Final response J. Landgraf

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.

Concerning the overall approach of the study I doubt that it is sufficient to evaluate the algorithm performance only on the convergence behavior of the code. To use an extreme, a very strict regularization will cause only little changes in the retrieved profile of two consecutive retrievals and thus eases to achieve a convergence as defined in Section 3. However the quality of the fit may be poor and in turn the retrieved ozone profile is of low quality.

> We feel that this remark is only relevant for Section 7, where different climatologies are compared. In the previous sections, algorithm adaptations are done with the same climatology. Here, improvement of convergence statistics is a direct consequence of the adaptations and can not be attributed to a different regularization. Regarding the comparison of the climatologies: 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.

To my opinion it would be very useful and also would emphasize the strength of the presented methods if the overall quality of the retrieved data is discussed as well. For example the SAA filtering in section 4 improves on the number of iterations needed in the SSA zone. But is the quality of the remaining data sufficient? Although this point is probably hard to address due to the lack of validation points in this area, it is crucial to judge the overall benefit of such a filtering.

> Filtering the spikes in the SAA area means discarding information from the measurements, but on the other side it enables a valid retrieval based on data which otherwise would have been lost. Like the reviewer points out, validation with ground-based measurements is complicated, and we consider it outside of the scope of this study. However, we have added to the paper information on the spectral residuals by looking at the goodness of fit of the forward model: "Filtering the measurements improves the goodness of fit of the forward model in the SAA region considerably: the reduced chi-square for converged retrievals drops from 52 to 6.5. Outside the SAA region the converged retrievals fit with $\chi^2_{red} = 1.3$."

In general the authors give proper credit to related work. Only in Section 5 I miss a short overview on previous work. I did not follow the most recent discussion on this subject but I remember the work of Lui et al. 2005 and van Diedenhoven et al., 2007, which both proposed a retrieval approach to deal with clouds of GOME ozone profile retrieval. The authors of the manuscript choose for a different approach which is probably motivated by the computational cost of the algorithm. I would appreciate here a more thorough discussion of the literature and also a validation of the OPERA ozone profiles for cloudy conditions.

In the new version we added a short overview of alternative cloud parameter retrieval methods such as the retrieval of cloud pressure and fraction in the O2-O2 band, and the mentioned method by Van Diedenhoven et al. We also added an overview of the extensive validation study of OPERA profile retrievals based on GOME measurements by De Clercq et al. (2007). They find that the agreement in the tropospheric part not so much depends on clouds, but mainly on the latitude. At high latitudes, where information content is at its maximum, agreements within 10% are found with ground-based stations; in the tropical zone, a bias of 35% and additional seasonal oscillations are introduced.

The use of Section 6 is not clear to me. To my opinion it mainly confirms findings of Liu et al. in the context of the OPERA algorithm. It is not clear to me if the presented validation with microwave measurements provides new insides respect to the work of De Clerq et al and Liu et al. If the authors think that this section is really needed then I suggest to shorten this section significantly and to put emphasize on the significance of this improvement for a global data set in relation to other error sources.

Bad ozone cross sections result in forward model biases, which in turn result in convergence problems. Section 6 can be considered as a confirmation of the work of Liu et al., which according to reviewer #2 is relevant within the framework of this study. The section has been abridged in the new version, as suggested. The validation study with the microwave instrument is dropped; instead an analysis on the improvement of the spectral fit is added.

The methodology presented in Section 4-6 is clear and well described. I have some difficulties with Section 7 where the authors discuss the effect of different choices of prior information coming from three different ozone climatologies. The ozone climatology is used as a first guess to start the iterative approach and as a side constraint to solve the ill-posed inversion problem by regularization. Here the retrieval is effected by the mean prior profile and the corresponding prior covariance matrix. The intercomparison (section 7.4) is based on the convergence behaviors and the DFS of the retrieval which raises two main questions to me:

First is the convergence behavior of the algorithm mainly governed by the first guess assumption or by the regularization i.e. the covariance matrix.

We have added additional results showing that the value and shape of the first guess scarcely influences the retrieved profile and the number of non-convergent retrievals, although a first guess taken closer the true profile results in less iteration steps. We have changed the comparison of the different climatologies: we will not take the first guess from the a priori, but from the previous retrieval instead. This guarantees more resembling sets of starting points.

Second, is a large DFS an indication for a better retrieval or only for a weaker regularization?

The DFS is a good measure for the quality of the retrieval, assuming that all climatologies use a realistic error covariance. In the new version of our study the analysis of the comparison has been rewritten. We replaced the TOMS climatologies at three different fixed relative errors by one with a more realistic error, based on the variability of the climatology with respect to ozone sonde measurements from the WOUDC database. Comparing retrievals with FK, MLL and this TOMS, one can see that better convergence is coupled to lower DFS. In this sense, the low DFS of MLL is an indication of the stricter regularization of this climatology.

I am convinced that in this context a validation with independent measurements can improve the analysis.

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.

Page 1172 line 15. The effect described in the text is hard to see in the corresponding fig 5c. A zoom-in would be very helpful.

A new image is made for Figure 5c, for which we feel that the dust storm from the Sahara is sufficiently distinguishable.

Figure 7: typo in caption 1018 molecules -> 1E18 molecules

Typo in Figure 6 is corrected.

Figure 8: What is the reason for the larger number of iterations for the TOMS climatology at the SAA area?

The mean number of iterations for TOMS at 20% is comparable with FK, but is larger than for MLL. Due to smaller *a priori* error of the latter, more confidence is attributed to the *a priori* and convergence is reached more easily. The referee's observation has been included in the analysis of the comparison: "The high convergence rate [of MLL] is facilitated by the small variability of the climatology, which forces the retrieval towards the a priori. Because less weight is put to the measurements, MLL shows also better convergence in the SAA."