Referee 3

Major comments

We recommend that the authors:

. first compare CALIOP-SODA and CALIOP V4 AOD as well as lidar ratios for all CALIOP V4 aerosol types before they classify their analysis by CALIOP V4 aerosol types. As they know, CALIOP V4 is likely to miss tenuous aerosols or misclassify aerosols. Starting by classifying CALIOP-SODA lidar ratios per (likely misclassified) CALIOP V4 aerosol type is confusing and slightly circular.

R: We appreciate the reviewer's comments and suggestions for improving our analysis. Our responses are highlighted in blue. While comment 1) is indeed relevant, it is less applicable to our work, as explained in the following. We agree with the reviewer in that aerosol misclassification and under-detection are issues to be taken into account in any CALIPSO-centric analysis. However, the reviewer would agree that addressing these issues in detail from a global-scale perspective is unfeasible, and that assessments of lidar ratio irrespective of aerosol type does not address the main issue tackled in our study, which is intercomparing CALIPSO-SODA lidar ratios with the V4 lookup table for specific aerosol types. Instead, our approach is more pragmatic: we first start by considering that CALIPSO aerosol typing is providing meaningful information which does not necessarily overlap with other typing definitions obtained from suborbital and ground-based lidar ratios (e.g. using HSRL or Raman lidar) but yield aerosol clusters with measurable characteristics. So, our objective here is characterizing the lidar ratio of CALIPSO aerosol type, and provide maps that can be guide the selection of lidar ratios for a future CALIPSO version. We would like to emphasize that our main objective is not to recommend specific refinements to the aerosol classification to achieve more consistency relative to other studies that make use of more advanced lidar measurements. In sum, we are interested in finding ways to select more adequate lidar ratio that can yield better aerosol extinction coefficient and aerosol optical depth. Regarding the misclassification and under-detection, our methodology intends to minimize these uncertainties by using an AOD threshold and performing the analysis when only one aerosol type is observed in the aerosol column.

. clarify why CALIOP does not consider the possibility of polluted dust within the PBL

Polluted dust is defined for layers with base above 2. 5 km because it was found in previous HSRL studies (e.g., Burton et al., 2013) that polluted dust in Version 3 has characteristic lidar ratios of 35 sr, suggesting that aerosols that were actually mixture of dust and marine aerosols were misclassified by V3 as polluted dust in the boundary layer. This finding motivated the inclusion of a new aerosol type in V4: dusty marine. To identify dusty marine aerosol mixtures, polluted dust was redefined as being confined to the free troposphere (>2.5 km), to reduce the number of samples misclassified as polluted dust in the boundary layer. For more details, we refer the reviewer to Kim et al. (2018).

. clarify their filtering method and technique as a bullet list or a table (includes 1L vs 2L techniques, cloud masking, altitude selection etc)

We appreciate this suggestion. In the revised manuscript, we have added the following table to summarize the methodology

Table 2: Summary of methodology applied to CALIPSO-SODA and V4 intercomparison.

Condition	**Application**
AOD threshold	SODA AOD > 0.05
Aerosol types	Over the ocean, with only one aerosol type throughout the column (excluding clear)
CALIPSO-SODA 1L assumption	applied to dust, smoke, and polluted smoke
CALIPSO-SODA 2L assumption	applied to clean marine, marine dust, polluted dust
CAD score	> |-50|
Cloud coverage	cloud free over the 5-km horizontal resolution

. clearly suggest which CALIOP product is accurate, which one is not. And proposes fixes to the algorithm or additional filters to be applied by the users moving forward.

Section 5 already provides a summary with the information required by the reviewer. The results are dependent on multiple factors including spatial resolution and geographical location and, thus, it is challenging to say that product A is more accurate than product B. While we did provide recommendations for refining the classification of dusty marine and polluted continental smoke aerosols, improving the typing scheme requires more dedicated efforts beyond the scope of our work. As our work deals with assessing prescribed lidar ratios and motivates future work for refining lidar ratios in a future CALIPSO version, we do not have specific suggestions for CALIPSO users.

Detailed comments:

. line 14 – "This implies that the CALIPSO classification scheme generally categorizes aerosols correctly" is too strong of a statement. Please consider watering it down.

We modified the sentence to read: “This implies that the CALIPSO classification scheme generally categorizes specific aerosols types correctly over regions where they are abundant”

. line 21 – “value”

Corrected, thanks.

. abstract – the authors point out issues with CALIOP V4 polluted continental/ smoke, dust, dusty marine and clean marine aerosol types. They should suggest some future fixes and possible filtering

For the sake of conciseness, we have left the abstract as is. Discussion is provided throughout the manuscript.

. line 105 – why not talk about HSRL instead of saying “airborne lidar observations”?

The sentence was slightly modified to: “airborne HSRL observations”

. Figure 1 – can you briefly specify how reliable is GEOS-5 ML?

McGrath-Spangler and Molod (2014) compared several methods for computing planetary boundary layer depth using GEOS-5. They found that the bulk Richardson number method (which is adopted in our paper) provides the best match with radiosonde-based estimates over land. Von Engeln and Teixeira (2013) noted that the bulk Richardson number method yields heights well below the inversion height in cloud-topped marine boundary layers, suggesting that the estimated MBL is more closely related to the cloud base height. The revised manuscript briefly summarizes this information.

. line 226 – I don’t understand the sentence. Please rephrase.

We have rephrased the sentence to read:

“Lastly, polluted dust resembles the spatial distribution of elevated smoke, which reflects the influence of biomass burning emissions (especially in the South Atlantic) and that these are the only two aerosol types-defined for aerosol plume elevations above 2.5 km a.m.s.l. when the depolarization ratio is below 0.2 (Table 1).”

. line 240 or eq. (1) – explain i=1 to N

N denotes the number of samples. This information is included in the revised manuscript.

. line 249 – largest mean AOD

The sentenced was modified accordingly.

. line 221 – consider replacing “modest” by low and then “particularly low” instead of “negligible”

Done

. Figure 5 & 6 – consider saying “density plot” instead of histograms

We appreciate the reviewer’s suggestions. We believe that the concept of bivariate histogram is more accurate than density plot.

. Figure 7 – add (a), (b) and (c) on graphs

The labels were already included in the original submission

. These Figures and Tables are redundant, consider consolidating/ simplifying – Fig 9 and Table 4 as well as Fig. 7-8 and Table 3

We have tried several combinations of tables and formats, as well as reducing the number of figures. However, it is quite challenging to create compact tables/figures that clearly show the results that we want to highlight.

. Line 275 to 277 – aerosol variability depends on the environment. Please refer to e.g., Shinozuka and Redemann (2011).

We appreciate drawing our attention to Shinozuka and Redemann. We have modified the sentence to read: “This could be in part caused by mixing of different aerosol plumes at such large horizontal scales especially near emission sources (e.g. Shinozuka and Redemann, 2011)…”

. Line 305 – there is a repeat in the sentence. Please rephrase

The sentence was rephrased to read:

“Interestingly, polluted continental lidar ratio for AOD > 0.15 reaches values near 72 sr, in good agreement with the value used in V4 (70 sr).”

. Line 303 to 307 – can we make sure to say that the lidar ratio for a specific aerosol type should be independent of the AOD? The fact that it varies with AOD points to some issues in the CALIOP algorithm and I suggest describing them.

Lidar ratio is an intensive parameter that does not correlate with AOD. Certainly, uncertainties in AOD can yield bias in lidar ratio for the same attenuated backscatter profile. However, variations of lidar ratio with AOD observed in our study are primarily related to the fact that optically thicker aerosol layers can be better identified by the typing algorithm, implying that lidar ratios will better match the prescribed lidar ratios for higher AOD.

. Line 432 – assumed by the CALIPSO

corrected

. first bullet in the conclusion – consider explaining the reasons and implications

We have added the following sentence: “Classification issues for 80-km averaged samples are likely, as spatial averaging are performed to increase the SNR for tenuous aerosol layers, rendering more uncertain retrievals than its 5-km and 20-km counterparts.”

. Line 443 – within +-10sr of those

corrected

. Line 460 – is a repeat of the sentence above

The sentence was slightly modified to read: “Namely, we attribute substantial differences between estimated lidar ratio and the prescribed value in V4 for polluted continental/smoke to aerosol classification, as the retrieved lidar ratios are 30 sr smaller than the one used in V4. ”

Shinozuka, Y. and Redemann, J.: Horizontal variability of aerosol optical depth observed during the ARCTAS airborne experiment, Atmos. Chem. Phys., 11, 8489–8495, https://doi.org/10.5194/acp-11-8489-2011, 2011.