Response to anonymous reviewer #1

First of all we would like to thank the referee for his/her thorough review of our manuscript and the helpful comments. Below you will find a point-by-point response (in black) to the reviewer's remarks (in blue).

MAIN COMMENTS

1. Please explain and discuss the selection of the horizontal resolution 1°x1°. Is it based on the characteristic structure of the ozone field?

The horizontal resolution of 1°x1° has been selected according to the user requirements defined for the ESA-CCI total ozone ECV product (User Requirement Document, 2011). These requirements are based on the ozone requirements of GCOS (Global Climate Observing System), CMUG (Climate Modelling User Group), IGACO (Integrated Global Atmospheric Chemistry Observations), and WMO (World Meteorological Organization). For total ozone a horizontal resolution of 20-100km is requested so that a 1°x1° grid should be a reasonable selection. This explanation has been added to Sec. 2.1.

Will sampling uncertainty be reduced if you select a coarser resolution (e.g., 2°x2°)?

The magnitude of the sampling uncertainty will be reduced using a coarser resolution (e.g., 2°x2° or 5°x5°) because the number of available measurements per grid cell increases. On the other hand the structures of the errors reflecting the spatial sampling of the satellite sensors remain.

Please discuss also the influence of different horizontal resolution of GOME, SCIAMACHY and GOME-2 Level 2 data on actual horizontal resolution of daily Level 3 maps.

The horizontal resolution of the L3 data (1°x1°) is significant coarser than the resolution of the single satellite L2 ground measurements. Therefore there is a minimal influence of the different sensor resolution on the L3 maps.

2. Merging strategy

a) Please explain the selection of GOME-2 as a reference instrument (the most stable? If yes, provide the corresponding references or analyses)

GOME (instead of GOME-2 as noted by the referee) has been used as a reference instrument because it completely overlaps (in time) with the other two sensors until June 2011. This enables us to calculate the correction factors for SCIAMACHY and GOME-2 for their entire periods from measurement start (2002 and 2007, respectively) until June 2011 which marks the end of the GTO-ECV CCI data record. Furthermore, GOME was found to be the most stable instrument over its lifetime before the application of the soft calibration scheme (Lerot et al., 2014).

b) Please explain the rationale for using only one instrument at a time. Would the alternative approach, using the data from all instruments in their overlapping period, reduce the sampling uncertainty?

We decided to use only one instrument at a time because
(1) for the period from 2002-2006, when GOME and SCIAMACHY simultaneously measure total ozone, GOME lost its global coverage in June 2003. Therefore GOME's advantage of a better sampling (global coverage within three days instead of six days for SCIAMACHY) was lost or rather limited to the European and North Atlantic region. Although data coverage was consecutively increased in the subsequent years by adding more ground-based receiving stations, we think that it won't be beneficial to include data from such limited regions.
(2) for the period from 2007-2011, when SCIAMACHY (sparse sampling) and GOME-2 (densest and most uniform sampling) simultaneously measure total ozone, only a small increase of 3-10% in data coverage per day is expected if SCIAMACHY data would be added.

3. Validation of Level 3 data.
From my point of view, the analysis presented in Section 3 is somewhat misleading. First, you explain that the representativeness of the ground-based network is insufficient in many places, and therefore cannot be used for monthly mean comparisons. Despite this, you then compared the monthly mean satellite and ground-based data and - as expected! – found deviations in some cases. Such comparison cannot provide information about the real quality of the GTO-ECV data.

The message of this section was in fact that the spatial representativeness error is avoided by co-locating at the level of individual grid cells, so not comparing zonal averages. However, we agree that the graph was confusing as it contained both the spatial and temporal representativeness error. We redid the graph so that it contains only the spatial representativeness error and clarified this further in the text. Due to our constraint of a minimum of 10 measurements per month for an accepted co-location, the temporal representativeness error is less of an issue here (and not so easily dealt with).

From my point of view, it would be much more informative to compare daily Level 3 data with coincident ground-based measurements in 1°x1° bins. The validation results will be, probably, close to those of Level 2.

We agree with the reviewer that a comparison of daily level 3 data with coincident ground-based data might lead to smaller differences between the level 2 and level 3 comparison because temporal sampling issues would be reduced/minimized. However, the main aim of this paper is to present and validate the official ESA-CCI Climate Research Data Package monthly level 3 data record product. From the resulting comparisons we feel that this aim has been achieved quite well within the unavoidable limitations of the validation process.

Furthermore, it is important to demonstrate that the applied correction improves the long-term stability of the data set, compared to original data sets. This important analysis is missing in the paper.

According to the companion papers by Lerot et al. (2014) and Koukouli et al. (2015), the long-term behavior of all three sensors is similar. Therefore, the time component of the correction factors was mainly applied in order to account for short-term fluctuations in the differences (e.g., for GOME-2 induced by the second troughput test in late 2009, see Figure 1).
However, Koukouli et al. (2015) found that the drift per decade w.r.t Dobson data is 0.2±0.2%, -0.7±0.0%, and 0.0±0.5% for GOME, SCIAMACHY, and GOME-2A level 2 data, respectively (their Table IV). With respect to Brewer data the drift per decade is 0.3±0.0%, 0.4±0.0%, and -0.6±0.4%, respectively.
For the combined level 3 product the drift per decade is -0.12±0.12% and 0.17±0.11% w.r.t Dobson and Brewer data, respectively, in the Northern Hemisphere (see Table 4), indicating an

excellent long-term stability of the data record.
We have added some more quantitative results related to the long-term stability in the text.

If you want to compare monthly mean data, then only the bins with uniform coverage by both satellite and ground-based data should be used. This would avoid speculations on the influence of non-uniform sampling.

Parts of this issue were already tackled through the application of the so-called cut-off values for latitude as a function of the month (see Sec. 2.1, p.4613, ll.15-22, and Table 2) which avoids monthly means based on non-uniform coverage in high latitudes close to the polar night.

In short, I suggest revision of the analyses presented in Section 3, and corresponding text and figures (and including only the results that illuminate the quality of GTO-ECV).

We agree with the reviewer and have revised Section 3.

4. Terminology and related issues.

a) “The standard error” (Sect. 2.3) is not a good term, because it is estimated as a sampling bias, and thus should be named properly. It should be also clarified: what parameter should be used as uncertainty of the monthly mean data?

The “standard error” introduced in Sec. 2.3 has been calculated as the standard deviation divided by the square root of the number of included measurements and multiplied by a factor accounting for spatial and temporal sampling issues of the sensors on a grid cell basis. Thus the error is in fact a “standard error” at the grid cell level.

b) “Seasonal variability” (Sect. 3.2.2) is also not a proper term for seasonal and latitudinal dependence of biases with respect to the Dobson network data.

We agree with the reviewer and have altered the title of the section accordingly.

c) Section 3.2.4. As explained above, I think that any conclusions about the data quality should be based on data with removed contribution of sampling uncertainties.

We agree with the reviewer that sampling uncertainties hinder whatever conclusions may be reached safely. However, we believe that the main take-home message of the validation part of this paper should be that the 1°x1° merged satellite product is of equal quality as the individual level 2 satellite products on a global scale, apart from a small number of outliers which definitely do not introduce statistically significant uncertainty issues.

It seem that SD (page 4622, line 8) has a different meaning that the parameter with the same name introduced in Sect. 2.3

We have changed “SD” to “standard deviation” in this sentence.

I have not understood the parameter “monthly mean variability”. Is this based on biases? Variance of a parameter x is var(x) = $<(x - \bar{x})^2 >$, but your unit is %, not $\%^2$. Perhaps, a formula would help.

The “monthly mean variability” was calculated as the standard deviation of the standard deviations of the monthly mean values, and not as the variance; apologies for this mistake.

It would be advantageous to rename “seasonality” into “seasonal variations of biases” and “latitude” into “latitudinal variations of biases”.

Changed as suggested.

5. The 2d histogram presented in Fig.14 indicates relatively large differences between GTO-GDP and GTO-CCI in some cases. Please discuss this in more detail and indicate the reasons for the differences.

We have analyzed the differences between GTO-ECV GDP and GTO-ECV CCI in more detail. Thereby we noticed a bug; the results of the comparison presented in Fig. 14 were based on a slightly extended (i.e. wrong) time period. We redid the analysis for the correct period (03/1996 – 06/2011) and found a new mean difference of 0.3±1.7 % (old comparison 0.2±1.7%).

The standard deviation of σ=1.7% indicates that 99.7% of the differences are within 0.3±5.1%. Hence, the large differences seen are extremely rare outliers. We think that the presentation of the differences as a 2-d histogram might be a bit misleading so that we would like to replace the existing plot with a new one showing only the mean difference ± 1, 2, and 3σ, respectively.

The large differences seen in Fig. 14 occur in spring (September to November) in the Southern Hemisphere and in early spring (February and March) in the Northern Hemisphere. In order to find the reason for the differences we compared monthly as well as daily mean data from the GDP and the CCI record. We analyzed SCIAMACHY data before they have been adjusted to GOME in order to exclude differences induced by the correction. Furthermore, we used L3 data sets which have been obtained using the same L3 gridding algorithm in order to exclude differences induced by differences in the L3 method. The remaining differences are of the same magnitude as seen in Figure 14. We conclude that the differences - and in particular the latitudinal structure of the differences - are mainly the result of the differences in the level 2 data records.
On the other hand, the application of different level 3 gridding methods (using all level 2 data per day vs. using only one measurement per day) leads to differences of up to ±4% in regions where two or more orbits per day overlap each other.

We have replaced the figure and changed/extended the text in Sec. 4.1.

6. P.4625, A discussion on “a small but significant time difference” (lines 15-20) looks not good without an illustration. Can this difference be seen in Fig.16 (bottom)?

We have deleted this sentence because the time dependence of the differences in total ozone between GTO-ECV CCI and SBUV-MOD has been presented already in Chiou et al. (2014) and McPeters et al. (2013); their Figures 4a and 6, respectively. A time dependence of the differences in standard deviations between GTO-ECV CCI and SBUV-MOD is not visible (see Fig. 16, bottom panel).

DETAILED COMMENTS (small clarification, rephrasing, technical corrections)

P.4610, L. 28. “the third article” Please indicate directly what are first two papers.

Done.

P.4616, L.11: explain the abbreviation SD

Done.

P.4616, L. 24-26: “left... right panel” -> “top... bottom”, respectively

Changed.

P.4618, L.1 “IFS-MOZART” explain the acronym

Done.

Figure 7 is illustrative. However, it should be specified which datasets are included to “ground-based measurements”. Since the comparison is performed with the each type of ground-based instruments separately, it would be reasonable to perform such analysis for each type of ground-based measurements.

The graph is based on all ground-based reference instruments used in the validation work. Spatial representativeness errors for only a single type of instrument would be even larger, but since these errors are avoided in the validation work by the co-location at grid-cell level, we consider these graphs of limited interest and prefer not to include them.

P.4618, L.21: “Any level 2 comparisons” - >”All level 2 comparisons ”?

Changed.

P.4622, L. 5-6: “(left)...(right)” . Please indicate also number(s) of Figure(s).

Done.

P.4624, L. 14 “Both data records”, please indicate them directly here.

Done.

Figure 16 (bottom). Please put a gray background for separation of small values and areas with no data.

Done.

P.4627, L.10-11. Perhaps, the sentence with explanation of CCMs is not needed.

The sentence has been deleted.

References:

Chiou et al., 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.

Koukouli et al., Evaluating a new homogeneous total ozone climate data record from GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, under review for J. Geophys. Res. Atmos., 2015.

Lerot et al., Homogenized total ozone data records from the European sensors GOME/ERS-2, SCIAMACHY/Envisat, and GOME-2/MetOp-A, J. Geophys. Res. Atmos., 119, 1639-1662, doi:10.1002/2013JD020831, 2014.

McPeters et al., The version 8.6 SBUV ozone data record: An overview, J. Geophys. Res., 118, 8032-8039, doi:10.1002/jgrd.50597, 2013.

User Requirement Document (URD), Issue 2.1, http://www.esa-ozone-cci.org/?q=webfm_send/37, 2011.