Authors' response to Anonymous Reviewer 1's comments on “Validation of the GRAPE single view aerosol retrieval for ATSR-2 and insights into the long term global AOD trend” by Thomas et al.

The Authors would like to thank the reviewer for their efforts and constructive suggestions, which have resulted in a much improved paper. The reviewer's specific comments and corrections are addressed below and are included in italics for reference.

> *General: The authors present the validation of results from the GRAPE algorithm applied to ATRS-2 to retrieve aerosol properties: AOD at 550 nm and effective radius are the primary parameters. I do not understand how these properties are retrieved. The authors refer to other publications where the methods are explained, however, they should provide sufficient information to understand the current paper; one of the references is a book chapter which may not be readily accessible. The authors do provide a summary in section 2.3, bujt the actual retrieval is not explained. In particular, they use variable mixtures of aerosol types and in DISORT to create LUTs, how are these used? What criteria are sued to determine the optimum aerosol model, especially because only one single wavelength is used? Furthermore, part of section 2.3 is on retrieval over land, but the rest of the paper shows and discusses only results over ocean. Why is retrieval over land discussed if not used? Why are the over land results not validated? The reader would be much helped with a statement in the beginning that this paper focuses on results over ocean, so there is no expectation for results over land. Furthermore, the results are presented without much explanation or critical analysis. For instance, the authors show results on the seasonal variation of AOD, wind speed and chlorophyll, but do not attempt to correlate these. Sea spray aerosol production is known to vary with wind speed to the power of approximately 3, so there should be a strong correlation between wind speed and AOD, as shown in recent publications by Mulcahy et al. or Glantz et al. Chlorophyll has been shown to be correlated to the chemical composition, especially there appears to be much more organic material in the sub-micron fraction in the summer (biologically active water with high Chl concentrations). This is also the optically active size range and organics likely change the hygroscopic and optical properties of sea spray aerosol as opposed to sea salt aerosol. There are quite some typos and there are references to Thomas et al 2009a and b, in the references list there are 2 for 2009 without the a or b.*

The section detailing the retrieval method has been significantly lengthened to try and provide a more complete picture of how the retrieval works; further details are given in response to the reviewers detailed comments below. The authors' feel a great deal of detail is not required here however, since this is already available in the literature (in particular Thomas 2009b is freely available online).
Also, it has been made clearer from the beginning of the paper that only AOD over the ocean is considered in this work. The reasons for this are made clear.
Regarding the correlation between maritime AOD and wind speed and chlorophyll concentrations, the authors acknowledge that one would expect a strong correlation between the wind speed and AOD. This is taken as read in this paper, although the relationship between these variables is not as straight forward as the reviewer makes it seem, especially when one considers that much of the ocean is influenced by aerosol transported from continental sources. To provide a valid correlation between wind speed and AOD, and even more so for Chlorophyll and AOD, one must first be certain the the AOD is free from the influence of continental sources. The Chlorophyll and wind speed time series included in this paper are solely to show that there was no strong trend in either variable over the time span of the GRAPE dataset. The authors feel detailed analysis of the correlation between these variables and the AOD lie outside the scope of this paper.
The references Thomas et al 2009a and b are now distinguished as such in the

bibliography.

Detailed comments: (page nrs indicated by last 2 digits, refer to (page,line))

Introduction: satellite remote sensing is presented as the solution to obtain global measurements. Although I agree with that, at the same time should be mentioned that there are many shortcomings and even results from dedicated instruments often do not agree over large regions and there are substantial discrepancies over certain areas. I suggest to add a few sentences on the immaturity of satellites.

The authors acknowledge that remote sensing of aerosol from satellites is an inherently under constrained problem, as is already discussed in the second paragraph of the introduction. An explicit statement that different satellite products often show disagreement has been added to this paragraph.

The authors don't feel that the satellite sensors themselves can be described as immature – it is rather a problem of constraining a very complex atmospheric/surface system using the limited amount of information available from passive remote sensing.

(84, 17)'The Along Track Scanning Radiometer series of instruments1 are: indicate what these series are

The Along Track Scanning Radiometer instruments are a series in their own right, starting with ATSR-1 (launched in 1990), continuing with ATSR-2 (1995 – 2008), AATSR (2002 -) and SLSTR (to be launched on the Sentinel 3 platform in 2013). They are a series in the same way the AVHRR instruments represent a series.

Section 2.1: (84,8) The instrument ceased operation in 2008, which instrument? ATSR- 1 or 2?

This has been clarified in the text.

Section 2.3: (86,15) The aerosol types and geographical distribution used in GRAPE are shown in Fig. 1.: I suggest to add a table with the aerosol microphysical properties to help the reader understand without going back to OPAC

This table has been added.

(86, 18) the size distribution of the aerosol types are perturbed by varying the mixing ratios of the different components which make up each 20 aerosol type from the values prescribed in OPAC: How is the mixing ratio determined? I see only an AOD at 550 nm

Each OPAC aerosol class (as is common for models of the microphysical properties of aerosol) is defined in terms an external mixture of mono-modal components. OPAC (and others) provide a number/mass mixing ratio of each component of the class. In the retrieval these mixing ratios are permitted to move from the prescribed values, which is parametrised as the resulting change in the effective radius of the class as a whole. This has been made clearer in the extended section 2.3 (see next point).

(87, 8), if other wavelengths would be included there would be info on the Ångström coefficient to fit

The algorithm description section of the paper (2.3) has been extended to more explicitly explain the optimal estimation approach, the parameters retrieved by ORAC, the measurements used by the retrieval and how the LUTs are constructed.

The retrieval forward model uses the aerosol microphyisical properties provided by OPAC and the DISORT radiative transfer model to model the TOA reflectance at 0.67, 0.87 and 1.6 microns (channels 2, 3, 4 of ATSR-2). The OE retrieval fits all reflectances simultaneously, using the parameters of AOD, effective radius and the magnitude of the surface reflectance (its spectral variation is fixed at the a priori value). The 0.55 micron, channel 1, of ATSR-2 is not utilised in GRAPE, because it is the most severely effected by the low-data-rate modes used by the instrument to conform to the telemetry limitations of the ERS-2 satellite. The change in aerosol component mixing ratio used by the retrieval to provide an effect radius retrieval could have equally parametrised by changes in the Ångström coefficient. Effective radius was chosen as it is a more microphysically meaningful quantity, and because it is the size parameter used by the cloud algorithm the aerosol retrieval was based on.

(87, 4) As the MODIS BRDF product is only available from 2000 onwards, data for the 5 equivalent date and location from 2002 are used to provide the surface reflectance. This should be better explained. Are 2002 BRDF used for the whole data set?

Yes, 2002 BRDF data are used to provide the a priori land surface reflectance for the entire dataset. This is one of the primary reasons that land AOD has not been validated. It is important to understand the the GRAPE aerosol product is essentially an addition to the cloud properties dataset, which fills in the gaps where there are no cloud properties. This has been clarified in the text.

(87,6) Errors resulting from this approximation are a major limiting factor to the accuracy of the GRAPE aerosol product over land Can this be shown?

The authors feel that it is self evident that, since the clear-sky TOA signal depends strongly on the land surface reflectance and it is obvious that the surface reflectance will vary from year to year in many areas (particularly where land use is subject to change, but also from variation in weather patterns etc), assuming a fixed annual cycle of surface reflectance will result in significant errors in the derived AOD.

(87, 9) the algorithm allows small changes in the overall surface reflectance, How is that done, how big are these changes and what determines them? Why is the spectral shape of the surface fixed? Is this over ocean or land? Why is NDVI used over ocean? Over land the scheme by Birks (2004) is used, please summarize what this scheme implies. Next par: (87, 22) why is it not possible to use the dual view? What is the implication of assuming a Lambertian surface that does not allow to use the dual view? Why could it be used for GlobAEROSOL (provide reference) and not here? Above was said that the single view MODFIS BRDF was used. ORAC has limited sensitivity to effective radius, so how can this be a primary product as stated a few par above (87,8) ?

Again, retrieval of surface reflectance in the GRAPE aerosol product is an example of the flexibility of the OE retrieval framework. Each retrieved parameter (AOD, effective radius and surface reflectance) has an associated a priori value, including an uncertainty on that prior value. The retrieved state is essentially a weighted least squares fit between the satellite measurements and the forward-modelled TOA reflectance, plus the retrieved state and the a priori. The smaller the uncertainty is on the a priori values, the more the retrieval will be constrained to lie close to these values. In the case of AOD, the a priori representative of background conditions, with a very large uncertainty, and the measurements provide the dominant contribution to the solution. For effective radius, the a priori is that provided by the OPAC specification of that particular aerosol class, again with a large uncertainty. In the case of surface reflectance, the a priori is given by the sea surface reflectance model over the ocean and the MODIS BRDF over land and is tightly constrained by a small uncertainty.

The retrieval is set up in this way to allow for small errors in the surface a priori surface reflectance. Due to the limited information content (degrees of freedom) provided by the 3 TOA reflectance measurements used by the retrieval, the tight constraint on surface reflectance is needed to prevent the retrieval becoming numerically unstable and converging to highly non-physical solutions. The spectral shape is fixed at the a priori for the same reason, over both ocean and land.

The authors have expanded this section to make it clearer. However, all of this is detailed

in the Thomas et al. 2009b paper referenced in the manuscript and thus we feel it does not need to be repeated in detail in this paper.
The NDVI is a useful tool for distinguishing cloud from ocean surface.
A brief outline of the Birks cloud flagging scheme over land surfaces has been given.
As mentioned above GRAPE is primarily a cloud properties product. Due to the difficulties of collocating inhomogeneous 3D structures like clouds from two different viewing angles, the cloud retrieval only uses the nadir view of ATSR-2. Since the aerosol retrieval was run within the cloud retrieval framework, it too only made use of the nadir view. In addition, the surface treatment used in GRAPE essentially assumed the surface acted as a Lambertian reflector – this would imply an identical surface reflectance in the two views of ATSR, which is not, in general, the case. It should be noted that future versions of the ORAC algorithm will combine a dual view aerosol retreival with the cloud retrieval, however.
Regarding effective radius, although the algorithm does retrieve it, it is relatively poorly constrained and is quite sensitive to errors in the a priori surface reflectance and assumed aerosol properties (again see Thomas et al. 2009b). In addition, there is very little ground truth data available to validate it against.

Section 3:
Bullits on top of p. 89 should be better explained, it is not at all clear what procedure is followed here. The closest "ground-pixel" of the satellite instrument (i.e. 4 km grid box) is identified. What does that mean? Closest to what? Ground pixel of AERONET? How does that work for a coastal site which is compared to over-ocean retrieval? Certainly there cannopt be an exact collocation

The closest "ground-pixel" refers to the retrieval pixel closest to the AERONET station in question. Only pixels with lie within ~20 km of this pixel are then considered for comparison with the AERONET station. For coastal sites, this means that approximately half of these pixels will lie over the sea, thus allowing us to compare the AERONET AOD with over-ocean retrievals. The authors acknowledge that this could well introduce a bias into the AOD estimates from each measurement, but it is the best available validation for these data. It should also be noted that it is impossible to get an exact collocation between a satellite retrieval and an AERONET measurement: the sampling provided by each system is entirely different. A valid comparison can really only be made by averaging over space (satellite) and time (AERONET) in an attempt to overcome these sampling differences.

2nd bullit: I don't have a clue what is done here: there is an AERONET pixel from bullit 1, at the coast, and there is supposed to be an over sea pixel within 5 pixels? How is that verified? What is flagged? What area? Why is it important that the area is of similar size as Ichoku et al?

The location of each AERONET station is known, as is the location of each retrieval pixel and the location of the coast-line. Hence we can tell which retrieval pixels are both over the ocean and within 5 retrieval pixels (~20 km) of each AERONET site. The reference to the area used by Ichoku is to indicate that this comparison is fairly equivalent to that done in previous work.

3rd bullit: what is a 'typical aerosol transport speed'? Isn't that just the wind speed? 40 km / hour is about 11 m/s which is quite high wind speed, even over most parts of the open ocean.

This is correct. The manuscript has been modified to reflect the fact that, at an average speed of ~20 km/hour (which is more reasonable average wind speed), an AERONET station will sample approximately 20 km of aerosol loading in the wind-direction. This corresponds to approximately ½ the radius sampled by the satellite, but on average only half of the total area sampled will lie over the ocean.

All three bullet points have been reworded and extended to make the methodology more clear.

(89, 17) How were sites determined to be representative? I believe that Kinne is an authority, but what has he said, and why? Are coastal stations indeed representative for open ocean conditions? Below is said that they are not? Please clarify here what you mean. Why was MAN (http://aeronet.gsfc.nasa.gov/new_web/maritime_aerosol_network.html) not used?

A reference has been added to a paper by Kinne, which gives an overview of how sites are selected based on how representative they are of their region.

Coastal stations are unlikely to be representative over open ocean conditions, however we are not comparing open ocean AOD retrievals to them, but adjacent retrievals which are over water.

MAN data is not available for the time period covered by the GRAPE dataset analysed here. Even if it were, the temporal and spatial coverage of the MAN data is, so far, too poor to provide a significant number of direct comparisons with a dataset like GRAPE.

(89,20) Looking at Fig 2, I see at least 10 (out of 22) that are not in North America or Europe. . ..

This is true, but there is nevertheless a strong concentration of AERONET sites in North America and Europe. The manuscript has been modified accordingly.

(89, 23) What is in table 1: quality control criteria? Are data rejected when they do not meet these criteria?

Correct. This has been explicitly stated in the caption to the table

(89, 26) Which are the conditions in which the retrieval works poorly? Please specify

These conditions for which each of the quality control criteria are testing are explained in table 1.

(90, 1) How were standard deviations in either satellite data or AERONET determined?

The standard deviation is of the samples included in the temporal average for AERONET, or the spatial average for the retrieval. This has also been clarified in the manuscript.

(90, 18) assumed spectral shape of the surface reflectance is typically fairly poor Why is spectral shape needed if AOD is only retrieved at 550 nm?

See explanation given above.

(90, 25) However, since the majority of the AERONET comparisons are coastal, many of them will contain some retrievals using the aerosol type assigned to the neighbouring land mass. Here an explanation is needed how aerosol types are assigned, what are the implications?

The type, or class, of aerosol used at each location is determined by the distribution shown in figure 1. Over the oceans the OPAC “maritime clean” class is used everywhere, except in the polar regions. For coastal AERONET sites, however, some pixels may well be assigned the “continental average” or “desert dust” classes used in the neighbouring land mass (the map of aerosol types has a low resolution of 5x5 degrees). Thus, it is likely that in any single comparison to AERONET, retrieval results using two different sets of assumed aerosol properties will be included in the averaged satellite AOD. Although there is in general no reason to expect one class to be more appropriate than another (since, neglecting any localised sources, the aerosol loading at a coastal site will be largely determined by the prevailing wind, which may be on or off shore), this will usually result in an increase in the spread of AOD values used in the satellite average. The paragraph in

question has been lengthened to make this point more clearly.
It is obvious that the use of a fixed distribution of a small number of aerosol types, such as used in GRAPE, is not an ideal approach to capturing the complex variability in the properties of real aerosol. However, the small number of measurements used in the retrieval greatly limits its ability to distinguish different types of aerosol, meaning that some predefined distribution of aerosol type is needed. The Thomas et al. 2009b paper provides an analysis of the errors introduced by the use of inaccurate aerosol properties. These points have also been added to the description of the retrieval algorithm (section 2.3).

Section 4: AVHRR has been validated
(92,1). Below appears that there are substantial differences between AATSR-GRAPE and AVHRR. If both are validated, how can there be such differences? To appreciate these differences, it is necessary that not only the GRAPE validation results are presented but also those for AVHRR. Furthermore, the authors should discuss the accuracy and reproducibility of AVHRR.

It is a awkward fact, as pointed out by the reviewers general comments, that there is generally poor agreement between different satellite aerosol datasets, even when they have been very extensively validated using the same ground truth data. There has been, and continues to be, a great deal of detailed study into why satellite AOD products differ: see, for example, Kokhanovsky et al. 2010, Myhre et al. 2004, Myhre et al. 2005 (both referenced in the paper). The differences between the GRAPE and GACP-AVHRR are no worse than would be expected for two separate aerosol AOD products, especially considering the relative simplicity of the retrieval assumptions used in both products.
A sentence has been added summarising the findings of validation of the GACP product. However, the reviewer seems to be suggesting that detailed results of GACP validation need to be included in this paper: the Authors do not feel this would be appropriate. Such work has already been published in peer reviewed journals (and, one can assume, future validation activities of AVHRR aerosol products will also be published) by the team who produced the products and performed the validation.

Since a comparison is made with AVHRR, wouldn't it be useful to briefly describe the pros and cons of AVHRR retrieval, so the reader can appreciate the differences?

Further details of the GACP algorithm have been added to the description of the product.

Figure (4): comparison is difficult to see especially since the figures are very small I strongly recommend that the authors show the difference between GRAPE and AVHRR rather than separate maps for each instrument. Such difference map will reveal the good and weaker points, before the causes are discussed.

This is partially due to the on-screen formatting of ACPD – the situation will be improved in the final publication, where Fig. 4 will be a full journal page. An additional figure showing the requested difference plots has also been added to the paper.

(94, 4) Why are marine aerosol model used in areas that are clearly influenced by other aerosol types?

See the explanation of how aerosol types are selected earlier in this document.

(94, 1) The par starting with 'The GRAPE product also shows' is very qualitative and speculative. I suggest to add substance.

This section has been reworked to include more quantitative measures of the comparison between the two datasets. However, as is stated in the paper, it is difficult to be more than speculative about the reasons for these differences based solely on monthly mean values from the GACP dataset.

Section 4.2: In the above, uncertainties and biases were discussed for GRAPE. Next the authors use GRAPE products to compare with AVHRR products and infer conclusions on time series and trends, but the uncertainties and biases in AVHRR are not discussed, nor are they accounted for any of the two instruments in the comparison. How is that justified? How representative are the GRAPE and AVHRR data?.

Error statistics are not readily available for the GACP data, so cannot be included in the discussion. All that can be done is to use the statistics of the available data to infer a range of internal variability in the dataset, which has been done to provide confidence intervals when fitting functions to the time series. To provide a fair intercomparison the GACP and matching GRAPE 1x1° monthly product have been treated identically in the analysis presented in sections 4.1 and 4.2.

The weighted averaging used to produce the GRAPE monthly product is necessary to correctly account for the influence of the a priori used in the GRAPE retrieval. The values were used in the analysis presented in both section 4.1 and 4.2. Furthermore, the GACP product isn't produced by an OE retrieval scheme, and thus the reasoning behind the application of the weighted averaging to GRAPE doesn't apply to it (even if the authors had produced the GACP 1x1° themselves).

The authors don't understand what the reviewer means by “how representative” the two products are.

p. 98: Bullits discussing Figure 7: Please help the reader by identifying what you are discussing: which figures are for Northern regions (98, 12) (etc.)? As regards the cycle (98, 15): GRAPE does not use Chl as explained above, how dies that affect the retrieval? Chl has a seasonal cycle. What could cause the difference in phase?

The specific abbreviations (NPO, NAO etc) have been added to the descriptions to clarify them, where appropriate.

It is true that neither GRAPE or GACP use chlorophyll values in their description of the surface reflectance, but variations in chlorophyll will still have a potential impact on the true surface reflectance and aerosol loading (through the release of DMS and organic compounds from the sea surface) over the remote ocean. Hence, a trend in observed chlorophyll could potentially explain an otherwise unexpected trend in observed AOD. The same argument applies to the surface wind strength.

One would not really expect a directly proportional relationship between AOD and chlorophyll concentration, as AOD is influenced by so many other factors, so a difference in the seasonal cycle is likewise unsurprising. The details of this interrelationship lie outside the scope of this paper.

(98, 21) Bullit on tropical oceans: the biomass burning season is clearly identified. Why is it not accounted for in the retrieval?

See discussion on the selection of aerosol type earlier in this document.

Figure 8 shows the seasonal wind speed and Chl, are these in phase with AOD? I'd expect a strong correlation with wind speed since sea spray production goes with the cube of the wind speed as mentioned above.

By observation, it can be seen that seasonal cycle in the wind speed shown in Fig. 8(a) and the AOD in the southern oceans (Fig. 7 SIO, SAO and SPO) are approximately in phase, as would be expected.

(98, 20) What is the solar zenith angle limit?

The limit for GRAPE is 80°. The AVHRR limit is likely to be similar. This information has been added to the text.

(98, 22) Figure 9 shows the latitude limits of both datasets throughout the period of

comparison. In the case of the GRAPE dataset, the actual limit lies equatorward of the defined 60_ limit for November–January in the Northern Hemisphere and for May–July 25 in the Southern Hemisphere, with a maximum discrepancy of 12_. What are the implications of these observations?

Clearly, this will introduce a sampling bias to the regional averages at higher latitudes, as is discussed in the paragraphs in the paper following on from the one in question.

(98, 26) Due to the changes in AVHRR overpass times, the GACP dataset shows a more complicated story, with an increase in both the size of the truncation of the the high latitude limits of the northern and southern regions, and in the number of months a year so effected, as the lifetime of each individual AVHRR instrument progresses Not clear what is meant with this statement or the next: please elaborate

(99,6) Given the band of elevated AOD observed at around 60_ S in the GACP results, a decreased sampling of the lower boundary of the southern ocean regions can explain the negative AOD trend observed in these regions during the period covered by the GRAPE data. This also explains why no such trend is observed in the GRAPE results.

These explanations have been more explicit in the text.

Figure 10 (99, 22): Although not temporally or spatially coincident with the majority of GRAPE time-series, this cycle is in phase with that seen in the GRAPE data, suggesting that the cycle seen by GRAPE is real. According to Figure 2 these stations were used in the validation; how is that possible if they are not spatially or temporally coincident with GRAPE? Are there other stations that are and would illustrate the temporal trends and phase?

The data presented in Fig. 10 are not spatially or temporally collocated with either of the satellite datasets, they are simply the seasonal variation in AOD observed at the AERONET stations in question. If the AERONET data was limited to be coincident with ATSR-2 overpasses, the would not be enough matches to make the seasonal variation apparent. An explanation of this has been added to the manuscript.

(00, 6) Last par of Section 4: what are these valuable insights'? Please list them so the reader gets these insights too.

This sentence has been removed from the paper.

(00, 15) Section 5, 1st par: sampling differences between the two instruments: explain. Could it be in the retrieval?

This point has been expanded upon in the text.

(00, 20) The GRAPE data also show noticeably higher AOD in regions affected by transportation of heavy aerosol loading from the continents. Conversely, the band of elevated AOD seen in the southern oceans in the GACP data is not apparent in the GRAPE data I don't understand this: SAO and SPOI show the best agreement.

The two points are not contradictory: The GACP and GRAPE dataset show the best overall agreement in the southern ocean, but there is a distinct feature apparent in the GACP data which is not present in the GRAPE product.

(00, 27) Showed the origins of the disagreement: was this indeed shown?

The authors feel that in the case of the global trend present in GACP, but not in GRAPE, this analysis has indeed revealed new information. The reviewer makes a good point though: the detailed reasons for the differences are not clear. This sentence has been removed.

(01, 15) Not only wind-generated aerosol determines the AOD, DMS plays a role as well

True. The text has been modified accordingly.

(01, 26) Why are the GLOBaerosol, the SWANSEA and the GRAPE data sets not compared instead of AVHRR?

This analysis was performed before the full GlobAEROSOL and Swansea products were available. A larger intercomparison with these new products is certainly worth considering for future work however.