Comment on "Cloud Aerosol Transport System (CATS) 1064 nm Calibration and Validation" by Pauly et al.

Tyler J. Thorsen
NASA Langley Research Center; Hampton, VA
tyler.thorsen@nasa.gov

With several recent CATS science-related publications, this is a timely manuscript on the inner working of CATS and its retrievals. The wide-reach of CATS-related science applications certainly makes this manuscript appropriate for publication in AMT, eventually. However, in its current form, serious revisions are needed as there is a general lack of rigor and substantial ambiguity in the writing and analyses.

Show the full variability of the coefficients

The objective of this paper is to calculate calibration coefficients and their uncertainties, yet only a select 120 days of nighttime coefficients are shown in Fig. 3. I strongly encourage the authors to showcase the result of all their hard work: plot the entire 2.5+ years of calibration coefficients (night and day) along with their uncertainties. Just showing 120 nighttime coefficients over some unknown time period leaves several open questions in a reader's mind. How stable is the calibration? Are then trends/drifts in the calibrations and their uncertainties that a user should be aware of? Does the relationship with the cold plate temperature hold over the entire mission?

In addition to a time series of the coefficients/uncertainties, compositing these values by local time would be extremely useful information for the science community. The sampling throughout the diurnal cycle is one of CATS' most unique aspects that has, and will continue to, attract interest from those investigating diurnal cycles. However, any potential conclusions from these analyses needs to be tempered by the increased calibration uncertainty (and other uncertainties) during the daytime. Plotting and discussing the dependence of the calibration coefficients and their uncertainties with respect to local time would provide valuable context for those wishing to use CATS to study the diurnal cycle.

Showing how the calibration uncertainties vary is essential as they propagate into every aspect of the downstream science data products. The authors tend to focus on the effect of calibration uncertainties on the in-aerosol / in-cloud attenuated backscatter. But, before they are relevant there, they impact the detection thresholds. Somewhere, the authors should comment on how the calibration uncertainties and their diurnal variability impacts their feature detection. For example, presumably the increase in daytime calibration uncertainty necessitates more conservative detection thresholds to avoid any false positives. Have estimates been made of how many features may go undetected from this? This could be sussed out by imposing an artificial increase in the nighttime detection thresholds to see what features go undetected.

Proper CALIOP comparisons

The CATS and CALIOP comparisons in Section 2.1 and Fig. 1 need to be refined and expanded upon. First, the altitude range in Figure 1 should be extended to include both the CATS and CALIOP calibration range (i.e. up to 39km): the SNRs at these high altitudes are the relevant ones for calibration. In addition to comparing SNR in the entire column, as is done in Section 2.1, the text should be expanded to include a discussion of the SNR difference in each lidar's respective calibration regions. Additionally, since CALIOP performs its calibration at 532nm, the SNR profiles for CALIOP at 532nm should be added in as well and compared with CATS.

Despite the wavelength difference, comparing CATS 1064nm to CALIOP 532nm is more of an apples to apples comparison since they are the respective workhorse wavelengths for each lidar. These are the wavelengths for each lidar that are calibration, where feature detection is performed and the most accurate optical properties are available for. Therefore, comparing CATS 1064nm to

CALIOP 532nm is the most relevant comparison to those using the data products. However, as the authors do discuss, it is important to also point out CATS' superior SNR at 1064nm during nighttime for those whose particular investigations would benefit from this.

Error analysis

Section 3 is very unclear. Words like "overall" and "typical" are used to describe the various numbers and ranges given without any explanation of what they correspond to. Please be more precise when giving these numbers. Are these the average calibration uncertainties? Are the ranges interquartile ranges? Minimums and Maximums? Standard deviations? What is "var" in Eq. (13)? The authors seem to refer to this as "variability", do they mean variance? Please also indicate at what significance level the uncertainties presented here and in the data products are given for.

The uncertainty in the assumed backscatter color ratio does not appear to be included in the error analysis. Its value can vary substantially for (e.g. *Burton et al.*, 2012), which should be accounted for in the error budget. Alternatively, this could be avoided by just using the CALIOP 1064nm scattering ratios directly (see my comment below on this) and replacing the *Kar et al.* (2018) 532nm CALIOP calibration uncertainty with the 1064nm CALIOP calibration uncertainty given in *Vaughan et al.* (2019).

I was disappointed that the correlation between the nighttime calibration and the cold plate temperature was not exploited more (although it is not clear if relationship holds outside this 120 day period, see my comment above). I think the authors are missing out on an opportunity to explore improving their daytime calibration using this regression.

I would also suggest adding a short paragraph to the end of this Error Analysis section comparing the calibration uncertainties to other work that has performed a similar normalization, specifically MPLnet and CALIOP. Both of these where mentioned in the Introduction as forming the basis and background for this current study. Some brief context relative to MPLnet/CALIOP would make a nice connection back to your initial motivation and help give perspective to the readers that are more familiar with MPLnet/CALIOP than they are CATS.

Validation

I appreciated that the validation of this calibration tough: HSRL/Raman techniques aren't feasible at 1064nm, so all your left with is comparisons to other lidars who also need to calibration to a molecular signal. Because of this, one cannot treat CPL and EARLINET as absolute truth. Therefore, I suggest re-framing the discussion in section 4 around comparing the two profiles in the context of each instrument's calibration uncertainty (e.g. add uncertainty bars to the profiles Figs. 5-9 that correspond to each instrument's respective calibration uncertainty). That would put these comparisons within the proper context. Without this, it is easy to read too much into apparent absolute agreement as the authors themselves do on page 9 lines 29-30 where the agreement is called "surprising". The CPL/CATS agreement is NOT "surprising" after considering that the CPL was scaled by an assumed scattering ratio of 1.27! This large factor is quite uncertain and one could chose many reasonable values for it that would strongly impact the comparisons in Figs. 5 and 6. Showing the uncertainties involved would help avoid one reading too much into any agreement/disagreement.

From a sample size perspective, EARLINET is the authors' best bet for a comprehensive comparison. I encourage the authors' not to forgo this opportunity and go beyond only comparing eight nighttime overflights. I encourage the authors to also include daytime comparisons and a large enough sample size to make meaningful statistical comparisons.

For the CALIOP comparison, what is the motivation for comparing attenuated backscatter in cirrus? Since this study is concerned with calibration, why not compare CALIOP and CATS attenuated backscatter profiles as was done in the CPL and EARLINET comparisons? Adding the complication of cloud detection and multiple scattering into this seems unnecessary and out of scope for this study.

Aerosol scatting ratios

The text describing the scattering ratios and their presentation in Fig. 2 is confusing and in need of revisions. First, as Reviewer 1 points out, it is not really fair to call this a "molecular" normalization technique since, as Fig 2. shows, aerosol comprises anywhere from 30–50% of the signal you're calibrating! This is a huge challenge/limitation that the authors aren't very up front about (see my comments in the next section concerning this). Considering the need for these scattering ratios and their large contribution to the overall uncertainty, the authors need to be more precise in describing how they are incorporated into the algorithm and build confidence that these values are accurate.

It is unclear how exactly these scattering ratios are applied. On page 4 the authors state that "the CALIOP data is used to estimate the spatially and temporally varying 1064 nm scattering ratio at these altitudes (Fig. 2)." In Fig. 2 zonal mean scattering ratios are plotted in 4 different months. Are these zonal means what is meant by "spatially and temporally"? Why are only 4 months plotted in Fig. 2? What scattering ratios are used for the months not plotted? More specifics are need here. Additionally, if zonal means are used, I would recommend putting standard deviations on the curves in Fig. 2 and discussing the amount of variability that is neglected by using mean values (this would need to be included in the error analysis as well).

For Eqs. (2) and (3): why not just use the CALIOP 1064nm scattering ratios directly instead of assuming a backscatter color ratio? As I mentioned above, the uncertainty in the assumed backscatter color ratio is likely larger than just using 1064nm CALIOP data directly. Plus, the error in 1064nm CALIOP scattering ratios has already been characterized (*Vaughan et al.*, 2019) which would make the authors' uncertainty analysis more straightforward.

Limb sounding instruments will, by far, give the highest accuracy scattering ratios at these altitudes. Did the authors explore any other alternatives to using CALIOP for getting the aerosol scattering ratios? At the very least, the CALIOP scattering ratios should be compared to the climatology of SAGE II, SAGE III, GOMOS, etc...

Don't oversell the approach

I've touched on this throughout my comments above. There are several statements throughout the paper that are misleading considering the large uncertainty in having a significant amount of aerosols in the calibration region and the reliance on CALIOP to account for this. The authors state in the abstract that "Overall, CATS has demonstrated that direct calibration of the 1064nm channel is possible from a space based lidar using the molecular normalization technique". But this statement is only half true because the CATS calibration relies on having another, already calibrated, lidar in space (CALIOP). You can't characterize this as a "direct calibration" if 30–50% of your calibration (i.e. Fig. 2) relies on inter-calibrating to CALIOP! In essence, the authors follow a similar approach as has been done in previous work: derive a 1064nm calibration from calibrated 532nm backscatter.

The are several instances of the authors being cagey about this. For example, on page 3 lines

30-31: “CATS exhibits high nighttime 1064 nm SNR, enabling 1064 nm attenuated total backscatter (ATB) direct calibration without any dependence on the CATS 532 nm signal”. This statement is very misleading. Yes, you have no dependence on the CATS 532nm signal, but you do depend on the CALIOP 532nm signal and, on top of that, a CALIOP 532nm signal that is conveniently already calibrated for you. If the CATS 532nm signal was of sufficient quality, you would have certainty used the CATS 532nm signal instead!

Another example: page 13 lines 24-29. Here the authors do say the aerosol loading is higher in the CATS calibration region than for CALIOP. Instead of waiting for the conclusion, the authors should mention this in the introduction and then again when discussing Fig. 2. Additionally, it is important to convey, quantitatively, the difference between the two: CALIOP has aerosol scattering ratios that are less than 1.02 in its calibration regions (*Kar et al.*, 2018, their Figure 2b), CATS has values between 1.4-2.0 (Fig. 2). That is a very significant difference.

In spite of all these difficulties, the general pathway the authors have taken to calibration CATS is likely the best approach. But, the authors need to chose their words carefully and convey that their approach is not going to be widely applicable to other 1064nm lidars since they do not demonstrate an *independent* calibration of 1064nm backscatter. Speculating on a way to truly do so is good fodder for the conclusion. The authors do some of this already, but I would encourage them to expand that discussion a bit. Have the authors considered the precision/accuracy trade offs between a lowered rep rate and increasing the altitude of the calibration range? If the CATS measurements weren't limited to between 51S-51N how would polar stratospheric clouds impact the calibration in polar regions?

More minor comments/edits

The CATS 532nm channel is mentioned in the abstract and a few other places, but it is never explained why the Mode 7.2 532nm data isn't used. Why is its SNR so much lower? Does it not have the same photon counting detectors? Is the laser outputting less energy at 532nm? Is an attempt made to calibrate it at all? It is used for any data products?

I would suggest shortening the abstract. Some of the content is too specific for an abstract and difficult to understand without reading the main text first. Make sure the abstract acts as a stand-alone description of the paper.

page 1 line 15: “range-resolved”, “vertical” and “profiles” are all redundant, choose one adjective here

page 2 line 6: change “Scientists have used various methods for calibrating” to “Various methods have been used for calibration”. Also, this statement needs a citation(s): what are the various methods being referred to?

page 2 lines 9-11: change “Sometimes, as is the case for MPLNet,” to “Since MPLNet”. Also change “sites. In these cases, the aerosol optical depth” to “, the aerosol optical depth”

page 2 line 17: change “to differences in the” to “the weaker”

page 2 line 37: what wavelength

page 3 lines 37-38: any idea why this is? It is surprising a mere 1km would make such a difference.

page 5 lines 12-17: where did these criterion come from? The Liu et al (2004) study? If not, explain how they were chosen.

page 5 lines 18-22: these few sentences are confusing

page 5 line 33: give the thresholds used

page 5 line 37: does "file" = "granule"? These appear to be used interchangeably in a few other places as well.

page 5 line 37: "On average, 60-70%..." This is a range, not an average. Give the average value or explain what the range corresponds to.

page 7 line 26: don't give the total uncertainty until after discussing the individual contributions

page 7 lines 30-31 and Eq. (5): why not just use an updated model for molecular scattering?

page 6 line 17: replace "noise introduced by the solar background" to "solar background noise"

Figure 3: calibration spelled wrong in the caption

Section 3: make clear that you're deriving the nighttime uncertainty first in this section.

page 6 line 36: change "is shown" to "computed"

Figures 5 and 6: also show the CATS curtain in these figures

page 11 line 39: if this is a concern, why not just collocate them within some distance/time tolerance?

page 12 line 24: change "essentially" to "approximately"

page 12 line 5: remove "a detailed discussion of"

References

Burton, S. P., R. A. Ferrare, C. A. Hostetler, J. W. Hair, R. R. Rogers, M. D. Obland, C. F. Butler, A. L. Cook, D. B. Harper, and K. D. Froyd (2012), Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, *Atmos. Meas. Tech.*, *5*(1), 73–98, doi:10.5194/amt-5-73-2012.

Kar, J., M. A. Vaughan, K.-P. Lee, J. L. Tackett, M. A. Avery, A. Garnier, B. J. Getzewich, W. H. Hunt, D. Josset, Z. Liu, and et al. (2018), CALIPSO lidar calibration at 532 nm version 4 nighttime algorithm, *Atmos. Meas. Tech.*, *11*(3), 1459?1479, doi:10.5194/amt-11-1459-2018.

Vaughan, M., A. Garnier, D. Josset, M. Avery, K.-P. Lee, Z. Liu, W. Hunt, J. Pelon, Y. Hu, S. Burton, and et al. (2019), CALIPSO lidar calibration at 1064 nm version 4 algorithm, *Atmos. Meas. Tech.*, *12*(1), 51?82, doi:10.5194/amt-12-51-2019.