Response to Reviewer 3's comments on “The GRAPE aerosol retrieval algorithm” by Thomas et al.

The authors thank the reviewer for their comments and recommendation, which we believe we have addressed, as detailed below. The reviewers comments are included in italic text, with our responses in regular font.

Thomas et al. present an aerosol retrieval algorithm and examine its performance using artificial data. The particularly interesting feature of the algorithm is that it provides an error estimate.

The manuscript is well written, well structured and concise.

I have five more general and a few specific critics to the manuscript:

1) It is somewhat disappointing that no retrievals on real data are compared to reference data. It is understandable, though, that the authors intend to do so in a forthcoming paper. They might consider publishing this as a “Part II” to the present article.

The follow up article, giving a validation of the algorithm applied to the full ATSR-2 record has been submitted to Atmos. Chem. Phys.

2) The acronyms are somewhat confusing. If indeed “GRAPE” and “ORAC” both need to be used, the difference between the two names needs to be clarified. “GRAPE” is used in the title, but most of the paper deals with “ORAC” (if both are exchangeable, GRAPE would be preferable since “ORAC” and the also used acronym “OPAC” are quite similar).

The use of the ORAC and GRAPE acronyms has been made clearer in the paper. The term ORAC is now used exclusively to describe the collection of retrieval schemes which share the basic algorithm. GRAPE is used to describe the project under which the retrieval under discussion was developed, and to refer to the product itself. The term “GRAPE retrieval” is now used to describe the particular version of ORAC used in the GRAPE project.

3) On p983, l17 it is stated that there are better aerosol retrieval algorithms available in ORAC. A discussion of this statement would be helpful. Why are they more advanced? Importantly, why should then the algorithm presented in this paper be applied?

The manuscript has been modified to give more details about the providence of the algorithm and explain its configuration.

4) A crucial limitation of the algorithm is the choice of the a priori aerosol characterisation. Why do the authors limit their retrieval to just five types (or, rather, three types considering that “Arctic” and “Antarctic” types are used only in the respective regions, where the retrieval likely anyway gives highly uncertain results due to the high surface albedo)? Well-defined aerosol climatologies exist with much more detailed information about spatio-temporal distributions of aerosol types.

The selection of appropriate aerosol types is a difficult choice in the development of all aerosol retrievals. More complicated aerosol climatologies were trialled in the development of the GRAPE product, but the extra complexity was found to offer no significant improvement in the quality of the products (vs AERONET for example). Indeed, such schemes tended to introduce obvious spatial artefacts into the results. An explanation along these lines has been added to the manuscript.

5) The important feature of the algorithms is to provide error estimates. If it can be shown that these estimates are meaningful, then the product would be highly useful even if erroneous. Thus, it would be very helpful to show in Figures 3, 4, and 6-13 as a third column the error estimate *minus the "true" error in order to identify where the error estimate works well (and also, where it is exaggerating the error substantially).*

It would be helpful to show (by hatching or outlining) regions where the retrieval is consistent with the reality (where the truth is within the error bar of the retrieval).

The authors investigated adding such plots, but feel they don't really add to much to the paper. If the underlying assumptions of the retrieval are incorrect (eg. the wrong aerosol class is used), then the retrieval generally gives erroneous results, but with uncertainties similar to those of a retrieval using correct assumptions. If the assumptions made are grossly inaccurate, the forward model will no longer be able to match the measured radiances and the retrieval will either fail, or converge with a high cost. In either case, the retrieved uncertainties don't tell you how accurate the retrieval is, just how well constrained the minima in the cost function is. Plots of the retrieved error compared to the true error for situations where incorrect assumptions are made reflect this: essentially, the retrieved errors are under estimates for most states, except where the assumptions made in the retrieval are correct.

The authors have modified the manuscript to make this point more clearly.

Specific comments:

p986 l6: scattering

Corrected

p987 l7: You have done the retrieval on ATSR2 already. Can you estimate how often radii larger or smaller than the extreme values in the LUT occur?

The frequency of this occuring in real retrievals has been commented on in the manuscript (it is rare).

p988 l10: perhaps "!I is given as the average"

Agreed

p992 l19: are these relative errors (fractions)?

No, they are absolute. This has been clarified in the manuscript.

p992 l19: The choice how to characterise the measurement uncertainties seems ad-hoc. Can they be justified? Isn't here a discrepancy to the quantification on p1000

These errors are somewhat ad-hoc, however determining accurate figures for the measurement noise on such measurements is not easy. The value of 2% quoted on p1000 wasn't definitive: getting a definitive value for the noise on satellite measurements is not an easy task. P1000 has been modified to represent the range of possible values from Smith et al. (1 – 5%). The discussion of the measurement errors on p992 has also been expanded to provide some justification.

l25?

The explanation of this has been expanded and an explanation of the method of including forward model error in the measurement error covariance matrix has been added to the previous section.

p993 l1: Why would you retrieve the effective radius also in the log space? Have you investigated what would change if you did it otherwise?

Effective radius is retrieved in log-space for similar reasons as the optical depth:

1) Retrieval in log space prevents negative states from being retrieved, without introducing a potential positive bias, as can be introduced if a hard cut-off at 0 is used (such a bias is the reason small negative AODs are allowed in MODIS collection 5 retrievals, for example).

2) Examining the distribution of aerosol effective radius (from AERONET Almucantar retrievals, for example) reveals that their distribution is not symmetrical, but has a long tail towards large radii, somewhat like a log-normal distribution (the lack of such a tail towards small particles is, of course, because radius must be positive). Since the retrieval implicitly assumes a symmetrical Gaussian PDF for the retrieved state, retrieval in log-space is appropriate.

The authors have not experimented with retrieving effective radius on a linear scale.

p993 l4: What is the justification of using such a low uncertainty in surface albedo?

This is a pragmatic decision: using a more realistic (for land surfaces) constraint on the a priori surface reflectance produced poorly constrained retrievals of AOD and effective radius. It should also be noted that a 0.01 error on the reflectance of the (dark) Ocean surface is not unreasonable.

p993 l15: Have the radiances been produced by the same model (DISTORT)?

Yes, they were. This has been clarified in the manuscript

p993 l18: What is the correct surface albedo value?

This as been stated explicitly in the manuscript (0.02).

p995 l18: maybe "error" rather than "precision"

The authors disagree. See the response to the similar point raised by Reviewer 2.

p1000 l2: beginning

Corrected.

p1000 l12: How often is "most"?

This has been quantified in the manuscript

p1000 l12: It would be quite useful to break these error estimates down to the individual sensitivity studies. Could you perhaps add a few rows to Table 1 showing for each sensitivity study the fraction of cases with errors

The authors feel that the reviewer has misunderstood how this sensitivity study has been done. The range of values given in Table 1 essentially define a 5-dimensional hyper-cube of potential values from the forward model. It is not really possible to show the variation due to each individual variable given in Table 1, since the error in each one is dependant on the value of all of the others.

p1001 l13: class

Corrected

p1004 l11: drop “haves”

Done

p1009 l11: against

Corrected