Final Response Anonymous Referee #1

We would like to thank the referee for his comments. Most have been incorporated in the newest version of the manuscript. Below our response to each comment.

This study seems to emphasize too much on the use of convergence behavior to evaluate the retrieval performance.

> In the new version we rephrase the purpose of paper: using the convergence behaviour of the algorithm as a diagnostic tool to isolate (and solve) different causes of non-convergence. This results in a faster algorithm with more successful retrievals. To be clearer about this we changed the title to "Improving ozone profile retrieval from spaceborne UV backscatter spectrometers using convergence behaviour diagnostics"

Furthermore, the use of DFS to evaluate the retrieval performance due to different climatologies is not justified because the DFS (same measurements and measurement errors, but with different climatologies) merely indicates the strength of the a priori constraint instead of the retrieval performance. Instead, other than the improvement of convergence behavior over special conditions, they should use other correlative measurements to evaluate and determine the choice of a priori constraint, which can have significant impacts on the quality of the retrievals for different altitude regions. Please see specific comments for more detail.

> In the new version of this paper the TOMS climatology has been expanded with a more realistic error estimate, based on sonde measurements, instead of three fixed relative errors. The results and interpretation will be restated. Our main point of the intercomparison is not to select the climatology with the best convergence statistics, but to show that the variability of the FK and MLL climatology is not always large enough to allow for successful retrievals in ozone anomalies such as the ozone hole in October 1998. By taking the total ozone column as an extra parameter in the climatology, a more appropriate a priori can be selected, solving these convergence problems.

Specific comments:

1. Page 1165, line 25, miss reference by Liu et al., 2005, JGR, 110 (D20307, doi:10.1029/2005JD006240)

> Reference has been added.

2. Page 1166, line 5, it would be very useful to summarize the quality of GOME OPERA retrievals based on the validation work of De Clerq et al. (2007) either here or in section 2.

> A summary has been added in Section 1: "Using an older version of the OPERA algorithm, the quality of the retrievals from GOME data has been validated extensively by De Clercq et al. (2007) by comparing them with sonde, lidar, microwave and analyzing the averaging kernels. The validation shows that from 1996-1998 instrument degradation issues are least, and, consequently, retrievals are best. The stratospheric part of the profile typically agrees within 10% with the correlative data. The agreement in the tropospheric part depends mainly on the latitude of the ground-based station. At high latitudes, where information content is at

its maximum, agreements within 10% are found; in the tropical zone, a bias of 35% and additional seasonal oscillations are introduced."

3. Page 1167 line 22, it says version "4.01", Table 1 says "GDP 3.02", so which version is used?

GDP version 4.01 is used throughout this study. Table 1 has been corrected.

4. In Table 1, what measurement error is assumed for level 1b data? Please provide this information since it is also very critical to retrievals just as the a priori covariance matrix and it affects the magnitude of the DFS.

Radiance and solar irradiance errors are taken from level 1b data. Table 1 has been adjusted.

5. Page 1170, in the equation before line 10, how is the reflectance error assumed/calculated?

Radiance I and solar irradiance I_0, and their errors σ_I , σ_0 are taken from level 1b data. Given reflectance $R=I / I_0$, the corresponding error is calculated from $\sigma = R \,\mathrm{sqrt}((\sigma_I / I)^2 + (\sigma_0 / I_0)^2)$. Information added to Table 1.

6. Page 1170, line 2, is the SAA filter only applied in the SAA region? At high latitudes, radiances at shorter wavelengths might also be affected by high energy proton particles (although much less frequent) or similarly affected by NO gamma and metal emission lines. The SAA filter might also be applied to these conditions.

The filter is applied for all measurements; this is better explained in the new text. As can be seen in Figure 4b, the filter is mainly active within the SAA region. Outside this region, where no spikes are detected, all measurements pass the filter.

7. In section 5, it was suggested that the retrieved negative surface albedo over dust conditions is mainly caused by the overestimate of FRESCO small cloud fractions. It should also be mentioned that the neglect of very absorbing dust aerosols (in the radiative transfer simulation) might also partly cause this occurrence of negative surface albedo values. I guess that negative values can still occur under heavy dust loading conditions even assuming no clouds.

The effect of aerosols is discussed in the new text: "The presence of absorbing aerosols, which are not included in the atmospheric model, cause similar radiance closure problems at the top of the atmosphere. The proposed workaround also solves these convergence problems, as can be seen from the dust outbreak event in February1998 flowing out from West Africa towards South America (Figure 5c). The dust cloud absorbs radiation in UV, lowering the reflectance measured in this regime. At the same time, FRESCO attributes the increased reflectance around 750 nm due to the presence of the dust cloud over a dark ocean to an increased effective cloud fraction. These two effects will force OPERA to retrieve the surface albedo below zero, which can be avoided by assuming a cloud-free model atmosphere."

8. Page 1176, line 16, a more appropriate reference for the TOMS V8 climatology is: Bhartia, P. K., and C. G. Wellemeyer (2002), TOMS-V8 total ozone algorithm, in OMI Algorithm Theoretical Basis Document, edited by P. K. Bhartia, Greenbelt.

Reference is changed.

9. In section 7, the use of DFS as an indicator to evaluate the retrieval performance due to different climatologies is not useful and inappropriate here, which has also been suggested by reviewer 1. Higher DFS here mainly indicates weaker constraint and weaker constraint/larger DFS is achieved at the cost of larger retrieval errors. According to the optimal estimation, we should use the actual climatological a priori error so as to optimally combine the information from our a priori knowledge and measurements. So by using a fixed a priori error of 20 percent, which can be too tight for the UT/LS but too loose for middle stratosphere, the inversion is strictly speaking not optimal estimation any more. Although the convergence behavior is improved especially under some extreme conditions (due to the poor representativeness of the FK and MLL climatologies for these conditions), the overall retrievals might not be improved. As matter of fact, the choice of a priori error can have a significant impact on the quality of the retrievals.

To determine a more realistic variability of the TOMS climatology, we have compared its profiles with all ECC sonde measurements done in 2000-2008 available at the WOUDC. Most sonde data include the total ozone column, integrated from its measurements up to balloon burst and extrapolated with climatology, or taken from parallel ground-based or space-borne total column measurements. Based on this total column value we pick a profile from the TOMS climatology, which is interpolated to date, latitude, and total column value. From the difference between sonde and TOMS profile, the variability is determined for 5 latitude bands and 11 atmospheric layers. Due to balloon burst, not sufficiently statistical data is available above 8 hPa. Here a fixed 15% is taken, corresponding with the typical relative variability in and above the ozone bulk of the MLL climatology.
Also added to the new manuscript: "The variability of TOMS shows that by including the total column as an extra parameter the base of the ozone bulk is better defined. For mid-latitudes the RMS error is ~30% in the upper troposphere and lower stratosphere, while *a priori* profiles taken from FK or MLL typically show an RMS error over 70% for this region."

I also agree with reviewer 1 that the authors should use correlative measurements to determine which choice of climatology is more appropriate to improve the retrievals.

An extensive validation study to assess the global improvement of the algorithm is considered to be outside the scope of this article. In this study, we emphasize on the use of convergence statistics as a diagnostic tool to evaluate possible (perhaps otherwise undetected) problem areas of the retrieval algorithm.

Figure 9 may be deleted because it does not really provide much useful information to evaluate the retrieval performance.

The analysis of the comparison of ozone climatologies have been rewritten. Figure 9 has been deleted.

10. In the last paragraph of section 7.4, the authors suggest choosing a priori error based on the application needs (e.g., speed, maximum information from the measurements). In addition to the comments about DFS mentioned above, the speed (due to different climatologies) is also not a critical issue here. The differences in No. of iterations between 15-25 percent is about 3-5 percent, which is too small to be concerned for operational processing. What matters a lot more is the quality of the retrievals. Therefore, I suggest deleting this paragraph.

We agree with the reviewer. Therefore, the analysis of the comparison of ozone climatologies have been rewritten. The mentioned paragraph has been deleted.

11. Page 1180, line 2, I think that it is always good if you can carefully remove those spikes. Including them will cause larger retrieval errors. The nominally large DFS (when including these pixels) is mainly due to the underestimate of measurement errors for those pixels affected by spikes.

We agree. The problem is to design a selective filter which only removes spikes, and not the unaffected measurements. Added to the text: “Alternative, more elaborate SAA filter schemes can be implemented; note however that filtering too much of the unaffected measurements will cause unwanted information loss.”

Technical corrections (...)

Typos are corrected.