text to:

“By combining the satellite data and model knowledge using data assimilation and by using the ground data as a reference the aim was to obtain the best possible data set for the total ozone in the atmosphere by combining all available ozone column information. It is however difficult to draw conclusions about the quality of individual dataset, since our work is based on the assumption that the ground observation are on average without mean bias, seasonal bias or trends.”

More specific comments:

The data set is called a multi-sensor-reanalysis although it includes not just different sensors but also different retrievals from the same sensor. What was the motivation for the choice of the retrievals and *how are possible correlations in the retrievals from the same sensor treated in the assimilation scheme?*

We have collected all freely available and complete retrieval datasets. We preferred not to choose between different retrieval data sets of the same satellite instrument. Therefore it was investigated how correlated the errors of these observations were. Their correlation was low enough (about 20%) to include both data sets. To explain this better we added in the text:

“Sometimes two level 2 data sets from the same instrument are available. However, these data sets from the same instrument can be seen as independent measurements since their errors are not more correlated than the errors within a single data set. In order not to show any preferences, both data sets have been used.“

and for verification of this statement we added:

“Based on the OmA statistics the average correlation of the algorithm pairs TOGOMI/GDP, TOSOMI/SGP and OMDOAO3/OMTO3 were estimated as, respectively, 0.22±0.10, 0.20±0.02 and 0.22±0.03 for April 2007 and 0.34±0.17, 0.30±0.09 and 0.14±0.03 for the month December 2008. From this low correlation we conclude that using the observations of the mentioned algorithm pairs as independent data sets in the data assimilation is a valid choice.”

In the data assimilation the observations are treated as independent observations, which is allowed since the correlation between the retrievals is low. The observations of different retrievals have different variances and therefore the super-observations consist of satellite observations weighted with the inverse of their variances as was mentioned in the text.

How does the MSR data set differ from other long-term data sets such as CATO (Brunner et al. 2006), ERA40 (Dethof and Holm, 2004) or, more recently, Kiesewetter et al, 2010(JGR115, D10307) in terms of bias correction and satellite data usage. What is the novelty aspect of MSR?

At the time writing our manuscript the paper of Kiesewetter et al. had not yet appeared, however we have added references to both Kiesewetter and Brunner. Dethof and Holm was already mentioned. After mentioning them in our introduction we conclude the following:

“Several assimilation-based long-term ozone records exist, and have been reported in the literature. Dethof and Holm, 2004, report on the quality of a 45-year reanalysis (ERA40) of ozone performed with the ECMWF model. This work is based on previous data versions of TOMS (version 7) and SBUV (version 6), and no additional corrections were applied to account for drifts and offsets against surface data. The 45 years can be divided in four different time windows, depending on the availability of satellite data. The quality of the analyses differs from one window to the next, which complicates the use of this reanalysis for trend studies.
Brunner et al., 2006b provide a quasi-3D ozone time series for the period 1979-2004, based on the NIWA total ozone dataset constructed from TOMS/SBUV v8 and GOME total ozone observations (Bodeker et al., 2001, 2005). As in the MSR dataset, Dobson ground observations were used to remove offsets and drifts in the satellite data. However, the MSR uses both Dobson and Brewer ground observations for correcting the satellite observations, and these corrections are based on dependencies of parameters directly related to retrieval errors. Important in correcting satellite data is the fitting against the effective ozone temperature, which is a critical parameter for total ozone retrievals. This is

a novelty in the MSR data set. The analysis by Brunner et al., 2006b is based on a two-dimensional equivalent latitude – potential temperature Kalman filter approach. Brunner shows that realistic 3D ozone distributions can be reconstructed on the basis of 2D ozone column information only. In the MSR the analyses are performed with a full 3D model, which avoids errors that may result from the 2D-3D mapping. However, the data presented are restricted to the 2D ozone column field, because this is strongly determined by the satellite observations.
Compared to existing long-term datasets, the MSR data is based on more satellite data sets by including all available retrievals of the recent satellite instruments GOME, SCIAMACHY, OMI, and GOME-2."

The effect of the correction is shown in a reduction of RMS in Table 2 and 3. It would be interesting to get a picture (similar to Figure 1 or 2) of the corrected time error time series. Could the seasonal variation of the differences be corrected?

We have taken this suggestion to heart by added a new Figure (Fig. 5 at the end of this file). It shows that seasonal variations disappear by applying our corrections.

The reduction in RMS (RMS1 vs. RMS2 or RMS3 vs. RMS4) by the correction scheme seems to be very small. It is mostly smaller than 5 %. Does this justify the whole effort of the correction scheme? Would it be a valid conclusion that the chosen correction scheme did not significantly reduce the RMS of the differences? Could the RMS reduction be bigger with a different approach?

It was our purpose to reduce offsets, trends and long-term variations in the satellite data, so that the data can be used as input to the assimilation scheme without biases and with known standard deviations. The RMS values decrease, which indicates the success of our correction. Note that the total RMS consists of a combination of biases and errors that behave randomly. The total RMS is in fact dominated by the RMS error of the ground observations, the representativity errors and the variance of the observations. We think we are close to the lower-limit of the final RMS values by removing the systematic biases. Note also that in this paper all RMS values have been calculated with the mean value of the data as reference, thus a correction of the overall offset will not improve the RMS.
We have adapted the text in section 3.3 to reflect this:

"For the purpose of data assimilation it is relevant to reduce offsets, trends and long-term variations in the satellite data, so that the data can be used as input to the assimilation scheme without biases and with known standard deviations. The satellite data set corrections are based on a few relevant regression coefficients fitted for the overpass time series of all stations together. By fitting all data together regional biases that may be caused by offsets of individual ground instruments are avoided. From here on the offset per WSI reported above will no longer be used. The relevant regression coefficients, i.e. those that reduce the RMS (Root Mean Square) between satellite and ground observations significantly, have been calculated and are shown in Table 3. Note that in this paper all RMS values have been calculated with the mean value of the data as reference, thus a correction of the overall offset will not improve the RMS. The total RMS is in fact dominated by the error of the ground observations, the representativity errors and the variance of the observations which are larger than the bias terms. A small decrease of the RMS is expected as a result of the correction."

The use of the offset is unclear. Is this the actual total bias between the satellite retrieval and the ground-based observation or a site specific value? It would be worthwhile to more information in the *paper. Does the offset contribute to RMS? Why is RMS3 (Table3) higher than RMS1 (table 2). I would assume it should be the same RMS before the correction.*

The offset as fitted for Table 2 is site specific (150 offset values), but this is only shown for illustrating its effect. For the remainder of the paper we used a total bias (one value) fitted for all ground observations at the same time. This is exactly the reason that RMS3 is higher than RMS1. We have tried to make this more clearly by explicitly adding in the text:

"RMS3 and RMS4 are higher than respectively RMS 1 and RMS 2 since the RMS in Table 2 is calculated by fitting an offset for each station and the RMS in Table 3 is calculated using a single offset for all stations."

Given the apparently small impact of the correction, has it been tested what would happen if the un-corrected retrievals were presented to the assimilation system. Small biases might be tolerated by the assimilation apparatus and the analysis would perhaps not differ much from MSR. A check with independent data could help to identify which of the analyses would be better.

As shown in the appendix and in Fig. 1 the corrections can be quite significant. This is also shown in the new Fig. 2 at the end of this file. So we do not agree to the statement that the correction apparently has a small impact. The MSR analysis closely follows the level 2 data set and would reflect the trends and seasonal biases present in the data. Using two biased datasets simultaneously may lead to rejection of data in the quality screening and jumps in the analysis.

As the bias correction was important, have other theoretical assumption such as the Gaussianity of the increments be tested before the assimilation.

This has been tested and is discussed in Eskes et al., 2003. The reference to this paper has been given and therefore it is not further discussed in this paper.

For the interpretation of the OmF and OmA statistics it should be made clear how long ago the forecast was updated with the previous analysis. This should be put in relation to the ozone lifetime.

Indeed, some explanation on the timing of the forecast was omitted here. It is a relevant point since this is for a data assimilation of multiple data sets different than from a single sensor data assimilation. We have added the text:

"In the data assimilation the forecasts are calculated in sequential steps of half an hour. For assimilation of a single sensor, such as TOMS or OMI the observations at a certain location are typically once a day and therefore the OmF and OmA reflect a time step of 1 day. For the assimilation of SBUV only (e.g. beginning 1995), accounting for a correlation length of 500 km, the revisit time is typically 1 week. For data assimilation of multiple sensors this is different, time steps between observations can range anything between half an hour and 1 day. This means that the OmF and OmA, and therefore the data assimilation results, are more restrained by the observations. Typical OmF and OmA behaviour that has been checked are: i) In general the OmA has to be smaller than the OmF, ii) No geo-location or geo-parameter dependencies have to be visible, iii) the RMS will mainly reflect the error distribution of the observations."

The lifetime of ozone in the lower stratosphere is in the order of months. This is orders of magnitude longer then our analysis time step. This makes the ozone lifetime irrelevant, except in ozone hole conditions. The TM3-DAM model has a simple heterogeneous chemistry module to deal with this situation.

It is common practise to spell out acronyms of the sensors, institutions such ESA or NASA or RMS at the first ocurrence.

We have added the meanings of the acronyms ESA and NASA. RMS was already spelled out in the text.

Response to Reviewer #2 M. Weber (original review text in italic)

2 Figures
Figure 1 shows monthly mean anomalies between two satellite data sets and De Bilt ground station. The differences in the seasonal behavior of the anomalies for the two data sets is not discussed at all in the main text. The question arises if the two predictors used in the correction ("effective ozone temperature" and "SZA") may have a similar seasonal behavior and may not be independent. Is it possible that depending on the phase of the seasonal behavior in the anomalies, the effective ozone temperature or SZA become the more dominant predictor?

We have added this to the text so that the discussion of Fig. 1 becomes:

"As an example Fig. 1 shows the monthly averaged anomalies (defined as satellite measurement minus ground measurement) over the Netherlands as a function of time. It is clear that either the ground station data and/or the satellite data contain a seasonally dependent error. A study of all satellite products for this station (Brewer MKIII, De Bilt) shows that a seasonal effect like this is fairly typical, but the amplitude and phase differs from one satellite product to the other. The seasonal cycles of the anomaly of the two satellite products shown here are even in anti-phase."

It has been checked that effective ozone temperature and SZA are almost independent and have added this information to the text. However, we have not further elaborated on the analysis of their dependency. Note also that in most cases only the dominant one of the two parameters is used to correct data.

Figure 2 provides a good summary of the performance of the individual data set with respect to the ground data, but only comparisons to the de Bilt station is shown. It would be nice to see similar comparisons in the polar region, tropics, and SH midlatitude. One could show anomalies for all overpasses within a given zonal band. This figure could be improved by showing anomalies with respect to ground data rather than the MSR level 2 (or corrected satellite data). A second figure could show then differences between the corrected satellite data (MSR level 2) and ozone temperature corrected ground data. This would document how well the corrections work for each of the data sets.

We have taken over this suggestion by replacing Figure 2 by a new Figure 2a and 2b (shown at the end of this file). Figure 2a and 2b show the globally averaged deviations between satellite data and ground observations without corrections for the satellite data in Fig. 2a and after correcting the satellite data in

Fig 2b. We have made these Figures also for various zonal bands with similar results and conclusions; therefore we have only shown the globally averaged results.

Figures 4 and 5 are described in the main text in two sentences and they do not provide any value to this study apart for (colorful) advertising of the assimilated data set.
Terms like ozone mass deficit is not explained nor the trend in the ozone mass deficit explained. I strongly recommend to drop both figures and substitute them with additional figures as discussed above.

Figure 4 was meant to demonstrate the resolution of the MSR data and Figure 5 shows an application for the MSR data, the monitoring of the ozone hole statistics. However, we have removed Figure 3 in the same section. Furthermore we have added more text to discuss Figure 4 and 5:

"One example drawn from the MSR ozone analysis data set is shown in Fig. 4, which shows the MSR ozone field derived for 15 April 1992 and 24 September 2002 at 12.00 UTC. These examples illustrate the high resolution of the data set, which allows monitoring events like the split of the ozone hole in 2002 with 6-hourly time steps. No discontinuity is visible in the images, which would have appeared in gridded level 2 data. The 6-hourly instantaneous and monthly mean ozone fields are available on the TEMIS web site, http://www.temis.nl/. For UV radiation studies the daily ozone fields at local noon are also made available on this web site. In Fig. 5 the average ozone mass deficit over Antarctica in the period 21-30 September is shown for the period 1978-2008. The ozone mass deficit is defined as the total amount of ozone needed to fill the ozone columns below 60º South to a level of 220 DU. The period covered by the MSR data set shows the beginning of the ozone hole in the eighties when CFCs entered the stratosphere and the more or less stable period afterwards when the Montreal protocol was endorsed. Exceptional year was 2002 when the ozone hole broke up into two areas (see Fig. 3), where after one part was quickly dissolved. Record holder was the year 2003 with a mass deficit of $42.2 \cdot 10^9$ kg on 26 September."

Figures 7 and 8 show observation minus forecast (OmF) and observation minus analysis (OmA) as a function of various parameters. The authors should explain in more detail in the main text what are the separate roles of OmF and OmA in the "quality control". Why do you need both? Also exactly define what is "analysis" (model value at +0h just before assimilating observations?) and what is "forecast" (+12h, +24h?). It would be helpful to properly define OmA and OmF and their different roles in the quality control in the beginning of Section 5.

Since the OmF and OmA are often used in the literature about data assimilation we have neglected to mention their meaning and purpose. We propose to add the following text:

"An important source of information for the evaluation of any data assimilation is the observation-minus-forecast (OmF) and the observation-minus-analysis (OmA) statistics produced by the TM3DAM analysis system ("forecast" is the model result just before assimilation and "analysis" is the assimilation result). This mechanism allows detection of sudden changes in the data quality and provides error estimates for the total ozone retrieval as well as the model performance. The analysis uncertainty is reported as a two-dimensional field, part of the analysis product. In the data assimilation the forecasts are calculated in sequential steps of half an hour. For assimilation of a single sensor, such as TOMS or OMI the observations at a certain location are typically once a day and therefore the OmF and OmA reflect a time step of 1 day. For the assimilation of SBUV only (e.g. beginning 1995), accounting for a correlation length of 500 km, the revisit time is typically 1 week. For data assimilation of multiple sensors this is different, time steps between observations can range anything between half an hour and 1 day. This means that the OmF and OmA, and therefore the data assimilation results, are more restrained by the observations. Typical OmF and OmA behaviour that has been checked are: i) In general the OmA has to be smaller than the OmF, ii) No geo-location or geo-parameter dependencies have to be visible, iii) the RMS will mainly reflect the error distribution of the observations.”

3 Other major issues
p. 11402, l.11: Here and other places the term stratospheric temperature or effective temperature is used. It is suggested to use a single term like effective ozone temperature or simply ozone temperature to make it clear which temperature is considered here.

We have changed everything to effective ozone temperature.

p. 11404, l. 28: Multiple level 2 data sets from the same instrument are sometimes used since their errors are not highly correlated. I do not understand what is meant here. Later in the text it is shown that the correlation between different retrievals are on the order of 0.5 (p. 11418, l.6), which means there is significant correlation. Please clarify. What values do the estimated correlation coefficients (p. 11418, l. 8) between data sets from different instruments have. (This important since these numbers are used in the error covariance matrix as described in Section 4). Is the calculated correlation coefficient between different instruments much lower than those between different data sets from a single instrument? Another question comes then in mind: do instruments for which only one data set is available get less weight in the data assimilation than instruments with multiple data sets available or does the estimated correlation coefficients causes a more appropriate weighting.

When using two data sets from the same instrument it is important to know if these datasets are not too much correlated, otherwise it would just mean that we are using the same observation twice and it get more weight. In our case each algorithm pair of the same instrument had a correlation coefficient of about 0.2, which means we have two almost independent observations from the same instrument. However, for the data assimilation another correlation coefficient is important: the correlation between errors of all observations within a superobservation. This depends on the spatial ozone variability within a grid cell and is estimated as 0.5. This number is an average and is determined of single-algorithm data sets. We admit the text was quite confusing about this issue at p. 11419. We decided to limit our analysis to the first correlation coefficient and for the second we refer to earlier work (Eskes et al., 2003 and references within). Then the new text becomes shortly:

“Based on the OmA statistics the average correlation of the algorithm pairs TOGOMI/GDP, TOSOMI/SGP and OMDOAO3/OMTO3 were estimated as, respectively, 0.22±0.10, 0.20±0.02 and 0.22±0.03 for April 2007 and 0.34±0.17, 0.30±0.09 and 0.14±0.03 for the month December 2008. From this we conclude that using all observations of the mentioned algorithm pairs in the data assimilation is a valid choice.”

p. 11406, l. 19. A better source for discussion of differences between ground instruments can be found in Staehelin et al., 2003 (Staehelin, J., Kerr, J., Evans, R., and Vanicek, K.: Comparison of total ozone measurements of Dobson and Brewer spectrophotometers and recommended transfer functions, Tech. Rep., WMO, World Meteorological Organization Global Atmosphere Watch (WMO-GAW) Report 149, 2003.)

We have added the reference of Staehelin et al. here.

p. 11407, l. 18. Why are different maximum collocation radii for overpasses are allowed for each satellite data set (50-200 km) to determine overpasses?

Here we tried to approximately reflect the pixel size with the different collocation radii. 50 km seemed a good maximum in general, but for example for satellite observations with pixel sizes of 320 km (for GOME) a higher maximum is a better choice. We now mention:

"This number is typically 50-200 km depending on the ground pixel size, see details in Table 1"

p. 11409, l. 8. Table 5 is mentioned here before Table 2. Change numbering of Tables.

We have removed the early reference to table 5, to avoid this problem. We did not change the table numbers.

p. 11409, l. 20. It should be mentioned here that the standard algorithms for Brewers as well as Dobsons assume a fixed effective ozone temperature (-46 C). This explains the offset used in Eq. 1 (l. 27) (please add an eq. number).

We have checked the temperature correction published by Kerr (2002) with the WOUDC data. We found that this correction is essentially correct, and used Kerr's formulation to correct the Dobson data. Speculation on the reasoning behind this correction is beyond the scope of our paper.

Eq. number has been added.

p. 11409, l. 27. Are the values 0.0013 and 46.3 in this Eq. are derived by fitting all 21 stations having simultaneous Brewers and Dobsons, or are these numbers from the literature. Please clarify.

The numbers are from the literature and later confirmed by our own analyses. We have adapted the text accordingly.

p. 11412. There are two type of corrections applied to the satellite data. In a first step all predictors and individual station biases are fitted (Table 2), but for the data to be used in the data assimilation the corrections using only significant predictors and a single bias for all stations is applied (Table 3). It is not explained why the second correction (with less degrees of freedom) is favored for the data assimilation input.

For finding the essential predictors (Table 2) it was useful to correct the individual station biases. However, for correction of the satellite data no ground station dependent bias can be used. Even if this was possible, using the offsets of all stations would imply that the satellite observations would need corrections with regional biases as shown in the new Fig. 6 at the end of this file. It is to be expected that the satellite observations have a global bias or dependencies from SZA, temperature or time, but such spatial variation in the biases cannot be explained physically. To emphasize this, we have added the text:

"For the purpose of data assimilation it is relevant to reduce offsets, trends and long-term variations in the satellite data, so that the data can be used as input to the assimilation scheme without biases and with known standard deviations. The satellite data set corrections are based on a few relevant regression coefficients fitted for the overpass time series of all stations together. By fitting all data together regional biases that may be caused by offsets of individual ground instruments are avoided. From here on the offset per WSI reported above will no longer be used."

p. 11412, l. 12: Do you have an explanation why OMDOA3 and OMTO3 have opposite temperature dependence in the anomalies?

Unfortunately we do not have an explanation. It would need a thorough analysis of both algorithms to explain this. For discussion we have added:

"It is remarkable that all "American" data (TOMS, SBUV, OMTO3) require a negative temperature correction, while most "European" data (GOME, SCIAMACHY, GOME2) do not need a temperature correction. The OMDOAO3 dataset is the only one showing a positive correction. As there has been a trend in stratospheric temperatures in the 30 year period (Randel et al., 2009), it is conceivable that this trend shows up as a spurious trend in the ozone data, if the satellite products are not corrected for effective ozone temperature."

p. 11412, l. 27: Please explain here why RMS3 is higher than RMS1 as shown in Table 2, I would have expected them to be equal (satellite data before corrections) or RMS3 should be even lower than RMS1 when using a single bias correction in the former. In general it is noticeable that the RMS after corrections is in general only slightly lower than before corrections (see comment by other reviewer). Would the improvement in RMS more significant when looking at monthly mean anomalies rather than daily anomalies due to large day-to-day variability dominating the RMS?

The offset as fitted for Table 2 is site specific (150 offset values), but this is only shown for illustrating its effect. For the remainder of the paper we used a total bias (one value) fitted for all ground observations at the same time. We have tried to make this more clearly by explicitly adding in the text:

"RMS3 and RMS4 are higher than respectively RMS 1 and RMS 2 since the RMS in Table 2 is calculated by fitting an offset for each station and the RMS in Table 3 is calculated using a single offset for all stations."

It is to be expected that the RMS improvement will be more significant for monthly mean anomalies.

p. 11415, l. 9: Some data were already rejected as part of the merging and correction procedure described earlier. Please describe what additional screening is explicitly done during data assimilation.

We have added the following information about the additional screening during the data assimilation:

"A quality screening is implemented to reject unrealistic ozone observations (i.e. below 50 DU or above 700 DU) or unreliable ozone observations (measured with a solar zenith angle higher than 85 degrees). Observations that deviate from the model forecasts more than either 3 times the observation uncertainty or 3 times the model uncertainty are also rejected. Because of this restriction and the assumption of unbiased observations, the satellite data has been corrected as described before."

No other screening was done.

Fig. 1: Please state in caption if satellite minus ground or ground minus satellite is shown.
Fig. 2: Indicate units of y-axes (DU).

We have added the information the captions.

4 Minor issues
An annotated file with additional suggestions for improving texting/spelling is provided.

We thank the reviewer for his help in improving textual aspects.

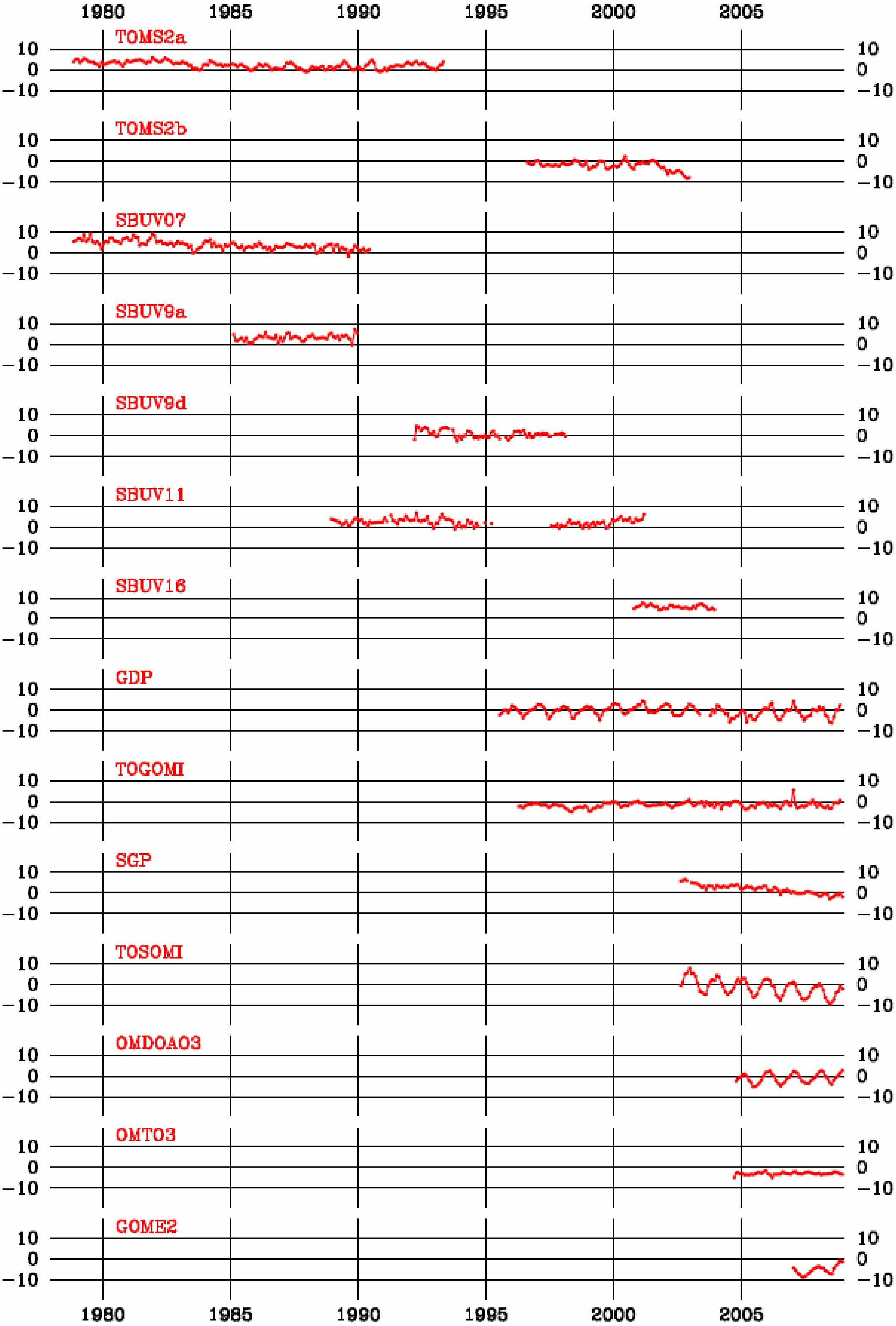

Fig. 2a Monthly averaged and globally averaged difference of the satellite ozone observation minus the ground observation in units of DU, for all satellite data sets used as function of time.

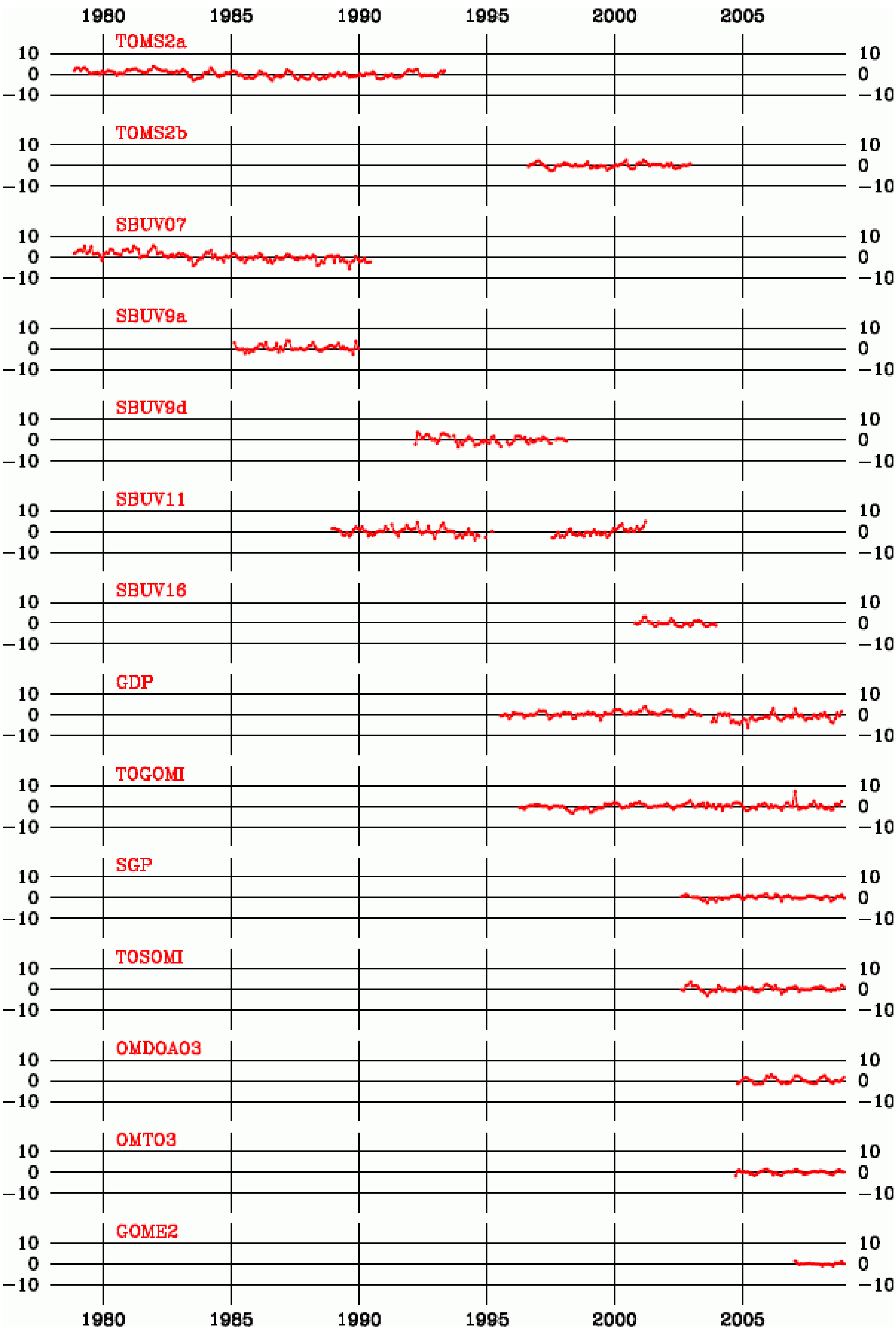

Fig. 2b Monthly averaged and globally averaged difference of the corrected (as described in Sect. 3.5) satellite ozone observation minus the ground observation in units of DU, for all satellite data sets used as function of time.

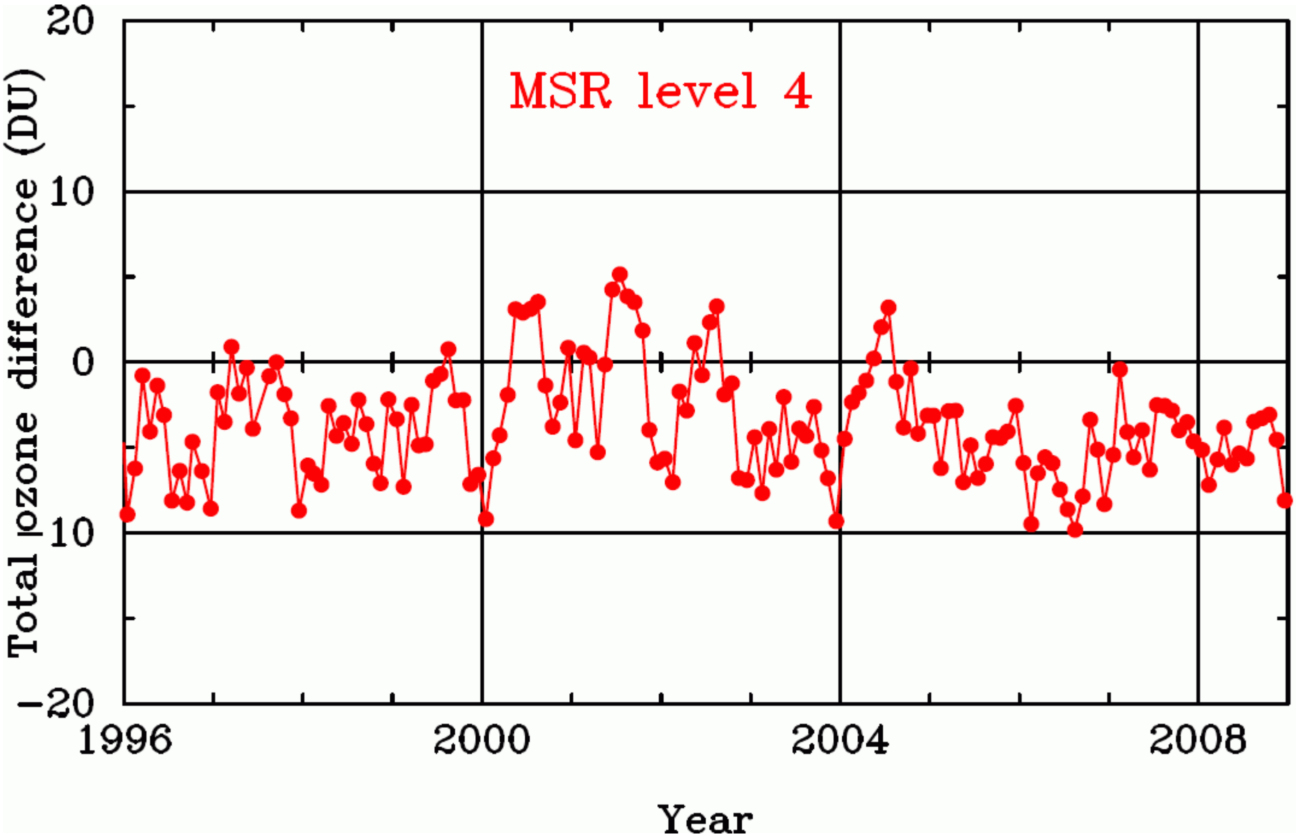

Fig 5. Monthly averaged anomalies (satellite measurement minus ground measurement) for the overpass data of the MSR at the ground station De Bilt (5.18º E, 52.1º N) in the Netherlands.

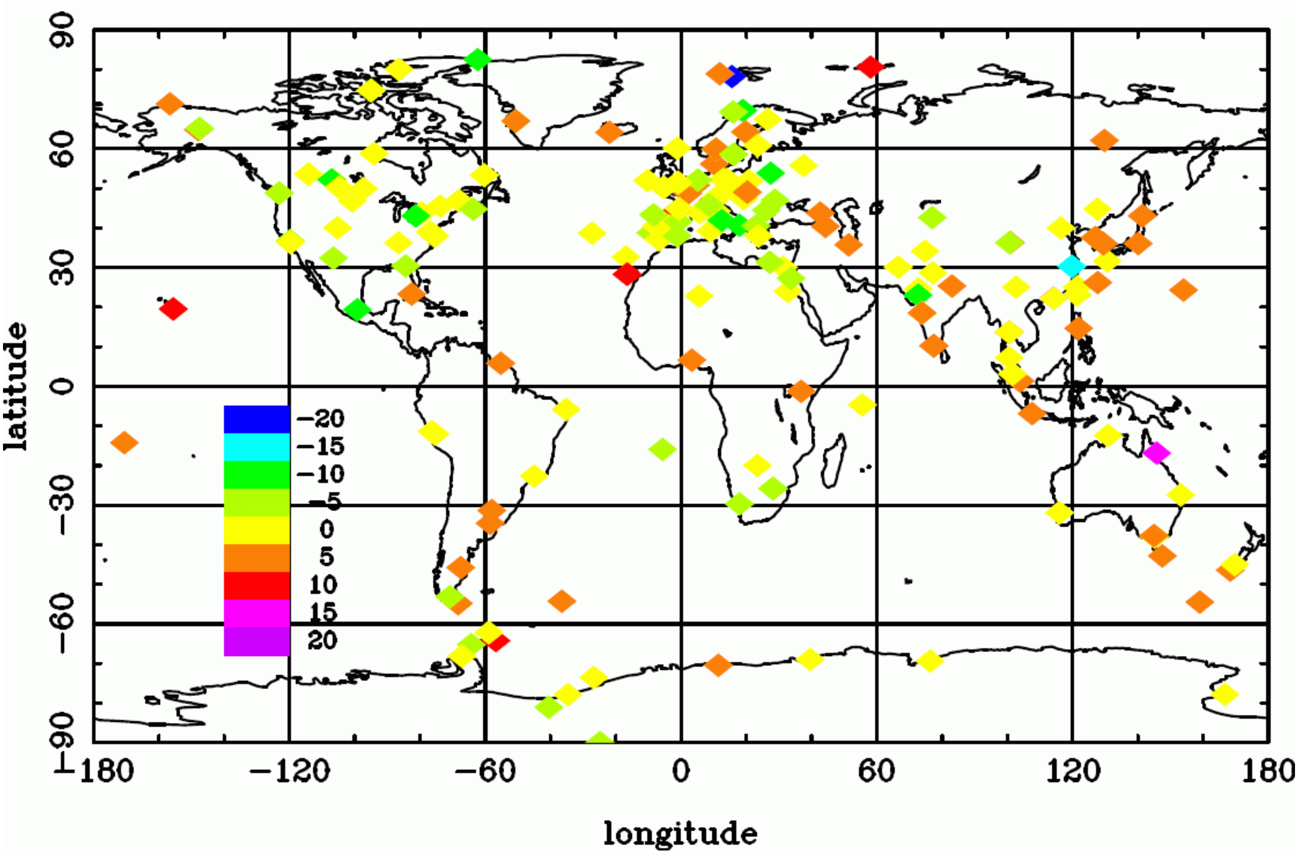

Fig 6. Fitted offset (MSR minus ground) between the MSR level 4 data and all selected ground measurement time series of the WOUDC in the period 1978-2008.