Atmos. Meas. Tech. Discuss.,
doi:10.5194/amt-2018-44-SC1, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Atmospheric Measurement Techniques Discussions

Open Access

Interactive comment on "Validation of MODIS 3 km Land Aerosol Optical Depth from NASA's EOS Terra and Aqua Missions" *by* Pawan Gupta et al.

A. M. Sayer

andrew.sayer@nasa.gov

Received and published: 14 March 2018

This Short Comment concerns the linear least-squares regression results presented in the manuscript (in e.g. Table 1). I am posting it here after discussion with the authors in person.

While it is a commonly-used technique, unfortunately AOD data of this type are generally not suitable for the use of ordinary least squares linear regression. The technique requires certain assumptions about the nature of the data to be able to provide quantitatively meaningful regression characteristics (and uncertainties on those characteristics), and these assumptions are all questionable or violated in the case of remotely-sensed AOD data of this type. For example, assumptions of linearity, independence

Printer-friendly version

Discussion paper

of data points, existence of a single population, Gaussian behaviour of residuals, and scale-independence of AOD uncertainties. The result is that the output numbers are not meaningful in the sense that we want to use them. It is not a matter of the results being noisy; they can be systematically biased or in some cases meaningless.

I acknowledge that it is a commonly-used technique but that should not in my view be a valid justification for doing something which is statistically inappropriate in a scientific journal. It is best for us to stop doing it and in this way hopefully spread good practice more broadly through the community.

The reason least-squares linear regression is a popular choice is it gives us two parameters (intercept and slope) with which we can say something about what biases/offsets are in the limiting cases of low-AOD and high-AOD regimes. The question then is what is the best way to convey this type of information in a more statistically-appropriate way?

Fortunately the authors have largely already done so. Since we typically frame our retrieval performance in terms of fraction within expected error (EE), the authors' inclusion of summaries of what proportion of matchups are below, within, and above the EE is one welcome step. Another is with the binned type of plots seen within e.g. Figure 6 (which incidentally already shows that the relationships are overall not linear). The values of the offset for the low-AOD bins provide an indication of typical biases in low-AOD conditions. And the relative magnitudes of the offset for the high-AOD bins provide an indication of typical biases in high-AOD conditions. Or if there is no apparent AOD-dependence then you can just state that the offset appears invariant with AOD. I suggest that the authors remove least-squares slope and intercepts results from the paper. For the same reason, ideally Pearson's linear correlation coefficient could also be replaced with Spearman's rank correlation coefficient. If the authors wish to include replacement information instead of slope/intercept to summarise the global statistics, I suggest adding something like the magnitude and sign of absolute bias as seen in the low-AOD bins, and the relative magnitude of the bias from the high-AOD bins.

CC BY

For example, eyeballing from the bottom-left panel of Figure 6 (Terra, defined relative to MODIS AOD), when MODIS retrieves AOD in the range -0.05 to 0 it looks like the typical offset is about -0.05. When MODIS retrieves AOD above about 0.4, it looks like the bin mean/median bias are positive and about 20%. So in this case you might say that the typical biases are around -0.05 in the cleanest conditions and +20% in high-AOD conditions. Or if you take the top-left panel (Terra, defined relative to AERONET AOD), it looks like the bias it looks like the typical bias is around 0.05-0.1 regardless of AOD. In my view those numbers are more appropriate and more useful statistics to report than the regression slope/intercept.

Interactive comment on Atmos. Meas. Tech. Discuss., doi:10.5194/amt-2018-44, 2018.

Printer-friendly version

Discussion paper