Responses to Reviewer #1 – John Yorks (amt-2018-206-RC1.pdf)

1. <u>Diurnal variability of aerosols</u>: On page 2 line 15, you discuss the assumption that the aerosols within the calibration transfer region are diurnally invariant. This is again discussed on page 5 lines 18-21. Please provide evidence that this is true or cite a paper that makes this claim to support this assumption.

   This is discussed more in detail in section 3.2. In this section we make the argument that a region which is atmospherically stable and is decoupled (as much as one can) from the lower atmosphere would not have a difference in the aerosol loading between day and night. As noted in that section, using an isentropic surface to isolate atmospherically stable regions is fundamentally a more correct approach than using tropopause heights (i.e., as we did in V3). Hoskins (1991) develops a concept called the 'Overworld', which is a region defined by isentropic surfaces that do not cross the tropopause. This increased stability acts to cap motions from the lower troposphere, with the exceptions of forced events (strong convection, forced lifting, volcanic events, etc…). This paper is cited on line 20 of page 8.

   To further confirm our assertion, we discussed the diurnal variability of aerosols with Dr. Larry Thomason, who has expert knowledge of the spatial and temporal distributions of stratospheric aerosols derived over many decades from the analysis of data from multiple sensors. Dr. Thomason confirms that over his career he has never observed a diurnal variation in background stratospheric aerosol measurements that was not ultimately traced to some kind of instrument artifact. In fact, in a recently published review paper (Thomason et al. 2018), diurnal variation of stratospheric aerosol loading was not mentioned because (a) background stratospheric aerosols are not significantly photochemically active and (b) diurnal changes have been not observed in the historical measurement record (Larry Thomason, personal communication).

2. <u>Limitations of V3</u>: On page 2 line 21, you say "First, the altitude of the V3 calibration transfer region was too low, and hence the assumed." Was a paper published already that shows this? If so, please cite it here.

   This is discussed in more detail in Section 2 (Version 3 532 nm daytime calibration), in which to compensate for the inherent differences between day and night we had to introduce a scaling to correct the scattering ratios at the lower altitude.

   Additional filtering and smoothing are applied to mitigate outliers. In particular, a minimum nighttime scattering ratio of 1.03 is used to compensate for diurnal differences in aerosol loading in the troposphere. Selection of this offset was based on observational analysis during development of the V3 algorithm, where it was noted that zonal distributions of attenuated scattering ratios in the calibration transfer regions fell below 1.0 in the tropics.

   There was no V3 532 nm daytime algorithm paper per se, but Kathy Powell did describe this in detail in her ILRC papers and presentations from 2008 (new citation) and 2010 (already cited elsewhere). The addition of Kathy's 2008 paper is noted below (changes in red).

   Calibration algorithms used in the version 3 (V3) series of L1 data products (Vaughan et al., 2018), released beginning in June 2009 are described in Hostetler et al., 2005, Powell et al., 2008, Powell et al., 2009, Powell et al., 2010, and Vaughan et al., 2010. Over the intervening years since the release of V3, several shortcomings have been identified in the 532 nm daytime calibration algorithm. First, the altitude of the V3 calibration transfer region was too low, and hence the assumed….

   To further clarify this the proposed change to text (in red) is below, in which we also cite Kathy's ILRC presentation and ask the reader to look at Section 2.

First, the altitude of the V3 calibration transfer region was too low, and hence the assumed diurnal invariance for the 532 nm daytime calibration was often not satisfied, as noted in Powell et al., 2010 and described further in Section 2.

3. <u>Latency</u>: Somewhere in Sections 2 and 3, please discuss the total latency of the 532 nm daytime calibration (V3 and V4). On page 3 lines 10-14 you discuss the need to use previous nighttime granules and on page 4 lines 14-15 you say clear air scattering ratios are accumulated for 7 days. Was the latency in V3 7 days? Did the changes for version 4 add additional time?

<u>Version 3</u>:

Seven days specifies the ideal time period over which the V3 algorithm accumulates all of the scattering ratios needed to derive the V3 calibration correction factors. However, 7 days is not the total latency of this product. The 7-day averaging window works like a FIFO queue; as new data is accumulated, the oldest data is discarded, and the averages are recomputed. Our <u>standard</u> V3 data products are generated within 2 to 4 days from downlink, partially due to this calibration approach and also because we require a number of inputs that are not available at the time of down-link.

Our <u>expedited</u> V3 products are generated with 24-36 hours from downlink. We use a less robust multi-day mean calibration, as well as estimates on some of the other required information (e.g., estimated attitude and ephemeris). This are more tailored for near-real-time processing applications and not for rigorous science – which is noted on the CALIPSO website.

The proposed change in the text below, found on page 3, lines 12 – 14, explains that the mean correction factors are built from a sequence of correction factors created over several days, and that more specificity (the question you brought up) will be provided later in section 2.

For any daytime granule, the 532 nm daytime calibration coefficients were then computed as the product of ~~these~~ a mean correction factor~~s~~ built from an accumulation of several days' worth of scattering ratios (discussed further in this section) and the mean 532 nm calibration coefficient from the previous nighttime granule (Powell et al., 2010).

<u>Version 4</u>:

Like version 3, the version 4 approach also accumulates the scattering ratios over some fixed period of time. But, instead of using the previous 7 days, the V4 approach accumulates data over a fixed number of orbits. The number of orbits required is different for day and night, but can be denoted generically as 'N'. Calibration estimates are derived using data acquired over ±½N orbits about the current orbit. This strategy automatically introduces a 3-4 day latency in the ability to create the daytime L1 product. However, because V4 uses MERRA2 reanalysis data rather than the GMAO-FPIT (Forward Processing for Instrument Teams) data used in V3, the actual CALIOP data product latency is on the order of 6 to 10 weeks from downlink. GMAO typically delivers the MERRA2 product one month at a time, usually the middle of the next month. So, for instance, CALIPSO would get October 2017 no earlier than November 10-20th, 2017.

Section 4, page 9 lines 12-15 (below) was updated to clarify that there was not a minimum window needed as in V3.

Applying standard propagation of errors techniques to the daytime calibration equations shows that an averaging period of 105 consecutive orbits (i.e. over 7 days), centered on the orbit to be calibrated, should be sufficient to derive calibration coefficients with acceptably low random uncertainties. Unlike V3, which requires a minimum of 4 days to accumulate the required scattering ratios to build the mean correction factor, the V4 approach does

not have a set minimum number of orbits. Any reduction in the number of orbits used to generate the calibration coefficients will be reflected in the associated uncertainties.

In addition to the changes cited above, an entirely new section (Section 3.9) was added to more completely describe the latencies of both the V3 (standard and expedited) and V4 (standard) products.

3.9 Data latencies

Data latencies – i.e., the times between data acquisition and data product delivery – have also changed between V3 and V4. The CALIOP V3 standard data products are generated within 3 to 5 days from downlink, partially due to the V3 calibration approach but also because the analyses require a number of ancillary inputs that are not immediately available (Winker et al., 2009). The latency for the V4 standard products is considerably longer. Because V4 uses the MERRA-2 meteorological data rather than the GMAO FPIT products, V4 standard products are typically not available until 6 to 10 weeks from downlink. However, the V3 expected products continue to be available with 24 – 36 hours from data acquisition (i.e., ~12 hours from data downlink). The expedited processing uses a faster (albeit less robust) calibration strategy, as well as estimates for some of the other required information (e.g., platform attitude and ephemeris). The expedited products are tailored specifically for near-real-time processing applications, whereas the standard products are designed for rigorous scientific analyses.

4. <u>Polar Clouds</u>: On page 8, lines 6-27 you discuss the new calibration transfer region. This is a good discussion, but I found myself wondering what impact PSC's have in the Polar region. Later, I read nice discussions on this in Sections 4.2 and 4.3. I suggest adding a sentence on page 8 stating that PSCs can introduce some uncertainty to the V4 532 nm daytime calibration constants and more details are discussed in those later sections.

We have added the following sentence (in red) to the paragraph on page 8, lines 23-27.

Two additional safeguards are used to avoid possible contamination of the clear-air attenuated scattering ratios. First, to both guard against features intruding into the lower stratosphere, and because the algorithm uses a climatological monthly mean 400K surface as the lower limit, , an additional altitude offset of 2 km is applied to further elevate the base of the calibration transfer region. Secondly, since the stratosphere is not entirely devoid of features, the algorithm employs a 1064 nm feature detection technique, as discussed in Sect. 3.4, to exclude cloud and aerosol layers from the calibration averaging scheme. In particular, the presence of undetected polar stratospheric clouds (PSCs) in the calibration transfer regions can introduce high biases into the calibration coefficient estimates. The potential impacts of feature contamination of the calibration transfer regions are discussed in detail in Sections 4.2 and 4.3.

5. <u>Grammar errors</u>: On page 8, line 15 the phrase "The work product from this study" is a bit confusing to me. I believe you mean that "The final result of this study…" On line 25 of that same page, you have two commas in a row. On page 23, line 13 there does not appear to be a space between "(2018)," and "which".

We have changed page 8 line 15 to change the phrase as you suggested.

~~The work product from~~ The final result of this study is a comprehensive set of lookup tables derived from 5 years of V3 L2 5 km cloud profile data and the corresponding GEOS 5 FP-IT (Forward Processing for Instrument Teams) Version 5.91…

The grammatical issues noted for *(1) line 25 of page* 8 and *(2) line 13 on page 23* have been corrected.

6. <u>1064 nm feature detection</u>: I think this is a really good idea. One concern I have though is that you are only identifying layers that are > 1 km thick. Certainly, you can get some very thin volcanic or smoke plumes in the UTLS (I've seen them in CALIOP and CATS data). How much do you think these types of layers contribute to what you see in Figure 9? Please add a sentence or 2 on this subject to Section 4.3.

    Consider the right column of Figure 9, in which the 1064 nm failed to identify a layer in the UTLS but the 532 nm did. Of the layers we failed to detect at 1064 nm, only relatively small fractions have vertical extents of 1 km or less. Segregated according to the horizontal averaging required for layer detection, these fractions are 13.3%, 21%, and 12% at averaging resolutions of, respectively, 5 km, 20 km, and 80 km. Figure xx (below) shows 1064 nm IAB (y-axis) versus 532 nm depolarization ratio (x-axis) for these 'less than 1 km' layers, scaled as in Figure 9.
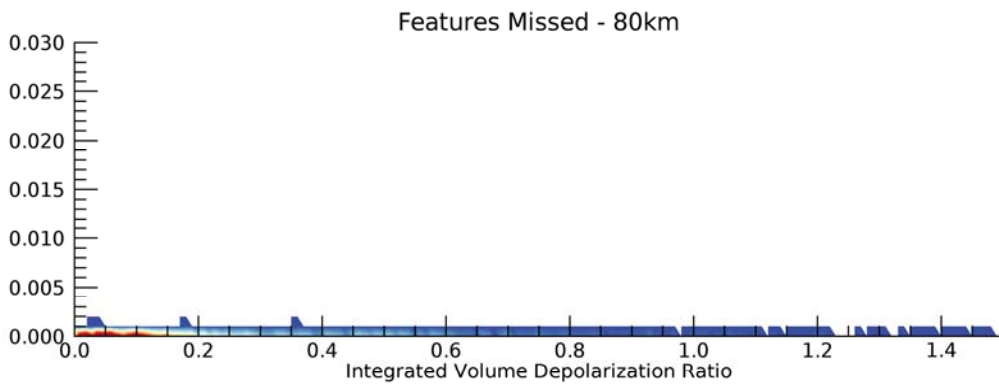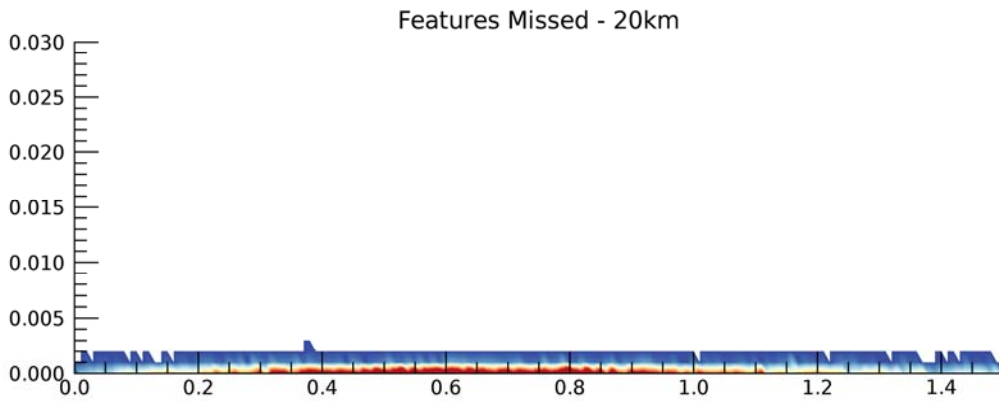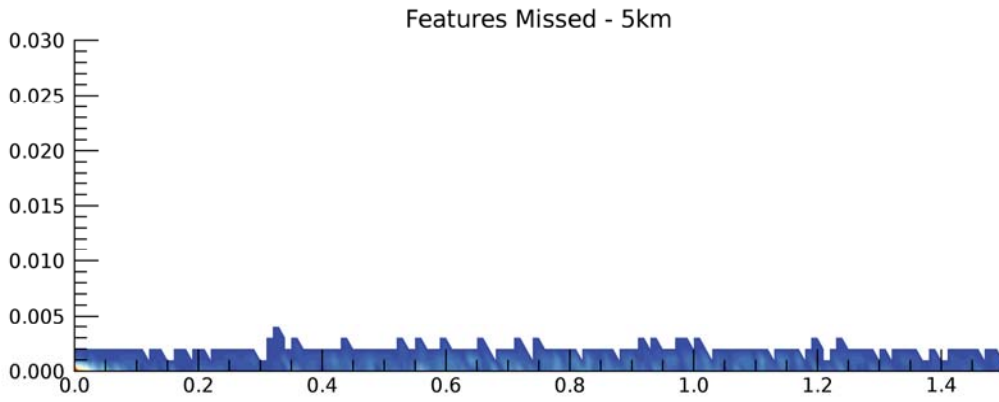
Figure XX: Identical approach used to generate Figure 9, but only using layers with a depth of less than 1 km

We have added the following sentence (in red) to the paragraph on page 20, lines 8-15.

The preponderance of the 1064 nm detection failures is seen in the bottom left corners of the plots in the right hand column of Fig. 9. These features, for which both γ'_{1064} and δ_v are very low, are likely to be faint and perhaps even persistent aerosol layers that consist of small particles that are not readily detectable at 1064 nm. For other features having low values of γ'_{1064}, the likelihood that the L2 algorithm has identified a false positive increases sharply as δ_v rises above ~0.7. That is not to say that all are false positives. As noted in Section 3.4, features are identified when the 1064 nm signal is above a threshold for 1 km or more. This will likely exclude the detection of thin features (< 1 km vertically) which may be strongly scattering but have small particle sizes (e.g., volcanic ash or elevated smoke). Also, while depolarization ratios approaching ~0.7 have been reported for contrail cirrus (e.g., Sassen and Hsueh, 1998) and have been occasionally observed in nighttime observations of polar stratospheric clouds (Pitts et al., 2009), values of this magnitude are highly unlikely for the bulk of the features that form in the stratospheric regions searched during the 532 nm daytime calibration procedure.

7. <u>Interpolation of missing data</u>: As I read Section 3.7, I found myself wondering how this interpolation may reduce the calibration accuracy. I later found a nice discussion on page 16. Please add a sentence to Section 3.7 that says something like "the implications of this interpolation on the accuracy of the calibration constant is discussed in Section 4.1 and Figure 5".

We have added the following sentence (in red) to the paragraph on page 14, lines 8-17.

Calibration coefficients derived from the day-to-night ratio of attenuated scattering ratios can only be calculated within those portions of an orbit in which both day and night observations are acquired. Because of the illumination patterns in the polar regions during the solstice seasons there are no matching daytime and nighttime samples near the poles in summer and winter. This seasonal high latitude lack of matching day and night data is accounted for by anchoring the ends of the neighboring daytime and nighttime calibration coefficient curves and interpolating between these end points. Where there are neither daytime and/or nighttime samples, the 532 nm daytime calibration coefficient and coefficient uncertainty curves are linearly interpolated as a function of orbital elapsed time anchored to the nearest neighboring 532 nm nighttime calibration. Figure 4 shows an example. In this orbit from July 2010, the day-to-night terminator occurs at ~60° N, and thus no corresponding nighttime measurements are available over the final ~1100 seconds of the daytime granule. The impact of this interpolation on the accuracy of the calibration coefficients and uncertainty estimates is discussed further in Section 4.1.

8. <u>Figure 9</u>: Something that is confusing me about this figure: Is the white color where the frequency equals 0 and blue is non-zero (1, 2, etc.). Or is the blue zero? The color bar would suggest the latter. Please try to clarify this

Figure 9 uses IDL color table 72, which is Color Brewer Scheme Red-Yellow-Blue. White is where the frequency is 0, while the minimum of the blue distribution is 1. In order to better convey all intended information, the figure was recreated with a different color bar that goes from blue to yellow, a re-scaling of the x and y axis, a single color bar at the bottom of the image, and a change from sampling counting to sample frequency. Sample frequency, derived independently for each of the 6 plots, is computed by dividing samples by the maximum sample. These changes do not impact the discussion contained in the section related to this figure.
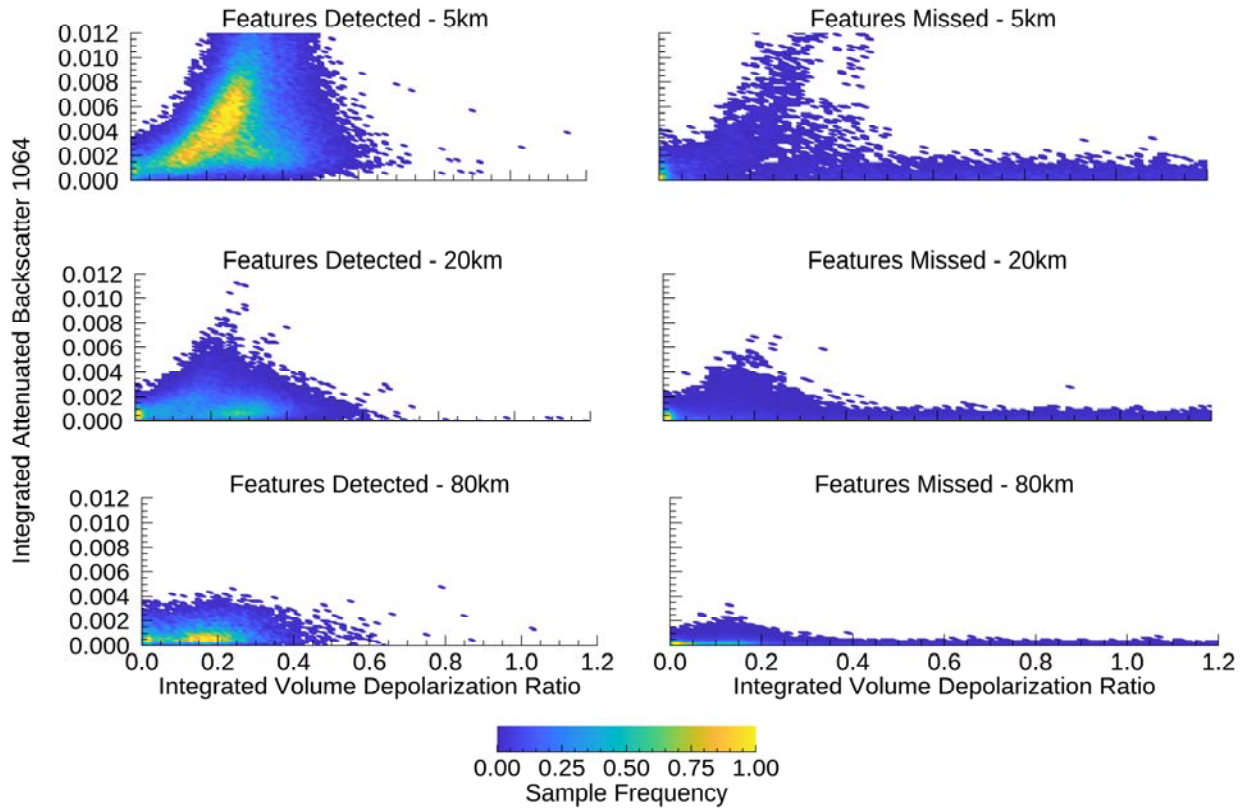
References:

Thomason, L. W., Ernest, N., Millán, L., Rieger, L., Bourassa, A., Vernier, J.-P., Manney, G., Luo, B., Arfeuille, F., and Peter, T.: A global space-based stratospheric aerosol climatology: 1979–2016, Earth Syst. Sci. Data, 10, 469-492, https://doi.org/10.5194/essd-10-469-2018, 2018.