AMT-2013-175 Määttä, et al.: OMI aerosol model uncertainty

Anonymous Referee #1

Response to reviewer 1's comments

We like to thank the Referee for extensive and constructive comments and recommendations. We have tried to take them into account when revising the manuscript. Our point-by-point answers are found below.

General comments and recommendation

> *This paper introduces a method to quantify the error in aerosol optical thickness (AOT) retrieval arising from the use of discrete aerosol microphysical models in the retrieval algorithm. The methodology is general, although it is introduced in the specific context of the OMI OMAERO algorithm. The topic is interesting (and important) and relevant to AMT. The specific quantification of uncertainty from aerosol microphysical model is something which, to my knowledge, has not been directly addressed in studies thus far, although many relevant studies from several groups (see specific comments), which are not discussed at all in the present study, have used statistical optimisation techniques to construct an overall error budget for aerosol retrievals. The overall structure and content of the paper are mostly ok. However there are some areas which I think should be extended and/or need clarification (as some things are not entirely clear at the moment). The main point for clarification is when the authors are talking about uncertainty due to only the aerosol microphysical model, and when they are talking about uncertainty due to the forward model as a whole (to include e.g. calibration uncertainty, surface albedo uncertainty, LUT interpolation uncertainty, as well as the aerosol microphysical model effects). The introduction is weak and skips over almost all the existing research in the field of AOT retrieval uncertainties (although this study, as mentioned above, has novel aspects not covered by those previous studies). The analysis also needs a little more substance, and the conclusion could benefit from more discussion of practice use of the results. I therefore favour revision to address the specific comments below. I would be happy to review the revised version, and don't think that these revisions should take too long to implement. I forsee that eventually this will be a strong paper which should be published in AMT, but it is not there yet.*

The Referee raises several important issues. First, we have tried to clarify the different sources of uncertainties we are dealing with and the terminology used. Second, we have strengthened the Introduction and added many of the references the Referee were kind enough to provide us. We are very thankful for these, and feel that this has considerable improved the Introduction. Third, we have added one new case study to add more substance to the analyses.

Specific comments

> *Title: In light of the fact that not just the microphysical model contribution to error in AOT retrieval is discussed, but also the 'model discrepancy error' which, I think, covers other factors too, perhaps a change in title would help make the paper's purpose clearer. Perhaps "Quantification of uncertainty in aerosol optical thickness retrieval arising from aerosol microphysical model and other sources, applied to Ozone Monitoring Instrument (OMI) measurements".*

This is a good point. We find the suggested title more in line with the actual contents or our paper and have changed the title in the new revision.

Introduction: this is fairly light on references in the opening paragraphs (in that there are none). Interestingly, in the list of instruments in the second page of the introduction, the authors list several but then omit two of those which have had statistically rigorous retrieval methodologies applied to them: the ATSRs and SEVIRI. POLDER is mentioned although, again, no references are given. The references below describe AOT retrievals for these sensors in which there is an end-to-end propagation of uncertainties from measurements and forward model through to retrieved quantities. These are pretty relevant to the study at hand so should really be discussed (at least in brief):

Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanré, D., Deuzé, J. L., Ducos, F., Sinyuk, A., and Lopatin, A.: Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observa- tions, Atmos. Meas. Tech., 4, 975-1018, doi:10.5194/amt-4-975-2011, 2011.
Govaerts, Y. M., S. Wagner, A. Lattanzio, and P. Watts (2010) "Joint retrieval of surface reflectance and aerosol optical depth from MSG/SEVIRI observations with an optimal estimation approach: 1. Theory", J. Geophys. Res., 115, D02203, doi:10.1029/2009JD011779
Sayer, A. M., Thomas, G. E., and Grainger, R. G.: A sea surface reflectance model for (A)ATSR, and application to aerosol retrievals, Atmos. Meas. Tech., 3, 813-838, doi:10.5194/amt-3-813-2010, 2010.
Sayer, A. M., G. E. Thomas, R. G. Grainger, E. Carboni, C. Poulsen, R. Siddans, Use of MODIS-derived surface reflectance data in the ORAC-AATSR aerosol retrieval algorithm: Impact of differences between sensor spectral response functions, Remote Sensing of Environment 116 (2012) 177-188, doi: 10.1016/j.rse.2011.02.029
Thomas, G. E., Poulsen, C. A., Sayer, A. M., Marsh, S. H., Dean, S. M., Carboni, E., Siddans, R., Grainger, R. G., and Lawrence, B. N.: The GRAPE aerosol retrieval algorithm, Atmos. Meas. Tech., 2, 679-701, doi:10.5194/amt-2-679-2009, 2009.
Wagner, S. C., Y. M. Govaerts, and A. Lattanzio (2010), Joint retrieval of surface re- flectance and aerosol optical depth from MSG/SEVIRI observations with an optimal es- timation approach: 2. Implementation and evaluation, J. Geophys. Res., 115, D02204, doi:10.1029/2009JD011780.

The above papers mostly discuss Optimal Estimation; the Dubovik study is a similar statistical methodology. The Thomas paper is an overall algorithm description while the two Sayer references go more into the specific details of the error budget calcula- tions and goodness-of-fit checks (similar in concept to what's done in this work). The Govaerts and Wagner studies are another implementation of this same basic idea.

This recent paper gives a particularly inventive application involving not only MODIS but also the GEOS-Chem model:
van Donkelaar, A., R. V. Martin, R. J. D. Spurr, E. Drury, L. A. Remer, R. C. Levy, and J. Wang (2013), Optimal estimation for global ground-level fine particulate matter concentrations, J. Geophys. Res. Atmos., 118, 5621–5636, doi:10.1002/jgrd.50479.
There may be others that I am not so familiar with, but the above are a good start.

We have added most of the suggested references.

In a more general sense, I think the introduction needs a lot more work. The initial paragraphs need some useful references, or if it is felt that the material is too general and well-known to need references, then it is in my view too general to need to write down in such length, so things like e.g. the list of sensors (which has other omissions e.g. AVHRR, SeaWiFS, MERIS) or reasons for being interested in aerosols could be trimmed down or removed entirely. This is a technical paper in a technical journal, anyone reading it is already going to know something about aerosol remote sensing. So I would prefer that the introduction instead be focussed on existing attempts at AOT retrieval uncertainty quantification, whether through the statistical methods discussed in the references above, or the more empirical methods applied to things like MODIS or MISR. That would be more interesting for the reader and more relevant to the topic at hand. Some *recommended references for the empirical (i.e. non-theoretical) AOT validation and error budgets include:*

Kahn, R. A., B. J. Gaitley, M. J. Garay, D. J. Diner, T. F. Eck, A. Smirnov, and B. N. Hol- ben (2010), Multiangle Imaging SpectroRadiometer global aerosol product assessment by comparison with the Aerosol Robotic Network, J. Geophys. Res., 115, D23209, doi:10.1029/2010JD014601.

Levy, R. C., Remer, L. A., Kleidman, R. G., Mattoo, S., Ichoku, C., Kahn, R., and Eck, T. F.: Global evaluation of the Collection 5 MODIS dark-target aerosol products over land, Atmos. Chem. Phys., 10, 10399-10420, doi:10.5194/acp-10-10399-2010, 2010.

Sayer, A. M., N. C. Hsu, C. Bettenhausen, and M.-J. Jeong (2013), Validation and uncertainty estimates for MODIS Collection 6 "Deep Blue" aerosol data, J. Geophys. Res. Atmos., 118, 7864–7872, doi:10.1002/jgrd.50600.

We want to thank the Referee for this list of references on aerosol retrievals with uncertainty analysis. As suggested, we have added the given references together with some discussion about the topic.

In the new revision, we have added missing instruments, together with references to work on aerosol retrieval algorithms.

Page 8511, line 10: this is one example of a statement which needs supporting by references. Arguably none of the sensors were really designed primarily for aerosol remote sensing!

We agree. We have removed this sentence from the revised manuscript as this statement is not relevant to this study.

Page 8511, line 29 and throughout: it would be better to state 'aerosol microphysical model' rather than just 'aerosol model'.

This is a good suggestion that improves the readability of the paper. We have replaced 'aerosol model' with 'aerosol microphysical model' in the revised manuscript.

Page 8515, line 10: this says that sigma is the uncertainty 'given as the standard deviation'. This is unclear and needs to be clarified. Standard deviation of what? No spatial averaging is mentioned. Or do the authors mean to say it is the magnitude of the error, as opposed to its variance (which would be the square of this value)? That is what I would guess from Equation 12 (a lot later in the text), but it should be made clear here. If so, is this value of sigma solely some instrument calibration uncertainty (which is my guess because of later discussion), or does it include e.g. forward model uncertainties too (see previous references)? I assume from the notation that this is uncorrelated spectrally? More and clearer information would be welcome here.

This sentence was poorly written. In the new revision we write: "standard deviation uncertainty in the measured reflectance which is assumed known". In our study, we originally used the measurement uncertainty estimates as calculated by the operational OMAERO algorithm and were provided in intermediate result files of the Level 2 processor. These were based on estimated spectral noise in the Level 1 data. Later we used a simple signal-to-noise (SNR) ratio based estimates for the observed reflectance uncertainty. These latter values are used in our reposted analyses. This SNR based approximation is mostly in agreement with the operational values.

Page 8516, equation 3: This is the equation for top of atmosphere reflectance above a Lambertian surface, yet page 8514 mentions an ocean bidirectional reflectance model, the use of which would not be consistent with this equation. Is a different formulation from Equation 3 used over ocean, or is the bidirectional reflectance converted to a Lambertian albedo (or assumed to be equivalent to one)? Again, more and clearer information would be welcome here.

You are right. The introduction of Eq. 3 is misleading and needs specification. We have experimented our uncertainty quantification scheme over land pixels only, and used Eq. 3 to simulate the TOA reflectance. We have corrected the text to be more accurate with this respect.

> *Page 8517, section 3.1: This first paragraph is a nice introduction to the section. I feel a reference or two for further reading would be helpful to the reader here, though.*

We added a reference to Gelman, et al. (2003).

> *Page 8518, line 1: I think here you should explicitly say "aerosol microphysical model m" rather than just "model m", to make the distinction between the aerosol microphysical model and the overall retrieval forward model clear. Or are you, here, lumping the total forward model error (which also includes e.g. calibration uncertainty, surface albedo uncertainty, LUT interpolation uncertainty) in with aerosol microphysical model effects?*

You are right. As we consider here the aerosol models from LUT, it is more precise to say "aerosol microphysical model". We have done the changes to the revised version of the manuscript.

> *Page 8518, line 8: "This value is sometimes called evidence." Please provide a reference? I'd also appreciate an extra sentence or two about this. It is not immediately clear to me how to calculate this constant for a given situation (perhaps it is already contained within the paper, but if so, it is not explicit enough to be obvious).*

The term evidence refers to the denominator in the Bayes' formula. It is the probability of the observations given the model. The "inverted" probabilities, i.e. the probability of a model given the observations must be calculated using the evidence. When the prior probabilities for individual models are equal we can work directly with the evidence, however, as then the relative evidences are the same as the relative posterior probabilities. The relative evidence between two models is sometimes called the Bayes factor in statistical literature.
In general, the evidence, or the model posterior probability, can be difficult to calculate. In our case, as we have only one unknown parameter, the evidence can be calculated from equation (5) by numerical integration. We have tried to make this clearer in the new revision. We added a reference, also.

> *Page 8518, line 17: see prior comment about what the authors mean by 'standard deviation' here. Also, for the specific example of OMI, the authors should state somewhere what these uncertainties are. It looks like a SNR calculation is assumed on page 8525, is that what's normal for OMI? That should ideally be stated earlier. And what about calibration absolute uncertainty (which is worse than the radiometric noise level for many sensors?).*

As mentioned above, the measurement noise standard deviation $\sigma(\lambda)$ used in the examples is from simple SNR approximation and not directly from the operational algorithm. This is now explicitly stated in the new revision.

> *Page 8520, section 4: again it would be good in the section title and text to be clear about the distinction between aerosol microphysical model uncertainty and total forward model uncertainty, and what is being discussed. From the text in this section itself, if looks as though they are dealing with the total forward model uncertainty (although evaluated on a per-microphysical-model basis), in which case I think this should be made clearer throughout the*

paper. In a more general sense, this analysis of residuals in this section seems to make sense. But the authors should be aware that there are ways of propagating uncertainties from measurements and forward model through to the end, including accounting for correlations between them, and you can then use the residuals as a check of how consistent your solution, forward model and assumptions are with the measurements themselves, making use of prior information if you would like to. See the references I suggested for the introduction. This technique (Optimal Estimation), as applied in most of those references, is very powerful. For example, these other techniques dynamically determine the appropriate weight for each measurement independently (maximising the use of the information content for that specific individual retrieval state), while the method presented here minimises based on (it seems) instrument noise only and then weights solutions after the fact. These two approaches are not mathematically equivalent. It is not clear to me why the authors here choose to use a more empirical technique based on analysis of residuals rather than something like Optimal Estimation? Perhaps because it is easier (which may be a fair enough reason), and maybe you don't need to have an idea about what your uncertainties are before the fact? This should be discussed somewhere in the text.

The first version of the manuscript was not very precise on different sources of uncertainties and the terminology used. Now explicitly consider aerosol microphysical model choice and model discrepancy, for which we include all forward model uncertainties, both systematic and non-systematic. Section 4 is dealing with total forward model uncertainty, which we call model discrepancy following Kennedy(2001). We have changed the title of Section 4 to be: Model discrepancy. We have now defined this term more carefully to include all uncertainties not attributed to the measurement noise. We have also tried to stress this distinction in the revised version of the manuscript.

The case of OMI AOT retrieval is quite a simple inverse problem as it is one dimensional with respect to the unknown parameter. Optimal estimation as a general method is usually expressed in terms of including a priori regularization or smoothness constraints to the solution vector and then solving the Gaussian approximation to the posterior distribution by numerical optimization methods. As such it is equivalent to the Bayesian parameter estimation we are performing. Our additional components are the model selection and statistical modelling of the model discrepancy by Gaussian processes. In addition we do full posterior analysis instead of Gaussian approximations.

One other point to discuss in this section (which the authors briefly mention earlier by acknowledging the limited set of aerosol microphysical models used in OMI processing): strictly, I think that the posterior probability distribution for AOT should be computer over the space of all conceivable possible aerosol microphysical models. Obviously this is not possible in practice, although something numerically similar can probably be obtained (because of degeneracy of some possible models and near-zero probability of other possible models) provided the space of aerosol microphysical/optical properties is explored adequately by the set of OMI models. I understand this paper is framed in the context of the operational OMAERO product, and as such it may not be possible to add more candidate aerosol microphysical models. But this issue should be discussed, and I think it would be useful to add a table or something summarising the key microphysical (e.g. size distribution parameters) and optical (e.g. SSA, asymmetry parameter) properties for each model.

In our study, all the available 50 aerosol models are equally likely as prior assumption, thus in the first step we fit all the models to the observations. After the model selection step only best fitting models are left according to their relative evidence. The number of best models is restricted to be ten for conveniences. The number of remaining best models can be less than 10 if the sum of relative evidences exceeds 80%. These are conventional choices, only. The Bayesian model averaging is done over these best models providing the combined posterior distribution of AOT. As suggested, we have included a new table (Table 1. in the revised version) where the size distributions parameters and wavelength dependent SSA from the used aerosol models are listed.

Section 5: This section could do with a few expansions. The basic idea (show posterior AOT distribution based on model likelihood for different cases) is a good one. What I would really like to see, though, is something we can relate to visually. As OMI is on the Aura platform, I'd like to see MODIS true-colour images for these cases (maybe MODIS AOT too, although that's not essential), to provide a nice visual point of reference. I would also like to see maps of the OMAERO AOT for these cases. This will help put the posterior AOT distributions (which if I understand correctly are for single pixels) in a better context. The authors could even run their methodology through the whole OMI orbit (rather than just one pixel) and so also produce maps of, say, lower and upper confidence intervals of AOT. These suggestions would be easy, and help make the application of the presented methodology directly viewable in an intuitive way (but also retain the individual-pixel posterior AOT distributions shown at present). I strongly suggest the authors do this. As it is, presenting results for a single pixel without any real context makes it difficult to judge how realistic the AOT estimates are (save for the Moscow case where AERONET is present).

We have added a new example case, where we process several pixels and compare the results to both the operational OMAERO product and to AERONET values. However, we were not able to include the MODIS true colour images in this revision.

Comparisons with any available AERONET data for the case studies (currently shown for Moscow only) would also be welcomed. There are AERONET sites around many of the regions discussed; if there were no AERONET measurements on the specific days used in this analysis, why not pick other days instead? There are about 9 years of OMI data to choose from and no obvious reason to pick the specific days which were used in the initial submission.

We have implemented this suggestion by selecting a day when there are AERONET measurements available. We have done this for the new example case discussed in previous comment.

Also, for these sections, why is the evidence from only a limited number of aerosol models shown (e.g. only two for the Saharan case)? Did the others have very small evidence, or were they manually discarded? This should be stated clearly in the text. Figure 5's caption suggests that only those models with a strong contribution were shown, but this should be mentioned in the text.

Yes, the Figures show only models that have highest evidence. In the Sahara case there were two models that had together 100% posterior probability of being the right model. First, all the 50 models are included. Then up to 10 most likely models are selected according to the relative evidence. The legend box in the figures shows the relative evidences (%) (i.e. relative posterior model probability percentage values) for the best models. We have modified the text to be more specific on the actual procedure.

Conclusions: This is ok, although I'd like to see more quantitative discussion about the results. For example, when including the residual-based analysis the posterior AOT distributions get wider, which is what's expected because more error sources are included. Can you make some statements about whether the relative magnitudes of the errors, e.g. in what situations is the aerosol microphysical model is the biggest single contribution to the error or are the other factors always dominant? I'd also like to see more discussion of the practical application of this method. As the authors state, it could be applied in OMI processing. Is there some plan to do this?

We have now added discussion about the relative magnitudes between the different sources of uncertainties. However, we admit that these numbers are quite approximate based on our limited experiments. This is an interesting topic and would be an important subject for further study. Likewise, we have added some discussion on practical implementation of the method.

> *Acknowledgements: The AERONET data use policy requests that AERONET PIs/site managers for the sites used are acknowledged. As the authors have used data from Moscow, they may wish to consider this courtesy.*

The Referee is absolutely right. We will add the courtesy to AERONET to Acknowledgements.

> *Figure 2: I don't like this figure, although appreciate what the authors are trying to show. We don't really learn much other than there are lots of patterns the residuals may take, and that they tend to be correlated spectrally. There are lots of repeated colours. Also, what are the error bars, standard deviation or standard error? I think this information could be better presented in a different way. Maybe you could show residuals for a smaller set of cases, similar to your later case studies. So maybe show some residuals for pixels over the Sahara, over biomass burning regions, and over clean oceans and label them accordingly, rather than this (unclear, in my view) mass of lines.*

As suggested, we have changed the content of Figure 2 such that there are shown residuals from smaller number of pixels representing predominant aerosol types. We don't show error bars, as we do not see them to be relevant for the residuals. We have modified the caption text for Fig.2 to clarify this, as suggested by Anonymous Referee #3, also.

> *Figures 5-8: I think the thick red line (overall posterior AOT) needs to be further distinguished. Maybe make it thicker still, or a dashed line? At the moment it does not really stand out from the others, and it would be clearer if it did.*

We agree. We have changed the line style of averaged posterior curve to be dashed black.

Typographical issues

> *Page 8510, line 4: 'the' focus or 'our' focus.*

Done. ('the' chosen)

> *Page 8510, line 16: 'desert' not 'dessert'.*

Corrected.

> *Page 8512, line 3: account 'for' the aerosol model selection uncertainty.*

Corrected.

> *Page 8517, line 24: I think you can just say "Bayes' formula" rather than "the Bayes' formula".*

You are right. Done.