Review of Toledano et al., "Assessment of Sun photometer Langley calibration at the high-elevation sites Mauna Loa and Izana"

General comments

In general, analyzing long and high quality time series of different instruments operated independently in two different networks is valuable work for the atmospheric physics community and appropriate for publication in ACP.

The manuscript describes the environmental variables at Mauna Loa and Izana, compiles climatologies of aerosol optical depth and finally aims at an analysis of the Langley calibration uncertainty for these two sites.

So the manuscript covers a wide scope, however, the scientific impact is weakened by a lack of a rigorous, in-depth analysis. In particular, the statistical (uncertainty) analysis includes several issues. A general indicator for this weakness is that uncertainty and accuracy are often used synonymously, systematic and statistical errors are not treated separately. Without any additional and more detailed discussion, accuracy should be replaced by uncertainty throughout the manuscript.

Specific comments

Section 2.1

P3, L4. "…because the AOD is very low and stable". In fact, I believe, the AOD variability is the actual criterion rather than just low AOD (although typically, both are correlated). This slight misconception appears again later in the manuscript.

P4, L3. Just as a question, I wonder why the data in Table 1 are not displayed as e.g. a bar chart? This would probably even save space and convey the information much easier. Then again, I would argue that cloudy periods are mutually exclusive from AOD measurement periods, so cloud statistics do not add any information for the conclusions here, if later, statistics on Langley days are shown anyways. One interesting insight from cloud information could be the probability of suitable Langley conditions in cloud free conditions.

Section 2.2

P5, L10. Table 2. Again, just a suggestion, but I believe that visual timelines of the instrument deployment (e.g. in the style of a Gantt chart) would be a lot more efficient than just printing numbers in Table 2. Note that "table" should be capitalized when followed by a number.

Section 3

P5, L27pp. If a quantitative comparison of time series is the goal here then the same period should be compared (rather than 1994-2016 versus 2000-2014), otherwise the discussion about other causes for differences is problematic. Regarding the cloud screening methods, as far I am aware of, at least part of the cloud screen for the PFR is based on Smirnov's method. What systematic difference can be expected from the differences of the methods, i.e. is Wehrli's method more stringent and therefore filters more data points (possibly with a bias of higher AOD)?

P6, L21. In fact, it looks more like a bimodal distribution, rather than log-normal.

P7, L3. Why is the AOD in Fig. 2 not shown for 500 nm, as in Fig. 1? Also, what is the reason of using a log scale here and not in Fig. 1? In addition, I would like to suggest that for Fig. 2 histograms would be better suited to reveal the distributions.

Section 4.1

P7, L28. Surely the criteria also affect the number of suitable Langley plots and hence are relevant in the "climatological sense".

Section 4.2

This complete section should be improved by reducing confusing and irrelevant sentences and sharpening the statistical argumentation.

P8, L17. Of course there is no physical measurement without uncertainty. Is the "noise caused by changes in atmospheric transmittance having a hyperbolic (…) dependence" mainly due to residual AOD variations, which affect the slope and/or y-intersect of the Langley plots?

P8, L30. This a confusing paragraph. The sentence "Should the instrument degradation…" can be safely omitted. What is the significance of the sentence "…instrumental issues can be discarded…"? In fact, the linear trend is small (but detectable) and has been correctly taken into account.

P9, L3, Fig. 5. For the y-label, change "density" to "N" and also a heading would help like in Fig. 2, indicating site and instrument. Also, why has the analysis been done for Mauna Loa for 14 years and 3 instruments, while for Izana, only for 4 years and 1 instrument?

P9, L14. The concept of "adding statistical uncertainty" is statistically confusing and the representation in Fig. 6 is suboptimal in many ways.

First, to avoid this confusion, I believe, simply the absolute uncertainty should be considered and plotted here. Also, how does the uncertainty of a one day Langley plot (as shown in Fig. 5) increase to "1% in total". Please clarify.

Second, the statistical uncertainty is generally expected to decrease with square root N, the number measurements, in this case number of days. So the data would ideally be plotted in log-log scale to be able to compare it to a linear slope of -0.5. A deviation from that slope indicates additional error sources (short term drift of the instrument or changes in the signal).

Third, the region between 1 and 10 days seems important, so more data points would be beneficial.

P9, L24. Please explain why suitable days get reduced.

P9,L28. Please clarify the "strong requirement" and include the variability of the AOD, rather than just low AOD.

P9, L35. Again, it is the lower variability of AOD and the wavelength dependence is caused by the Angstrom exponent >0.

Section 4.3

For a “deep assessment” a lot more factors should at least be mentioned. E.g. gas absorption of ozone at 500nm, how is the ozone considered, climatological values?

Or, e.g. what is the effect of different definitions of air mass? As mentioned, it becomes important for large air masses.

P10, L11, Fig. 7. 401 nm is not relevant here, so it should be omitted for clarity of the figure.

Considering the standard deviation for MLO in Fig. 5 of 0.3% it is not surprising that variations at the 0.4 % level are not significant and that there are “no correlations”. Plotting error bars or bands for the Cimel AOD in Fig. 7, may visually reduce the expectation to detect correlations.

Also, why not use a 2-day moving average? Maybe the dip around the 31.10. would actually be significant in the 500 nm Cimel calibrations.

P10, L17. “…averaging over several weeks”. From Fig. 6, it looks like averaging more than 10 days does not significantly reduce the uncertainty.

P10, L18. Could the authors please explain the physical reason why turbulence at 12 km altitude and variations in the refractive index should have an effect on the AOD? Surely this would affect the imaging of stars, but does the blurring effect cause direct solar radiation to be scattered outside the FOV of the sun photometers?

P10, L27. Would a Brewer really be better suited for this study? Is the sensitivity of the instrument an issue here, or the stability?

Section 5

P11, L23. At least in the conclusion, the statistical uncertainty should be clearly specified as 1-sigma standard deviation for a one day Langley plot. From Fig. 5, this was estimated to 0.3% (Mauna Loa) and 0.5% (Izana) . “…a single Langley plot will be typically within 1% of the mean”. What exactly does the 1% signify? 95% confidence interval? Why should the averaging be replaced by the temporal linear fit? Fitting a straight line would be a generalized method, including averaging as a special case with a line with slope zero.

P11, L32. The discussion about the subtropical jet was not really conclusive.

Technical corrections

There are different rules about capitalization, but I think in the context of e.g. “direct sun measurements”, “sun photometry” etc., the common practice is to not capitalize “sun”.

P1,L11. “…this uncertainty being smaller…”

P3, L15. “…it is the reference observatory…”

P4, L18. “…direct sun signal…”

P6, L2. “Smirnov”

P8, L9. “a certain”, or even better “a slight”?

P11, L14. “…we find a climatological average...”

P12, L4. “signal losses”