

Review of paper:

**The over-land aerosol optical depth set for data assimilation by filtering, correction and aggregation of MODIS collection 5 optical depth retrievals**

*Hyer et al.*

***General comments***

The paper is long, which is a bit discouraging to start it off. However, the major points are nicely summarized. The error understanding refers to location and times of AERONET samples, which do only represent a sub-sample of MODIS events, as demonstrated for the cloud and possibly also the snow contamination. The regional analysis provides a lot of useful insides in retrieval assumptions and deficiencies.

It is known that MODIS retrievals have problems over South America also in terms of seasonality. The biomass model, as I recall, is tied to African (Savannah) biomass observations and therefore much too absorbing in type. This leads to AOD overestimate early in the biomass season (overestimates). However, large AOD events in the later part of the biomass season over South America (possibly as high AOD are co-located with clouds) are missed so that the then retrieved AOD values are too low. This 'artificially' advances the biomass season forward in time and should be responsible for a large RMS. For the central Africa discussion I miss comparison at Ilorin, which has a strong biomass signal at year's end. I like also in that context that the AERONET sites are introduced with lat lon ... but for some sites (e.g. Taipei, Xianghe) this info is missing and should be added.

The surface albedo bias correction approach (using only AERONET\_MODIS deviations at low AOD values) is interesting but at the same time also a bit confusing. Are we not throwing everything in one basket with a single global parameterization, while there are strong seasonal and regional differences in surface albedo strengths and ratios? Maybe corrections could be stratified by vegetation type and season?

While it is nice to see that overall compliance improved with all these (innovative) efforts it is a bit disappointing that the improvement seems rather marginal. On the other hand the real measure of improvement may be somewhat hidden by the limitation to compare to AERONET samples.

With respect to the figures, I like the pdf-quartile lines in Figure 1a, I would prefer white fonts for the site names on the dark blue background in Figure 5a, the colors ('green'?) in 5c are difficult to see, for curiosity I also would look at slope/correlations sub-samples with  $\tau_m$  (not  $\tau_a$ )  $> 1.4$ . I also wonder id we could increase data-volume/coverage with MODIS deep blue data?

Overall this is a great paper. Congratulations!