AEROSAT pixel-level uncertainties working group: 2019 update

Thomas Popp, Andrew M. Sayer, Yves Govaerts, Pekka Kolmonen, Antti Lipponen, Marta Luffarelli, Tero Mielonen, Falguni Patadia, Adam C. Povey, Kerstin Stebel, and Marcin L. Witek

What's new for 2019?

- Completed analysis of current pixel-level uncertainties with 7 participating satellite teams
  - ADV, BAR, CISAR, Dark Target, Deep Blue, MISR, ORAC
  - None are perfect but all have some value
- Submitted paper to AMT (waiting on editor & initial handling)
  - Review of existing uncertainty estimates
  - Framework to evaluate them
  - Results for the above teams
- A concrete, explicit AEROSAT paper – **well done team!**

A review and framework for the evaluation of pixel-level uncertainty estimates in satellite aerosol remote sensing

Andrew M. Sayer[1,2], Yves Govaerts[3], Pekka Kolmonen[4], Antti Lipponen[4], Marta Luffarelli[3], Tero Mielonen[4], Falguni Patadia[1,2], Thomas Popp[5], Adam C. Povey[6], Kerstin Stebel[7], and Marcin L. Witek[8]

[1]GESTAR, Universities Space Research Association, Columbia, MD, USA
[2]NASA Goddard Space Flight Center, Greenbelt, MD, USA
[3]Rayference, Brussels, Belgium
[4]Finnish Meteorological Institute, Atmospheric Research Centre of Eastern Finland, Kuopio, Finland
[5]Deutsches Zentrum für Luft-und Raumfahrt e. V. (DLR), Deutsches Fernerkundungsdatenzentrum (DFD), 82234 Oberpfaffenhofen, Germany
[6]National Centre for Earth Observation, University of Oxford, Oxford, OX1 3PU, UK
[7]NILU - Norwegian Institute for Air Research, NO-2007 Kjeller, Norway
[8]Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Drive, Pasadena, CA 91109, USA

Correspondence: Andrew M. Sayer (andrew.sayer@nasa.gov)

Abstract. Recent years have seen the increasing inclusion of per-retrieval prognostic (predictive) uncertainty estimates within satellite aerosol optical depth (AOD) data sets, providing users with quantitative tools to assist in optimal use of these data. Prognostic estimates contrast with diagnostic (i.e. relative to some external truth) ones, which are typically obtained using sensitivity and/or validation analyses. Up to now, however, the quality of these uncertainty estimates has not been routinely assessed. This study presents a review of existing prognostic and diagnostic approaches for quantifying uncertainty in satellite AOD retrievals, and presents a general framework to evaluate them, based on the expected statistical properties of ensembles of estimated uncertainties and actual retrieval errors. It is hoped that this framework will be adopted as a complement to existing AOD validation exercises; it is not restricted to AOD and can in principle be applied to other quantities for which a reference validation data set is available. This framework is then applied to assess the uncertainties provided by several satellite data sets

Key concepts (1)

- Uncertainty and error are different
  - Uncertainty is an *expression of level of confidence* in the result, of expectation of the error distribution
  - Error is a *realisation* drawn from the uncertainty distribution
  - Analogy with rolling a die, *expectation* is 3.5 (cf. uncertainty) but *result* is 1, 2, 3, 4, 5, or 6 (cf. error)
- When we provide a level 2 uncertainty estimate, suggest we agree to provide a one standard deviation confidence interval around the solution

Key concepts (1)

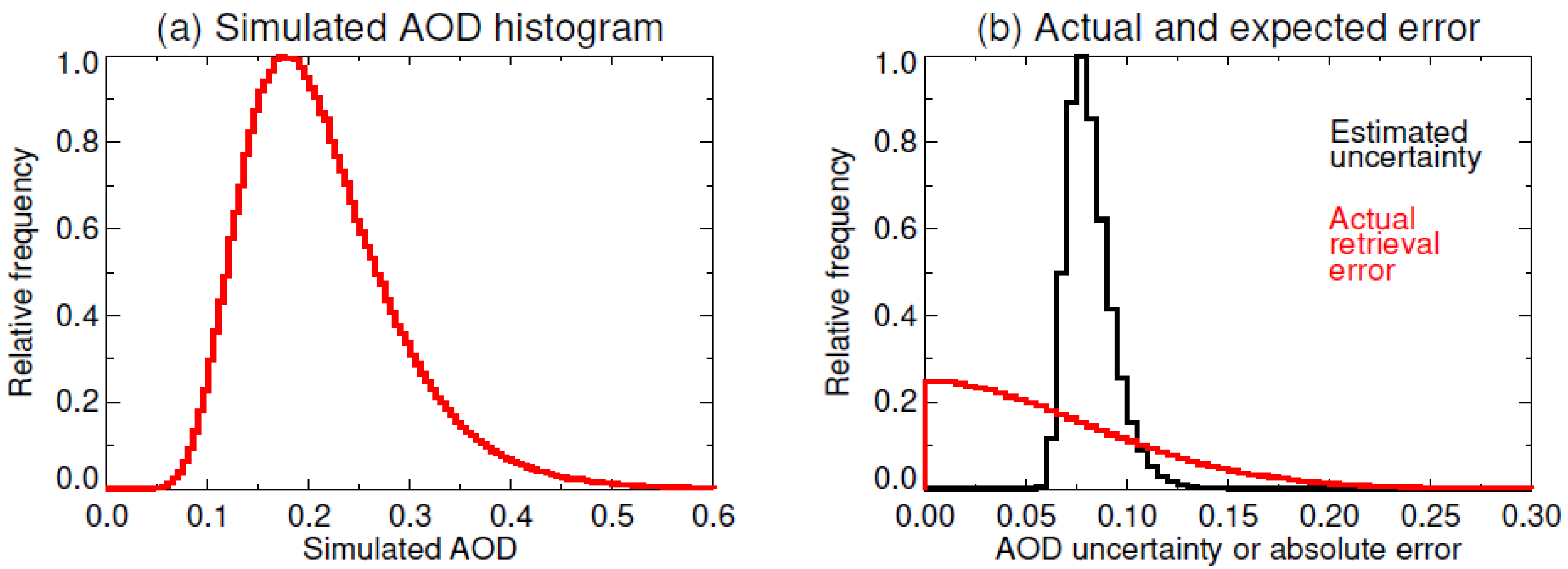


Figure 1. (a) Sample AOD histogram drawn from a lognormal AOD distribution with geometric mean 0.2 and geometric standard deviation 0.35. (b) Distribution of (black) estimated retrieval uncertainties and (red) actual absolute retrieval errors obtained if error characteristics followed the MODIS DT land model, $\epsilon_S = \pm(0.05 + 0.15\tau)$.

Uncertainty and **error** distributions do, and should, have different shapes

Key concepts (2)

- Uncertainty estimates may be *diagnostic* (relative to some known truth) or *prognostic* (predictive)
- Prognostic are, if reliable, more useful for real-time applications (e.g. data assimilation)
  - These can be full formal error propagation (**BAR**, **CISAR**, **ORAC**), empirical (**Deep Blue**), or in between (**ADV**, **MISR**)
- Uncertainties may be evaluated using expected statistical properties of ensembles of normalised errors
  - i.e. error divided by uncertainty
  - Account for uncertainty on the reference in the normalisation

How can we evaluate uncertainty estimates? (1)

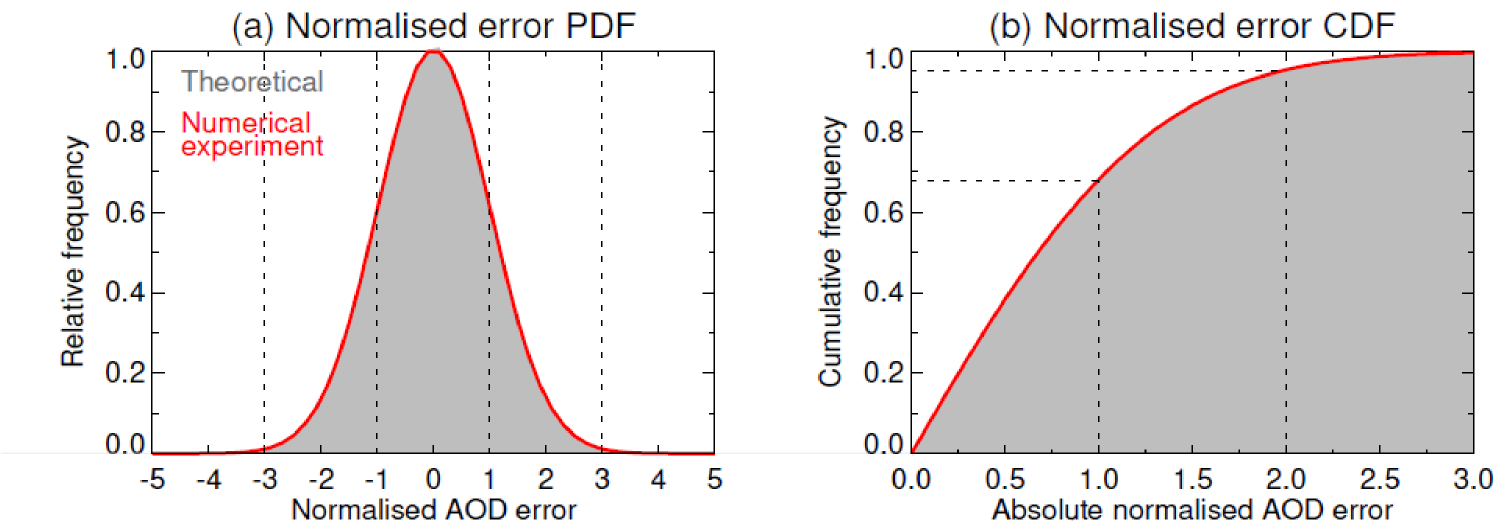


Figure 3. (a) PDF and (b) CDF of normalised error distributions drawn from the numerical simulations in Figure 1; theoretical (grey shading) and simulation (red) results lie on top of one another. Note the CDF is of absolute normalised error. Dashed lines indicate various well-known percentage points of Gaussian distributions.

- Look at PDFs and CDFs of normalised error, using a reference such as AERONET
- Tells you about the bias in the retrieval, and the overall magnitudes of the error, with respect to expected uncertainties

How can we evaluate uncertainty estimates? (2)

- Look at percentiles of binned absolute retrieval error as a function of estimated uncertainty
  - Tests how well-calibrated the uncertainty estimates are, i.e. are larger uncertainties associated with larger errors?
  - A tougher, test of skill than PDFs and CDFs
  - A difficulty for practical application: limited data volume for some satellite data sets

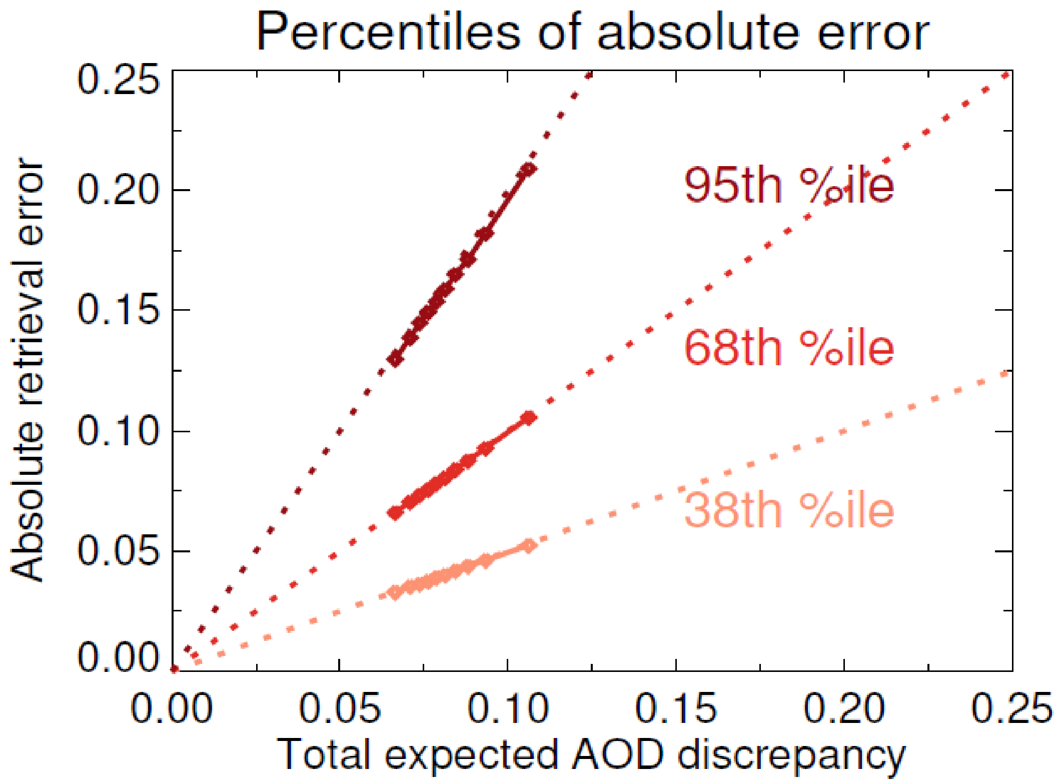
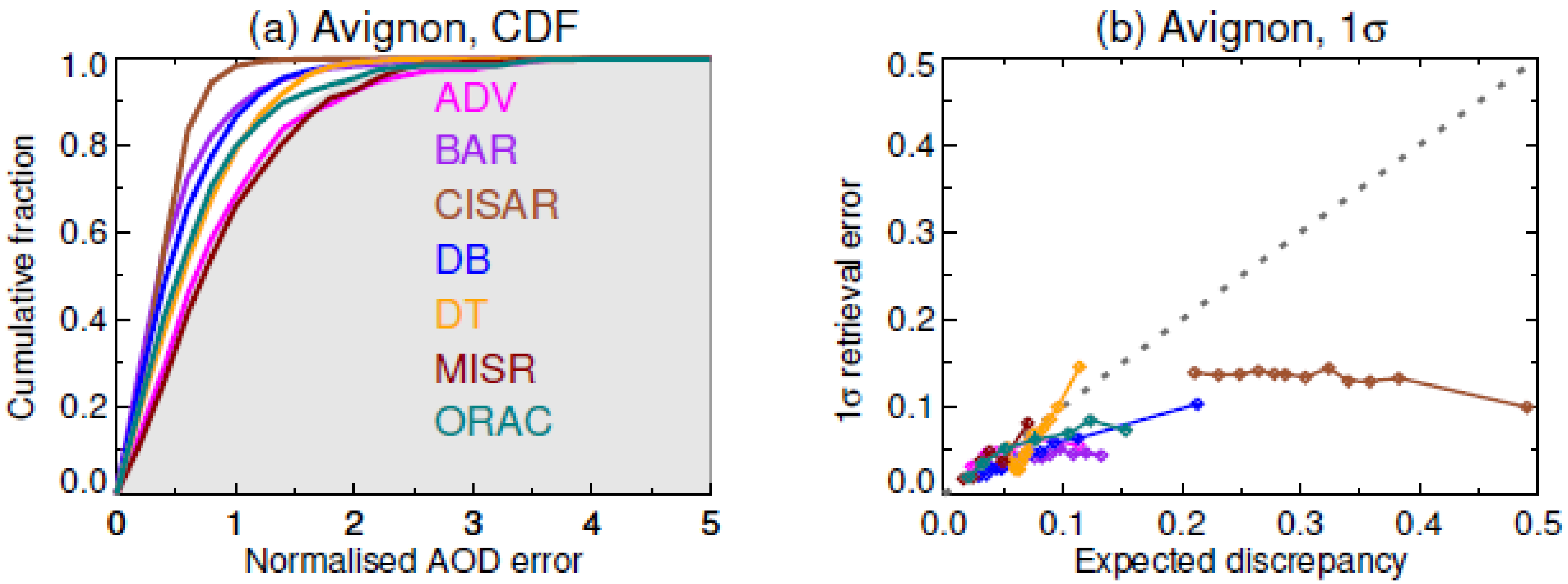


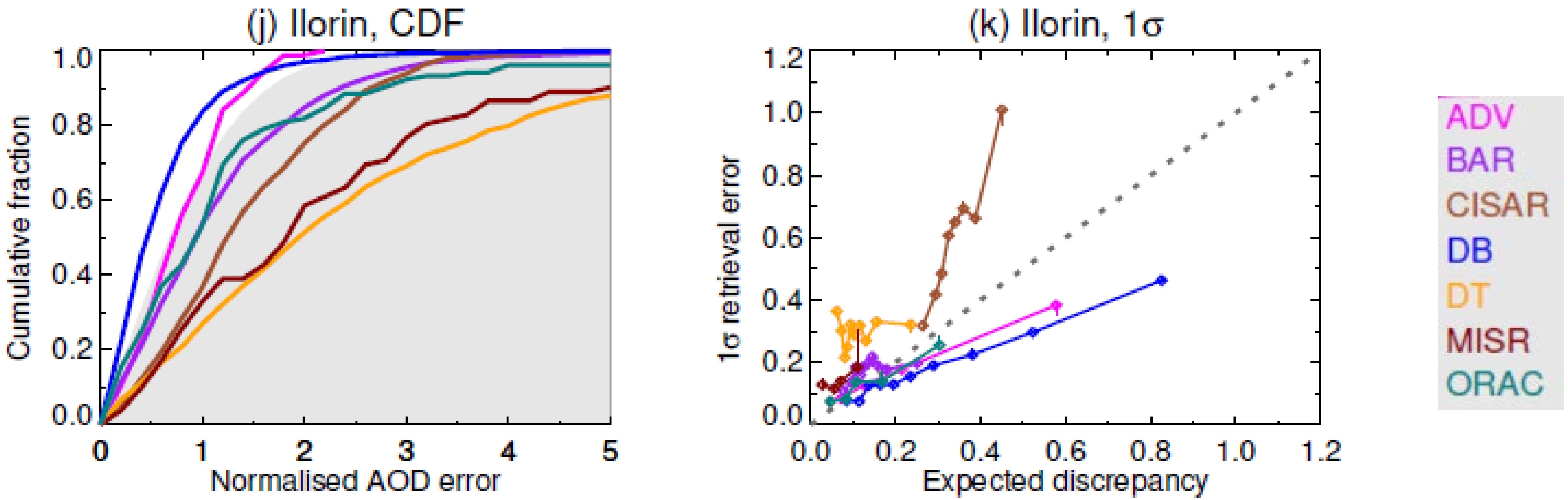
Figure 4. Expected AOD discrepancy against percentiles of absolute AOD retrieval error. Symbols indicate binned results from the numerical simulation; within each bin, paler to darker tones indicate the 38th, 68th, and 95th percentiles (approximate 0.5σ, 1σ, 2σ points) of absolute retrieval error. Dashed lines (0.5:1, 1:1, 2:1 respectively) show theoretical values for the percentiles of the same colour.

Example real results (1)



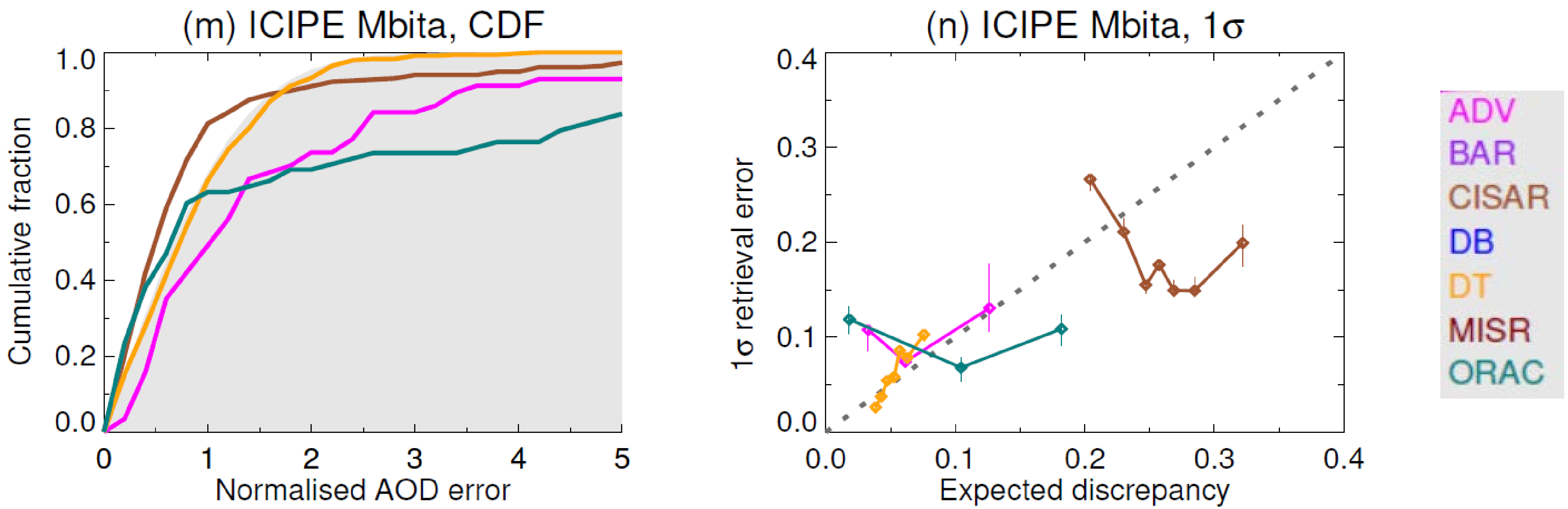
- Left: At sites often thought easy to retrieve (e.g. flat, densely-vegetated land, moderate SSA), algorithms tend to overestimate uncertainty
- Right: Some skill in discriminating between relatively lower and higher-uncertainty regimes

Example real results (2)



- When confronted by a situation outside of family of retrieval assumptions (e.g. SSA at Ilorin is significantly lower than DT, MISR algorithms include), uncertainty estimates tend to fail

Example real results (3)



- In complicated terrain (e.g. potential mixed land/water pixels), some algorithms provide retrievals which have larger errors than expected, while others provide fewer/no retrievals at all
- Is it better to report a possibly-bad retrieval, or provide nothing at all? Likely application-dependent. How should these decisions be communicated in

Possible next steps within AeroSat

- Individual teams to perform larger-scale evaluation/refinement of uncertainty estimates?
  - **MISR** team have recently done this for v23 dark water
  - Incorporate dispersion between AOD retrievals for different optical models into estimates (**Dark Target** ocean, **ORAC**)?
- Can we extend to other aerosol properties?
  - Paper gives some suggestions; quantities like FMF, SSA, AE often bounded and/or have large uncertainties on reference data
- Can we move towards propagation of L2 uncertainties into L3 data?
- Study how often, relative to expectations, algorithms do vs. do not provide retrievals?

Andy can't be here but would love to lead/collaborate on any of the above! Thomas, thanks for presenting! Everyone, thanks for comments on this work over the past several years!