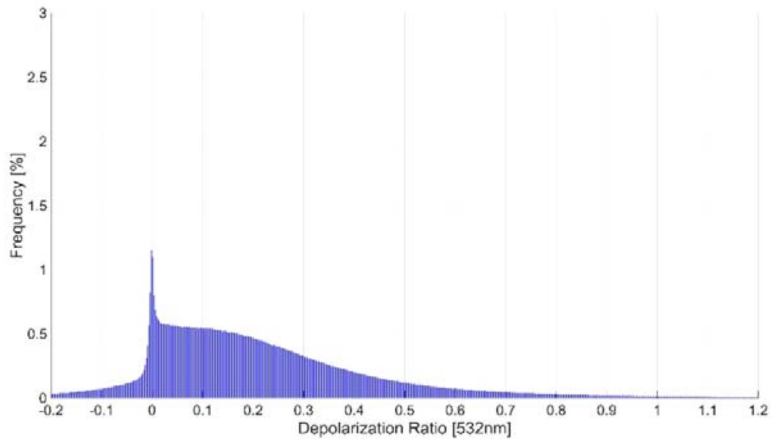
Dear authors,

While your study is of high interest, I have the feeling that you miss some important literature on the same subject and moreover, many of the aspects mentioned should be revised, focusing specifically on the following points that I see from a first reading:

1. Differences between the CALIOP and MODIS global DODs are large. Is there any explanation about this discrepancy? Please note that the MODIS-derived global DOD is substantially higher than those reported in most of the recently published works (e.g., Ridley et al., 2016; Voss and Evan, 2020; Gkikas et al., 2021). A description is needed on how the global averages have been computed for both sensors. Do you acknowledge any weighting factors based on the grid cell surface area? According to Levy et al. (2009), the approach for the calculation of the global DOD is quite critical (see Fig. 5). Summarizing, I recommend including a table providing the corresponding global DODs given by relevant studies (relied either on observations or models) in order to check (and discuss) the consistency of your findings.
2. The manuscript could greatly benefit by previous studies that have performed similar analysis. For instance, the authors mention the climatological and conditional dust products, which have been introduced for the first time in Marinou et al., (2017) and then applied on Proestakis et al., (2018). No discussion or comparison is presented in the manuscript. Moreover, the separation methodology used in the manuscript has been extensively implemented in the framework of EARLINET (e.g. Tesche et al., 2009, 2011; Ansmann et al., Ansmann et al., 2011). Furthermore, Amiridis et al., (2013) introduced for the first time the depolarization-based separation methodology on CALIPSO. However, there is no reference or discussion on this study as well! Given that all the aforementioned studies are available in the literature, which are the innovative aspects of the present study?
3. Lines 105-109: Please update the information based on the final paper version of Gkikas et al. (2021) in which the MODIS-Aqua Collection 6.1 data, over the period 2003-2017, have been used.
4. Lines 251-264: A short description of the applied techniques for the derivation of DOD is needed, based on MODIS, over continental and marine regions. How much feasible is to discriminate mineral particles from sea-salt over oceans relying only on size parameters? It is not clear to me how you can separate dust from sea-salt over land using a very high single scattering albedo (almost equal to 1; similar to those recorded for sea-salt particles) and ignoring its spectral variation. Moreover, how much reliable the Ångström exponent is above land (see Section 4.4.5 in Levy et al. (2013))? Are you using only Deep Blue retrievals over land? In this case, how do you discriminate dust aerosols from other types when the Dark Target algorithm it is applied?
5. Section 3.1: Since you are using CALIPSO and Aqua retrievals, you can collocate them in order to eliminate the impact of the different sampling between the two satellite sensors which are flying in the A-Train constellation. Taking advantage of the almost coincident observations you can assess the assumptions made in Lines 394 – 407.
6. Trend analysis: I cannot understand why you put so much focus in EAS and NWP without discussing other regions of the planet (e.g. Middle East).
7. Uncertainty analysis: It would be important to present global maps of the DOD uncertainty both for CALIOP and MODIS in order for the reader to better understand how uncertain the obtained DOD averages are.
8. Lines 619-627: I don't agree with this statement. It is true that it is not easy to evaluate DOD retrievals against AERONET because the sun-photometric measurements are representative for the entire atmospheric column. Nevertheless, you can select either sites (even though are few of them) in desert areas (the contribution of other aerosol species is minor or negligible), or to set appropriate coincident thresholds on AOD and Ångström exponents (see for example Basart et al. (2009)) or to rely on almucantar retrievals (Gkikas et al., 2021) or to follow the approach that you are mentioning in your manuscript (Pu and Ginoux, 2018). In any case, an evaluation analysis it is needed in order to support the reliability of the satellite DODs (see also Schuster et al., 2012; Amiridis et al. 2013).

9. Table 1: Are you using the spectral SSAs or only the values at 470 nm?
10. In the manuscript, dust is distinguished from non-dust aerosols based on particle shape information (i.e., the use of particulate depolarization ratio) for CALIOP. However, the particulate depolarization ratio in L2 is too noisy, showing values for dust, dusty marine, polluted dust aerosol subtypes from negative up to 1.0 and above (see figure below).



Moreover, approximately 11% of all dust, dusty marine, polluted dust aerosol subtypes have particulate depolarization ratios < 0.05. Since in the methodology the dust, dusty marine, polluted dust aerosol subtypes are assumed mixtures of dust and non-dust components, how do the authors treat the negative and larger-than-one particulate depolarization cases in their Quality Assurance procedure? Do the authors consider the dusty aerosol mixtures of particulate depolarization ratio lower than 0.05 as non-dust mixtures? Which are the uncertainties introduced in the final dust product by these values? Please quantify.

11. The authors provide a CALIPSO-based dust product, based on the particulate depolarization ratio, applied to L2 backscatter coefficient profiles. Based on the manuscript it is not clear whether the methodology is applied only on the dust, dusty marine, and polluted dust aerosol subtypes, and not at the other types (e.g. elevated smoke, marine, …) at the 60m aerosol layer. Or whether an average over consecutive 60m layers is computed to remove noise. Please provide more in-depth description of the selected methodology. Moreover, which is the effect of the identified aerosol subtype misclassification on the dust product? Many important studies are mentioned by the authors (e.g. Burton et al., 2013), however the effect of the misclassification on the dust product needs discussion and quantification.
12. Based on the methodology, the dust, dusty marine, and polluted dust aerosol mixtures are distinguished into a dust and a non-dust component. Thus, at the end, there are three types of backscatter coefficient: (1) the initial backscatter coefficient of non-dust mixtures (e.g. elevated smoke, …), (2) the dust backscatter coefficient of the separated dust component, and (3) the remaining backscatter coefficient of the separation, the non-dust component. According to my understanding the extinction coefficient of (1) does not change since the methodology is not applied to non-dust mixtures. Regarding the case (2), a uniform global Lidar Ratio (LR) is implemented to calculate the dust extinction coefficient. However, the authors do not discuss the case three (3), regarding the remaining backscatter coefficient of the non-dust component. For the calculation of the non-dust extinction coefficient component, the authors should identify the non-dust aerosol subtype in the dusty aerosol mixture, in order to assign a proper LR. The authors have not provided a detailed explanation. Since the AOD is then computed by the integration of the extinction coefficient profile, the authors should either provide a solid justification of the non-dust aerosol-subtype assignment including quantification the corresponding uncertainties, or to avoid using the new AOD and the corresponding Sections, after the intermediate dust separation.

13. It is not properly discussed, how the averaging extinction coefficient procedure is computed, prior to integration for the DAOD. According to Amiridis et al. (2013) and Tackett et al. (2018), the methodology should follow first a “per-overpass” averaging within a specific grid, and accordingly integration of the mean profile, calculated by all overpasses in the grid. However, the methodology followed by the authors is not clear in this point. Please discuss, and in case a different methodology is provided justify the selected approach or revise accordingly.
14. The manuscript would greatly benefit by introducing tables of the Quality Assurance procedures, applied to both CALIPSO and MODIS, including the corresponding literature related to each filter.
15. What I am missing in the study is a validation intercomparison against ground reference lidar instruments to validate the profiles acquired (e.g. EARLINET/ACTRIS), or even an intercomparison against dust models.
16. The uncertainty analysis is not performed in-depth. Many aspects, such as the effect of non-uniform global Lidar Ratio, the presence of highly polarizing pollen, the presence of volcanic particles or the effect of depolarizing marine particles (in Low RH), the effect of topography and orography (e.g. weighting effects on the mean profiles due to mountains), negative or high positive backscatter values and how they are treated (including references) are not discussed and quantified through a proper error-propagation analysis and an estimation of the uncertainties.