Reply to comments to the paper "A comparative analysis of UV nadir-backscatter and infrared limb-emission ozone data assimilation" (R. Dragani)

I thank the Reviewers for their comments and suggestions. These have reported below (highlighted) with my reply (in normal text) and addressed in the current version of the manuscript where appropriate, unless indicated otherwise.

A modified version of the ACPD paper with tracked changes is attached at the end of this document.

Reply to Reviewer # 1:

General comments:

Reading back through Dragani 2011, MIPAS ozone (different version) and GOME (ERS-2, not the same GOME) were already assimilated in ERA Interim. I know it's not the same data but I think it should be mentioned somewhere. How do the results presented here compare with ERA Interim? Does the CCI MIPAS ozone bring anything new compared to the version that was used in ERA-I? I'm not asking for any extensive comparisons, just a comment.

The Reviewer is right in saying that ERS-2 GOME ozone profiles and ENVISAT MIPAS ozone profiles were both assimilated in the ERA-Interim production. In particular, it is noted that:

- The CCI algorithm used to retrieve the GOME-2 ozone profiles used in the present study is a development of the Rutherford Appleton Laboratory (RAL) retrieval scheme that was used to retrieve the ERS-2 GOME ozone profiles assimilated in the ERA-Interim reanalysis. This point has now been added to the paper.
- Regarding the MIPAS data, there are three differences (in the spectral characteristics of the Level 1b data, in the data processing, in the L2 algorithm) between the CCI product and the one used in ERA-Interim. These have now been discussed in the paper.

Specific comments:

- P2 L24: McCarty et al. is still not finished. At this point I suggest changing this reference to Bosilovich et al. (2015). The ozone chapter contains the same information and this tech memo is already published and citable.
  Reference replaced as advised.

- P4 L 5-14: What is GOME's footprint?
  For the period used in this study (2008), the GOME-2 typical footprint in the forward scan is 80 km × 40 km (across track × along track). This aspect has now been added to the paper. It is noted here that after the launch of MetOp-B, a change in the orbit swath (from 1920 km to 960 km) was applied to MetOp-A (July 2013) that led to a smaller footprint of 40 km × 40 km.

- P4 L26: The link is old and redirects to http://cci.esa.int/. You may want to update it.
  The link has been updated, as advised.

- P6 L7: ‘the ozone continuity equation is expressed as a linear relaxation...’ Hmm, there’s more to the continuity equation than just chemistry. How about ‘contains’ instead of ‘is expressed as’? Maybe I misunderstood something.
  The sentence has now been rephrased.

- P9 Last paragraph of Section 5.1: Is this because with the stratosphere constrained by MIPAS the analysis increments arising from total column data are distributed differently? You talk about this later on (P13) – how about ‘this will be discussed in Section 5.2.2’? Also, see my comment to P13 L6.
  A reference to section 5.2.2 has been added, as advised.

- P10 L1: ‘version 3.04’, is this correct? As far as I know the recent ‘official’ versions are 2.2, 3.3, 3.4 and 4.2.
  The Reviewer is right, the reported MLS version was incorrect. From the original hdf files, the MLS data used in the study is version 3.3. This has now been corrected throughout.

- P10 L19: A bit more about how the ‘degrading’ is done. Is it by interpolation from the two nearest pressures or the average within the layer onto which you are interpolating? This probably makes little difference for MLS comparisons but I found that for ozonesonde data, with their high vertical resolution (many sonde measurements per model layer) it’s better to integrate than to interpolate between the two nearest points. This is because the model/DAS layer ozone values represent the layer averages, whereas sondes provide point measurements.
  The model ozone and the sonde ozone profiles are on two different pressure level grids at the start of the procedure. As it was mentioned in the paper, the comparison is performed using a vertical interpolation on the coarsest vertical grid, which for the ozone sondes is represented by the model levels. Please note that this interpolation is only performed within the region of the atmosphere covered by both profiles, i.e. no extrapolation is performed as part of the procedure, and only applied to the sondes that reach at least 40 hPa. Once the model and sonde profiles are on the same vertical grid, integrated columns are computed for both. This vertical integration is only applied for plotting purposes to facilitate the visual inspection. This point has now been clarified in the paper.

- P11 L20-29: It would help to see some percent values. Not necessarily in the figure but in the text.
  Percent values have been added to the text, as advised.

- P13 L6: I agree that this is the most probable reason why the MIPAS analysis results are improved below 400 hPa but it is not really shown here – it is just stated. The reader may wander if vertical transport between the observed and unobserved layers wouldn’t also play a role.
  I would exclude that the vertical transport could be in part responsible for the positive impact the assimilation of MIPAS has in the lower troposphere. Figure 8 panel d) in the paper presents the relative differences between the Exp/MIPAS and Exp/Ctrl analyses. It clearly shows changes in the ozone analyses below 400hPa down to the surface. This difference can only be triggered by the element(s) that differ in the two experiment setup, and that is the assimilation of MIPAS data, which has a vertical coverage down to 400 hPa at best. If the ozone assimilation was based on a multi-variate system (i.e. completely interactive) within a long-window 4D-Var, then perhaps a

change in the stratospheric ozone could have modified the circulation and produce an impact in the lower troposphere. But here, ozone is univariate. It means it has no impact on the rest of the system in general, and on the winds in particular (neither horizontal winds nor the vertical component). The vertical velocity in the upper troposphere down to 400hPa has values of up to 2-3 hPa/hour, which means the vertical transport could explain less than 50hPa vertical displacement within the 12 hour assimilation window, but not changes below 500hPa. A comment has been added to the paper.

- P15 L3: Again, I'm confused about the MLS data version (3.04 or 3.3?).
  This has now been corrected to version 3.3, thank you. See also reply above.

Technical comments:

- P10 L3: '82S to 82N' ➔ '82°S to 82°N'.
  Corrected as advised, thank you.
- P11 L32: 'southern than' ➔ 'south of'.
  Corrected as advised, thank you.
- P12 L25: 'what does it happen in the lowermost troposphere' ➔ 'what happens'?
  Corrected as advised, thank you.

Reply to Reviewer # 2:

General comments:

In this paper I see the name of the satellite instrument and the retrieved ozone product(s) being used interchangeably a lot when they should be kept separate in my opinion. For example: P4, L1-3: In these sentences the two are starting to get intermixed.

I do agree with the Reviewer that for simplicity the name of the instrument is often used to identify the dataset. In the example mentioned (namely P4, L1-3), I agree the two should not have been used interchangeably, and it has now been corrected. The paper has been reviewed for similar cases.

While the instrument is unique, there are several implementations of retrieval algorithms for a particular satellite instrument's data. For example, for GOME-2 the ozone retrieval algorithms / products known are Miles et al (O3-CCI), Cai et al, and Hassinen et al (O3MSAF). The retrieval algorithms may all have different behaviour, which makes it harder to make general statements like 'GOME-2 data is....'. The same holds for MIPAS, the instrument (and L1 data) is not the same as the ozone product coming out of a retrieval algorithm. Please check the manuscript, and identify where you really mean the instrument, and where you refer to the ozone retrieval product.

The Reviewer is correct in saying that for both the GOME-2 and the MIPAS instruments several algorithms have been developed. In the paper, having used the CCI datasets for both instruments, it felt safe to eventually omit the reference to CCI and simply identify each dataset by the corresponding instrument name. To avoid confusion, a note has now been added in section 2.

Specific comments:

- P1 L23: You mention the warming/cooling of the air in the atmosphere, and then mention its (long term) effect on climate. A warming / cooling of air has a more immediate effect on the atmosphere: a temperature difference leads to a density difference, which leads to a pressure difference, which in turn leads to flow of air. In that way, the global ozone distribution can affect the dynamics of the atmosphere.
  The point I was trying to make is that ozone can have an impact on different time scales, affecting the dynamics of the atmosphere as point out by the Reviewer, but also on longer time scale behaving as a green-house gas in the troposphere, thus impacting the Earth's climate. I rephrased it to try and make the point clearer.

- P4 L 12: Do you mean resolution, or sampling? The sampling of the instrument is usually defined as the distance between detector pixels (in nm), but the resolution of the instrument is also affected by the width of the instrument's slit function (which may be wider, and span more than one detector pixels).
  I indeed meant spectral resolution being between 0.2-0.4nm, e.g. from EUMETSAT note at www.eumetsat.int/website/home/Satellites/CurrentSatellites/Metop/MetopDesign/GOME2/index.html.

- P7 L22-25: The author starts off with mentioning that O3-CCI/GOME-2 has the largest difference in the 4 month period. While this is true in the beginning, it would be more insightful for the reader if the discussion on the differences would be split into the first two months and the last two months, as is the case in the later sentences where the O3-CCI/MIPAS differences are split into Jul/Aug and Sept/Oct. Given that the behaviour of the difference changes with time I feel that giving a range of the average difference is more representative than a single value over the four month period.

In that sentence, I was pointing out that on average over the four month period, the analysis departures from the GOME-2 data showed a positive mean value around 5DU while those for MIPAS are on average close to zero. The discussion of figure 4 is done already by looking at the two periods separately as suggested by the Reviewer, from the paper: "*The MIPAS measurements indicate that during July-August the global mean ozone analyses are about 5DU too high (they were 10DU too low based on the GOME-2 data). Here, the discrepancy between the two instruments is very likely related to different coverage of the two instruments, particularly over the high latitudes in the SH, as shown in Figure 1. During September-October, the O-A residuals for the two instruments are more similar and they both indicate an underestimation of the ozone analyses of about 5DU above 100hPa*". If I correctly understood the Reviewer comment, I believe the discussion is already in line with their suggestion, so no change has been made to the text.

- P8 L8-11: Using the larger provided uncertainty to compensate for the fact that vertical correlations (by means of Averaging Kernels (AK's) and covariance matrices from the ozone retrieval) are not used in the assimilation systems is risky. Can the author give an estimate whether the larger provided retrieval uncertainty is similar in the value and the sign of the vertical correlations?

  Regarding the use of the AKs (see also my reply to Reviewer # 3), in IFS the use of the AKs has not yet been implemented, thus using them is not an option. In the MACC/CAMS IFS (please note this is a different system from the operational IFS) some tests have been performed. However, the results obtained when assimilating the ozone data with the AKs compared to the same assimilation without AKs are at best neutral (A. Inness, personal communication). The use of the full error covariance matrix has never been tested, neither in IFS nor in the MACC/CAMS IFS. I would like to stress that the observation uncertainties used in the assimilation runs were those provided by the data providers, and that they were not inflated. What the paper shows is that the provided uncertainty seems to be larger than that diagnosed with the Desrozier method, which is based on first-guess and analysis departures. Also, please note that assimilating data with an inflated observation uncertainty is not risky. On the contrary, inflation is a standard practice used in data assimilation for safety and to be conservative, as with a larger uncertainty one can limit the level of changes the observations produce on the resulting analyses.

- P8 L18: The author states that no corrections are applied for GOME-2 O3 nadir profile data above 5hPa but does not mention what kind of corrections are applied below. It leaves the reader in doubt on what happens. Only at P8-L32-34, at the end of the paragraph that discusses MIPAS, the reader finally finds that no corrections are applied to either the nadir or limb retrieval product. Please make this clear in the paragraph that discusses O3-CCI/GOME-2.

  As stated in the paper, data assimilation is normally performed in a conservative way, meaning observations are never given more weights than their uncertainties would imply. In the case of GOME-2, the comparison between the provided and estimated uncertainty shows that the former is larger than the latter below 5hPa, implying that the data assimilation would give a smaller weight (importance) to the observations than we would have anticipated from the uncertainty estimate, thus here there is no need to further e.g. inflate the observation uncertainty. As stated in the paper, "*…The only consequence of such an overestimation would be to limit the impact of the corresponding data*". An explicit note has now been added to the paper.

- P9 L5-6: The author mentions that the behaviour in the tropical region is different for O3-CCI/GOME-2 than for O3-CCI/MIPAS. From figure 4 it is clear that the largest differences seem to coincide with the ITCZ, which is a clouded region in the tropics with high cloud tops. In P5-L9+10 the author states that the MIPAS data has been carefully screened for clouds, while I see no such statement for the O3-CCI/GOME-2 data. This could explain the large differences, when the nadir ozone profile retrievals are affected by ozone ghost columns, as the clouds block the observation of ozone below the cloud top. Would it be possible to investigate the effect of removal of pixels with a large cloud fraction from the O3-CCI/G2 data set on the global distribution of differences?
  I found in the literature a mention to the fact that in the retrieval scheme, cloud radiative transfer is not modelled explicitly. Instead an effective Lambertian surface albedo is co-retrieved. This is thought to have consequences below the cloud top (Miles et al, 2015). This information and the possible issues associated to not having an explicit cloud model have been included in the paper. The pixel cloud fraction is not available so it is not possible to check and remove cloud-affected pixels. I am not an expert in retrieval and it is difficult for me to say what impact modelling the clouds could have on the vertical profile, but I would think that the impact can mostly affect the lower troposphere up to the cloud top, and only marginally, if at all, the stratosphere. Figure 4 shows the global analysis departures over the layer 0.05-100 hPa (essentially the stratosphere and lower mesosphere).

- P10 L19: As far as I understand the comparison of two satellite retrievals, both instruments should be brought to a common grid and the AK's should be cross-applied. See Calisesi, et al (2005), Regridding of remote soundings: Formulation and application to ozone profile comparison, J. Geophys. Res., doi:10.1029/2005JD006122. Would this be a feasible approach for this study? For comparisons using reference data with a very high vertical resolution, such as ozone sondes, the transformation in Calisesi is not required because the ozone sonde's 'averaging kernel' peaks only near the measurement altitude (as it is a very localised measurement of the air it passes through).
  The comparisons discussed in the paper are not between two satellite retrievals, but rather between observed profiles (from MLS and ozone sondes) and co-located ozone analyses resulting from the assimilation experiments. For what I can see, the MLS data does not come with averaging kernels, so it is not possible to convolve the analyses with those in the comparisons. It is also noted that MLS, being a limb instrument, also has by design averaging kernels that strongly peak near the measurement altitude.

- P11 L29-31: Both instruments show reductions of the standard deviation (Fig 9). The one from O3-CCI/GOME-2 occurs over a wider vertical range than the one from O3-CCI/MIPAS, wheras the latter seems to have stronger localised reductions. Which of the two would be preferable for the assimilation as a whole and why?
  It is a tricky question to answer and deciding on which one of these two situations is preferable might depend on other considerations as well, for instance on the region where the improvement is found. They are both a sign of an improvement. In general the more one can reduce that standard deviation the better the fit to the independent data is. Here, the important point here is

not to say which one is better than the other but trying to identify where and how much the data from these two viewing geometries can be expected to make an impact.

General question:

The author demonstrates that the comparison with MLS and ozone sondes improves when GOME-2 and MIPAS based ozone profiles are assimilated, but if it is not too far out of scope of this paper, it would be interesting to get an indication of the change in the skill of the IFS in general as a result of the assimilation of the additional ozone input. E.g.: the effect on wind vectors or temperature.

This aspect was not considered in this work for two reasons. The first reason is that the focus was to exclusively look at the impact on ozone produced by the assimilation of limb and nadir data. The second reason is that in the current system, ozone is implemented as a univariate variable. This means that in general the assimilation of L2 ozone products has no impact on the rest of the system. In my experience, this is normally the case. In rare situations, however, a measurable impact can indirectly be triggered on some meteorological fields if the assimilation of the Level 2 (L2) ozone data is able to modify the data usage of the ozone-sensitive radiances (IR/O3) measured by instruments like IASI, AIRS, and CrIS. For this to happen, the L2 ozone product needs to have a sufficiently high number of observations and be able to modify the IR/O3 data assimilation in the UTLS (i.e. where these radiances have the highest sensitivity) and their data usage. If that occurs, a chain reaction can be triggered within the 4D-Var data assimilation. In an attempt to fit all data at once, 4D-Var can modify the assimilation of other high-impact observations (e.g. from sensors like the microwaves) that can, in turn, lead to a measurable impact on meteorological fields. However, this is not the norm, nor in my understanding is the assimilation of the IR/O3 radiances by all other Numerical Weather Prediction centres, thus adding that element in the paper could be misleading. For this reason, it is preferred not to discuss it.

Typographical comments:

- P2 L2: signature → ... the signing of an international treaty.
  Corrected as advised, thank you.
- P3 L5: greatly → very.
  Corrected as advised, thank you.
- P4 L28: **The** CCI ...
  Corrected as advised, thank you.
- P5 L10: verified? You may mean 'present'.
  The verb "verified" refers to the (LTE) conditions. I would prefer to keep the verb verify instead of using the verb "present".
- P5 L32: satellites (plural).
  Corrected as advised, thank you.
- P6 L16: "An example of **a** background error profile and **a** vertical correlation..."
  Corrected as advised, thank you.
- P6 L18: Introduction of acronym TCO3 without prior explanation (also not in table 1).
  The acronym TCO3 has been added to table 1. Thank you.
- P12 L24: IS equivalent.
  Since the sentence was formulated as a question, the auxiliary was already included at the start.
- P12 L25: what → why

The use of “why” in this case would not be correct. I really meant to ask what mechanism is or could be responsible for the improving the quality of the Exp/MIPAS ozone analyses at levels below the availability of the MIPAS observations. This was left unchanged.

- Figure 12: The plots are small and the black and blue are sometimes hard to distinguish with this line width. Would it be possible to provide larger plots with thicker line? One could try a 2-1-2 panel ordering instead of the current 3-2.
  The panels in figure 12 have been rearranged as asked. The line thickness has only marginally been increased as larger lines would have made the results for two experiments almost overlap when they do not.

References:

- Miles et al (2015) double doi.
  It has been removed, thank you.
- Munro et al (2006): Please check initials of Munro, it seems that there are spurious letters, as the other reference has an 'R.' only.
  The initials were correct, but the “and” between the first and second co-authors was misspelled as “amd”. Thank you for noticing it. It has now been corrected.

Reply to Reviewer # 3:

I thank Dr S. Chabrillat for his comments and suggestions, but it is noted that the location of specific text from the paper provided in his comments is always incorrect and that some of the minor comments asking for corrections on text were not find in the ACPD version of the manuscript, neither in my own local copy nor in the submitted version available on-line. I wonder if this review is based on the actual ACPD paper. I tried to address the points raised at the best of my ability all considering.

General comments:

I believe that this paper is a useful and valuable contribution to the field of ozone data assimilation but it fails to consider related work, and many appropriate references are missing. Hence while there is no need for any additional assimilation experiment, the text should still undergo major revisions.

I agree that many publications have focussed on ozone assimilation. The introduction refers to a number of studies performed with MIPAS (Dethof 2003, Wargan et al 2005, Geer et al 2006, Dragani 2013), MLS (Jackson 2007, Feng et al 2008), and MLS+OMI (Stajner et al 2008). I might not have covered the whole available literature but I do not agree it failed to consider related work. That said, I have now added a few more references.

1. This paper gives a false impression that ozone data assimilation is still in its infancy. There is a whole community working on this topic for a long time but none of its previous work is mentioned in the introduction nor considered in the discussion. I think that the introduction should be extended to provide proper context, and that this context should be used in the discussion.
   I agree with the Reviewer that there is a vast literature on ozone data assimilation, including some review papers, and as mentioned above the introduction has been extended. However, I do not agree that the paper lacks of context. As stated in the paper, the study started from two considerations: the first is the decision made by NASA in generating the MERRA2 reanalysis, in which the SBUV nadir profile assimilation was completely replaced by that of MLS (limb profiles) and OMI total columns. This contrasts with the normal trend in (NWP/reanalysis) data assimilation of using as many observations as possible. The second consideration was the lack of plans for additional limb instruments (neither on operational nor research platforms) in the foreseeable future. I believe this, as context, has been covered in the introduction and the discussion has been tailored to it.

2. The datasets assimilated in this study were developed for the ESA project O3-CCI. That project led to several validation papers which discuss extensively the uncertainties and information content of the corresponding datasets. Since a good evaluation of observational uncertainties is paramount for data assimilation, such prior work is highly relevant for this paper. Hence the O3-CCI validation papers should be at least cited in section 4, and the choices made for the present assimilation study should be discussed in this context.
   The revised version of the manuscript has been extended to include additional references from the existing literature.

3. The averaging kernels of GOME-2 nadir profiles are still not taken into account by IFS. This is a serious limitation of the present study, because many other assimilation systems now take properly into account such vertical smoothing errors. This limitation should at least be clearly stated in the conclusions and abstract of the paper: "This study demonstrates the potentials and limitations of each dataset and instrument type" – but only in the context of data assimilation with the current IFS at ECMWF.

   Regarding the use of the averaging kernels (AKs) in the assimilation, the IFS version used here (e.g. the same run for the operational weather forecasts) does not include them at all (meaning they have not been implemented yet). Some preliminary tests have been performed, however, with the C-IFS version (now used as part of the Copernicus Atmosphere Monitoring Service). Results so far have shown at best a neutral impact of using the AKs in our ozone assimilation (A. Inness, personal communication), thus not really a limitation in practice. This has been commented in the paper.

Specific comments:

1. The introduction does not mention the results obtained in previous projects about data assimilation of stratospheric ozone, giving a false impression that ozone analyses are available only in two meteorological reanalyses (i.e. ERA-Interim and MERRA-2).

   This is not a reanalysis paper, thus I cannot justify a review of the ozone analyses in all available reanalysis, for which readers can refer to the literature (including information provided at www.reanalysis.org) and projects such as the SPARC Reanalysis Intercomparison Project (S-RIP). A comment has been added.

   Yet simultaneous assimilation of limb and nadir ozone datasets was reported and discussed as early as 2002 (Struthers et al., 2002). Nearly ten years ago, Lahoz et al. (2007) were already able to review this field. Even considering only the European projects, I believe that it is not possible to ignore such prior work as the ASSET intercomparison (Geer et al., 2006), the developments for the PROMOTE project (Viscardy et al., 2010) or the numerous results obtained for the MACC series of projects (see e.g. Inness et al., 2013; Inness et al., 2015).

   I agree that some of these papers discussed assimilation of nadir and limb data in various combinations, and a number of studies were referred to in the paper already. The introduction has been further extended as appropriate, but note my remark below and in point 5 regarding the MACC IFS ozone, as well as the fact that the MIPAS assimilation in IFS discussed in Geer et al, 2006 is partly based on the Dethof (2003) work cited in the paper.

   The absence of any citations about ozone assimilation in MACC is especially strange, because the MACC projects were coordinated by the same Institution as the author (ECMWF) and relied on a version of the same model (IFS). MACC allowed an intercomparison of the ozone analyses delivered in Near Real Time by four different systems assimilating nadir and/or limb datasets (Lefever et al., 2015). Even though the assimilation experiments were very different, this earlier study reached a very similar conclusion with a very similar Data Assimilation System (DAS):"IFS-MOZART is able to deliver realistic analyses of ozone both in the troposphere and in the stratosphere, but this requires the assimilation of observations from nadir-looking instruments as well as the assimilation of profiles, which are well resolved vertically and extend into the lowermost stratosphere".

   I intentionally avoided to mention the MACC system. The similarities between the MACC and operational IFS are related to the meteorological part. There are a number of differences between the two ozone analysis systems and not always a result obtained with one system can be extended to the other (e.g. the assimilation of the ozone-sensitive radiances was found beneficial in IFS and operationally assimilated since November 2011 but to produce neutral to slightly negative impact in the MACC-IFS system - where they are not assimilated yet; on the other hand the MACC system uses MLS, which was found to degrade the ozone analyses in the UTLS region when tested with the operational IFS – please see also my reply to point 5 below). The Lefever et al. (2015) paper has now been referred to.

Overall it is necessary to extend significantly the introduction in order to provide the missing context, and to take prior work into consideration in the discussion of the results (section 5.2.2).

Again, I suspect the Reviewer is not referring to the current version of the paper. That said and as mentioned above, the introduction has been extended as appropriate, see also my reply throughout this document.

2. P.1, line 24: the concern for the ozone decline is primarily due to the expected increase of Ultraviolet radiation at the surface. This should be mentioned in the introduction, along with a general reference about the issue.
   An explicit sentence has been added to the current paper.

3. The description of the assimilated datasets (section 2) and the data quality analysis (section 4) both fail to consider the extensive validation work realized for the O3-CCI ozone datasets. At least three papers investigate the quality of the MIPAS and GOME-2 datasets which are assimilated here. Hassler et al. (2014) present an overview of stratospheric ozone profile measurement data, document measurement techniques, spatial and temporal coverage, vertical resolution, native units and measurement uncertainties; Laeng et al. (2015) and Keppens et al. (2015) compared the available retrieval algorithms for MIPAS and GOME-2, respectively, explaining the choice of the algorithms selected for the O3-CCI datasets. These studies about observational uncertainties should be used in the description of the assimilated datasets and could be useful for the discussion of the results. Miles et al. (2015) should be cited, not only as a reference for the assimilated GOME-2 dataset, but also for its specific validation results.
   As mentioned above, the paper has been extended as appropriate.

4. The limited vertical resolution of the GOME-2 dataset should be explained more extensively, citing a specific paper (e.g. Keppens et al., 2015) in addition to the overarching reference (Rodgers, 2000). Since GOME-2 profiles have "between 5 and 6 degrees of freedom", figure 2 does not show their vertical resolution. It shows instead the vertical grid of the retrieved product. This confusion could be seriously misleading for the novice reader. While it is less of a concern thanks to its limb-viewing geometry, MIPAS does not have perfect vertical resolution either (von Clarmann and Grabowski, 2006; Laeng et al., 2015). This should also be mentioned in section 2.
   Correct, figure 2 shows the product vertical grid. As matter of fact the legend states: "*Schematic of the vertical coverage and vertical resolution provided by the CCI GOME-2 (red lines) and the CCI MIPAS (blue lines) retrieval algorithms*". I have rephrased the original legend as follows: "*Schematic of the two ozone products' vertical coverage and vertical resolution as provided….*". That said, as also pointed out by Reviewer #2, the main text could be misleading, and it has been changed from *"... the two instruments used in the present study offered different horizontal (Figure*

1) and vertical (Figure 2) coverages" to "...*the two sets of retrievals used in the present study offered different horizontal (Figure 1) and vertical (Figure 2) coverages*".

5. Description of the DAS (section 3): what is the IFS version number ("cycle") used here? How does it compare with the versions used in ERA-Interim (Dragani, 2011) and the MACC reanalysis (Inness et al., 2013) as far as ozone assimilation is concerned?
   ERA-Interim was run with a 2006 IFS version (CY31R2), the current experiments with a much newer version CY40R1 (2013), the MACC reanalysis with cycle CY36R4 (2010). The experiments here benefit from higher horizontal and vertical resolutions than the two analyses. In addition and specifically on ozone, there are many differences in terms of assimilated ozone data, ozone bias correction and associated anchor, ozone data quality control, and forecast model/chemistry. Because of these differences, a one-to-one comparison between the three corresponding ozone analyses is not immediate. These have now been added to the paper in the main text and a new table 3.

6. The modelling of ozone in IFS is not properly described, again leading to a lack of context for the discussion of the results. How is ozone photochemistry represented in the forecast model? Assuming that the parameterization by Cariolle and Teyssèdre (2007) is used here, this is not an explicit modelling of ozone photochemistry. So what does the sentence (p.5, line 15) "In this forecast model and analysis system, ozone is fully integrated (Dethof and Hólm, 2004)" mean exactly? The parameterization by Cariolle and Teyssèdre has some limitations which should be stated as they could explain some of the assimilation results.
   As said in the paper, the parametrization follows the Cariolle and Déqué (1986) [CD86] for the homogeneous chemistry. An additional term to parametrize the heterogeneous chemistry was added to the original CD86 formulation, as discussed in Dethof and Holm (2004). The scheme has been used in the same formulation since. However, the coefficients of this linear regression are updated regularly (typically every two years). These are provided by Daniel Cariolle and collaborators. The above mentioned Cariolle and Teyssèdre (2007) paper describes how they are calculated. Regarding the limitation of this scheme, I do agree it is not perfect, for example the coefficients are produced with a 2D model that does not include explicitly the heterogeneous chemistry, reason why an additional term had to be added in the IFS. Work is on-going to address these points, and preliminary results are encouraging (though on the medium and long forecast range). A comment has been added.
   The expression "ozone is fully integrated" means that ozone in IFS is a prognostic variable like, for instance, temperature, and not a climatology. The sentence has been changed from "fully integrated" to "prognostic variable".

7. P. 5, lines 23 -26: "accounting for the vertical sensitivity of any retrieved product as provided by the data averaging kernels (AKs) is currently not possible in the IFS". Please provide a reference for this limitation of IFS. Many other DAS now do take AKs into account, implementing a straightforward approach (explained e.g. by von Clarmann and Grabowski, 2006). Hence this limitation of IFS is a key caveat for this study because it limits the applicability of its findings to other DAS (see third major comment).

As mentioned above, I believe the review is not based on the current ACPD paper, as matter of fact P. 5, lines 23 -26 of the manuscript under review does not refer to the mentioned sentence, which is instead at p. 6, lines 12-15.

Regarding the specific point (see also my reply to point 3 above), in IFS the use of the AKs has not yet been implemented, thus using them is not an option. In the MACC/CAMS IFS (please note this is a different system from the operational IFS) some tests have been performed. However, the results obtained when assimilating the ozone data with the AKs compared to the same assimilation without AKs are at best neutral (A. Inness, personal communication). A comment has been added to the paper.

"With such an approximation, the vertical spread of the ozone information provided by the assimilated ozone observations depends on the background error variances and covariance (B) for ozone". Please provide a reference about this.

By design, the location where 4D-Var increments are placed, i.e. where an observation has the largest impact on the analysis, depends on both the background error and where the data show sensitivity, i.e. the AKs in the case of retrievals. A good proxy is represented by the maximum of the convolution of the background error and AKs. With the box-car approximation, each AK function is assumed to be 1 over the layer it refers to and 0 otherwise, thus the impact of the data only depends on the background error and localized where this is maximum. Han and McNally (2010) showed an illustration of it in the case of IASI radiance assimilation, using the Jacobians in place of the AKs. This part has been explained more plainly.

This approximation also fails to properly take vertical smoothing errors into account, and may constrain the analysis with a priori information contained in the retrievals. For example, in some viewing geometries the GOME-2 retrievals do not contain any usable information close to the surface.

Correct. As it is, the system assimilates the level 2 product consisting of the information provided by the measurements as such, and by the a priori used during the retrieval scheme, noting that:

- The a priori is, in general, information and as such there is nothing wrong with assimilating a product that includes it as long as that a priori does not misrepresent the ozone state variability. (I would also argue that if the information in a retrieval - whether from the measurements or the a priori – is not usable, this should also be reflected in the observation uncertainty.)
- The assimilation also relies on the model background, on the physical consistency with other variables (temperature, winds, etc…), and in complex systems - like those used in NWP – also on the simultaneous assimilation of other observations. Different assimilation techniques, e.g. KF-based, could be more sensitive to the way the observations are used, thus allowing one to better exploit these additional pieces of information than 4D-Var does. In the KF, for instance, the model is integrated forward in time, and as a new observation is available it is used to reinitialize the model before continuing the integration. 4D-Var is an initial condition problem, the aim is to find the initial condition that gives a trajectory that best fits all observations in the assimilation window ("all" here means all data used to constrain any variable included in the state vector) and the model background. For that reason, demonstrating that the assimilation of 'purged' retrievals in a 4D-Var NWP system is better than that of conventional retrievals can be very difficult (see also my comments on the use of the AKs).

8. Figures 5-6 could be quite interesting for the retrieval and validation communities who are not familiar with the estimation of observational errors allowed by data assimilation (i.e. the method by Desroziers et al., 2005). The attempt to explain this method (p.6, lines 28 - 32) is quite unclear, it should be re-written and expanded.

   The text has been rephrased. However, the reader should refer to the literature for more detailed explanation.

   P. 7, lines 18 -19: “the differences between provided and estimated uncertainties appear to be rather large”. This is an important result for the aforementioned communities (even though these uncertainties “only represent up to about 4% of the observation values”). Hence it should be shown, i.e. figures 5-6 should be expanded with similar latitude - pressure cross-sections showing the provided uncertainties and using the same color scale.

   The figures have been modified to include also the differences relative to the observations, as advised. The text has also been modified accordingly.

9. P. 7, lines 1-4: paragraph is unclear, please re-write. “The reason for this is still under Investigation at the time of writing”: indeed, this is not expected from the comparisons of ozone total columns between SBUV and GOME (Chiou et al., 2014)

   Please note the figure shows the differences between GOME-2 and the control analyses (co-located to the observations). These analyses are yes constrained by the SBUV retrievals, but also By the SCIAMACHY TCO3, in addition to IR/O3 radiances, and affected by the way the assimilation is performed (e.g. GOME-2 is assimilated as a profile, SBUV as a 6-layer partial column product) and by the model background. It is also noted that the v8 NRT SBUV product is assimilated here while the Chiou et al paper is based on the v8.6 reprocessed SBUV profiles. Without assessing the impact of each of these elements individually, one can only speculate on the element responsible for those differences. This point has been highlighted in section 4 of the paper.

10. P. 7, lines 9-11: “The first-guess check implemented in the IFS discards all observations that, after successfully pass the data quality control, show discrepancies from the background of 30DU or more over the column”. Please re-write (e.g. “after having successfully passed”).

    Actually the sentence in the manuscript under review reads as *“…, after they successfully pass the …*”. No change has been made to the paper.

    “Figure 4 shows that the observations from both instruments are well within such a threshold”. Figure 4 cannot be used to justify this Background Quality Check (BGQC) because it shows a global mean of the O-A departures while the BGQC is applied to individual observations.

    In the paper under review, the sentence reads “*Figure 4 shows that on average the observations from both instruments are well within such a threshold, although it is noted that individual observations might have shown residuals from the background larger than 30DU”.* No change has been made to the paper.

Minor Comments:

- P. 3, line 23: remove extra closing parenthesis.
  I could not find any extra closing parenthesis in the manuscript under review.
- Legend of figure 1 : state the year plotted here (i.e. 2008).
  The year has been added to the legend.

- Figure 3: right panel would be much nicer as a color-coded contour plot.

  This plot (as stated) is from Dragani and McNally (2013). I appreciate the suggestion, but unfortunately I no longer have the original data to reproduce the plot in colours.

- P. 7, line 6: “instruements” → “instruments”

  I could not find this typo in the version under review.

- P. 7, line 9: “that, after successfully pass the data quality control” – please re-write

  In the ACPD manuscript under review, the sentence actually reads “*The first-guess check implemented in the IFS discards all observations that, after they successfully pass the data quality control, show…*”.

- P. 7, line 18: “differencies” → “differences”

  I could not find this typo in the version under review.

- Table 1: Acronyms “BUV” and “ODS” are missing. I suggest to list first the satellite instruments, followed by the other acronyms

  ODS is defined in the main text (page 2 line 3). BUV, which is first mentioned in page 2 line 13, is instead defined in table 1 as advised in page 2 line 4 “*acronyms not defined in the text can be found in table 1*”. Table 1, as just said, collects all the acronyms that, if defined in the text, could have made the sentences too long. Considering that new acronyms were added during the review process, table 1 was split into two tables, one for satellite platforms and instruments, and the other for all other acronyms.

- Figure 7: color scale does not work well. Use same red-blue scale as for figures 5 and 6.

  The colour scale from figure 7 onwards had to be changed because when we tried to define 10 colour variations for the positive values from the light pink to the deep red, the differences between some of the shades were so little that made it difficult to distinguish them in the plot. As a consequence we had to change the pink-red positive part of the colour scale to use a yellow-orange-red one.

A comparative analysis of UV nadir-backscatter and infrared limb-emission ozone data assimilation

Rossana Dragani

European Centre for Medium-Range Weather Forecasts, Shinfield Park, RG2 9AX, Reading, UK

Correspondence to: Rossana Dragani (rossana.dragani@ecmwf.int)

Abstract. This paper presents a comparative assessment of ultra violet nadir-backscatter and infrared limb-emission ozone profile assimilation. The Meteorological Operational Satellite A (MetOp-A) Global Ozone Monitoring Experiment 2 (GOME-2) nadir and the ENVISAT Michelson Interferometer for Passive Atmospheric Sounding (MIPAS) limb profiles, generated as part of the European Space Agency Climate Change Initiative, were individually added to a reference set of ozone observations and assimilated in the European Centre for Medium-Range Weather Forecasts (ECMWF) data assimilation system (DAS). The two sets of resulting analyses were compared with that from a control experiment, only constrained by the reference dataset, and independent, unassimilated observations.

Comparisons with independent observations show that both datasets improve the stratospheric ozone distribution. The changes inferred by the limb-based observations are more localized and, in places, more important than those implied by the nadir profiles, albeit they have a much lower number of observations. A small degradation (up to 0.25 mg/kg for GOME-2 and 0.5 mg/kg for MIPAS) is found in the tropics between 20 and 30 hPa. In the lowermost troposphere below its vertical coverage, the limb data is found able to modify the ozone distribution with changes as large as 60%. Comparisons of the ozone analyses with sonde data show that at those levels the assimilation of GOME-2 leads to about 1 Dobson Unit (DU) smaller root mean square error (RMSE) than that of MIPAS. However, the assimilation of MIPAS can still improve the quality of the ozone analyses, and - with a reduction in the RMSE up to about 2 DU - outperform the control experiment thanks to its synergistic assimilation with total column ozone data within the DAS.

High vertical resolution ozone profile observations are essential to accurately monitor and forecast ozone concentrations in a DAS. This study demonstrates the potential and limitations of each dataset and instrument type, as well as the need for a balanced future availability of nadir and limb sensors, and long-term plans for limb-viewing instruments.

1 Introduction

Since the discovery of its global decline in early 1980s (Farman et al.,1985, Solomon et al.,1986), ozone has attracted the interest of both the scientific community and policy-makers (e.g. WMO, 2014a, b, and earlier assessments), as well as of the general public. Such an interest is driven and justified by the crucial role ozone plays in the chemistry and in the thermal structure of the atmosphere: a change in the amount of ozone could lead to a warming/cooling of the Earth (depending on the altitude where the change occurs), it could affect the Earth's climate (e.g. McLinden and Fioletov, 2011), ~~as well as~~ and human life - both as consequence of its stratospheric decline, which would lead to increased ultraviolet radiation reaching the surface (discussed in a number of quadrennial assessments and progress reports of the United Nations Environment Programme, UNEP, e.g. UNEP, 2012), and for its role as pollutant (e.g. EPA, 2015; Madronich et al., 2015).

The concern for the ozone decline - at an observed rate for the global total column of 2.5% between the 1980s and early 1990s (WMO, 2014a) - led to the ~~signature~~ signing of an international treaty (the 1997 Montreal Protocol and subsequent amendments) to regulate the release in the atmosphere of Ozone-Depleting Substances (ODSs, e.g. SPARC, 1998; Randel and Wu, 2007). The 2014 WMO (acronyms not defined in the text can be found either in table 1, if they refer to satellite platforms and instruments, or in table 2) assessment of ozone depletion stresses that although recent studies agree that the start of the 21^{st} century represented a turning point in the global total column ozone trend, which now sees a slow increase in its abundance, it is not yet clear whether the current global increase can be attributed to a reduction in the amount of ODSs (WMO, 2014a). Projections of the ozone future evolution seem to agree that the ozone amount will recover towards the stratospheric levels registered before the 1980s by the end of the current century (e.g. Barnes et al., 2014). However, they do not agree on when such a recovery will be achieved (Velders and Daniel, 2014). Thus, the attention in closely monitoring the ozone evolution remains high.

Satellites have been critically important in delivering valuable information to continuously monitor this important atmospheric gas over the last four decades. The first satellite ozone measurements date back to the early 1970s when the nadir-pointing BUV instrument was launched on board of the Nimbus-4 satellite (Heath et al., 1973) followed by a successful number of similar instruments (SBUV and SBUV/2) on Nimbus-7 in 1978 and several NOAA platforms (Bhartia et al., 1996) that stretch up to present. In addition, a large variety of instruments to measure ozone have been launched since 1978 and many more are planned for the next ten to fifteen years. A rather complete account of the past and planned missions able to deliver ozone measurements can be found at http://www.wmo-sat.info/oscar/gapanalyses?view=108. This wealth of ozone observations offer the opportunity to derive a record of over forty years to study the ozone variability and changes, as well as derive trends. However, they are highly inhomogeneous in viewing geometry (nadir, limb, occultation, or a combination of them), in observed spectral range (UV, visible, near-infrared, thermal infrared, microwave) and spectral resolution, in spatial coverage (GEO vs. LEO platforms) and spatial resolution (ranging from a few hundred kilometres to a few kilometres).

A powerful way to integrate and exploit such a wealth of observations in a way that is consistent with their uncertainties is within a Data Assimilation System (DAS) with many successful examples covering the medium-range weather forecasting and reanalysis available in the literature (e.g. Daley, 1991; Courtier et al., 1994; Veersé and Thépaut, 1998; Rabier et al., 2000; Kalnay, 2003; Dee et al., 2011). Examples of ozone assimilation date back to the late 90s, when NWP centres seized the opportunity of exploiting these observations to improve the radiance assimilation, and constrain the wind analyses (e.g. Derber and Wu, 1998; Holm et al., 1999; Jackson and Saunders, 2002; Struthers et al., 2002). Polavarapu et al. (2005) presents an overview of some of the challenges of middle atmosphere data assimilation, in general, and ozone data assimilation, in particular. The latter is also reviewed by Lahoz et al. (2007). To exploit data synergy, guarantee redundancy, thus resilience to sudden changes in the observing system, and to be less sensitive to limitations of a particular instrument design, the general tendency in both NWP and reanalysis productions is to assimilate as many observation types as possible

~~(e.g. Uppala et al., 2005; , 2011; , 2011)~~. An extensive literature exists for both types of production. The state-of-the-art in the use of observations, particularly from satellites, in NWP was presented during the ECMWF's 2014 Annual Seminar (*Use of Satellite Observation in Numerical Weather Prediction*, 8-12 September 2014, Reading, UK, proceedings available at http://www.ecmwf.int/en/learning/workshops-and-seminars/past-workshops/). In the context of recent reanalyses, an account can be found at http://www.reanalysis.org. Furthermore, within SPARC a Reanalysis Intercomparison Project (S-RIP, http://s-rip.ees.hokudai.ac.jp/) was started in 2012 aiming at comparing a number of recent reanalysis data sets (for various key diagnostics) to understand the causes of differences between them. Worthy of mention is a S-RIP special issue in *Atmos. Chem. Phys.* on the project assessments (Eds. Haynes, Stiller, and Lahoz), http://www.atmos-chem-phys.net/special_issue829.html. ~~However~~ In contrast to the NWP and reanalysis general tendency of using as many observations as possible, there are also cases where precise choices on data selection are made from the outset. For instance, in the latest NASA/GMAO reanalysis (MERRA-2), ~~McCarty et al. (2015)~~ Bosilovich et al. (2015) explain that their ozone analyses were constrained using the SBUV ozone products (version 8.6 McPeters et al., 2013) until 2004 when the ozone assimilation was switched to a combination of ozone products retrieved from MLS (Froidevaux et al., 2008) and OMI (Bhartia, 2002) both flying on board the Aura platform (Schoeberl et al., 2006). The former is a limb-emission sensor providing ozone profiles retrieved from microwave measurements; the latter is a nadir backscatter instrument providing total column ozone from measurements in the visible and ultraviolet spectral range.

An analysis of the Earth Observation capability planned for the next decade up to about 2025 shows a generally good temporal coverage from nadir-looking instruments with sensitivity to ozone, on either LEO or GEO platforms. Many of these instruments are or will be operated as part of operational missions thus ensuring long term data provision (for instance the forthcoming Sentinels 4 and 5). The same cannot be said for instruments with a limb viewing geometry. These satellite instruments are rarely operational and their combined data record show significant gaps. Yet, limb sounders are ~~greatly~~ very important when coming to ozone monitoring as their high vertical resolution can lead to significant differences when deriving trends. Furthermore, when designed to have a sensitivity to spectral ranges such as the infrared (like in the case of the ENVISAT MIPAS sensor, Dubock et al. (2001); Fischer et al. (2008)) or sub-millimitre spectral range (like in the case of the Aura MLS sounder), these instruments can provide measurements in night-time conditions thus overcoming one of the most severe limitation of UV-Visible measurements. Many studies have been published in the literature to assess the impact of limb ozone data, particularly from the MIPAS and MLS instruments. The results from the first assimilation trial using the MIPAS ozone profiles within the ECMWF DAS were discussed by Dethof (2003). In their results, MIPAS assimilation was found able to improve the quality of the ozone analyses at high latitudes in the winter hemisphere, in general, and provide a better characterization of the Antarctic ozone hole, in particular. These results were confirmed by later studies. For instance, Wargan et al. (2005) discussed the improvements in the ozone analyses, particularly in the polar night region and below the ozone maximum. Geer et al. (2006) were able to reproduce accurately the unusual event of the ozone hole split that occurred in September 2002 in the Met Office system. Using two configurations of the ECMWF DAS, Dragani (2013) , Dragani et al. (2015) discussed the assimilation of ~~repeated the Dethof study using~~ MIPAS using retrievals obtained after a severe instrumental problem triggered changes in the ~~MIPAS~~ instrument spectral specifications. Their results confirmed that the assimilation of MIPAS ozone data could substantially improve the quality of the ECMWF ozone analyses compared to a baseline only constrained by UV-Visible ozone data. A number of studies focused on the assimilation of the MLS ozone profiles recording improvements in the stratospheric ozone distribution (e.g. Jackson, 2007; Feng et al., 2008; Stajner et al., 2008). In particular, Stajner et al. (2008) also discussed how the assimilation of the MLS ozone profiles can be exploited synergistically with total column ozone from OMI (also on board of the NASA's Aura satellite) to improve the tropospheric ozone column analyses, in addition to the vertical ozone distribution. This conclusion is confirmed by a recent study. Lefever et al. (2015) found that a combination of ozone retrievals from nadir-looking UV-Vis instruments and the MLS could provide better constraint on the tropospheric analyses than that provided by MLS alone.

All these studies agree that limb measurements are valuable, yet the need for these observations does not match the long term availability and plans for limb instruments. Thus, do the results from the available studies and the above considerations, as well as the pragmatic decision taken in constructing the MERRA-2 reanalysis call for more sustained and longer term plans for limb measurements than currently available?

By presenting a comparative assessment of the impact on the ECMWF ozone analyses of assimilating either nadir or limb ozone profiles to the same ozone baseline, this study aims at addressing the above question. The present work focuses specifically on the point of view of data assimilation with potential implications for a large part of the ozone research spectrum spanning from NWP to climate reanalysis, from air quality to stratospheric trends and variability.

The paper is structured as follows: the observations used in the present study are presented in section 2; the DAS set-up is discussed in section 3. A preliminary assessment of the data quality prior the assimilation is discussed in section 4 while the assimilation results are analysed in section 5. Concluding remarks and recommendations are summarized in section 6.

2 The ozone datasets

This section focuses on the two ozone products that were used to address the question raised in section 1, namely datasets retrieved from measurements of the MetOp-A GOME-2 instrument and the ENVISAT MIPAS sounder. Additional ozone information was used to constrain all assimilation experiments. This is described in section 3.

Launched on board of the EUMETSAT MetOp satellites, GOME-2 (Callies et al., 2000; Munro et al., 2006) is one of the new generation European instruments. The first two GOME-2 instruments were part of the MetOp-A (October 2006) and MetOp-B (September 2012) payloads, respectively. A third one is scheduled to be launched on MetOp-C in 2017. This series of instruments continue the long-term monitoring of atmospheric ozone started by the ERS-2 GOME and ENVISAT SCIAMACHY instruments, whilst generally characterized by much smaller pixels than their predecessors to make their observations useful for air quality forecasting (e.g. Hao et al., 2014). Like its predecessors, GOME-2 is an optical spectrometer that measures the Earth's backscattered radiance and extraterrestrial solar irradiance, in the ultraviolet and visible part of the spectrum (240-790 nm). Its high spectral resolution (between 0.2-0.4 nm) permits to obtain an excess of 4000 spectral points from four detector channels per individual measurement. The first two channels (covering the 240-315 nm and 310-403 nm spectral regions) are important for ozone retrieval. In the period used in this study (2008), the GOME-2 instrument was characterized by an orbit swath of 1920 km, and a typical footprint in the forward scan of 80 km (across track) by 40 km (along track), allowing for a daily global coverage.

The second dataset was retrieved from the MIPAS measurements (Fischer et al., 2008). This was a Fourier transform spectrometer launched in 2002 on a Sun synchronous polar orbit on board of ENVISAT. The instrument measured thermal emission at the atmospheric limb in the mid-infrared spectral range between 4.15 and 14.6 μm (or 680-2275 cm^{-1}), thus permitting vertical profile retrieval of several minor atmospheric constituents. It was initially operated with a high spectral resolution of 0.025 cm^{-1} reduced to 0.0625 cm^{-1} in January 2005 to resume operations after instrumental problems occurred in March 2004. The reduced spectral resolution led to a proportional reduction in the measurement time from 4.5 sec to 1.8 sec that was exploited to increase the number of measured spectra in each scan in order to have a finer vertical limb grid in the UTLS region and an altitude range coverage from 6 to 70 km. The reduction in the measurement time was also exploited to improve the horizontal resolution between two contiguous limb scan measurements. The instrument was operated until April 2012 when communication with the satellite was lost.

This study focuses on the assimilation of the GOME-2 and the MIPAS ozone profiles retrieved by the ozone consortium (O3-CCI hereafter) created as part of the ESA Climate Change Initiative (CCI, Plummer, 2009, ~~www.esa-cci.org~~ http://cci.esa.int). It is noted that several algorithms have been developed over the years to retrieve ozone profiles from the GOME-2 and MIPAS measurements besides those implemented by the O3-CCI. As only the O3-CCI datasets are used here for both instruments, explicit references to the algorithm are omitted hereafter. For simplicity, the datasets will be usually referred to with the name of their corresponding instrument unless misleading. The CCI programme was established, on one hand, in response to the GCOS call for climate-quality satellite data, and, on the other hand, to realise the full potential of the ESA global Earth Observation (EO) archive. CCI aims at providing stable, long-term, satellite-based ECV data products to support climate modellers and researchers. Particular attention is paid in characterizing the observation uncertainty and providing comprehensive, fully traceable information on calibration and validation, long term algorithm maintenance, data curation and reprocessing.

The O3-CCI retrieval scheme for nadir ozone profiles was initially developed at RAL for the ERS-2 GOME instrument (Munro et al., 1998). These ERS-2 GOME ozone profiles retrieved with the RAL algorithm were found beneficial in improving the quality of the ERA-Interim ozone reanalysis in the middle Stratosphere (Dragani, 2011). ~~It~~ The retrieval scheme is based on an optimal estimation algorithm, which combines measurements and an *a priori* in a way consistent with their error covariance matrices (Rodgers, 2000). It uses three sequential steps to retrieve and improve the ozone information from the Hartley (266-307 nm) and Huggins (323-335 nm) bands. These provide altogether between 5 and 6 degrees of freedom for signal (Rodgers, 2000; Keppens et al., 2015)). The *a priori* is derived from the McPeters et al. (2007) climatology while the temperature and pressure profiles are taken from the ERA-Interim reanalysis (Dee et al., 2011). An empirical correction is included in the CCI product to address the instrument throughput degradation at the shorter UV wavelengths (Lang et al., 2009). The cloud radiative transfer is not modelled explicitly. Instead an effective Lambertian surface albedo is co-retrieved. Miles et al. warns that, in presence of clouds, this solution leads to a negative bias in retrieved ozone below the cloud top. Particular attention was paid in characterizing the various sources of uncertainty in the resulting retrieval. A detailed overview of this product is presented by Miles et al. (2015).

The O3-CCI retrieval scheme for limb ozone retrievals was jointly developed by the Institut fur Meteorologie und Klimaforschung (IMK) and the Instituto de Astrofísica de Andalucía (IAA) (von Clarmann et al., 2003, 2009). The scheme makes use of all the measurements within the 740-800 cm^{-1} and 1060-1110 cm^{-1} spectral ranges after filtering out cloud-contaminated spectra. It assumes local thermodynamic equilibrium conditions, normally verified in the troposphere and most of the stratosphere in the selected spectral regions (Echle et al., 2000). In this study we made use of the version fv0003 dataset. A full description can be found in Sofieva et al. (2013). An ozone profile product retrieved from the MIPAS measurements was also assimilated in the ERA-Interim reanalysis between October 2003 and March 2004. However, differences exist between that product and the one used here. The one assimilated in ERA-Interim was the near-real-time product retrieved with the operational ESA level 2 algorithm, named Optimized Retrieval Model (ORM, Ridolfi et al., 2000; Raspollini et al., 2006). It was retrieved from the measurements made before the instrumental problem of March 2004 that was overcome with a number of changes in the instrument set-up, including the spectral characteristics. The CCI product consists of reprocessed ozone data available from January 2005 onwards, thereby based on measurements using the modified set-up.

Due to differences in spectral ranges and viewing geometries, the two ~~instruments~~ sets of retrievals used in the present study offered different horizontal (Figure 1) and vertical (Figure 2) coverages. The geographical distribution of the GOME-2 data (top panel of Figure 1) varies with latitude and time with the daily data count that ranges from a few tens to a few hundreds over a 2° latitudinal band. In contrast, MIPAS offers a more homogenous data coverage (bottom panel of Figure 1) albeit a much lower data count than GOME-2 with a MIPAS:GOME-2 data count ratio that ranges from 1:2 up to 1:40.

The vertical coverage and vertical resolution of the two products are schematically shown in Figure 2. Overall, the O3-CCI nadir profiles (NPO_3 hereafter) are provided on 19 vertical levels spanning the atmosphere from surface up to 0.01 hPa. The O3-CCI limb ozone profiles (LPO_3 hereafter) are also derived on a fixed vertical grid consisting of 32 vertical levels spanning the region of the atmosphere from 0.05 down to 300 hPa.

3 The data assimilation system

The data assimilation system used here is a low resolution version of the ECMWF Integrated Forecasting System (IFS). IFS is a comprehensive atmospheric forecasting system that simulates the dynamics, thermodynamics and composition of the Earth's atmosphere and interacting parts of the Earth-system. It includes three components: a global spectral atmospheric model, an ocean wave model, and an ocean model that simulates the ocean circulation and sea ice.

At the time of writing, the global high resolution spectral model uses a resolution truncation of ~~T_l1279~~ $T_{co}1279$, which corresponds to a ~~linear~~ cubic octahedral reduced Gaussian grid of ~~16~~ about 9 km grid spacing, and 137 vertical levels spanning from surface to 0.01 hPa (corresponding to an altitude of about 80km). The data assimilation is performed using a four-dimensional variational (4D-Var) data assimilation scheme (Rabier et al., 2000) formulated in terms of increments (Courtier et al., 1994; Veersé and Thépaut, 1998) and with a 12-hour assimilation window. At the time of writing, over seventy different data sources were received and monitored daily from satellite alone. Variational quality control and first-guess checks are carried out for all assimilated data. A variational bias correction scheme (VarBC, Dee, 2005) is used to automatically detect and correct for observation systematic biases. VarBC is formulated for all observations as linear regressions of a number of bias predictors that are observation type, and sensor dependent, as well as geographically varying. The coefficients of those regressions are part of the state vector and computed during the 4D-Var minimization, thus updated every assimilation cycle. Detailed information on the ECMWF system is available at www.ecmwf.int/en/research/modelling-and-prediction.

In this forecast model and analysis system, ozone is ~~fully integrated~~ a prognostic variable (Dethof and Hólm, 2004). The ozone model is parameterized according to an updated version of the Cariolle and Déqué (1986) scheme (CD86 hereafter). In ~~the~~ CD86 ~~parameterization~~, the ozone ~~continuity equation~~ photochemistry is ~~expressed~~ parametrized as a linear relaxation towards a photochemical equilibrium for the local value of the ozone mixing ratio, the temperature, and the overhead ozone column. In addition, an ozone destruction term that depends on the equivalent chlorine content for the actual year is included to parameterize the heterogeneous chemistry. For simplicity, the CD86 scheme and the additional term for the heterogeneous chemistry will be referred to as *modified CD86*. The coefficients of the modified CD86 linear regression are updated regularly over the years (thanks to the collaboration with Daniel Cariolle). These are calculated as described in Cariolle and Teyssédre (2007). It is noted that the forecast model presents a number of limitations, for example the coefficients are produced with a 2D model that does not include explicitly the heterogeneous chemistry, which had to be included in the IFS via an additional term. Work is on-going to test alternatives that could address these points. The preliminary assessment shows encouraging results in terms of the impact on the medium and long forecast ranges (B. Monge-Sanz, personal communication).

The variational bias correction scheme originally introduced to automatically detect and correct for observation systematic biases in the radiances (Auligné et al., 2007) was later extended to retrieved ozone products (Dragani, 2009). For relevance in the discussion, it is noted that accounting for the vertical sensitivity of any retrieved product as provided by the data averaging kernels (AKs) is currently not possible in the IFS. Preliminary tests assimilating ozone retrievals with AKs were performed using a modified version of the IFS that was developed as part of the series of FP6, FP7 and H2020 funded projects GEMS (Hollingsworth et al., 2008), MACC Simmons (2010) and its follow-on projects, and here referred to, for simplicity, as the MACC-IFS. The results obtained from these tests compared to the same assimilation without AKs were at best neutral (A. Inness, personal communication), and thus no further pursued. ~~Thus~~ When the AKs are neglected, retrieved observations, including ozone data, are assimilated with a box-car approximation, in which each AK function is assumed to be 1 over the layer it refers to and 0 otherwise (i.e., with perfect AKs). By design of variational data assimilation methods, the location where an observation can be expected to have an impact on the analysis depends on both the background error and the region where the observation shows sensitivity, as expressed here by the AKs. As an approximation of what happens within the data assimilation, the largest impact can be expected in the region where their convolution is maximum. Han and McNally (2010) showed an illustration of it applied to the assimilation of IASI radiances where the IASI Jacobians were used instead of the AKs. With ~~such an~~ a box-car approximation, the vertical spread of the ozone information provided by the assimilated ozone observations depends on the background error variances and covariance for ozone. This is particularly the case of the assimilation of TCO_3. An example of background error profile and vertical correlation matrix obtained from Dragani and McNally (2013) is given in Figure 3. This figure shows that the ozone background error is largest in the stratosphere between 20 and 80 hPa (panel a), implying that the ozone increments generated by the assimilation of TCO_3 products is most likely spread over this region of the atmosphere. Similar considerations are less straightforward when assimilating profiles, as the region of the atmosphere where an impact can be made also depends on the observation error characteristics.

Because of the explicit mention, it is noted that differences exist between the ozone analysis system used here, and the one that was used in the ERA-Interim reanalysis discussed in detail by Dragani (2011), as well as the one that used for the MACC reanalysis (Flemming et al., 2009; Inness et al., 2013). These differences are summarized in table 3 for the period considered in this study. For other overlapping periods between the two ozone reanalyses, the reader is advised to refer to the corresponding literature.

3.1 Experiment set-up

Three assimilation experiments (a control, referred to as Exp/Ctrl, and two perturbation experiments) were run for the period July-October 2008 using a low resolution version of the IFS with an horizontal truncation of T_l511, which corresponds to about 40 km grid resolution, and 91 vertical levels spanning the atmosphere from surface up to 0.01hPa.

With the exception of observations sensitive to ozone, the three experiments assimilated the same set of observations consisting in those available in the ECMWF archive for the period under study. Regarding the ozone observations, the Exp/Ctrl was constrained with ozone sensitive radiances in the infrared (from HIRS, IASI, and AIRS sounders as described in Dragani and McNally (2013)) and ozone retrievals. The latter consisted of a total column ozone (TCO_3) product retrieved at KNMI from the ENVISAT SCIAMACHY measurements (Brinksma, 2004; Eskes et al., 2005; Antón et al., 2011), and ozone retrievals from NOAA-16, -17 and -18 SBUV/2 instruments (Bhartia et al., 2012). These observations were all taken from the ECMWF operational data archive. It is noted that the SBUV/2 data produced by NOAA as 21-level ozone profiles were converted into a six-layer product (Top-1 hPa, 1-2 hPa, 2-4 hPa, 4-8 hPa, 8-16 hPa and 16 hPa-surface) at ECMWF.

The two perturbed experiments were then run with exactly the same configuration and observing system of Exp/Ctrl plus the assimilation of either the MetOp-A GOME-2 NPO_3 product or the ENVISAT MIPAS LPO_3 dataset. In the remainder of this paper, the former perturbed experiment will be referred to as Exp/GOME2 while the latter will be referred to as Exp/MIPAS.

4 Data quality analysis

A preliminary data analysis was performed for both datasets prior to the assimilation. The purpose was to determine the level of agreement between observed and modelled ozone fields, and, due to their importance in data assimilation, provide a preliminary assessment of the observation uncertainties. This is motivated by the fact that NWP-based systems tend to be conservative when using new observations. This can be achieved either by being particularly selective on the data that are actually assimilated (for instance through first-guess checks that depend on the level of discrepancy between model and observations) or by limiting their impact by inflating their provided uncertainties, which determine the weight the observations themselves have on the analyses. Although significant progress has been made in understanding and characterizing most sources of error in the observations, as well as in following best practice, limitations in those estimates still exist. This is either because some of the sources of uncertainty are particularly difficult to characterize or because they are unknown. Of these two aspects, the former is investigated by examining the observation-minus-analysis (O-A) residuals (or simply analysis departures) that measure the discrepancy between the observations and co-located analyses. Here, the analyses were taken from Exp/Ctrl and thus not constrained by the observations under assessment. The latter aspect is instead investigated by comparing the assumed (i.e. provided) observation uncertainty with an estimate derived with the Desroziers et al. (2005) method. Under the hypotheses of optimality in the analysis and separation of scales between background and observation, the observation uncertainty (meant as the standard deviation) can be estimated using the first-guess and analysis residuals from the observations. This method is a simple consistency diagnostic, in which the observation error covariance matrix, R can be estimated using the first-guess and analysis departures from the observations ((O-B), and (O-A), respectively), as $\mathrm{R} = \mathrm{E}\{(\mathrm{O}-\mathrm{B})(\mathrm{O}-\mathrm{A})^{\mathrm{T}}\}$, where $\mathrm{E}\{\}$ and $()^{\mathrm{T}}$ indicate the expectation and the transpose operators, respectively. This diagnostic was derived in the case of an optimal analysis method and it is applicable under the assumption that the correlation scales of background and observation error are sufficiently different (Desroziers et al., 2009). This diagnostic has successfully been used to estimate R with operational observations (e.g. Bormann and Bauer, 2010; Bormann et al., 2010; Stewart et al., 2014).

Time series of the global mean analysis departures for both instruments are presented in Figure 4. To ease the comparison, these are provided in terms of integrated column over the region between 0.05 and 100 hPa, which is covered by both instruments. Despite the fact that the analyses used to derive the statistics are largely constrained by UV-retrieved ozone measurements, GOME-2 is the instrument that shows the largest mean discrepancy, with O-A differences being positive during the four month period assessed and a mean value of about 5DU (about 1.5% of the global mean ozone column over that region, estimated as 90% of a global TCO_3 value of 350DU). The reason for this is still under investigation at the time of writing. Possible reasons for this outcome could be: *i.* differences in the information provided by the UV-based retrievals (i.e. the GOME-2 NPO3 and those assimilated in Exp/Ctrl, the SBUV partial columns and the SCIAMACHY TCO3); *ii.* differences in the information provided by the GOME-2 NPO3 and the IR/O3 radiances; *iii.* the data assimilation itself (that is the model background and/or the way the ozone observations are treated within the DAS). Without assessing the impact of each of these elements individually, conclusions cannot be drawn. Regarding point *i.*, Chiou et al. (2014) compared total column products generated from the a newer, reprocessed version of the SBUV dataset (v8.6), and the GOME-2 NPO3 retrievals finding differences in the monthly zonal mean well within 1%. The MIPAS measurements indicate that during July-August the global mean ozone analyses are about 5DU too high (they were 10DU too low based on the GOME-2 data). Here, the discrepancy between the two instruments is very likely related to different coverage of the two instruments, particularly over the high latitudes in the SH, as shown in Figure 1. During September-October, the O-A residuals for the two instruments are more similar and they both indicate an underestimation of the ozone analyses of about 5DU above 100hPa. The first-guess check implemented in the IFS discards all observations that, after they successfully pass the data quality control, show discrepancies from the background of 30DU or more over the column. Figure 4 shows that on average the observations from both instruments are well within such a threshold, although it is noted that individual observations might have shown residuals from the background larger than 30DU.

The second aspect, which has a bearing in the level of impact of the two instruments on the ozone analyses, is the characterization of the observation uncertainty. An estimate for both instruments was derived using first-guess and analysis residuals from the observations as discussed in Desroziers et al. (2005). Figure The left panel of figure 5 shows the relative difference between the estimated and provided uncertainty for the GOME-2 NPO_3 product. The provided uncertainty appears to be larger than that estimated at all latitudinal bands from surface up to about 5hPa while in the upper stratosphere it is smaller than the estimated uncertainty, especially at high latitudes in the summer hemisphere (the Northern Hemisphere, NH, in this case) and at midlatitudes in the winter hemisphere (i.e. the Southern Hemisphere, SH, in this case). Despite the differences between provided and estimated uncertainties appearing to be rather large, they only represent up to about 4% of the observation values (~~not shown~~ right panel of figure 5). Keppens et al. (2015) also found an overestimation of a few percent in the comparisons of this product against ground based measurements (refer to their RAL product v2.1 in VMR). From a data assimilation perspective, an uncertainty overestimation, here seen at most levels from surface up to about 5hPa, is generally desirable to account for the effect of vertical correlations in the observation error that are normally neglected in most data assimilation systems, including the IFS. The only consequence of such an overestimation would be to limit the impact of the corresponding data, thus no correction is required in this region of the atmosphere. The underestimation of the observation uncertainty in the upper stratosphere is more of a concern. This is because underestimated uncertainties increase the impact of the corresponding observations and if that is associated with poor quality data the assimilation can lead to negative impact on the analyses and forecasts. As the ozone abundance above 5hPa substantially reduces with height, it is expected that if there was a detrimental effect on the ozone analyses as a consequence of assimilating these observations without any correction on their uncertainty, this should generally be negligible or at worst small. Thus, it was decided not to apply any correction to the GOME-2 observation uncertainties above 5hPa.

The vertical cross-section of the relative difference between the estimated and the provided uncertainty for the MIPAS LPO_3 is shown in Figure 6. In the lower and middle stratosphere, the MIPAS uncertainty appears to be generally overestimated compared with its estimated equivalent, with the exception of the high latitudes in the winter hemisphere (SH). The overestimation is typically within 4% relative to the observation values. This appears consistent with the residual variability estimate provided by Laeng et al. (2014) when assessing the MIPAS error budget using comparisons with several reference satellite datasets, for instance MLS. The diagnostic they use should be consistent with the one applied here for small differences in the provided and estimated uncertainties. The provided field is on average smaller than the estimated uncertainty in the upper troposphere, especially in the winter hemisphere, and in the upper stratosphere. The underestimation of the MIPAS uncertainty in the upper troposphere and in the upper stratosphere can potentially produce a negative impact on the analyses and forecasts. Using the same argument discussed above, any detrimental effect in the upper stratosphere is expected to be generally from negligible to small. Thus, no correction is made at those levels. The possible underestimation of the MIPAS uncertainty in the upper troposphere mostly affects the mid and high latitudes in the SH. It is noted that the period under consideration is July-October, thus winter and spring times in the SH. A data assimilation system strongly constrained by UV observations can show limitations in polar night conditions, while facing an important transition in springtime, when the UV data availability at these latitudes slowly restarts (as shown for instance in the top panel of Figure 1). Thus, it could be argued that the model-based estimate of the MIPAS uncertainty at high latitudes in the SH is less reliable than that provided at other latitudinal bands. A hint of a small underestimation of the uncertainty compared to its model equivalent in the middle troposphere in the southern midlatitudes can also be found for GOME-2 (Figure 5). Arguably, this could also point to problems in the assumed background errors in this region of the atmosphere and time of the year. Based on these considerations, it was decided to test the assimilation of both datasets without applying any correction on the provided uncertainties.

5 Assimilation results and discussion

5.1 Impact on the analyses

The comparison of the Exp/GOME2 and Exp/MIPAS analyses with those from Exp/Ctrl is shown in Figure 7. Both instruments contribute to an increase in the extra-tropics total column ozone, although with different amounts. In contrast, their impact differs in the tropical region. Here, the GOME-2 assimilation contributes to an overall reduction of up to 10 DU in the mean TCO_3 (top panel of Figure 7) while that of MIPAS shows a small increase of up to about 2 DU (bottom panel of Figure 7).

The mean vertical cross-section of the ozone analysis differences shown in Figure 8 also depicts a consistent picture between the two datasets at most levels and latitudes, particularly in the extra-tropics. To account for the difference in the order of magnitude of the analysis residuals in the troposphere and stratosphere, the vertical cross-sections were also scaled to a mean ozone profile computed from the Exp/Ctrl ozone analyses over the three month period under consideration and shown in the bottom panels of Figure 8. On average, both datasets tend to increase the ozone analyses in a thick layer between 20 and 70 hPa with differences of up to 2 mg/kg from the Exp/Ctrl. Over this layer, the largest difference in terms of impact between the two instruments can be seen at high latitudes in both hemispheres where the assimilation of the GOME-2 dataset seems to add more ozone with respect to the Exp/Ctrl than done by the assimilation of the MIPAS profiles. With a mean difference of about 1 mg/kg over the 20-70 hPa layer, the estimated mean column change produced by either dataset in the middle stratosphere adds up to an increase of around 24 DU.

In the upper stratosphere and in the UTLS region, the assimilation of either the GOME-2 or the MIPAS ozone profiles leads to a reduction of the ozone analyses of an amount that depends on the instrument and latitudinal band. The largest difference is noticeable in the tropics, in the region of the ozone mixing ratio maximum around 15 hPa. This could also explain the discrepancy found in Figure 7 in the tropics.

In the lower and middle troposphere, both the assimilation of MIPAS and that of GOME-2 profiles tend to remove ozone in the extra-tropics. This is also the case for the tropical troposphere in the case of MIPAS while the assimilation of GOME-2 leads to a tropospheric ozone increase in this latitudinal band.

An aspect that emerges from Figure 8 (bottom panels) is that the assimilation of the MIPAS ozone profiles can substantially modify the ozone distribution in the lower and middle troposphere with changes as large as 60%. Such a change occurs despite the fact that a limb instrument vertical coverage does not normally extend to pressure levels below about 300 hPa, and the observation uncertainty in the upper troposphere is normally larger than in the stratosphere, thus having a lower weight in the data assimilation system. Yet, the structure of the changes are consistent in sign and often in amplitude with those inferred by the GOME-2 nadir ozone profiles. This aspect will be further discussed in section 5.2.2.

5.2 Comparisons with independent data

To determine whether the changes introduced by the assimilation of either the GOME-2 or the MIPAS ozone profiles represented an improvement or a degradation, the quality of the ozone analyses from the experiments described above was assessed in terms of the analysis agreement with independent, unassimilated ozone observations. Ozone profiles retrieved from MLS (version ~~3.04~~ 3.3) and ozone sondes available at the World Ozone and Ultraviolet Radiation Data Centre (WOUDC) were used as independent ozone references for the stratosphere and troposphere/lower stratosphere, respectively.

The MLS observations offers a near global horizontal coverage from about 82°S to 82°N. The used vertical range spans from about 316 hPa up to 0.01 hPa, which also coincides with the model top, and a vertical resolution of about 3 km. Froidevaux et al. (2008) compared the previous v2.2 MLS ozone retrievals with matching ozone data from ground-based and satellite observations, and found that the differences are generally within 5% in most of the stratosphere, although residuals of 10-20% were found in the lower stratosphere. Livesey et al. (2013) reports differences between the v3.3 and v2.2 MLS ozone profiles typically within 1-2% in the stratosphere and lower mesosphere.

The ground-based stations that are included in the WOUDC archive are spatially inhomogeneous, with the highest availability over Europe and North America. The archive includes measurements performed with both electrochemical concentration cell (ECC), and Brewer Mast (BM) sonde types. The ECC precision was estimated by Komhyr et al. (1995) to be within $\pm 5\%$ in the vertical range between 200 and 10 hPa, between -14% and +6% above 10hPa and between -7% and +17% below 200 hPa. The same order of precision was found by Steinbrecht et al. (1998) for the BM sondes. A data check is used to discard all soundings with a pressure burst either at or below 40 hPa.

5.2.1 Methodology

The comparisons were performed in two stages. First, the 3D ozone analysis closest in time to the independent observation was interpolated at the observation location. This gives a temporal mismatch of up to 3 hours between the observation sensing time and the analysis valid time. The second stage takes care of the vertical interpolation. This is done by ~~degrading~~ interpolating the profile with the highest vertical resolution to the coarsest grid. Only the levels spanning the region of the atmosphere encompassed by both datasets are used, and in the specific case of the ozone sondes only sounding that reached at least 40 hPa were included. In the comparisons with MLS, the coarsest grid is represented by that of the MLS data with its vertical resolution of about 3 km while in the comparisons with the ozone sondes the model has the lowest vertical resolution. The comparisons with MLS are displayed as the vertical cross-section of the change in the analysis fit to the observations due to the addition of either GOME-2 or MIPAS to the reference observing system assimilated in the control experiment. Such a change, Δ, is defined as follows:

$$\Delta = \left|\mathrm{STAT}\left(\mathrm{MLS} - \mathrm{Analyses}^{(\mathrm{Exp/PERT})}\right)\right| - \left|\mathrm{STAT}\left(\mathrm{MLS} - \mathrm{Analyses}^{(\mathrm{Exp/CTRL})}\right)\right| \quad (1)$$

where $\mathrm{STAT}()$ can be either the mean or the standard deviation. For either statistics, a negative value of Δ means that the analyses from a given perturbed experiment, generically labelled above as Exp/PERT, fits MLS observations better than those from Exp/Ctrl, thus leading to an improvement. In contrast, a positive value of Δ is associated to a degradation in the ozone analyses.

The comparisons with the ozone sondes are instead shown in terms of mean RMS residuals, RMSE, between the sonde profiles and the co-located analyses from the three experiments (Exp/Ctrl, Exp/GOME2, Exp/MIPAS). Thus the smaller is the RMSE, the better is the analysis fit to the sonde measurements. For plotting purposes, the RMSE are computed and displayed in terms of integrated column quantities. The level of agreement between the three sets of ozone analyses and the ozone sondes is also discussed in terms of a Mean Common Area score Factor (MCAF). The Common Area score Factor (CAF, e.g. as proposed by Dragani et al., 2015) is an indicator of how well the analysis profile fits the observation profile. It is defined as

$$CAF = \frac{\int_p min([\mathbf{O_3}^{(observation)}(p), \mathbf{O_3}^{(model)}(p)])dp}{\int_p max([\mathbf{O_3}^{(observation)}(p), \mathbf{O_3}^{(model)}(p)])dp} \quad (2)$$

where $\mathbf{O_3}$ is the ozone vertical profile at the observation location. The integrals are computed over the observation and model common vertical pressure grid. The score varies between zero and one with a CAF value of one implying a perfect fit to the observations. The MCAF score is based on a similar idea as the CAF, but averaged over a number of profiles and stations, for instance here it is computed for all profiles available over given latitudinal bands (the high latitudes in the SH and NH; the midlatitudes in the SH and NH; the tropics). It is important to bear in mind that, due to the sonde vertical coverage, the CAF and MCAF scores are most representative of the part of the atmosphere spanning from the surface up to about 10 hPa at most.

In all comparisons, the ozone analyses were spatially co-located with the independent observations allowing a maximum of three hour time lag. To avoid any spin-up effect, only the results computed over the August-October 2008 period are presented.

5.2.2 Analysis and discussion

Figure 9 gives an indication of the improvement in the fit of the Exp/GOME2 and Exp/MIPAS analyses to MLS compared to that of the Exp/Ctrl analyses, based on equation 1 and presented in terms of both the mean (top panels) and the standard deviation (bottom panels). Despite the large difference in data counts, the two datasets produce similar large scale structures in the analysis fit to MLS at all latitudes for most levels in the middle to lower stratosphere in both the mean (top panels) and the standard deviation (bottom panels). By contrasting Figures 9 and 8, it results that both instruments generally lead to stratospheric improvements in the analysis agreement to MLS at most latitudes, particularly in the atmospheric layer between 10 and 70 hPa, with a reduction in the mean residuals from MLS of up to about 2 mg/kg (about 40%) in the middle stratosphere. In the tropical region, both instruments lead to a slightly degraded fit (up to about 0.5 mg/kg, about 15%) of the ozone analyses to MLS between 20 and 30 hPa. Such a degradation, which is more pronounced in the case of MIPAS than in that of GOME-2, is balanced by an improved fit in the tropical regions above and below the 20-30 hPa layer. With changes up to 1 mg/kg, the improvement in the tropics outside the 20-30 hPa layer is more important for the nadir instrument than for the limb. At high latitudes, the nadir instrument triggers a reduction in the level of agreement with MLS. Here, the changes are up to about 0.5 mg/kg (15%) in the NH (summer hemisphere) and about double (35%) in the SH (winter hemisphere). Negligible to small changes (typically less than 0.25 mg/kg, i.e. within 5%) are triggered in the upper stratosphere with a sign that depends on the latitudinal band. Overall, panels **a)** and **b)** of Figure 9 highlights a complementarity of the two observing systems in impacting the corresponding stratospheric ozone analyses. The bottom panels of Figure 9 show that the assimilation of either instrument leads to a reduction in the standard deviation in the extra-tropical middle stratosphere, although when assimilating MIPAS the impact is slightly larger (thus better) in places than in the case of GOME-2 assimilation. A small degradation can be found in the standard deviations when GOME-2 is assimilated between 20 and 40 hPa at latitudes ~~southern than~~ south of 60°S - where the fit to MLS was degraded also in the mean -, as opposed to a substantial improvement determined by the assimilation of MIPAS. This increased noise given by the Nadir, UV-based observations is an indication of the difficulty of this observation type to provide a strong constraint to the ozone analyses and forecasts at high latitudes in the SH during winter-spring time. Furthermore, in contrast to what is found in the mean changes, here the vertical coverage of the impacted region is lower for MIPAS than that for GOME-2. Those differences can, arguably, be related to the instrument vertical sensitivity associated with the different viewing geometry. Averaging kernels (AKs) that sharply peak at the level of maximum sensitivity for the limb sensor (e.g. Ceccherini and Ridolfi, 2002) as opposed to broad and highly overlapping AKs for GOME-2 (Miles et al., 2015) allow the former to generate much more localized changes than the latter can do.

The conclusions drawn by assessing the change in the level of stratospheric agreement between the analyses and the MLS profiles implied by the assimilation of either the GOME-2 or MIPAS ozone profiles are confirmed by the comparisons with ozone sondes. The MCAF scores are computed over five latitudinal bands and divided into two contributions for the stratosphere and troposphere. For simplicity, the 100 hPa pressure level has been used as the separation level for the two contributions. The stratospheric and tropospheric scores for the midlatitudes in the NH, characterized by the largest availability of sonde profiles, are presented in Figures 10 and 11, respectively. The results at the other four latitudinal bands generally confirm those at midlatitudes in the NH. Overall, both the MIPAS and GOME-2 ozone profiles produce stratospheric analyses that are in better vertical agreement with the ozone sonde profiles than those from Exp/Ctrl, with the Exp/MIPAS score being slightly higher than that of Exp/GOME2 (Figure 10). The MCAF scores are generally 20% lower in the lowermost part of the atmosphere up to 100 hPa (Figure 11) than they are in the region above (Figure 10). Although, both instruments lead to an improved agreement of the corresponding analyses to sonde profiles compared to the Exp/Ctrl, the differences with the Exp/Ctrl scores reduce significantly. In particular, both GOME-2 and MIPAS appear to produce similar levels of agreement, especially in the tropics (not shown). In the extra-tropics, Exp/GOME2 is marginally better in the NH (summer hemisphere) while Exp/MIPAS is slightly better in the SH (winter hemisphere, not shown).

Why are the scores of the two Exp/PERT analyses so close in the troposphere? Based on Figure 2, is good vertical coverage with low vertical resolution (like for GOME-2) equivalent to having a poor vertical coverage but with high vertical resolution where available (like for MIPAS) in the troposphere? If so, what ~~does it happen~~ happens in the lowermost troposphere, in particularly below the level of MIPAS availability of roughly 400 hPa?

An answer to these additional questions can be found by analysing the vertically resolved fit of the three sets of analyses to the ozone sondes (Figure 12). As explained in section 5.2.1, this is expressed in terms of the RMS difference between the sonde profiles and co-located analyses, thus the smaller such a difference is, the higher the level of agreement between the analysed and observed ozone profiles becomes. Figure 12 in general confirms the results over the stratosphere discussed above, including the difficulty of the limb data in the tropics at pressure levels within the 20-30 hPa layer shown in panel **b)** of Figure 9. In the upper troposphere (down to about 400 hPa), Exp/MIPAS still shows a higher level of agreement to the sondes than Exp/GOME2. In the lower troposphere (below 400 hPa), the Exp/GOME2 assimilation generally outperforms that of Exp/MIPAS. However, in this region the assimilation of MIPAS LPO_3 still improves the ozone analyses compared to the control experiment, despite a vertical coverage for the limb sounder only down to about 400hPa. Two possible mechanisms to explain the improvements in the lower troposphere: 1) a synergistic assimilation of limb ozone profiles and total ozone column data, and 2) the vertical transport. In the current system, ozone is used as a univariate variable, which means it does not affect the rest of the system in general, and the winds in particular. Moreover, with typical vertical velocity in the upper troposphere of typically 2-3 hPa/hour, the vertical transport could explain less than 50hPa vertical displacement of the information inferred by the assimilation of MIPAS data within a 12-hour assimilation window. With a vertical coverage provided by the MIPAS instrument down to about 400 hPa at best, the vertical transport alone cannot explain the changes in the lower troposphere. Based on these considerations, this ~~This~~ positive result ~~is~~ can only be a consequence of exploiting the synergy between the LPO_3 and the total column product that was assimilated in all experiments. Figure 3 (courtesy of Dragani and McNally, 2013) shows that the ozone background error variances are largest in the stratosphere between 20 and 80 hPa, implying that the ozone increments generated by the assimilation of total column products alone would most likely be spread over this region of the atmosphere. However, in this case, the combination of improved stratospheric ozone concentration obtained thanks to the assimilation of the LPO_3 observations, and the synergistic assimilation of the latter with TCO_3 products provides an indirect constraint on the ozone analyses at levels below the limb vertical coverage.

These results would suggest that the vertical resolution does matter particularly in a region like the UTLS. The ~~observation~~ retrieval vertical coverage, although important, matters to a less extent as an impact can be made in unobserved regions if the synergy with other observations can be exploited within the data assimilation system.

6 Concluding remarks and recommendations

The present study aims at providing a comparative analysis of UV nadir-backscatter and infrared limb-emission ozone profile data assimilation, and thus draw conclusions on how the viewing geometry impacts the ability of these two classes of sensors in constraining and improving the quality of the resulting analyses. The MERRA-2 reanalysis already represents an example where a precise, pragmatic choice was made from the outset in whether to assimilate limb ozone profiles in place of UV nadir-backscatter ozone profiles, despite the tendency typical in NWP and reanalysis of using as many observations as possible. This consideration called for a more quantitative assessment of the relative impact on and ability to improve the quality of the resulting ozone analyses of these two classes of observations. An analysis of the Earth Observation capability planned for the next ten to fifteen years shows a generally good temporal coverage from nadir-looking instruments measuring the backscattered solar radiation in the UV(-Vis) spectral range. Many of these instruments are or will be operated as part of operational missions thus ensuring long term data provision. In contrast, over the same time-frame the availability of measurements from limb-viewing sensors is, at best, poor, and normally confined to research satellites with limited long-term continuation, if any at all. Thus, the results are of primary interest to space agencies as input in shaping future plans for Earth Observation, and have implications for a large part of the ozone research spectrum spanning from Numerical Weather Prediction (NWP) to climate reanalysis, from air quality to stratospheric trends and variability.

The main findings from this study are:

- On average, the two datasets produce changes in the vertical distribution of the ozone analyses with similar large-scale patterns, in particular an increase of the ozone analyses up to 2 mg/kg compared to the control experiment at levels between 20 and 70 hPa and an ozone amount reduction in the extra-tropical lower and middle troposphere. Comparisons with independent observations indicate that those changes improve the agreement between the resulting ozone analyses and independent observations.

- In the middle stratosphere, the changes induced by the assimilation of the MIPAS limb ozone profiles are more localized than those implied by the assimilation of the GOME-2 nadir profiles.

- Differences in the impact were noticeable, especially in the tropics and at high latitudes. In the tropics, both instruments lead to degraded fit to MLS (up to about 0.5 mg/kg increase in the residuals) in the layer between 20 and 30 hPa, and to improved fit (up to about 1 mg/kg reduction in the residuals) above and below that layer. A reduction in the level of agreement between the ozone analyses and the MLS observations is also triggered by the GOME-2 assimilation at high latitudes where the mean residuals are increased of up to about 0.5 mg/kg in the NH (summer hemisphere) and about 1 mg/kg in the SH (winter hemisphere).

- The MCAF scores show that both datasets lead to improved ozone analyses compared to the control experiment (about 20% higher MCAF score in the stratosphere and about 10% higher MCAF in the troposphere than the MCAF from Exp/Ctrl). But while the assimilation of MIPAS is better than that of GOME-2 in the stratosphere, the scores in the troposphere (here simply referred to as the layer below 100 hPa) are much similar to each other despite the fact that the limb vertical coverage does not normally extend below about 400 hPa

- In the upper troposphere (down to about 400 hPa), the assimilation of MIPAS profiles outperforms that of GOME-2 based on the fit to the ozone sondes. The former shows RMSE values between 1 and 2 DU lower than those derived for the latter.

- In the lowermost troposphere, the assimilation of the MIPAS ozone profiles was found able to substantially modify the ozone distribution with changes as large as 60%. Comparisons with ozone sondes profiles show that, although not always better than that of GOME-2, the assimilation of the MIPAS ozone profiles improves the fit to the independent data by reducing the RMSE of up to 2 DU compared to that of the Exp/Ctrl analyses.

The results presented in this paper highlight the complementarity of the two observing systems and confirmed how the instrument characteristics (observed spectral range and viewing geometry) shape the observation ability in constraining the ozone analyses. Overall, both types of observations can improve the vertical distribution of the stratospheric ozone analyses that are too high above 20 hPa and too low below 25-30 hPa in the control experiment. The assimilation of the nadir observations proves to be more successful than that of the limb observations in the tropical stratosphere. In contrast, the assimilation of the limb ozone profile is essential at high latitudes and in the upper troposphere. Only small differences between the two perturbation experiments were seen in the lowermost troposphere that is not sampled by limb sensors but where nadir instruments, like GOME-2, can provide useful information (Miles et al., 2015). On one hand the limb limited sensitivity can only be realized here if their synergy with other ozone observations (in particular total column ozone products) can be exploited within the data assimilation system; on the other hand such a synergy cannot completely replace the improvement produced by the assimilation of nadir-based ozone profiles.

To accurately monitor and forecast ozone concentrations in a data assimilation system, high vertical resolution ozone profile observations are essential. With only total column ozone measurements, analyses would depend too much on the model background and the background error covariances with analysis increments that would be most likely located in the region of the atmosphere where the background error is largest.

Limb observations are essential to bring the improvements in the ozone analyses necessary for example to assess long-term stratospheric changes from reanalysis productions, especially at high latitudes, but also in the upper troposphere and across the tropopause. However, in the lowermost part of the atmosphere important for air quality, nadir-based profiles are still the most important source of information.

The present study has demonstrated the potential that each dataset and instrument type has in its own right. The findings discussed in this study would indicate the need for a more balanced availability between nadir and limb sensors than currently exists thus supporting revised EO plans to improve the availability and long-term continuation of limb viewing instruments in the forthcoming years.

Acknowledgements. The author was supported by the European Space Agency through the Climate Change Initiative / Climate Modelling User Group (CCI/CMUG) project. A special thank-you goes to Dr. Stephen English (ECMWF) and Dr. Roger Saunders (The Met Office) for comments and useful discussion. Anabel Bowen and Simon Witter (ECMWF) skilfully improved the quality of the figures. The ESA-funded Ozone-CCI consortium is thanked for the provision of the CCI GOME-2 and MIPAS ozone profiles. The MLS ozone data used in this study for the independent validation were the reprocessed (version 3.3) ozone profiles retrieved from the NASA Distributed Active Archive Center at the Goddard Space Flight Center. The ozone sondes were retrieved from the World Ozone and Ultraviolet Radiation Data Centre (WOUDC).

References

Antón, M., Kroon, M., López, M., Vilaplana, J. M., Bañón, M., van der A, R., Veefkind, J. P., Stammes, P., and Alados-Arboledas, L.: Total ozone column derived from GOME and SCIAMACHY using KNMI retrieval algorithms: Validation against Brewer measurements at the Iberian Peninsula, Journal of Geophysical Research: Atmospheres, 116, doi:10.1029/2011JD016436, 2011.

Auligné, T., McNally, A., and Dee, D.: Adaptive bias correction for satellite data in a numerical weather prediction system, Q. J. R. Meteorol. Soc., 133, 631–642, 2007.

Barnes, E. A., Barnes, N. W., and Polvani, L. M.: Delayed Southern Hemisphere Climate Change Induced by Stratospheric Ozone Recovery, as Projected by the CMIP5 Models, J. Climate, 27, 852–867, doi:10.1175/JCLI-D-13-00246.1, 2014.

Bhartia, P.: OMI Algorithm Theoretical Basis Document, Tech. Rep. version 2, NASA, available at http://eospso.gsfc.nasa.gov/sites/default/files/atbd/ATBD-OMI-02.pdf, 2002.

Bhartia, P., McPeters, R., Mateer, C., Flynn, L., and Wellemeyer, C.: Algorithm for the estimation of vertical profiles from the backscattered ultraviolet technique, J. Geophys. Res., 101, 18 793–18 806, 1996.

Bhartia, P. K., McPeters, R. D., Flynn, L. E., Taylor, S., Kramrova, N. A., Frith, S., Fisher, B., and DeLand, M.: Solar Backscatter UV (SBUV) total ozone and profile algorithm, Atmos. Meas. Tech. Discuss., 5, 5913–5951, 2012.

Bormann, N. and Bauer, P.: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. I: Methods and application to ATOVS data, Quarterly Journal of the Royal Meteorological Society, 136, 1036–1050, doi:10.1002/qj.616, 2010.

Bormann, N., Collard, A., and Bauer, P.: Estimates of spatial and interchannel observation-error characteristics for current sounder radiances for numerical weather prediction. II: Application to AIRS and IASI data, Quarterly Journal of the Royal Meteorological Society, 136, 1051–1063, doi:10.1002/qj.615, 2010.

Bosilovich, M. G., Akella, S., Coy, L., Cullather, R., Draper, C., Gelaro, R., Kovach, R., Liu, Q., Molod, A., Norris, P., Wargan, K., Chao, W., Reichle, R., Takacs, L., Vikhliaev, Y., Bloom, S., Collow, A., Firth, S., Labow, G., Partyka, G., Pawson, S., Reale, O., Schubert, S. D., and Suarez, M.: MERRA-2: Initial Evaluation of the Climate, Tech. rep., NASA TM-2015-04606, Availablefromhttp://gmao.gsfc.nasa.gov/pubs/tm/, 2015.

Brinksma, E.: Validation of SCIAMACHY TOSOMI ozone columns with ground-based data. Available from www.temis.nl/protocols/O3total.html, Tech. rep., KNMI, 2004.

Callies, J., Corpaccioli, E., Eisinger, M., Hahne, A., and Lefebvre, A.: GOME-2 - Metop's Second-Generation, Sensor for Operational Ozone Monitoring, ESA Bulletin 102, ESA, 2000.

Cariolle, D. and Déqué, M.: Southern Hemisphere Medium–Scale Waves and Total Ozone Disturbances in a Spectral General Circulation Model, J. Geophys. Res., 91, 10 825–10 846, 1986.

Cariolle, D. and Teyssédre, H.: A revised linear ozone photochemistry parameterization for use in transport and general circulation models: multi-annual simulations, Atmos. Chem. Phys., 7, 2183–2196, 2007.

Ceccherini, S. and Ridolfi, M.: Averaging Kernels for MIPAS near real time retrievals, Tech. Rep. TN-IFAC-OST0201, IFAC-CNR, available at http://www.ifac.cnr.it/retrieval/documents/AK_report.pdf, 2002.

Chiou, E. W., Bhartia, P. K., McPeters, R. D., Loyola, D. G., Coldewey-Egbers, M., Fioletov, V. E., Van Roozendael, M., Spurr, R., Lerot, C., , and Frith, S. M.: Comparison of profile total ozone from SBUV (v8.6) with GOME-type and ground-based total ozone for a 16-year period (1996 to 2011), Atmos. Meas. Tech., 7, 1681–1692, doi:10.5194/amt-7-1681-2014, 2014.

Courtier, P., Thépaut, J.-N., and Hollingsworth, A.: A strategy for operational implementation of 4D-Var, using an incremental approach, Q. J. R. Meteorol. Soc., 120, 1367–1388, 1994.

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, 1991.

Dee, D.: Bias and data assimilation, Q. J. R. Meteorol. Soc., 131, 3323–3343, 2005.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: Configuration and performance of the data assimilation system, Q. J. R. Meteorol. Soc., 137, 553–597, 2011.

Derber, J. and Wu, W. S.: The use of TOVS cloud-cleared radiances in the NCEP SSI analysis system, Mon. Weather Rev., 126, 2287–2299, 1998.

Desroziers, G., Berre, L., Chapnik, B., and Poli, P.: Diagnosis of observation, background and analysis-error statistics in observation space, Q. J. R. Meteorol. Soc., 131, 3385–3396, 2005.

Desroziers, G., Berre, L., and Chapnik, B.: Objective validation of data assimilation systems: diagnosing sub-optimality, in Proceedings of the ECMWF Workshop on diagnostics of data assimilation system performance, 15-17 June 2009, Reading, UK, 2009.

Dethof, A.: Assimilation of ozone retrievals from the MIPAS instrument on board ENVISAT, Tech. Memo. 428, Eur. Cent. for Medium-range Weather Forecasts, Shinfield Park, Reading, UK, 2003.

Dethof, A. and Hólm, E.: Ozone assimilation in the ERA-40 reanalysis project, Q. J. R. Meteorol. Soc., 130, 2851–2872, 2004.

Dragani, R.: Variational bias correction of satellite ozone data. Technical report R43.8/RD/0934, Tech. rep., ECMWF, available from R. Dragani (rossana.dragani@ecmwf.int), 2009.

Dragani, R.: On the quality of the ERA-Interim ozone reanalyses: Comparisons with satellite ozone data, Q. J. R. Meteorol. Soc., 137, 1312–1326, 2011.

Dragani, R.: Validation of the reprocessed MIPAS and SCIAMACHY retrievals using ERA-Interim, and one-year assimilation of MIPAS ozone profiles at ECMWF, Tech. rep., ECMWF, 2013.

Dragani, R. and McNally, A. P.: Operational assimilation of ozone-sensitive infrared radiances at ECMWF, Q.J.R. Meteorol. Soc., pp. 2068–2080, doi:10.1002/qj.2106, 2013.

Dragani, R., Abdalla, S., Engelen, R. J., Inness, A., and Thépaut, J.-N.: Ten years of ENVISAT observations at ECMWF: a review of activities and lessons learnt, Q. J. R. Meteorol. Soc., 141, 598–610, doi:10.1002/qj.2380, 2015.

Dubock, P. A., Spoto, F., Simpson, J., Spencer, D., Schutte, E., and Sontag, H.: The Envisat Satellite and Its Integration, ESA bulletin, 106, 26–45, available at http://www.esa.int/esapub/bulletin/bullet106/bul106_2.pdf, 2001.

Echle, G., von Clarmann, T., Dudhia, A., Flaud, J.-M., Funke, B., Glatthor, N., Kerridge, B., López-Puertas, M., Martín-Torres, F. J., and Stiller, G. P.: Optimized spectral microwindows for data analysis of the Michelson Interferometer for Passive Atmospheric Sounding on the Environmental Satellite, Appl. Opt., 39, 5531–5540, 2000.

EPA (Environmental Protection Agency): EPA Strengthens Ozone Standards to Protect Public Health/Science-based standards to reduce sick days, asthma attacks, emergency room visits, greatly outweigh costs. EPA News Release from Headquarters on 10/01/2015 (available at https://yosemite.epa.gov/opa/admpress.nsf/Press%20Releases%20from%20Headquarters?OpenView), last access on 8 June 2016, 2015.

Eskes, H. J., van der A, R. J., Brinksma, E. J., Veefkind, J. P., de Haan, J. F., and Valks, P. J. M.: Retrieval and validation of ozone columns derived from measurements of SCIAMACHY on Envisat, Atmos. Chem. Phys. Discuss, 5, 4429–4475, 2005.

Farman, J. C., Grewe, V., Hein, R., Schnadt, C., Bruhl, C., and Steil, B.: Large loss of total ozone in Antarctica reveal seasonal ClOx/NOx interaction, Nature, 315, 207–210, 1985.

Feng, L., Brugge, R., Hólm, E. V., Harwood, R. S., O'Neill, A., Filipiak, M. J., Froidevaux, L., and Livesey, N.: Four-dimensional variational assimilation of ozone profiles from the Microwave Limb Sounder on the Aura satellite, J. Geophys. Res., 113, D15S07, doi:10.1029/2007JD009 121, 2008.

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, 2008.

Flemming, J., Inness, A., Flentje, H., Huijnen, V., Moinat, P., Schultz, M. G., and Stein, O.: Coupling global chemistry transport models to ECMWF's integrated forecast system, Geosci. Model Dev., 2, 253–265, 2009.

Froidevaux, L., Jiang, Y. B., Lambert, A., Livesey, N. J., Read, W. G., Waters, J. W., Browell, E. V., Hair, J. W., Avery, M. A., McGee, T. J., Twigg, L. W., Sumnicht, G. K., Jucks, K. W., Margitan, J. J., Sen, B., Stachnik, R. A., Toon, G. C., Bernath, P. F., Boone, C. D., Walker, K. A., Filipiak, M. J., Harwood, R. S., Fuller, R. A., Manney, G. L., Schwartz, M. J., Daffer, W. H., and1 R. E. Cofield, B. J. D., Cuddy, D. T., Jarnot, R. F., Knosp, B. W., Perun, V. S., Snyder, W. V., Stek, P. C., Thurstans, R. P., , and Wagner, P. A.: Validation of Aura Microwave Limb Sounder stratospheric ozone measurement, J. Geophys. Res., 113, 2008.

Geer, A. J., Peubey, C., Bannister, R., Jackson, D. R., Lahoz, W. A., Migliorini, S., O'Neill, A., and Swinbank, R.: Assimilation of stratospheric ozone from MIPAS into a global general circulation model: the September 2002 vortex split, Q. J. R. Meteorol. Soc., 132, 231–257, 2006.

Han, W. and McNally, A. P.: The 4D-Var assimilation of ozone-sensitive infrared radiances measured by IASI, Q. J. R. Meteorol. Soc., 136, 2025–2037, 2010.

Hao, N., Koukouli, M. E., Inness, A., Valks, P., Loyola, D. G., Zimmer, W., Balis, D. S., Zyrichidou, I., Van Roozendael, M., Lerot, C., and Spurr, R. J. D.: GOME-2 total ozone columns from MetOp-A/MetOp-B and assimilation in the MACC system, Atmospheric Measurement Techniques, 7, 2937–2951, doi:10.5194/amt-7-2937-2014, 2014.

Heath, D. F., Mateer, C. L., and Krueger, A. J.: The Nimbus-4 Backscatter Ultraviolet (BUV) atmospheric ozone experiment: two years' operation, Pure and Applied Geophysics, 106, 1238–1253, 1973.

Hollingsworth, A., Engelen, R. J., Textor, C., Benedetti, A., Boucher, O., Chevallier, F., Dethof, A., Elbern, H., Eskes, H., Flemming, J., Granier, C., Kaiser, J. W., Morcrette, J.-J., Peuch, R. R. V.-H., Rouil, L., Schultz, M. G., Simmons, A. J., and Consortium, T. G.: Toward a monitoring and forecasting system for atmospheric composition: the GEMS project, B. Am. Meteorol. Soc., 89, 1147–1164, 2008.

Holm, E. V., Untch, A., Saunders, A. S. R., Bouttier, F., and Andersson, E.: Multivariate ozone assimilation in four-dimensional data assimilation, in: SODA Workshop on Chemical Data Assimilation, KNMI, De Bilt, The Netherlands, 9-10 December 1998, pp. 89–94, 1999.

Inness, A., Baier, F., Benedetti, A., Bouarar, I., Chabrillat, S., Clark, H., Clerbaux, C., Coheur, P., Engelen, R. J., Errera, Q., Flemming, J., George, M., Granier, C., Hadji-Lazaro, J., Huijnen, V., Hurtmans, D., Jones, L., Kaiser, J., Kapsomenakis, J., Lefever, M., Leitão, J., Razinger, M., A., Schultz, M. G., Simmons, A. J., Suttie, M., Stein, O., Thépaut, J.-N., Thouret, V., Vrekoussis, M., Zerefos, C., and the MACC team: The MACC reanalysis: an 8 yr data set of atmospheric composition, Atmospheric Chemistry and Physics, 13, 4073–4109, 2013.

Jackson, D. and Saunders, R.: Ozone data assimilation: preliminary system, Technical Report 394, Met Office, 2002.

Jackson, D. R.: Assimilation of EOS MLS ozone observations in the Met Office data-assimilation system, Q. J. R. Meteorol. Soc., 133, 1771–1788, 2007.

Kalnay, E.: Atmospheric Modeling, Data Assimilation and Predictability, Cambridge University Press, 2003.

Keppens, A., Lambert, J.-C., Granville, J., Miles, G., Siddans, R., van Peet, J. C. A., 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, Atmospheric Measurement Techniques, 8, 2093–2120, doi:10.5194/amt-8-2093-2015, 2015.

Kinnison, D. E., Brasseur, G. P., Walters, S., Gracia, R. R., Marsh, D. R., Sassi, F., Harvey, V. L., Randall, C. E., Emmons, L., Lamarque, J. F., Hess, P., Orlando, J. J., Tie, X. X., Randel, W., Pan, L. L., Gettelman, A., Granier, C., Diehl, T., Niemeier, U., and Simmons, A. J.: Sensitivity of chemical tracers to meteorological parameters in the MOZART-3 chemical transport model, J. Geophys. Res, 112, D20 302, doi:10.1029/2006JD007879, 2007.

Komhyr, W. D., Barnes, R. A., Borthers, G. B., Lathrop, J. A., Kerr, J. B., and Opperman, D. P.: Electrochemical concentration cell ozonesonde performance evaluation during STOIC 1989, J. Geophys. Res., 100, 9231–9244, 1995.

Laeng, A., Grabowski, U., von Clarmann, T., Stiller, G., Glatthor, N., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Lossow, S., Sofieva, V., Petropavlovskikh, I., Hubert, D., Bathgate, T., Bernath, P., Boone, C. D., Clerbaux, C., Coheur, P., Damadeo, R., Degenstein, D., Frith, S., Froidevaux, L., Gille, J., Hoppel, K., McHugh, M., Kasai, Y., Lumpe, J., Rahpoe, N., Toon, G., Sano, T., Suzuki, M., Tamminen, J., Urban, J., Walker, K., Weber, M., and Zawodny, J.: Validation of MIPAS IMK/IAA V5R_O3_224 ozone profiles, Atmospheric Measurement Techniques, 7, 3971–3987, doi:10.5194/amt-7-3971-2014, 2014.

Lahoz, W. A., Errera, Q., Swinbank, R., and Fonteyn, D.: Data assimilation of stratospheric constituents: a review, Atmospheric Chemistry and Physics, 7, 5745–5773, doi:10.5194/acp-7-5745-2007, 2007.

Lang, R., Munro, R., Livschitz, Y., Dyer, R., and Lacan, A.: GOME-2 FM3 Long-Term In-Orbit Degradation - Basic Signatures After 2nd Throughput Test, Technical Report EUM.OPS-EPS.DOC.09.0464, EUMETSAT, 2009.

Lefever, K., van der A, R., Baier, F., Christophe, Y., Errera, Q., Eskes, H., Flemming, J., Inness, A., Jones, L., Lambert, J.-C., Langerock, B., Schultz, M. G., Stein, O., Wagner, A., and Chabrillat, S.: Copernicus stratospheric ozone service, 2009-2012: validation, system intercomparison and roles of input data sets, Atmospheric Chemistry and Physics, 15, 2269–2293, doi:10.5194/acp-15-2269-2015, 2015.

Livesey, N. J., Logan, J. A., Santee, M. L., Waters, J. W., Doherty, R. M., Read, W. G., Froidevaux, L., and Jiang, J. H.: Interrelated variations of O_3, CO and deep convection in the tropical/subtropical upper troposphere observed by the Aura Microwave Limb Sounder (MLS) during 2004-2011, Atmospheric Chemistry and Physics, 13, 579–598, doi:10.5194/acp-13-579-2013, 2013.

Madronich, S., Shao, M., Wilson, S. R., Solomon, K. R., Longstreth, J. D., and Tang, X. Y.: Changes in air quality and tropospheric composition due to depletion of stratospheric ozone and interactions with changing climate: implications for human and environmental health, Photochem. Photobiol. Sci., 14, 149–169, doi:10.1039/C4PP90037E, 2015.

McCarty, W., Considine, D., Lee, T., Randles, C., Coy, L., Wargan, K., Bosilovich, M., and Gelaro, R.: Use of Satellite Observations in NASA Reanalyses: MERRA-2 and Future Plans, Technical report, NASA CGMS43 NASAWP06, v1, 2015.

McLinden, C. A. and Fioletov, V.: Quantifying stratospheric ozone trends: Complications due to stratospheric cooling, Geophys. Res. Lett., 38, L03 808, doi:10.1029/2010GL046012, 2011.

McPeters, R. D., Lebow, G. J., and Logan, J. A.: Ozone climatological profiles for satellite retrieval algorithms, J. Geophys. Res., 112, D05 308, doi:10.1029/2005JD006823, 2007.

McPeters, R. D., Bhartia, P. K., Haffner, D., Labow, G. J., and Flynn, L.: The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res. Atmos., 118, 8032–8039, doi:10.1002/jgrd.50597, 2013.

Miles, G. M., Siddans, R., Kerridge, B. J., Latter, B. G., and Richards, N. A. D.: Tropospheric ozone and ozone profiles retrieved from GOME-2 and their validation, Atmos. Meas. Tech., 8, 385–398, doi:10.5194/amt-8-385-2015, 2015.

Munro, R., Siddans, R., Reburn, W. J., and Kerridge, B. J.: Direct measurement of tropospheric ozone distributions from space, Nature, 392, 168–171, 1998.

Munro, R., Eisinger, M., Anderson, C., Cllies, J., Corpaccioli, E., Lang, R., Lefebvre, A., Livschitz, Y., and Albinana, A.: GOME-2 on MetOp, in: The 2006 EUMETSAT Meteorological Satellite Conference, Helsinki, Finland, 12-16 June 2006, ISBN 92-9110-076-5, EUMETSAT, 2006.

Plummer, S.: The ESA Climate Change Initiative - Description, Technical Report EOP-SEP/TN/0030-09/SP, ESA, available from http://46.19.32.34/files/ESACCIDescription.pdf, 2009.

Polavarapu, S., Shepherd, T. G., ROCHON, Y., and REN, S.: Some challenges of middle atmosphere data assimilation, Quarterly Journal of the Royal Meteorological Society, 131, 3513–3527, doi:10.1256/qj.05.87, 2005.

Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F., and Simmons, A.: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics, Q. J. R. Meteorol. Soc., 126, 1143–1170, 2000.

Randel, W. J. and Wu, F.: A stratospheric ozone profile data set for 1979-2005: Variability, trends, and comparisons with column ozone data, J. Geophys. Res., 112, D06 313, doi:10.1029/2006JD007339, 2007.

Raspollini, P., Belotti, C., Burgess, A., Carli, B., Carlotti, M., Ceccherini, S., Dinelli, B. M., Dudhia, A., Flaud, J.-M., Funke, B., Höpfner, M., López-Puertas, M., Payne, V., Piccolo, C., Remedios, J. J., Ridolfi, M., , and Spang, R.: MIPAS level 2 operational analysis, Atmos. Chem. Phys., 6, 5605–5630, 2006.

Ridolfi, M., Carli, B., Carlotti, M., von Clarmann, T., Dinelli, B. M., Dudhia, A., Flaud, J.-M., Hoepfner, M., Morris, P., Raspollini, P., Stiller, G., and Wells, R. J.: Optimized forward model and retrieval scheme for MIPAS near-real-time data processing, Appl. Opt., 39, 1323–1340, 2000.

Rodgers, C. D.: Inverse Methods for Atmospheric Sounding. Theory and Practice, Atmospheric, Oceanic and Planetary Physics, World Scientific Publishing and Co. Pte. Ltd., PO Box 128, Farrer Road, Singapore 912805, 2000.

Schoeberl, M. R., Douglass, A., Hilsenrath, E., Bhartia, P., Beer, R., Waters, J., Gunson, M., Froidevaux, L., Gille, J., Barnett, J., Levelt, P., and DeCola, P.: Overview of the EOS Aura mission, IEEE Trans. Geosci. Remote Sens., 44, 1066 – 1074, 2006.

Simmons, A.: Monitoring Atmospheric Composition and Climate, ECMWF Newsletter, 123, 10–13, 2010.

Sofieva, V. F., Rahpoe, N., Tamminen, J., Kyro, E., Kalakoski, N., Weber, M., Laeng, A., von Clarmann, T., Stiller, G., Lossow, S., Degenstein, D., Bourassa, A., Adams, C., Roth, C., Lloyd, N., Bernath, P., Hargreaves, R. J., Urban, J., Murtagh, D., Hauchecorne, A., Roozendael, M. V., Kalb, N., and Zehner, C.: Harmonized dataset of ozone profiles from satellite limb and occultation measurements, Earth Syst. Sci. Data, 5, 349–363, 2013.

Solomon, S., Garcia, R. R., Rowland, F. S., and Wuebbles, D. J.: On the depletion of Antarctic ozone, Nature, 321, 755–758, doi:10.1038/321755a0, 1986.

SPARC: SPARC/IOC/GAW Assessment of trends in the vertical distribution of ozone, no. 1 in SPARC Report, pp. 1–289, 1998.

Stajner, I., Wargan, K., Pawson, S., Hayashi, H., Chang, L.-P., Hudman, R. C., Froidevaux, L., Livesey, N., Levelt, P. F., Thompson, A. M., Tarasick, D. W., St"ubi, R., Andersen, S. B., Yela, M., K"onig-Langlo, G., Schmidlin, F. J., and Witte, J. C.: Assimilated ozone from

EOS-Aura: Evaluation of the tropopause region and tropospheric columns, J. Geophys. Res., 113, D16S32, doi:10.1029/2007JD008 863, 2008.

Steinbrecht, W., Shwartz, R., and Claude, H.: New Pump Correction for Brewer-Mast Ozone Sonde: Determination from Experiment and Instrument Intercomparisons, J. Atm. O. Tech., 15, 144–156, 1998.

Stewart, L. M., Dance, S. L., Nichols, N. K., Eyre, J. R., and Cameron, J.: Estimating interchannel observation-error correlations for IASI radiance data in the Met Office system†, Quarterly Journal of the Royal Meteorological Society, 140, 1236–1244, doi:10.1002/qj.2211, 2014.

Struthers, H., Brugge, R., Lahoz, W. A., O'Neill, A., and Swinbank, R.: Assimilation of ozone profiles and total column measurements into a general circulation model, J. Geophys. Res., 107, 2002.

UNEP, Environmental Effects Assessment Panel: Environmental effects of ozone depletion and its interactions with climate change: progress report, 2011, Photochem. Photobiol. Sci., 11, 13–27, doi:10.1039/C1PP90033A, 2012.

Uppala, S. M., Kållberg, P. W., Simmons, A. J., Andrae, U., da Costa Bechtold, V., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J.-F., Morcrette, J.-J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Quart. J. Roy. Meteor. Soc., 131, 2961–3012, 2005.

Veersé, F. and Thépaut, J.-N.: Multiple truncation incremental approach for four-dimensional variational data assimilation, Q. J. R. Meteorol. Soc., 124, 1889–1908, 1998.

Velders, G. J. M. and Daniel, J. S.: Uncertainty analysis of projections of ozone-depleting substances: mixing ratios, EESC, ODPs, and GWPs, Atmos. Chem. Phys., 14, 2757–2776, doi:10.5194/acp-14-2757-2014, 2014.

von Clarmann, T., Glatthor, N., Grabowski, U., Höpfner, M., Kellmann, S., Kiefer, M., Linden, A., Tsidu, G. M., Milz, M., Steck, T., Stiller, G. P., Wang, D. Y., Fischer, H., Funke, B., Gil-López, S., and López-Puertas, M.: Retrieval of temperature and tangent altitude pointing from limb emission spectra recorded from space by the Michelson Interferometer for Passive Atmospheric Sounding (MIPAS), J. Geophys. Res., 108, 4736, doi:10.1029/2003JD003602, 2003.

von Clarmann, T., Höpfner, M., Kellmann, S., Linden, A., Chauhan, S., Funke, B., Grabowski, U. andGlatthor, N., Kiefer, M., Schieferdecker, T., Stiller, G. P., and Versick, S.: Retrieval of temperature, H2O, O3, HNO3, CH4, N2O, ClONO2 and ClO from MIPAS reduced resolution nominal mode limb emission measurements, Atmos. Meas. Tech., 2, 159–175, doi:10.5194/amt-2-159-2009, 2009.

Wargan, K., Stajner, I., Pawson, S., Rood, R. B., and Tan, W.: Assimilation of ozone data from the Michelson Interferometer for Passive Atmospheric Sounding, Q. J. R. Meteorol. Soc., 131, 2713–2734, 2005.

WMO: Scientific assessment of ozone depletion: 2014, Tech. rep., Geneva, Switzerland, global Ozone Res. Monit. Proj. Rep. 55, 2014a.

WMO: Assessment for Decision-Makers: Scientific assessment of ozone depletion: 2014, Tech. rep., Geneva, Switzerland, global Ozone Res. Monit. Proj. Rep. 56, 2014b.

Table 1. List of acronyms and abbreviations for the satellite platforms and instruments used in this paper.

Acronym	Definition
AIRS	Advanced InfraRed Sounder
ENVISAT	ENVIronmental SATellite
ERS-2	European Remote Sensing 2
GOME-2	Global Ozone Monitoring Experiment 2
HIRS	High-resolution Infrared Radiation Sounder
IASI	Infrared Atmospheric Sounding Interferometer
MetOp	Meteorological Operational Satellite
MIPAS	Michelson Interferometer for Passive Atmospheric Sounding
MLS	Microwave Limb Sounder
OMI	Ozone Monitoring Instrument
SBUV	Solar Backscatter Ultra Violet
SCIAMACHY	Scanning Imaging Absorption Spectrometer for Atmospheric Cartography
TOMS	Total Ozone Mapping Spectrometer
UARS	Upper Atmosphere Research Satellite

Table 2. List of all other acronyms and abbreviations used in this paper.

Acronym	Definition
BUV	Backscatter Ultra Violet
CTM	Chemistry Transport Model
ECV	Essential Climate Variable
EOS	Earth Observing System
ESA	European Space Agency
EUMETSAT	European Organisation for the Exploitation of Meteorological Satellites
FP6	The Sixth Framework Programme
FP7	The Seventh Framework Programme
GCOS	Global Climate Observing System
GEO	Geostationary
GMAO	Global Modeling and Assimilation Office
GEMS	Global and regional Earth-system (Atmosphere) Monitoring using Satellite and in-situ data
H2020	Horizon 2020 Programme
IR	InfraRed
LEO	Low Earth Orbit
MACC	Monitoring Atmospheric Composition and Climate
MERRA	Modern Era Retrospective-analysis for Research and Applications
MOZART	Model for OZone And Related chemical Tracers
NASA	National Aeronautics and Space Administration
NOAA	National Oceanic and Atmospheric Administration
NWP	Numerical Weather Prediction
RAL	Rutherford Appleton Laboratory
SPARC	Stratospheric Processes and Their Role in Climate
TCO3	Total Column Ozone
UTLS	Upper Troposphere and Lower Stratosphere
UV	Ultra Violet
UV-Vis	Ultra Violet and Visible
VMR	Volume Mixing Ratio
WOUDC	World Ozone and Ultraviolet Radiation Data Centre

Table 3. Major differences between the ozone analysis system used in this paper and those of the ERA-Interim and the MACC reanalyses. Detailed information on the set-up of these two reanalyses can be found in the corresponding literature. For the ozone-specific differences, an indication of the region of the atmosphere where an impact is expected is indicated in the fifth column. In there, the **ST** and **TR** letters stands for Stratosphere and Troposphere, respectively. [+]The near real time (NRT), coarse vertical resolution profiles retrieved from NOAA-16, -17, and -18 SBUV/2 measurements. [$]Reprocessed, coarse vertical resolution profiles retrieved from NOAA-16, -17, and -18 SBUV/2 measurements. [*]NRT TCO3 from the SCIAMACHY measurements made in nadir-viewing geometry. [§]Infrared radiances from AIRS, IASI, and HIRS. [†]NRT TCO3 from OMI. [#]Reprocessed OMI TCO3. [‡]NRT profiles from MLS (v2.2), used down to 68hPa. [¶]Reprocessed profiles from MLS (v2.2), used down to 215hPa. [††]IASI channel 1088. [‡‡]AIRS channel 1585.

	Exp/Ctrl	**ERA-Interim**	**MACC**	**Region of impact**
Model cycle	CY40R1 (2013)	CY31R2 (2006)	CY36R4 (2010)	/
Hor. resolution	T_l511 (40 km)	T_l255 (80 km)	T_l255 (80 km)	/
Ver. resolution	L91	L60	L60	/
Top of Atm.	0.01 hPa	0.1 hPa	0.1 hPa	/
Used TCO3	SCIAMACHY[*]	SCIAMACHY[*], OMI[†]	SCIAMACHY[*], OMI[#]	ST
Used profiles	SBUV/2[+]	SBUV/2[+], MLS[‡]	SBUV/2[$], MLS[¶]	ST + TR
Used radiances	IR/O3[§]	N/A	N/A	UTLS
Bias correction (BC)	Yes	No	Yes	ST + TR
BC Anchor	SBUV/2, IASI[††], AIRS[‡‡]	N/A	SBUV/2, MLS	
Quality control	O-B < 30DU	N/A	O-B < 30DU	Mostly ST
Forecast model &	Modified CD86	Modified CD86		ST
Chemistry			MOZART-3 CTM	ST + TR
			(Kinnison et al., 2007)	

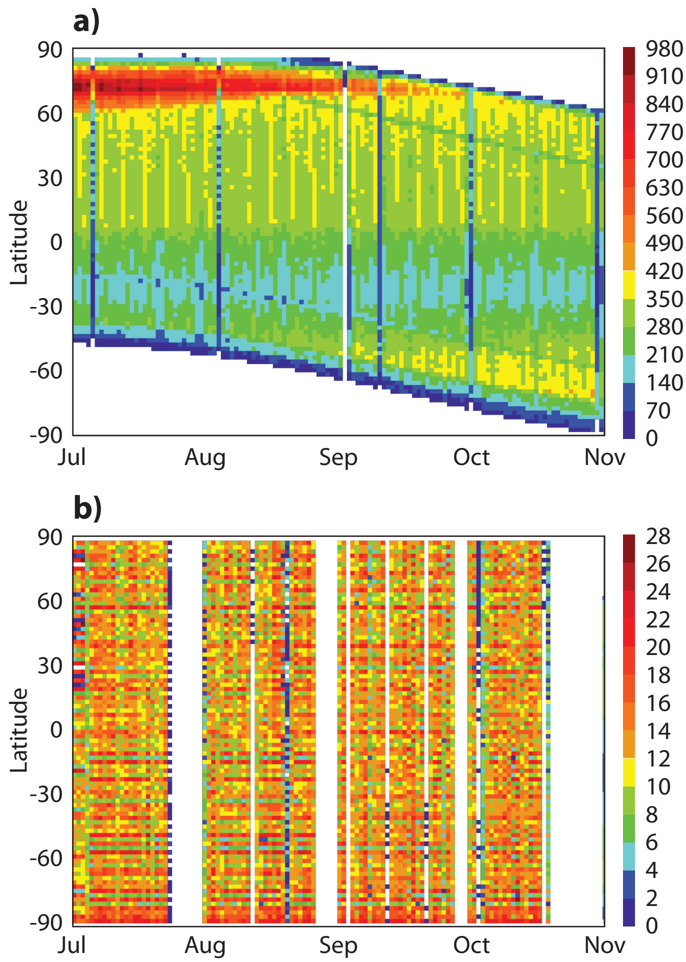

Figure 1. Data coverage and daily count binned in 2° latitudinal bands for the CCI GOME-2 (top panel) and the CCI MIPAS (bottom panel) data during the period July-October 2008. The two colour scales are different. The colour scale used for GOME-2 (panel **a**) varies from 0 (dark blue) to 980 (dark red) observations with a step of 70; the one used for MIPAS ranges from 0 (dark blue) to 28 (dark red) observations with a step of 2.

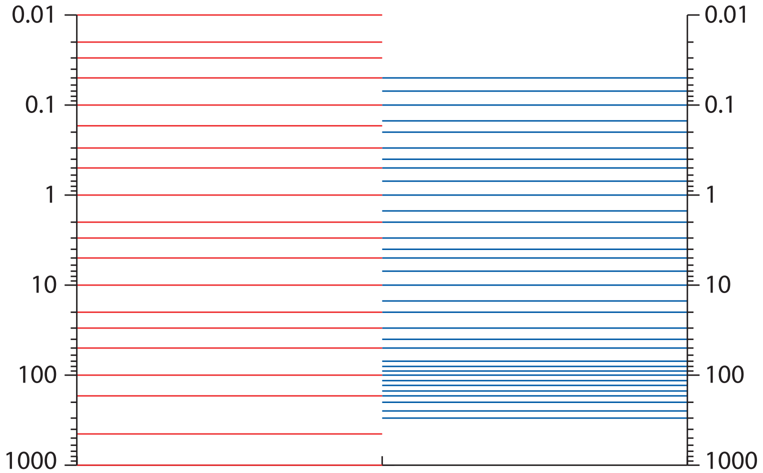

Figure 2. Schematic of the two ozone products' vertical coverage and vertical resolution as provided by the CCI GOME-2 (red lines) and the CCI MIPAS (blue lines) retrieval algorithms. The vertical axis represents pressure in hPa. The surface pressure level for GOME-2 is for illustration purposes drawn at 1000 hPa.

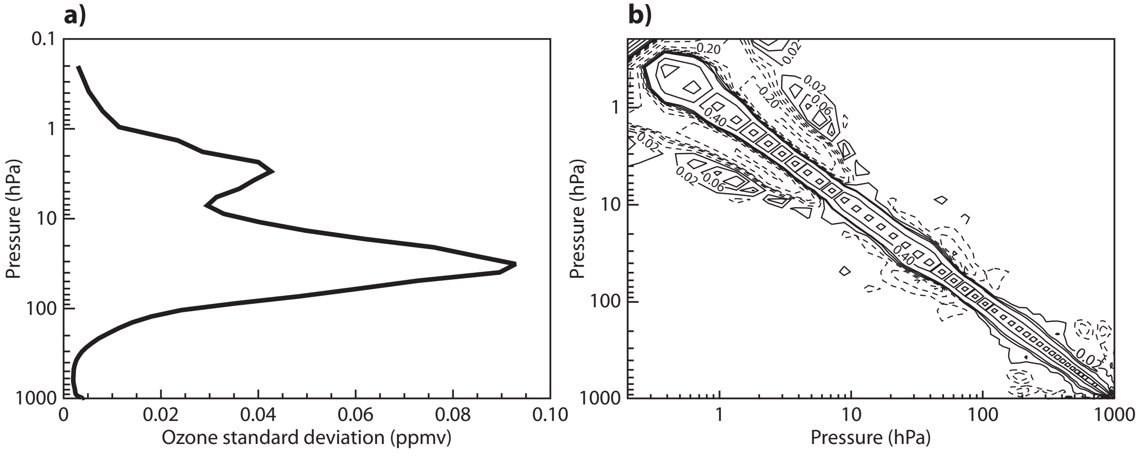

Figure 3. Example of ozone background error standard deviation profile (left panel) and vertical correlation matrix (right panel) for the ozone background errors. Data are in mixing ratio (parts per million by volume, ppmv). In the right panel, negative correlations are plotted as dashed lines, and the interval is 0.04 ppmv between -0.1 and 0.1 ppmv, and 0.2 for larger absolute values. Source Dragani and McNally (2013).

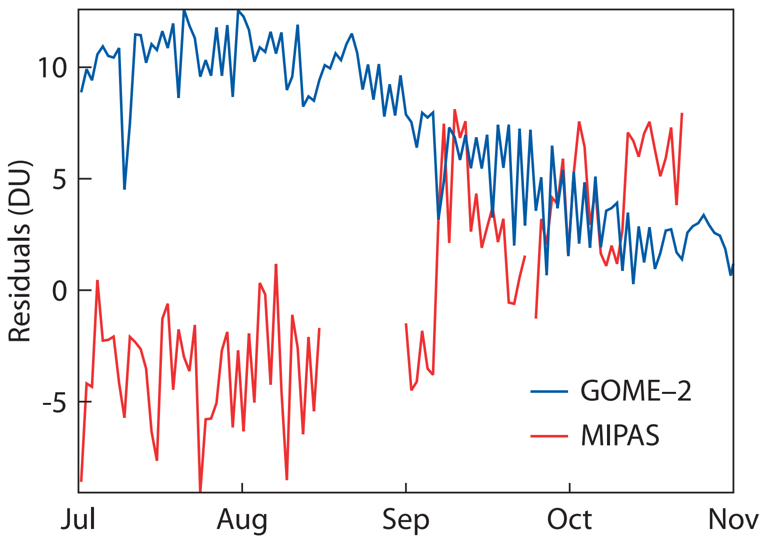

Figure 4. Time series of the global mean analysis departures for GOME-2 NPO3 (blue line) and MIPAS LPO3 (red line) over the period Jul-Oct 2008. Statistics are shown as integrated columns computed over the GOME-2 and MIPAS common vertical grid spanning from 0.05 to 100 hPa. Data are in DU.

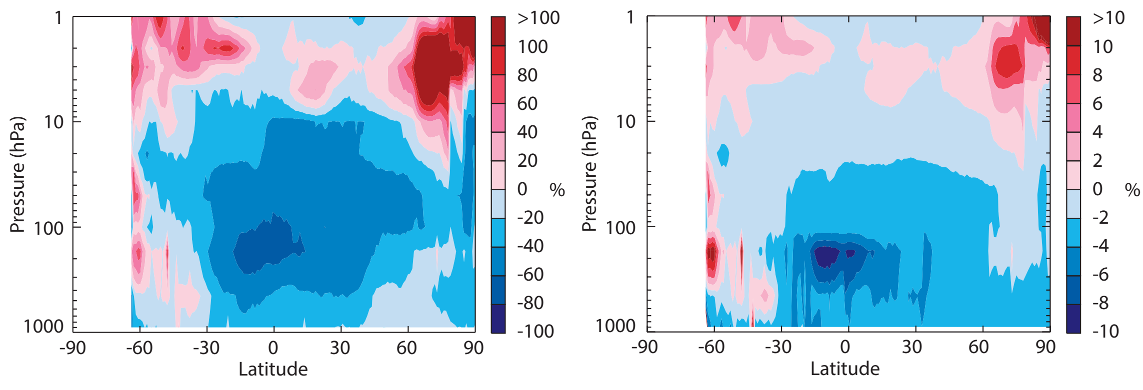

Figure 5. Left panel: Vertical cross-section of the difference between the estimated and the provided uncertainty relative to the provided uncertainty for GOME-2 NPO3 over the period Jul-Oct 2008. Negative (positive) values in blue (red) colours mean that the provided uncertainty is larger (smaller) than that estimated with the Desroziers et al. (2005) method. Data are in %.**Right panel:** As in the left panel but with relative difference computed with respect to the observation instead of its provided uncertainty. Data are in %.

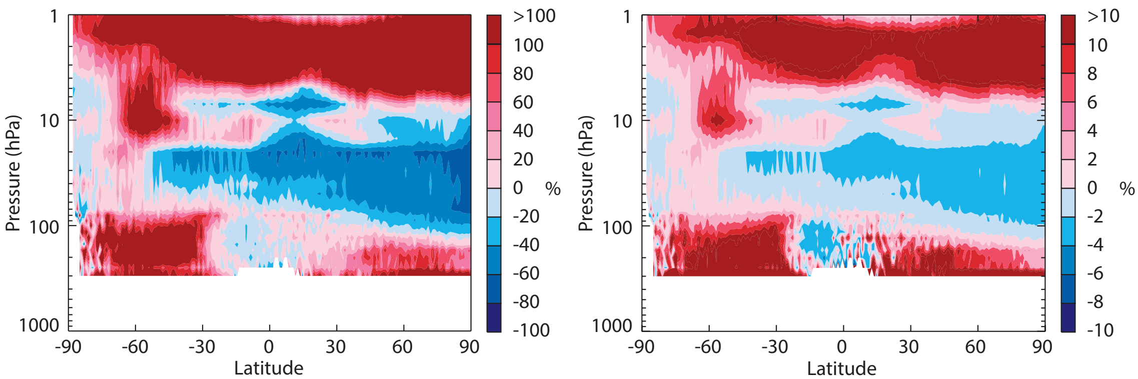

Figure 6. As in Figure 5 but for the CCI MIPAS LPO_3 data.

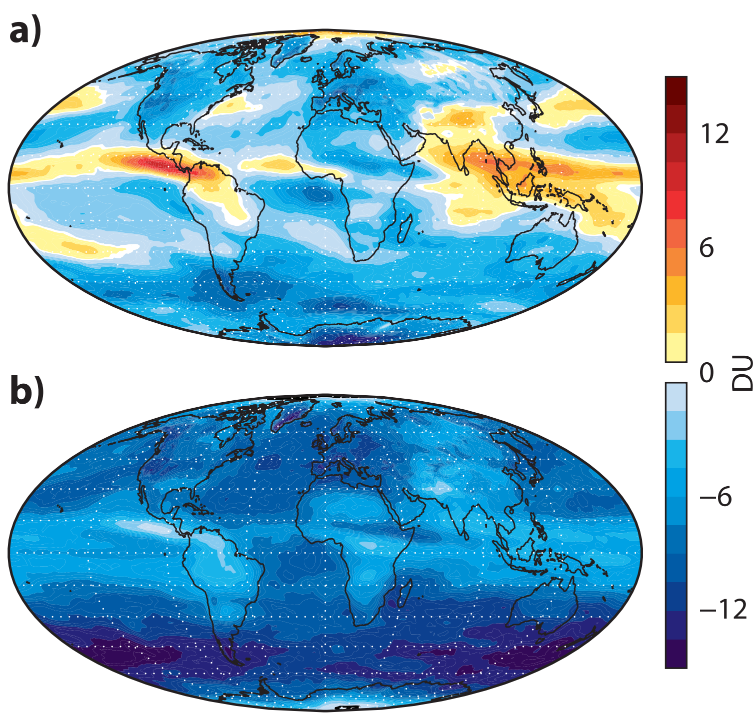

Figure 7. TCO_3 mean differences between the analyses from Exp/Ctrl and those from Exp/GOME2 (top) and Exp/MIPAS (bottom) computed for the period Aug-Oct 2008. Negative (positive) values in blue (red) colours mean that the additional instrument is increasing (decreasing) the TCO_3 amount compared to that obtained from the ozone datasets assimilated in Exp/Ctrl. Data are in DU. The colour scale ranges from -15 DU (dark blue) to +15 DU (dark red) with a step of 1.5 DU.

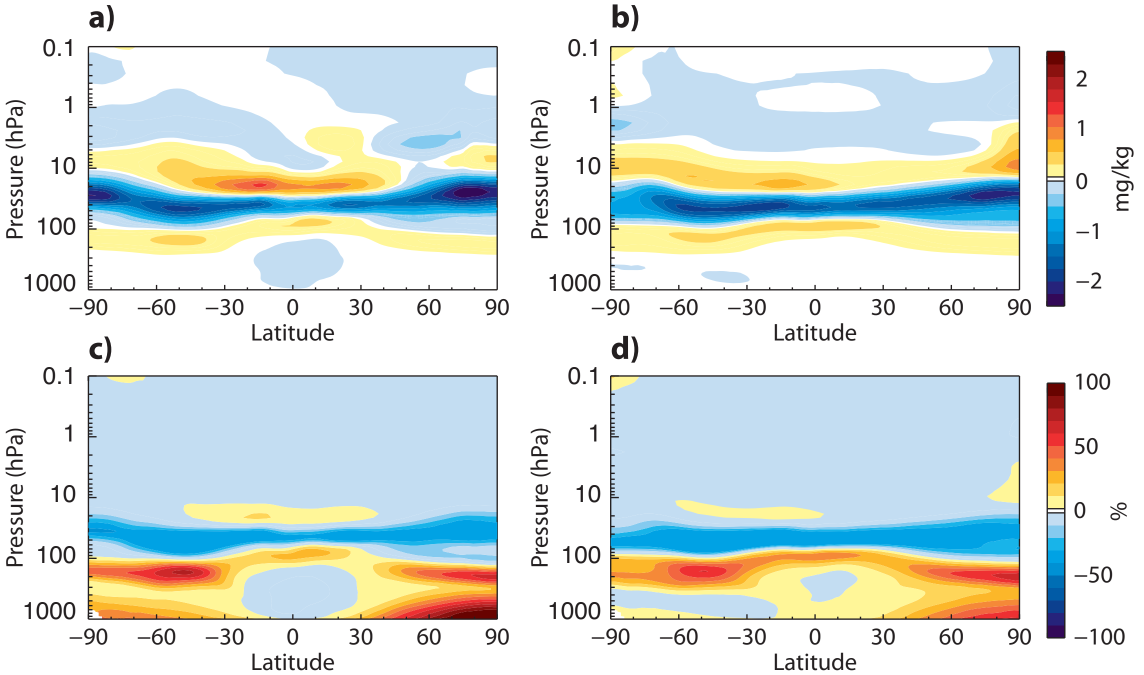

Figure 8. Top panels: Zonal mean temporal mean differences between the ozone analyses from Exp/Ctrl and the ozone analyses from Exp/GOME2 (top left panel) and Exp/MIPAS (top right panel) computed for the period Aug-Oct 2008. Negative (positive) values in blue (red) colours mean that the additional instrument is increasing (decreasing) the ozone mixing ratio amount compared to that obtained from the ozone datasets assimilated in Exp/Ctrl. The white regions are where the differences are within $\pm$ 0.1 mg/kg. Data are in mg/kg. The colour scale ranges from -2.5 mg/kg (dark blue colour) to +2.5 mg/kg (dark red colour) with a step of 0.25 mg/kg, with the exception of the yellow and pale blue that start at 0.1 and -0.1 mg/kg, respectively, and represent a step of 0.15 mg/kg. **Bottom panels:** Like in the top panels but for the relative differences computed with respect to a temporal mean global mean ozone profile obtained from the Exp/Ctrl analyses. Data are in %. The colour scale ranges from -100 % (dark blue colour) to +100 % (dark red colour) with a step of 10%.

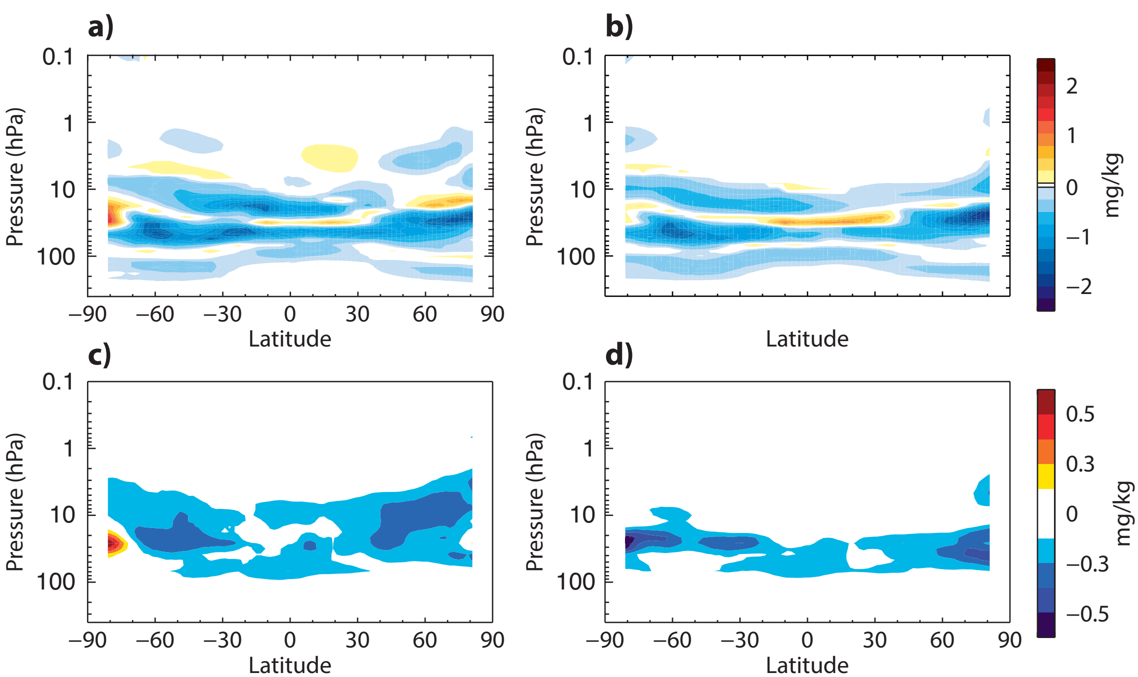

Figure 9. Top panels: Change in the zonal mean differences between the MLS retrievals and co-located ozone analyses from the perturbation experiment compared to the Exp/Ctrl for Aug-Oct 2008. The perturbation experiment is Exp/GOME2 in the left panel, and Exp/MIPAS in the right panel. Negative (positive) values in blue (red) colours indicate a decrease (an increase) in the mean in the Control and thus a degraded fit to MLS ozone profiles. Data are in mg/kg. The colour scale ranges from -2.5 mg/kg (dark blue colour) to +2.5 mg/kg (dark red colour) with a step of 0.25 mg/kg. **Bottom panels:** As a), but for the standard deviation of differences. The colour scale ranges from -0.6 mg/kg (dark blue colour) to +0.6 mg/kg (dark red colour) with a step of 0.1 mg/kg.

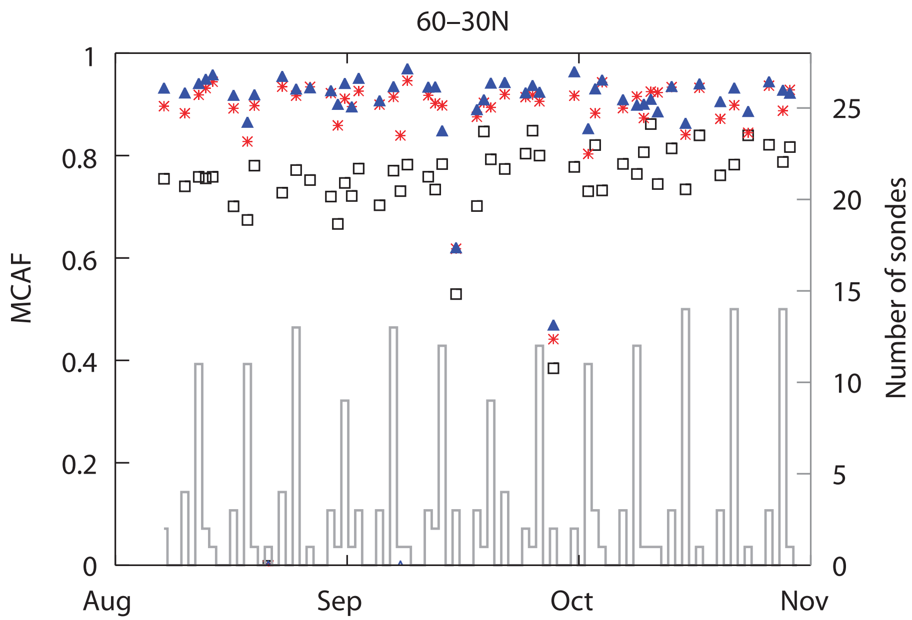

Figure 10. Time series of the stratospheric contribution (considered as that at pressure levels lower than 100 hPa) to the Mean Common Area Fraction (MCAF) score computed for different sonde profiles and the three sets of ozone analyses. The analyses were taken from Exp/Ctrl (black squares), and the two Exp/PERT experiments assimilating also the reprocessed GOME-2 NPO_3 (red asterisks) and MIPAS LPO_3 (blue triangles). The plot refers to the midlatitudes in the NH. The grey histogram in each panel shows the number of daily sondes included in the MCAF score and refers to the right hand side axis.

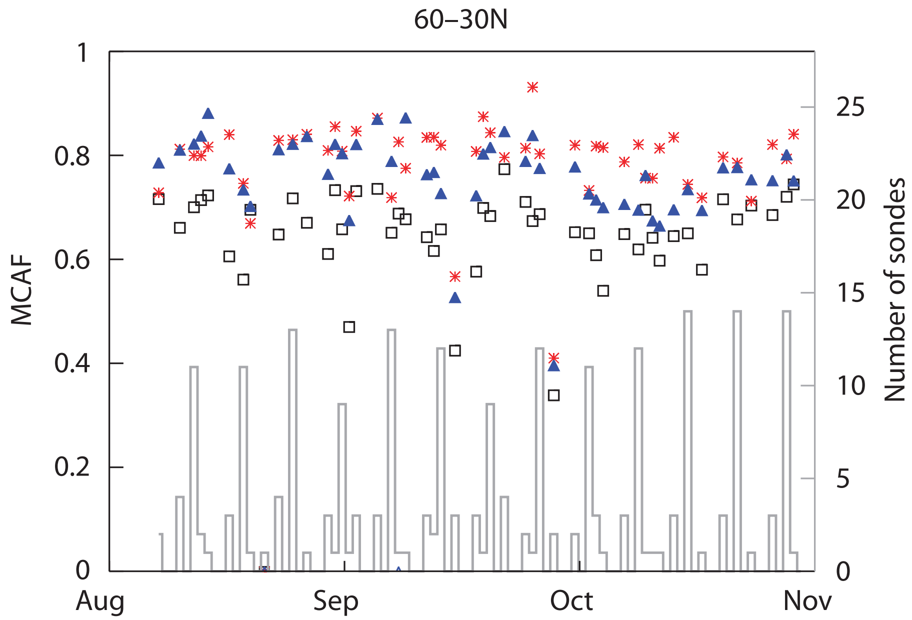

Figure 11. As Figure 10, but for the tropospheric contribution.

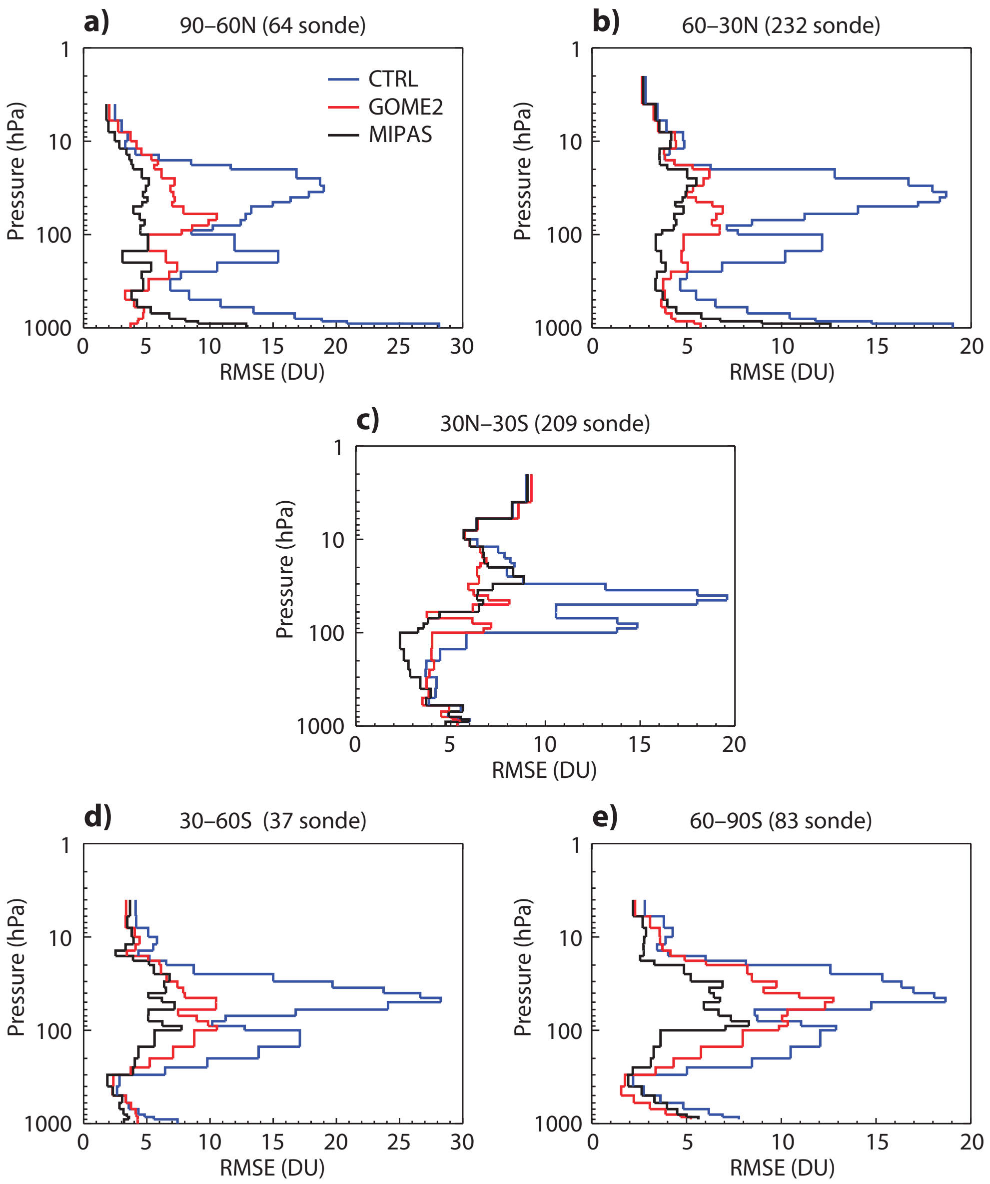

Figure 12. Fit of the ozone analyses from three experiments to ozone sondes given in terms of the RMSE over five latitudinal bands. The comparisons were computed by averaging over Aug - Oct 2008. The three analyses were taken from Exp/Ctrl (blue lines), and the two Exp/PERT experiments assimilating also the reprocessed GOME-2 NPO_3 (red lines) and MIPAS LPO_3 (black lines) from the O3-CCI. The latitudinal band each panel refers to and the number of ascents included in the average can be found in the corresponding panel title. Data are in Dobson Unit (DU).