1 Reply to A. Rozanov

Thank you for all the insightful comments and suggestions. The minor revisions listed below will be fixed in full in the revised manuscript.

2 Minor Revisions

- page 5067, lines 7-1: OMPS has to be mentioned and cited.
- page 5067, lines 19-20: ... or retrieval of aerosol parameters in addition to extinction coefficient - this statement is confusing because if you make an assumption about aerosol parameters (composition and particle size distribution in absolute units) the extinction coefficients are determined and do not need to be assumed additionally. I suggest to delete the end of the statement starting form in addition
- page 5068, line 3: As Odin is scanned... - confusing statement. First, the passive form seems to be unsuitable here. Second, I guess this is not Odin (satellite platform) which scans. It should be rather OS instrument. Otherwise IRI would be scanning as well which is not the case. The entire Odin satellite actually nods, scanning the IRI as well as the OS. However, due to the large vertical range of the IRI, the upper troposphere and stratosphere are visible for nearly the full duration of the scan.
- Page 5070, lines 3-4: Finally, the measurement vector is normalized by one or more high altitude measurements - a couple of lines below you write that you use multiple normalization altitudes. Why do you need to confuse reader by a multiple choice here?
- Page 5070, Eq. (4): Provide typical values of the N-th and (m-N)-th altitudes and of N.
- Page 5071, discussion of Fig. 1: It should be clearly stated that the discussion is valid under assumption that there is no diurnal variation of aerosols.
- Page 5072, line 15: this is similar to Eq. (6) - I guess you mean Eq. (4).
- Page 5073, lines 25-26: ... there is a solar scattering angle, or viewing geometry, dependence on particle size - The statement sounds quite messy to me. Please rewrite.
- Page 5076, line 2: How the matrix K is defined? I suppose this is a matrix containing all $K_k(r)$ but this is not explicitly stated in the text. Furthermore, the same notation is used at page 5070 for another matrix. This is confusing and needs to be changed.
- Page 5076, after Eq. (10): please provide typical noise values for OSIRIS measurement vector.
- Page 5078, line 1: due to the high altitude calibration - I guess you mean normalization.
- Page 5079, Sec. 4.2: Please provide the formula to calculate the smoothing error.
- Page 5080, Fig. 7: Please provide the percentage error.
- page 5082, Fig. 10: Zero lines are hardly visible.

3 Major Revisions

We have attempted to address the more substantial revisions below, however we feel some of them are outside the scope of this paper. We would like to clarify that this is a technique paper exploring the information contained within the OSIRIS measurements and the implementation of an improved retrieval. Although an error analysis and other satellite comparisons are presented to show reasonable results in typical cases, this paper was not intended to be a comprehensive validation study and error analysis of the new product.
• Eq. (3): At a quick look it is unclear to me how a normalization to model data can increase sensitivity of a measurement. Please justify the statement.

*Particularly at low altitudes, the bulk of the scattered signal is due to Rayleigh scattering. Normalizing the measurement by a Rayleigh signal (note that its a subtraction in log space) largely removes the Rayleigh contribution, creating a measurement vector much more dependent on the aerosol contribution. We will include this in the revised paper.*

• Page 5070, lines 7-8: ... normalization is chosen over an altitude range such that Eq. (3) is at a minimum within a chosen noise margin - the statement is totally unclear. What is a noise margin? How it is defined for such a complicated combination of values. Please provide values. What should be at minimum, is it $y_i$? If you look for a minimum $y_i$ which is still larger than some margin this will give you one boundary of the altitude range, where is the second one? Please rewrite the description making it easier to understand.

*The noise of the measurement vector is straight forward to calculate from the measurement error (Eq. 15), and is typically near 0.01 for normalization altitudes. All values that lie within 0.01 of the minimum measurement vector value are then normalization altitudes. This process is described in more detail in reference Bourassa et al. (2011b) but will be clarified in the revisions as well.*

• Page 5070, Eq. (4): The definition of matrix W is quite difficult to understand. It would be better if you could write a formula to calculate elements of the matrix.

*The weighting matrix helps to account for the coupling of the retrieved value at i, with measurements at multiple altitudes, j. The details of the MART technique are described in the references Bourassa et al. (2007, 2011a) and Degenstein et al. (2004, 2009), and are not important to the technique described here.*

• Page 5071, line 28: This represents the worst case for OSIRIS ... - what is the worst case? Are you talking about the shift in the local time or about the systematic difference in the extinction ratios?

Page 5072, line 1: ... lower altitudes and mid to high latitudes do not show this bias to nearly this degree - please explain why.

*Comparisons of version 5 ascending and descending nodes show the largest differences in retrieved extinction during this period. That is what was meant by “worst case”. This can be well explained by a few factors:*

1. *Due to the OSIRIS orbit scattering angles have the largest separation in the tropics.*
2. Low altitudes have higher contributions of multiple scattering, helping to minimize the effects of incorrect phase functions.
3. *High altitudes and latitudes likely have particle sizes more similar to the assumed particle size of 0.08 µm mode radius and 1.6 mode width as this distribution was based on the mid-latitude Deshler (2003) measurements.*

*Will add these factors to the revised manuscript. We have also updated figures 2 and 9 in the paper to show additional altitudes and latitudes. These are included as figure 1 and 2 below.*

• Page 5073, lines 2-5: please provide the parameters for the used particle size distributions.

Page 5073, line 11: ... the measurement vectors below approximately 800 nm provide almost no discrimination between particle sizes - Looking at Fig. 3 it is impossible to judge if this statement is true. It is obvious that the measurement vector per definition provides no sensitivity around 750 nm. However, below about 650 nm the curves are different. It is just difficult to judge whether the difference is significant or not. A relative difference plot could help.

Page 5073, Fig. 3: The small plot with size distributions is hardly readable.

Page 5073, Fig. 3: It would be interesting to see similar results for other scattering geometries.

*Figure 3 has been updated to more clearly show the size distributions and we will include distribution parameters in the revised manuscript, and are included here in Table 1. The relative differences between particle size cases is now shown as well. The updated figure is included as figure 3 below. We have also attached figure 4 below which shows the results for several geometries and more widely varying particle size distributions. During forward scattering conditions (SSA=60) short wavelengths show enough variation across particle size to distinguish between small particles (mode radii of 0.04 to 0.06 µm) but still show*
very little sensitivity to larger particles (mode radii $> 0.08 \mu m$). Geometries with larger multiple scattering contributions (larger scattering angles) show less distinction between particle sizes, even for small particles. This confirms the results from the figure shown in the paper, where for a robust retrieval capable of distinguishing large and small particles under a variety of geometries longer wavelengths are needed.

Table 1: Particle size distributions at 22.5 km used to create figure 3

<table>
<thead>
<tr>
<th></th>
<th>Fine Mode</th>
<th>Representative</th>
<th>Bimodal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Density (cm$^{-3}$)</td>
<td>4.0</td>
<td>3.0</td>
<td>3.2</td>
</tr>
<tr>
<td>Mode Radius ($\mu m$)</td>
<td>0.080</td>
<td>0.077</td>
<td>0.080</td>
</tr>
<tr>
<td>Mode Width</td>
<td>1.6</td>
<td>1.75</td>
<td>1.6</td>
</tr>
</tbody>
</table>

- Pages 5074, Eq. (8): This equation describes only the scattering term and thus seems to be incomplete. I would also expect to see a multiplicative extinction term describing the attenuation of the light coming from the Sun and traveling towards observer after the scattering event. This term should also contain the aerosol extinction coefficient. This makes however measurement kernels, K(kr) non-linear already for the single scattering (multiplication of two terms containing sums of contributions from different particles). Thus, I doubt, if the concept of measurement kernels is suitable for limb observations. Indirectly you recognize this fact by throwing K away in Sec. 4 and using the Jacobian concept instead. So why do not you skip the K-concept from the beginning and redo Sec. 3 based on Jacobians? This will at least eliminate any concerns about the validity of the used concept.

- Pages 5074-5075: It is unclear how the vertical inhomogeneity of the aerosol number density and size distribution is handled within the measurement kernel concept. Bearing in mind the complexity of the limb scattering geometry, it looks suspicious that the information content can be analyzed for each tangent height independently. There should be some assumptions behind the method which are not sufficiently discussed in the paper. If you decide to keep the concept please include more detailed discussion on the validity and possible limitations of the method.

Eq. (8) includes only scattering from the tangent point, and assumes both extinction and multiple scatter contributions are negligible. For an accurate retrieval these assumptions are certainly too crude, however, to help understand the information contained in a limb scatter signal they are a good place to start. This can be seen in Figure 4 in the paper where these very simple single scatter kernels are compared to the multiple scatter kernels which are calculated using SASKTRAN and incorporate all these effects. Despite the assumptions, agreement between single and multiple scatter kernels is quite good. These sensitivities are important in understanding the limits of limb measurements as well as the information contained therein. While the kernels are not used directly in the final retrieval algorithm, they are integral in its development, explaining the need for longer wavelengths as well as why very little information is left in the measurements at a second scattering geometry. It is also worth noting here that kernels used in the Sec. 3 information analysis are Jacobians, only with respect to particles of a given size. In fact, the Jacobians used in Sec 4 can be thought of as a subset of these kernels, where K has been convolved with a lognormal distribution.

Because Eq. (8) assumes all scattering occurs near the tangent point no vertical information is contained in the single scatter kernel. This is clearly not physically true, and our calculations of the multiple scatter kernel does include this altitude coupling. While this makes the kernel dependent on the vertical profile chosen for the analysis, the effect is minimal due to the relatively long path length through the tangent shell. This is also evident in the agreement between the single and multiple scatter kernels. A more thorough explanation of this will be included in the paper.

- Page 5076, Eq. (10): the much-greater-than sign $\gg$ used in Eq. (10) is not conform with the conclusions made below. This sign assumes usually that the left hand side of the equation is at least an order of magnitude larger than the right hand side. This means, however, that both cases result in one piece of information. Please reconsider the discussion or replace the sign by greater-than sign $>$.  

3
This will be clarified in the revised version.

- Sec. 3: How do the conclusions of this section change if you (i) add 470 nm wavelength, (ii) change the assumed mode width of the particle size distribution?

  (i) The implications of a short wavelength normalization were not considered in the information content analysis. This normalization causes the measurements vectors to be quite non-linear as a function of mode radius and mode width, creating substantial difficulties when searching the solution space. Sensitivity to smaller particles is also reduced. For these reasons only single wavelength measurement vectors were used in the analysis.

  (ii) Different mode widths (3 particle size distributions) were used in the wavelength analysis (Fig. 3). However, for the kernel analysis it is not necessary to assume any mode width, as particles at each monodisperse radii are considered individually.

- Page 5077, Eq. (12): It looks like you solve this equation for each altitude independently. Is it really the case? If yes it seems to be a major drawback with respect to Version 5 retrieval because the layer interaction is completely ignored. Is the MART technique still applied? Please provide a proper description of the retrieval method with respect to the vertical distribution of the parameters.

  Page 5077, lines 9-10: the method to calculate Jacobians is unclear from this description. Please add more details. What is actually the forward difference method?

  The equation is solved independently at each altitude to calculate the updated aerosol parameters, ignoring altitude coupling. This will cause a slight over or under estimation of the true Jacobian, increasing the number of iterations required for convergence. The reason for doing this is the computational cost of computing the full Jacobian at each altitude. For a scan with 10 tangent heights the full Jacobian will take 10 times longer to compute than one without altitude coupling considered. As the Jacobian computation accounts for the bulk of the processing time, it is much more efficient to spend a small amount of time on one or two extra iterations, than compute the full Jacobian.

  To improve convergence time the number density is initialized with the version 5 retrieval, which uses the MART technique. However, after this initialization the MART technique is not used further. The paper will be updated with a more descriptive explanation of the retrieval and forward difference method used to compute the Jacobians.

- I doubt if the norm of Jacobian is a suitable convergence criterion. A small norm tells you that you cannot move far away from your current guess but it does not mean the fit and solution are good. If no additional checks are done it is rather a criterion to throw the obtained result away.

  It was erroneously reported that the jacobian norm is used as a convergence criterion. In fact, only the mean squared residual and iteration limit are checked. If either is exceeded the retrieval is stopped, this will be updated in the revisions. In the analysis only data which has a mean squared residual < 4e-4 was used, as this is approximately the average noise margin of the measurements.

- Page 5078, Eq. (18): It is unclear why you are writing this equation because the matrix is not used further in the course of the paper. Furthermore, it is unclear how you translate the measurement error into the error of the retrieved quantities (statement at the end of Sec. 4.1). I cannot accept the reference to Rodgers as an explanation of the method because it is not clear which formula you use and if it is suitable to quantify the errors of your method. So please provide a description how you estimate the errors in the retrieved quantities mentioned at the end of Sec. 4.1.

  The error analysis presented here follows Sec. 3.2 in Rodgers (2000). First, gain matrices \( G = \frac{\delta x}{\delta y} \) are computed numerically by perturbing the measurement vector at a particular altitude and simulating a retrieval. Next, the error covariance matrix (Eq. 18), is used to compute the error in the retrieved quantities, \( S \), using \( S = GS_sG^T \). This will be explained further in the revised manuscript.

- Page 5079, Sec. 4.2: I guess the formula to calculate the averaging kernels is given by \( A = \frac{\delta x_{true}}{\delta x_{true}} \). This needs to be clearly stated in the paper. The averaging kernels depend possibly on the assumed typical values for the extinction coefficient and the mode radius (and may be also on the distribution width). This must be checked and the results need to be reported (at least in a form of a short discussion).
Yes, that was the formula used, and will be included. The averaging kernel will be dependent on particle size (and extinction) as well as scattering geometry. This will be discussed, however, the technique is applicable to any case, and it was not the intent of this paper to present results for all OSIRIS situations - only a typical one.

- Page 5079, Sec. 4.2: In many publications which analyze averaging kernels also such retrieval characteristics as measurement content and vertical resolution are provided. I would suggest to include these characteristics also in your paper.

The vertical resolution as computed from the full width half maximum of the kernels is now included in the figure.

- Page 5079, Sec. 4.2: The investigations (and corresponding plots) need to be done for more than one observation geometry. The effect of underlying clouds (i.e., elevated altitude of the reflecting boundary) needs to be investigated.

These are important studies that need to be performed for a full error analysis of the version 6 product. However, the technique outlined here is sufficient to answer these questions, and we think full results (which will require substantial computational resources) are outside the scope of this technique paper.

- Page 5080, Sec. 4.3: The errors shown in Fig. 6 disagree with the results presented in Secs. 4.2 and 4.3. For example, the albedo error is much lower and the statement In the retrieved quantities this translates to an error of 10% in the retrieved quantities near the peak of the aerosol layer. is not confirmed by the right panel of the plot. Please comment on these issues and provide plots for a couple of other observation geometries.

Page 5080, Fig. 7: Please provide similar plot for different particle radii.

The albedo error discussed in Sec. 4.3 is for a geometry with a substantial upwelling contribution (at least for OSIRIS geometries) and is meant as a near worst-case analysis. This is not the same geometry as shown in Fig. 6, which is a forward scatter geometry, and therefore less effected by albedo errors. The errors in the retrievals reach 11.4% near the peak of the distribution for mode radius and 10.1% for extinction. This will be more clearly articulated. In regards to including additional geometries, albedos and particle sizes the purpose of this paper is to showcase a technique for particle size retrieval and corresponding error analysis. We believe that we have shown the applicability of both, and multiple case studies are outside the scope of this paper.

- Sec. 4: An investigation of possible errors due to the assumed mode width of the particle distribution is clearly missing.

This is addressed in Sec. 4.5 and Fig. 8, where retrievals with incorrectly assumed size parameters (both mode width and a second course mode of particles) are performed for a wide variety of geometries.

- Page 5081, Sec. 4.5: It would be nice to see a couple of plots showing the initial guess, true, and retrieved extinction profiles for several typical observation geometries. The provided information does not allow to make any conclusion on how different these profiles are.

The mode radius is initially always assumed to be 0.08µm with a mode width of 1.6, with typical retrieved mode radii ranging from 0.04 to 0.15µm depending on altitude and latitude. Because the number density is initialized with the version 5 retrieval the extinction typically varies less, with variation between version 5 and version 6 visible in Figure 1 and 2 below.

- Page 5082, Fig. 9: Looking closely to the plot one sees that a significant improvement is achieved until the first half of 2004 while it is much less pronounced thereafter. Moreover the data points in the middle of 2007 are completely missing in Version 6. Please comment on this issue.

The improvement can be more clearly seen in the updated figure (shown here as Figure 2, which we will be included in the revisions. There was an instrument calibration issue with the infrared imager in 2007 which caused the loss of data and is why version 5 does not have this gap.
The polar orbit of SAGE III does not allow for tropical measurements, and the latitudes measured are well defined as a function of time - see Figure 5. Because of this comparisons as a function of latitude which include sufficient measurements for statistical analysis are difficult. The attached figure will be included to help illustrate this point.

We have attached Figure 6 here which shows a comparison of SAGE II and OSIRIS Ångström coefficients as a function of latitude and time at 22.5 km and will include this figure in the revised paper.

The Ångström coefficient as measured by SAGE II is systematically smaller than that from OSIRIS. Between 22 and 27 km OSIRIS measured an Ångström coefficient of approximately 2.8 to 3, while SAGE II produced an Ångström coefficient much closer to 2. The colour scales are different on the two figures which may be cause for confusion. This will be noted explicitly in the text. The attached figure 7, which we will include in the revised manuscript, shows modelled Angstrom coefficients for different particle size distributions assuming a mode width of 1.6. Typically, the Angstrom coefficients of the two wavelength ranges vary by between 0.5 and 0.75. This analysis will be dependent on the particular size distribution and parameters chosen, however values in this range are expected to be typical.

If the Ångström coefficients are that sensitive to the spectral range why do not you just compare the effective mode radii derived from both satellites? These should be independent of the wavelength range.

The effective radius measurement from SAGE II could be compared to an effective radius from OSIRIS, however this is unlikely to be a good comparison for two reasons:

1. SAGE II measured Ångström coefficients directly, while only producing effective radii as a higher level product. This adds another layer of error to the comparison, particularly due to the underestimation of particles below 100 nm (Wurl, 2010).
2. OSIRIS mode radius measurements are more prone to error than the Ångström coefficients, as shown in Sec. 4.5.

4 References

Figure 1: Comparison of the ascending and descending node OSIRIS version 5 extinction measurements for several altitude and latitude bands. The left column shows tropical measurements between 20N and 20S while the right column shows mid latitude measurements between 40 and 60N. The top row shows the solar scattering angle of the measurements. In the tropics measurements show a systematic difference depending on satellite track which correlates well to the solar scattering angle. In the mid latitudes much smaller differences are present.
Figure 2: Same as Figure 1, except using the version 6 data. Much smaller difference are now present between ascending and descending nodes in the tropics.

Figure 3: Panel A shows the 3 size distributions used in the simulations at 22.5 km, corresponding parameters are included in table 2. Panel B shows the measurement vectors as a function of wavelength for the three cases. Panel C shows the relative difference in the measurement vectors from the true bimodal case. The shaded grey area is the relative error in the bimodal measurement vector due to a 1% error in the measured radiance.
Figure 4: Same analysis as Figure 3 - panel (B), for a wider variety of particle sizes and geometries. All distributions tested here have a mode width of 1.6, with mode radii given in the legend, and ranging from 0.04 to 0.14 µm.

Figure 5: Latitude of the coincident SAGE III/OSIRIS scans as a function of time.
Figure 6: Angstrom coefficients as a function of latitude and time at 22.5 km. Top panel shows values retrieved by SAGE II with bottom panel showing OSIRIS results.

Figure 7: Angstrom coefficients as a function of mode radius for the SAGE II (red) and OSIRIS (blue) wavelength ranges given a lognormal distribution with a mode width of 1.6.