Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2017-289-RC2, 2017

Interactive comment on "New approach to the retrieval of AOD and its uncertainty from MISR observations over dark water" *by* Marcin L. Witek et al.

Anonymous Referee #1

Received and published: 21 September 2017

This paper describes a change to the MISR aerosol retrieval algorithm. They select an ensemble of aerosol types and, for each, compute the radiances that would be observed at a range of aerosol optical depths (AOD). Previously, ensemble members were evaluated separately so each gave an AOD and cost, which were then filtered and averaged to calculate the final product. This paper proposes minimising a single cost function (being the sum of the individual cost functions) to find the AOD and it's uncertainty. The technique is rationalised based on two months of observations and is shown to produce more believable uncertainties, on average, than the previous algorithm.

I recommend this paper for publication after minor revisions. The technique proposed is definitely a step in the right direction and the paper is superbly drafted. However, the technique and description thereof could be improved by a more statistical approach. The paper justifies itself with qualitative descriptions of global averages and internal metrics rather than any validation activity, which is common but always disappointing. Specific comments on the paper are listed below, with some minor details collected at the end. The notation PxLy refers to line y of page x.

- My experience is in optimisation. One defines a cost function and selects an algorithm to efficiently search the 'surface' of that function for its global minima. The uncertainty is a measure of the 'width' of that minima in multi-dimensional space (i.e. the magnitude by which a variable could be changed without significantly increasing the cost). The cost function is usually the RMS difference between some modelled value and a measurement. If the model is accurate and the measurement suffers only random noise (of known variance), the minimal value of the cost function will sample a χ^2 distribution, from which one can determine the probability that this measurement fit that model.

  To me, this paper essentially proposes that $f(\tau)$ is a probability density function (PDF) for AOD and that it is normally distributed. It follows that the most likely AOD is the τ that maximises f and the uncertainty is the function's width. The proposed ARCI threshold can then be understood as eliminating retrievals that are exceedingly unlikely. Describing the problem with these basic statistical concepts could vastly simplify the paper, avoiding awkward phrasing like P8L5.

- Because this is a fairly straightforward statistical problem, there exists a variety of tools to check that (a) f is in fact a good model of the PDF, (b) f is normally distributed, and (c) the selected aerosol models are an unbiased sampling of the complete state space of real-world aerosols. A brief discussion of some of those points could provide a standardised means to evaluate your assumptions and avoid qualitative judgements, such as the function 'closely resembles a Gaussian' (P7L12).

- Are you tabulating f as a function of linear or log τ? Figure 1 uses both as an x-axis, which is misleading. It should be logarithmic as AOD is log-normally distributed (which is clear from the asymmetry about τ_{max} in Fig.1(3)). If you're using linear space, you will underestimate the uncertainty and overestimate the mean.

- Why is there no validation of the new algorithm? It seems fairly substantial to move from averaging a few aerosol types per pixel to averaging 74. A few comparisons against AERONET or MODIS would be fine for a paper like this. A simple comparison of V22 vs. V23 would be a start, considering you did it for the uncertainty!

- In Sec. 3, you implicitly assume that the choice of aerosol type overwhelms any measurement error. Could Fig. 1 be adapted to show the sensitivity of a χ^2 curve to typical measurement error? I'd expect it to move the curve slightly, but much less than the spread between curves.

P10L24 I'm unhappy with this paragraph.

L27 I think this is trying to distinguish between a validation activity, which you sadly aren't doing, and an uncertainty estimate, which you are. By definition, uncertainty is a parameter describing the range of values that can be reasonably ascribed to the quantity that is being measured. I believe that provides a 'measure of how far the retrieved AOD deviates from the "truth"'. The distinction is that uncertainty is a prediction of that difference while validation is a direct calculation of it.

L30 It's good to be clear that the estimated uncertainty is sensitive to the way you solve the problem. However, you don't tell the user what to do with that information. I think a rational response at the moment is to avoid MISR data as it's more sensitive to your assumptions than the environment. I can think of three approaches to remedy this:

1. Give up and declare that your uncertainty values are uncalibrated, providing a pixel-by-pixel assessment of the relative reliability. (I'd recommend that you normalise the values to clarify that their magnitude is not inherently meaningful.)
2. Show that, despite the algorithm's theoretical sensitivity to your assumptions, the uncertainties you produce are an approximation of the true error. This would be done through a validation activity (e.g. the distribution of $(\tau_{MISR} - \tau_{AERONET})^2/\sigma^2_{MISR}$ is approximately normal).
3. Demonstrate that the sensitivity to your assumptions is small. The precise choice of types is a matter for another paper, but it's important to quantify the uncertainty's sensitivity to it. A straightforward way to do so would be re-running the retrieval with a few types removed at random.

- Sec. 4 argues that this method is good because it excludes high AOD retrievals. Could you provide some evidence that, for the two months of data you've considered, there were no large aerosol events?

P3L17 The spread of the MISR ensemble is providing a quantitative insight into the uncertainty in each retrieval *due to the assumptions made*. While the description of ensemble techniques at L9 is technically correct, ensemble techniques are used to estimate uncertainties that can't be accurately or efficiently calculated by other means. It's exceedingly rare to perturb more than one of the input data, auxiliary parameters, and underlying assumptions. Numerical weather prediction perturbs its input data in order to estimate the sensitivity of a chaotic system. Climate models perturb the auxiliary parameters because they are unknown. MISR perturbs the assumed aerosol type because the radiances available don't fully constrain the problem. MISR doesn't need to perturb the input data as the physics of remote sensing are sufficiently linear that error propagation does a reasonable job of estimating the uncertainty due to measurement error. Hence, I wouldn't agree that extending ensembles to 'all possible sources of error' would be overly useful. Ensemble techniques are used to quantify uncertainties due to poorly understood, poorly constrained, or exceedingly non-linear error sources.

P8L38 Within this paper, the only evidence that the cloud filtering is effective is showing that mean AOD is lower. MISR is on the same platform as a MODIS, so you have the ability to check if your cloud flagging spatially agrees with them. That would be rather more convincing than the distribution of a month's observations presented in Fig. 5.

Fig.3 (b) is rather concerning. Do the peaks in retrieval count correspond to the divisions of your LUT? Also, could 3(a) and (c) be shown as 2-D histograms with the mean overplotted? Your argument would be stronger if the decrease in mean AOD with increasing ARCI is due to a decreased prevalence of large AOD (the cloud-contaminated retrievals) while the variation with χ^2 is more uniform.

P9L2 This paragraph ascribes the variations in Fig. 3 at low χ^2 or ARCI to poor sampling. That implies that there should be retrievals there but you didn't see them. Very low χ^2 implies a very close fit to observations, which is unlikely, and very low ARCI implies a very unlikely fit, which should happen infrequently if the ensemble of aerosol types was well-chosen. Hence, I'd ascribe the sharp variations in Fig. 3 in those regions to scenes that are poorly suited to this retrieval.

P9L19 I wouldn't say that the trend in AOD is statistically robust. I'd say that the shape of 3(c) isn't evident in 3(d), so we don't ascribe the kink in the former to a change in frequency.

Fig.4 This is a superb figure and deserves more attention than Fig. 3. However, the caption is unclear if it is plotting the same data as in Fig. 3.

P10L12 Any idea why cloud contamination is a function of latitude? Does the ARCI threshold need to vary with latitude?

P12L19 You didn't provide a 'strong statistical foundation'. You justified the ARCI threshold by the shape of the distribution of AOD. Statistics would calculate a theoretically sensible value of ARCI based on typical noise and a very large ensemble of aerosol types.

- Finally, I would prefer it if the paper and any data files released clearly describe the retrieved product as 'ensemble mean AOD'. Evaluating a range of aerosol types is an excellent way to sample the unconstrained parts of state space (such as refractive index). Providing an ensemble of results to the user illustrates what the data constrains and what it doesn't. However, a combination of ensemble members doesn't necessarily have a physical meaning. To use an example from a related problem, a thick but high cloud can produce the same TOA thermal radiance as a thin and low one. Giving the user both results shows that both are possible. An ensemble mean, though, gives a medium-thickness layer midway through the atmosphere, which is inconsistent with the data.

A few more minor points:

P4L6 Perhaps 'The previous MISR dark water algorithm' would be a more informative title to someone skimming the paper?

P5L2 reflectance is defines defined as

P5L26 Considering you don't define them, and their precise definition is unimportant to this paper, perhaps remove specific references to the now neglected χ^2, parameters?

P6L34 'turns out to be' is rather colloquial. Perhaps 'and will be shown to produce superior results to the original algorithm'?

P7L4 If these are continuous functions of τ, you are presumably interpolating as the LUT is discrete. What are you interpolating — ρ, χ^2, or f?

P7L7 That distribution has a long tail on it to be called Gaussian. (I know what you're getting at, but there are distributions which are 'Gaussian but with a long tail'.)

Fig.2&5 Why does the scale go to 3? It washes out the global distribution of AOD while emphasising variations in retrievals that are implied to be wrong.

Fig.5 The degree symbol is underlined.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2017-289, 2017.