Response to Reviewer # 2

Thanks very much for your careful reading of the manuscript and for the useful suggestions. We list below the changes we have made in the manuscript in response to your comments.

1) Abstract: The abstract could be more informative, by listing the 5 main changes in the calibration that I have listed above, together with the main outcomes that I have also listed above. On the other hand, the text at lines 24-28 could probably be moved to the introduction.

We have added the following in the abstract:

"Due to the greatly reduced molecular number density and consequently reduced signal-to-noise ratio (SNR) at these higher altitudes, the signal is now averaged over a larger number of samples using data from multiple adjacent granules. As well, an enhanced strategy for filtering the radiation-induced noise from high energy particles was adopted. Further, the meteorological model used in the earlier versions has been replaced by the improved MERRA-2 model. An aerosol scattering ratio of 1.01 ± 0.01 is now explicitly used for the calibration altitude. These modifications lead to globally revised calibration coefficients which are, on average, 2-3% lower than in previous data releases."

2) P2 L18-19: can be computed → used to be computed in V3

This statement is true for both versions. For better clarity we have replaced "can be computed" by "are computed".

3) P2 L19 (atmospheric model): replace these words with a term that specifies the type of model (forecast, analysis, reanalysis, climatology, standard atmosphere?)

We have added "assimilation" before models.

4) P2 L19 (GMAO): with reference to the GMAO web site, where several products are listed (FP, FP-IT, Seasonal forecasts, MERRA-2, 7km-G5NR, SMAP L4), indicate which product is specifically being used in V3 (I can guess easily that some are not relevant, but I feel that it should be specified).

We now specify the GMAO data used for all V3 and V4 CALIOP data releases in the discussion immediately following equation 4 in the revised manuscript:

"CALIPSO versions 3.01 and 3.02 used GEOS 5.2 data. Versions 3.30 and 3.40 used the FP-IT near real time assimilation products (GEOS version 5.9.1 and 5.12.4). The initial release of the CALIOP V4 data products (version 4.00) used the FP-IT product built with GEOS 5.9.1. The current V4 release (version 4.10) uses the MERRA-2 reanalysis product (Molod et al., 2015; Gelaro et al., 2017), which has enhanced physics, including a new gravity wave drag parameterization that is capable of producing a Quasi biennial Oscillation (QBO), and spans the entire CALIOP data record, from April 2006 to the present."

5) P3 L5: are nearly absent –> are thought to be nearly absent

Done.

6) P5 L10: loading of _6-8% –> backscatter ratio of 1.06-1.08 (aerosol loading is an ambiguous term: it may suggest a percent expressed in terms of mass concentration)

7) P5 L11 (loading decreases to _1-1.5%): same comment as above

We have modified the sentences as below:

"Both the instruments show significant aerosol scattering ratios of 1.06-1.08 at 30-34 km at the tropics, decreasing to ~1.02 in the polar regions. On the other hand, at 36-39 km R decreases to ~1.00-1.02. GOMOS shows a low bias compared to SAGE II at both altitude ranges, with a scattering ratio of ~1.01 at 36-39 km."

8) Equation 2 suggests that the laser pulse energy and the amplifier gain are monitored continuosly and accounted for on a pulse by pulse basis; however this is not explained in the text (nor in P09). I would suggest to clarify this. Note that if they are not accounted for on a pulse by pulse basis, then probably their indication in equation 2 is unnecessary.

The laser pulse energy is monitored from pulse to pulse but the amplifier gain is changed twice in an orbit to account for day-night conditions. These are explained clearly in the CALIPSO Level 1B ATBD. Here we are only giving the outlines of the calibration calculations and thus added a brief clarification:

"E_0 is the laser pulse energy continuously measured on the platform, and G_A is the electronic amplifier gain adjusted for night and day operation."

9) P6 L5-6: the fact that the units of C are expressed in km3 sr suggests that S(z)/(E_0*G_A) is dimensionless. This should be clarified. I am inclined to think that the lidar signal S(z) will have some form of units (volts? photon counts? readings on an ATD converter?) which would reflect onto the units of C [note also that E_0 is an energy and should have units of J].

That is correct; we failed to account for the energy term. The units of C are km^3 sr J^{-1} count. ('count' comes from the ATD converter.) and have been corrected in the text.

10) P6 L10 (measured): these values are not "measured" because they are from a Model

Yes, we have deleted "measured" from the sentence.

11) Equation 4: I would suggest to give a plot of the two-way transmittance with z above 30 km, highlighting the contribution of ozone and of molecular scattering separately for a "average" conditions.

Tables 1 and 2 below show the mean two-way transmittances for molecules and ozone calculated from profiles averaged over 60^{o}N-60^{o}S for one granule (2008-02-15T11-42-36ZD) of V3 and V4 data. The values are very close to one for both the molecular and the ozone component in the altitude region above 30 km. Further, there is hardly any change going from V3 to V4, as can be seen in Figure 1 below. Because the differences are essentially negligible, we have not added additional plots in the manuscript.

Table 1: GEOS 5.2 data (CALIOP version 3.01)

Z (km)	$T_m^2(z)$	$T_{O_3}^2(z)$
35.9037	0.9995	0.9982
29.9760	0.9981	0.9916
0.0228	0.8005	0.9607

Table 2: MERRA-2 data (CALIOP version 4.10)

Z (km)	$T_m^2(z)$	$T_{O_3}^2(z)$
35.9037	0.9995	0.9982
29.9760	0.9981	0.9917
0.0228	0.8002	0.9610

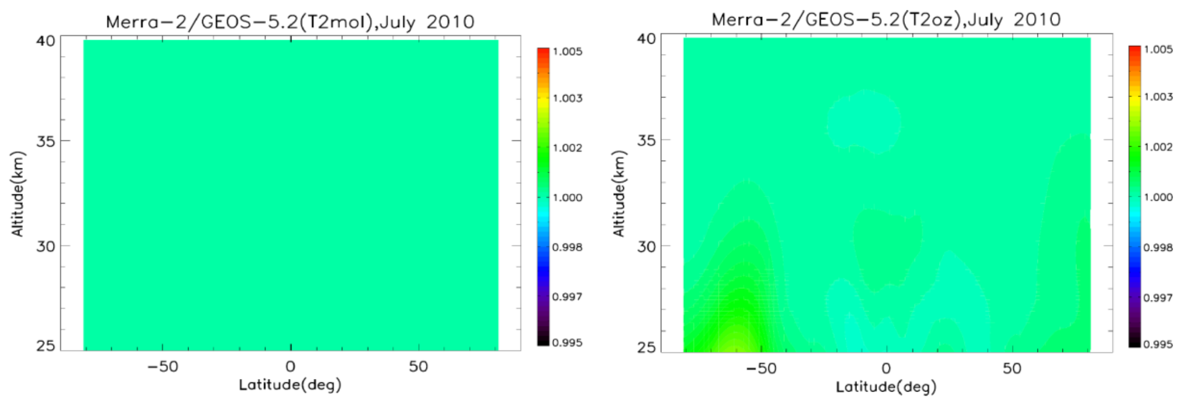

Figure 1: Height latitude cross section of zonal mean MERRA-2/GEOS-5.2 two-way transmittance ratios for July 2010.

12) P6 L21: provides –> is thought to provide

Done.

13) P6 L21-22: has any comparison been made of the beta_m from GMAO and MERRA-2? Could the difference between the datasets be described in one-two sentences?

Yes, we have done extensive comparisons between the two models before implementing the MERRA-2 model. We have added the following in the manuscript:

"As an example, comparison of CALIOP V3 (created using GEOS-5.2) and V4 (using MERRA-2) data for July 2010 in the calibration region for both V3 and V4, i.e., between 30-40 km (including all latitudes) indicates that the fractional difference (V4-V3/V3) in molecular density varies from zero to about 1.5%, with a mean difference of ~0.7%. The molecular backscatter coefficients between the two models will differ by the same amount. Fractional difference in ozone density (or absorption) varies from about -10% to 5% with a mean difference of ~ -1.7%. The resulting total two-way transmittance changes between GEOS-5.2 and MERRA-2 vary from about -0.01% to 0.03% with a mean difference of ~0.003%. These values can vary somewhat with latitude and season."

14) P6 L26-27: a few words could be spent to explain how you arrived at the value of R=1.01. Does it derive from any measurements? Is it simply the value that yields best CALIOP data?

This is explained in detail in section 2.1 in the paragraph immediately following Figure 2.

15) Figure 4: x-axes for figures use all different scales: latitude, along-track distance, granule elapsed time, etc. This may be confusing! It would be better to use a scale such as latitude, which does not need any particular explanation. If however you think that this is not the best way to represent the data, I suggest to indicate in each figure caption where the zero is for each granule (e.g. at the day-night terminator) and which way the satellite motion goes (e.g. North to South). See also Figure 6.

All these terms are part of standard CALIPSO terminology and help clarify different aspects of the data. Thus we prefer to retain these terms as such. Where necessary (as in Figure 5) we have given both "latitude" and "granule elapsed time" for clarification. In any case, per your suggestion, we have added information on the starting point of the scale in Figures 4, 5, and 6.

16) Figure 4: specify in the figure caption that these are V3 data calibrated at 30-34 km.

Done.

17) P9 L17-18 (PDACs for which all data points are rejected by this process are labeled as invalid): this should be better clarified. Do you mean that you would reject a PDAC that for instance has 0 data points, but would accept one with one data point or more? Would it not be safer to express the threshold as a percent? (e.g. invalid the PDACs that have less than 50% of the expected data points).

Yes, the algorithm was designed to accept a PDAC with at least 1 data point in each range bin in the calibration region. We have added the following text for clarification:

"At least one sample in each range bin in the calibration region for any PDAC is required. Otherwise, the calibration coefficient and its uncertainty for this PDAC are labeled as invalid, and excluded from further calibration processing. This is different from V3 where for each failed

PDAC, a historical estimate of the calibration coefficient (daily average of all valid calibration coefficients from the previous day) was used (see P09 for details)."

18) P9 L20: is 0.15% a global figure or does it refer to the % rejection in the SAA region?

This was done globally.

19) P9 L28 (radiation-induced noise): specify if you refer to the SAA and the impact of high energy particles on the measurements, or to a photodetector non-linearity effect.

Actually the text clearly states that we are talking about radiation induced noise at high latitude and not in SAA.

20) P10 L16-17: give size and date range for the test data, and indicate it also in percent of the whole dataset

This information was already there in lines 4-6 on the same page, P10. In any case, we have added the following sentence to clarify this,

"As mentioned above, the data set used for testing the filters encompassed the years 2007 through 2012 thus including more than 90% of the data available at that time."

21) P12 L7 (significantly lower success rates): it looks like if it approaches ZERO success rate in the SAA. Can you state the actual value reached at the minimum? This fact may deserve a comment in the text. In particular, re-state that in V4 the low calibration success can be overcome thanks to averaging over adjacient orbits. I do not quite understand, however, how in V3 you were able to calibrate in this area.

Yes, the minimum value of the success rate actually reaches zero over the SAA region. Whenever a particular PDAC failed any of the 3 filtering steps in V3, a historical estimate (daily average from the previous day) was used. We have added the following text to address this concern:

"The minimum value of the success rate within the SAA region reaches zero. In V3, historical calibration coefficient estimates (daily average from the previous day) were used whenever a PDAC would fail any of the 3 filtering steps, and these historical values were included in all subsequent averaging operations (see P09 for details)."

In this same paragraph, we had already mentioned that in V4 the low success rate is ameliorated by multi granule averaging. The details of filtering procedure for V3 including in SAA are given in section 3d of Powell et al. (2009).

22) Figure 8: show the V3 calibration too.

This has been done.

23) P14 L10-12: comment and possibly explain about the large reduction in the first year and the subsequent recovery.

We have added the following text to address this:

"The relatively rapid decay in C over the first year of the mission is attributed to a persistently increasing wavelength mismatch between the laser transmitter and the etalon in the receiver (largely corrected by the initial retuning of the etalon in March 2008), compounded by boresight misalignment (Hunt et al., 2009)."

24) P14 L22: comment and possibly explain why C is so much smaller over the South pole.

We have added the following text:

"The lower values of the calibration coefficient over Antarctica are due to thermal beam steering effects in the instrument that occur as the satellite first enters the sunlit portion of the orbits when approaching the night-to-day terminator (e.g., as seen in Figure 4)."

25) Figure 10 (bottom): plot also for V3.

There was an error in computing the calibration uncertainties in V3 and as such those uncertainty values in the V3 release are incorrect. Therefore we have not plotted the uncertainties from V3. We have added the following text to explain:

"We note, however, that there was a bug in the V3 code that caused the uncertainties reported in the L1 data products to be underestimated by a factor of 3 or more. For this reason, the lower panel of Figure 10 plots only the V4 uncertainties, and not the differences between V3 and V4 that are shown in the upper panel."

26) Equation 6: at high altitude, the attenuated scattering ratio R' should be identical to the backscatter ratio R unless there are clouds/aerosols above a layer. Is it worth mentioning?

We are not sure if this statement will add much to the discussion here and thus refrained from mentioning this.

27) P20 L16-17: a couple of peaks on the blue curve in Figure 15b could deserve a comment from the authors.

We have added the following text:

"At a couple of locations, the R' curves show significant deviations, which could be due to some real variations in aerosol loading or noise in the data."

28) Figure 16: a negative R seems to be present above the calibration range in V4 (right hand panels). How is the trend above 40 km? Is what I see in the figure just statistical noise, or is there a decreasing trend above this altitude? I suggest that this deserves to be commented.

We cannot comment on the trend above 40 km because we do not take measurements above 40 km. This 40 km limit is now called out in the abstract as well as in the caption for Figure 3.

29) P23 L7-9: there is still a flight to flight variability (+/- 5%), and I suggest that this fact could be commented.

We found this remark somewhat perplexing. As stated in the abstract, the body of the paper, and in the inset on Figure 17, the mean bias in the nighttime comparisons is 1.6% ± 2.4%. Assuming a Gaussian distribution of the measurements, observing that the “flight to flight variability” is on the order of ±5% is entirely consistent with the statistics we have presented; i.e., if we consider two standard deviations, the flight-to-flight variability is ±4.8%. Therefore, given that your observation is, in effect, a restatement of the statistics already presented in the manuscript, we have not commented further on this.

30) Conclusions L7 (two major changes): I believe that there are more than 2 changes. I did list 5 at the beginning of this review, based on what is described in the manuscript.

We have added the following text to address this:

“Among other important changes are an improved noise filtering scheme, adoption of MERRA-2 as the meteorological model, and the explicit accounting for the presence of residual aerosol in the calibration region.”

31) Conclusions: At the moment, this section is only an abstract/summary of the article. It could be expanded, by discussing with more detail and emphasis: (a) the repercussion of the V4 calibration on CALIPSO products; (b) the repercussion on major downstream users and on major scientific results that have made use of the CALIPSO mission (e.g. climate science applications); (c) a discussion of potential future work to improve the calibration even better. Any issues encountered and lessons learned could also be described here. The conclusions should put the paper into the wider science perspective.

We have added the following text to address this:

“In particular, the attenuated backscatter values increase by about 2-3% on average, which enables increased detection of tenuous layers by the level 2 algorithm, particularly in the stratosphere. The improvements in stratospheric aerosol retrievals will be invaluable for cross-validation of the stratospheric aerosol products from other instruments such as the Stratospheric Aerosol and Gas Experiment III on International Space Station (SAGE III-ISS), and are expected to lead to a better understanding of climate related issues.”

Thanks once again and we hope that you will be agreeable to these revisions.