RESPONSE to Anonymous Referee #1

We would like to thank Anonymous Referee #1 for his/her careful reading of our manuscript and the useful comments made for its improvement.

REPLIES to the Major comments

1. Section 2 is called "Data sets and methodology", but I could not find any methodology. In my opinion, it is necessary to include a subsection explaining how was performed the comparison. For instance, were only direct-sun Dobson/Brewer measurements used in the comparison?, do the authors work with daily TOC ground-based averages or averages around satellite overpass?. According to table 1, the satellite footprint ground pixel size of the instruments are very different, thus, which was the spatial collocation criteria used in the comparison exercise?. All this questions should be answered.

Paragraph added as requested.

2. In subsection 3.1.2, the authors must explain the possible causes of the clear seasonality shown in the evolution of the satellite-Dobson differences for all satellite algorithms except OMI_TOMS (Fig. 2). Is this behaviour related to the "obvious SZA dependency" seen in Fig. 4 (second row left)?. In my opinion, the authors should relate these two plots.

The fact of the matter is that all satellite DOAS algorithms [in our case, the GOME-SCIAMACHY-GOME2 and OMI_DOAS] datasets suffer from seasonality effects due to the fact that different effective temperatures are used for the ozone cross-sections whereas the ground-based measurements use a steady temperature effect. Furthermore, at high SZA, it is known that the DOAS algorithms cannot fit well the observed features, what with the SZA effect being both a seasonal and geometrical issue. What we are seeing in this Figure is mixture of both these effects. A relevant text was added to the article.

3. Brewer and Dobson instruments are calibrated using the Bass and Paur ozone cross-sections at a fixed temperature while the satellites use BDM. This must be commented in the text. Do the authors think that temperature dependence of cross sections could partially explain some of the differences reported in the manuscript?.

With the GDP 4.x family of algorithms, we do not use the BDM cross-sections but rather cross-sections measured with each instrument: GOME x-sections for GOME and GOME-2,

SCIA x-sections for SCIA. The exception being TOMS V8 which uses, to the best of our knowledge, the Bass & Paur x-sections. It certainly falls within the plans for the next generation of algorithms, namely, the GDP5.0 Direct Fitting algorithm, to harmonise the use of x-sections, however this will come at a later stage. On the impact of the temperature, the point as we understand it is more whether some of the reported differences could come from the fact that ground-based data sets use cross-sections at a fixed temperature (while temperature dependences are accounted for in satellite algorithms). We postulate think that the answer is most probably yes, and it probably explains part of the different behaviour between Dobson and Brewer instruments (Dobson measurements are more sensitive to temperature variations).

4. This reviewer would like know why for the Brewer comparisons (Fg.4 second row right) the plot is "a lot more homogeneous" than for Dobson comparison.

Relevant information added to the text as requested.

5. The analysis of the dependence on CTP and cloud fraction does not exit. Only a brief description of the plots is given (page 3033). A discussion of these results is need.

Relevant information added to the text as requested.

6. Previous validation papers are only briefly commented in the text. Although they qualitatively agree with the present results, they are not really compared. The authors should add to section 3, a new subsection with a detailed comparison between their results and the results of previous studies.

The relevant text was added to the discussion of Figure 1, as well as the discussion of section 2.1.

REPLIES to Specific comments

- Page 3021. Line 20. GDP must be defined here.

Amended as asked.

- Page 3022. I think that lines 12-13 and 24-26 refer to the same issue. Please rewrite lines 12-13 using the information given in lines 24-26.

Amended as asked.

- Page 3025. Lines 2-4. A brief explanation about this decreasing trend should be given here.

The issue is related to the well-known degradation effect in SCIAMACHY which generates time-dependent spectral features in the measured reflectances and introduces an artefactual trend in the ozone total columns. This problem has been under investigation in the SQWG, and level-1 experts are looking for possible improvements of the calibration methods to solve it.

- Page 3026. Lines 25-26. This information should be moved to Page 3023, line 15.

Amended as asked.

- Page 3027. Lines 15-16. The Brewer retrieval algorithm derives the total ozone column from four wavelengths, not five. An additional wavelength is used to retrieve SO2.

Amended as asked.