Response to Referee 3

The paper "Characterisation of dust aerosols from ALADIN and CALIOP measurements'' aims to assess the performance of ALADIN by comparing with CALIOP, and uses the synergy of both sensors to improve and increase information over an extreme dust transport episode. The manuscript falls within the scope of the journal. However, the presentation and discussion of the paper is not clear and the submitted study is subject to deficiencies. I would recommend publishing considering the major revisions and addressing the specific comments that follow. Furthermore, the text needs rearrangements, especially in terms of the structure in the presentation of the methodology, figures, and discussion in specific sub-sections (see comments below).

We greatly appreciate the review and detailed comments provided by the referee. Our responses to the specific comments are as follows.

Major comments:

- I don't see any direct evidence presented by the authors that supports the statement that "ALADIN is more susceptible to signal attenuation from CALIOP". I suggest removing that statement from the abstract and elsewhere.

Thank you for pointing out the need for further clarification regarding our statement about ALADIN's susceptibility to signal attenuation compared to CALIOP. This statement is derived from our analysis of the extinction coefficient retrievals from both ALADIN and CALIOP, as illustrated in Fig.6 and 7. We noted that ALADIN often has missing retrievals at the base of dust layers (2.4 – 3.4 km layer), primarily flagged for low SNR, indicating signal attenuation is a key factor. In contrast, CALIOP retrievals appear to be less affected by signal attenuation in the studied area. We trust this clarification further substantiates our statement.

- The authors state that their work paves the way for forthcoming spaceborne HSRL missions, particularly the ESA ATLID space lidar (set for a 2024 launch) and Aeolus-2. How does this paper do that?

We deleted this sentence from the abstract. For clarification, our paper showcases the capabilities of Aeolus as the precursor for the next generation of spaceborne HSRLs, like the forthcoming ATLID and Aeolus-2. While Aeolus's primary goal is wind observation, it has introduced the ability to independently retrieve aerosol backscatter and extinction coefficients from orbit, thereby directly measuring the lidar ratio. Despite Aeolus's limitation in spatial resolution, our work demonstrates that its direct lidar ratio measurements can offer improvements over previous technology, as evidenced by our corrected CALIOP extinction retrievals. By establishing the effectiveness of Aeolus's aerosol observation technique, this work paves the way for future spaceborne HSRL missions.

- In abstract and elsewhere, please mention the limitation of ALADIN/Aeolus on retrieving the total particle backscatter coefficient, otherwise the reader will get into confusion (for example, while the term "co-polar backscatter" is not a perfectly valid one, it could be used to distinguish from total backscatter).

The following sentence in the abstract has been rephrased: "…… ALADIN data can be used to estimate aerosol extinction and co-polar backscatter coefficients separately without an assumption of the lidar ratio". Details of ALADIN's limitation on retrieving the total backscatter coefficient has been discussed using Eq (1) – (3)., and this part has been moved to Section 2.1 in the updated manuscript.

- QC flags: The paper extensively utilizes CALIOP and ALADIN, however pre-processing and Quality-Assurance criteria applied on the data used for the comparisons of backscatter and extinction coefficients are not sufficiently presented nor discussed.

We have revised 'Section 2 Data and Methods' to enhance clarity regarding the QA criteria. In the subsection about Aeolus ALADIN data, we have included the following paragraph:

"The quality control of ALADIN's Level-2 SCAmb products involves several criteria: the validity of extinction and backscatter coefficient retrievals; the backscatter-to-extinction ratio (BER); Mie and Rayleigh SNRs; estimated errors in extinction and backscatter coefficients; and the accumulated optical depth. These criteria are comprehensively detailed in (Flamant et al., 2020b). ALADIN's L2A processing strategy has a high sensitivity to errors so that small errors in extinction propagate from one bin to the next, often leading to negative extinction coefficients. To mitigate this issue, an additional filtering step is used in this study to eliminate negative extinction coefficients."

Regarding the subsection of CALIOP data, we have revised the following paragraph:

In this study, the CALIOP Level-2 V-4.21 aerosol profiles APro (CAL_LID_L2_05kmAPro-Standard-V4-21) are used for comparison against ALADIN aerosol retrievals. The Level-2 APro data include a cloud-aerosol discrimination (CAD) score, which we use as a QC flag, selecting only aerosol retrievals with a CAD score less than -20.

- Spectral conversions: Furthermore, assumptions on 532-355 nm spectral dependencies on depolarization, lidar ratio, extinction and backscatter should be supported (preferably using as reference the paper of Floutsi et al., (2023) where averages for these properties using long-term ground-based lidar measurements are given).

Thanks for suggesting this recent reference. Based on the measurements from multiple experiments provided in this reference, we added an extra row of 355 and 532 nm lidar ratio into Table 1.

- For 4 & 5 (aforementioned comments) to be tackled, I would suggest that the authors extensively elaborate on the datasets and methodology sections, to clearly present the processing chains and assumptions leading to the study conclusions.

We have updated Fig.4 and 5 by adding a colourbar for gradient visualisation, changing the ALADIN mean profile colour to red for enhanced visibility, and including a separate plot on the right margins to display the number of valid retrievals at each altitude.

In response to a detailed elaboration on datasets and methodology, the methodological content initially in the first part of Section 4 has been relocated to Section 2. Additionally, we have restructured 'Section 2 Data and Methods' to enhance clarity regarding the QA criteria. In the subsection of “Aeolus ALADIN aerosol products” and “CALIPSO CALIOP aerosol products”, we have included a detailed paragraph about the QA criteria.

- Collocation: Section 2.3 “Collocation of Aeolus and CALIPSO”: The paper presents and discusses the following collocation criteria: (1) “3◦×3◦ grid, sets the maximum temporal disparity at 9 hours and the maximum spatial difference at 200 km”, (2) “a spatial distance under 1◦ and a temporal discrepancy not exceeding 24 hours, based on data between 30th June 2019 and 28th September 2021”, and (3) “between 30◦ N and 30◦ S, most collocated observations are within 4 hours and 100 km”. It is not clear at all the selected collocation criteria that are applied in the framework of the study. More important is the authors to discuss atmospheric homogeneity in terms of aerosols and clouds. How is it ensured that the two satellite sensors probe the same air masses? For example, the authors could provide a study on spatiotemporal representativeness in terms of the selected and applied criteria including literature - discussion on NA meteorology. Ensuring that the two systems probe the same air masses is fundamental for the follow-up intercomparison, non assessing it makes the outcome conclusions questionable.

We would like to clarify that the collocation criteria of “a spatial distance under 1° and a temporal discrepancy not exceeding 24 hours” originate from the database developed by Feofilov et al. (2022). While this database originally employed a temporal threshold of up to 24 hours, it allows researchers the flexibility to apply more restrictive temporal criteria as needed. In Fig.1, we have adapted this dataset to a narrower temporal constraint of 9 hours, and regridded the collocation data to a 3°×3° grid. This approach aims to provide a comprehensive global overview of CALIOP and ALADIN collocations, assisting researchers in comparing or integrating data from the two lidars. Specifically, our study employed collocation thresholds of 4 hours and 100 km, focusing primarily on the region between 30° N and 30° S. We have revised Section 2.3 by adding the following sentences for clarification:

“…… Although the dataset utilises a temporal disparity of up to 24 hours, it enables researchers to reduce the temporal threshold. Fig.1 is a representation of the global distribution of these collocated profiles when applying a stricter temporal threshold of 9 hours.”

In Fig.1 caption the following sentence has been rephrased:

“……In this plot the dataset is constrained to a temporal disparity of no more than 9 hours, and has been regridded to 3°×3° globally.”

We acknowledge that our analysis did not incorporate a trajectory model like HYSPLIT to ensure that both lidars observe the same air parcel. This decision was influenced by the difficulty of identifying the same atmospheric volume with valid retrievals from both lidars, particularly given the specific constraints of the dust event and time period. Our study focuses on comparing the entire vertical profiles of backscatter and extinction retrievals rather than specific layers, which necessitates certain assumptions when comparing profiles from the two lidars. The 4-hour temporal and 100-km spatial disparity criteria used in this study represent the best possible scenario under these conditions, and have been similarly employed in other studies comparing CALIOP and ALADIN, including research by Dai et al. (2022) and Flament et al. (2021).

- Cloud contamination: The dust transport event examined was extreme, however the extensive presence of clouds may affect the scenes examined. Here, with respect to Aeolus Cloud Filtering, three methods are presented and discussed, the (1) SEVIRI CLM cloud mask, (2) the CM SAF cloud mask, and (3) AEL-FM. However, it is not clear which - if not all of the aforementioned cloud-screening datasets are applied. Please elaborate on this aspect, including discussion of the quality of the applied procedures, assumptions, and uncertainties. With respect to CALIOP, which cloud filtering criteria are applied?

We appreciate your attention to the cloud contamination issue and apologise for the confusion regarding our cloud-screening methods. Initially, we examined both the SEVIRI CLM and CM SAF cloud masks but ultimately did not use them in our analysis due to their poor performance in dust areas. Instead, we employed the SEVIRI dust mask for flagging ALADIN observations: a profile is designated as a dust aerosol observation if 95% of the corresponding resampled footprints are flagged as dust in the relevant SEVIRI data. This decision and methodology have been elaborated upon at the end of Section 2.4:

"In the case studies presented here, the SEVIRI dust mask is used to identify dust-dominated profiles within ALADIN observations. As CALIOP Level-2 APro products already discriminate between aerosol and cloud features, they do not require additional cloud masking."

- In the conclusions section, the authors draw generic conclusions on Aeolus and CALIPSO, however their work is based on a single event, which is an extreme one, over a specific domain, and for a time period not exceeding a few weeks. Thus, the outcomes should not be treated as generic since the statistical study lacks the depth to support the argument.

We have revised sentences in the conclusion section to emphasise that our findings and statements are specific to this extreme dust event.

Specific Comments:

- A CALIPSO-based mean depolarization ratio profile is provided, reporting also mean particulate depolarization of 0.32. However, this depolarization is accompanied by a standard deviation ~ ±0.15 which translates to non-pure dust layers apparent in the atmosphere, resulting in particle depolarization values lower than 0.3. How do the authors treat those layers? Treating them as pure-dust layers and applying dust-related conversion factors contaminate the outcomes, so the authors have to address the aerosol mixtures accordingly. Moreover, please mention the pre-processing chains in terms of particle depolarization ratio profiles leading to the non-noisy profile in Fig.4. How do the authors treat larger than 1 and lower than 0 CALIPSO V4 L2 5km depolarization values?

The correction method accounts for aerosol mixing when calculating the conversion factor, and is specifically applied to layers identified as predominantly dust. In Fig.4(c), the standard deviation in depolarization ratios between 2.5 and 7 km altitude mainly arises from observational variations. Below 2.5 km, the decreasing mean depolarization ratio suggests increased mixing of dust with marine aerosols, which is also reflected in the lower lidar ratios at this altitude in Fig.10.

Consequently, the correction method utilises only ALADIN profiles above 2.4 km. This approach is outlined in the manuscript: "...... CALIOP retrievals use an average lidar ratio of 43.5 sr above 2.5 km — an area less impacted by maritime aerosols and regarded as the dust layer. For ALADIN retrievals, a selective filtering strategy has been implemented, maintaining only data within the 2.4 to 5.8 km altitude range that best characterises the dust layers. Within this particular altitude segment, the mean lidar ratio for dust layers stands at 63.5 sr....". The application of CALIOP extinction retrieval corrections is then confined to only dust aerosol layers as seen in Fig.8.

The following sentences have been added in Section 3.1 to elaborate the filtering of depolarization ratio: " ...... After omitting values below 0 and above 1, the depolarization ratio has an average of 0.32 between altitudes of 2.5 and 7 km".

- You can modify Figures 4a, 4b, 5b, and, 5d, to linear instead of logarithmic x-axis scales, add a colorbar for the gradient values, and a second axis reporting on the sample size of profiles resulting in the mean profiles.

We have now included colourbars to represent gradient density values in Figures 4(a), 4(b), 5(a), and 5(b). Additionally, the ALADIN mean profile colour has been changed to red to enhance visibility. The count of valid retrievals contributing to each mean profile has been added to the right margin of each subplot.

In response to your suggestion about the axis scales, we tested with presenting the backscatter coefficients on a linear scale as shown in Fig. S1. However, a linear representation can significantly compromise the visibility of lower backscatter coefficients as they span several orders of magnitude. For this reason, we have decided to retain the logarithmic scale for the figures in our manuscript, as it more effectively displays the full range of data, particularly the comparisons of smaller coefficients. This is consistent with the expectation that particle concentration decreases exponentially with height above the boundary layer.

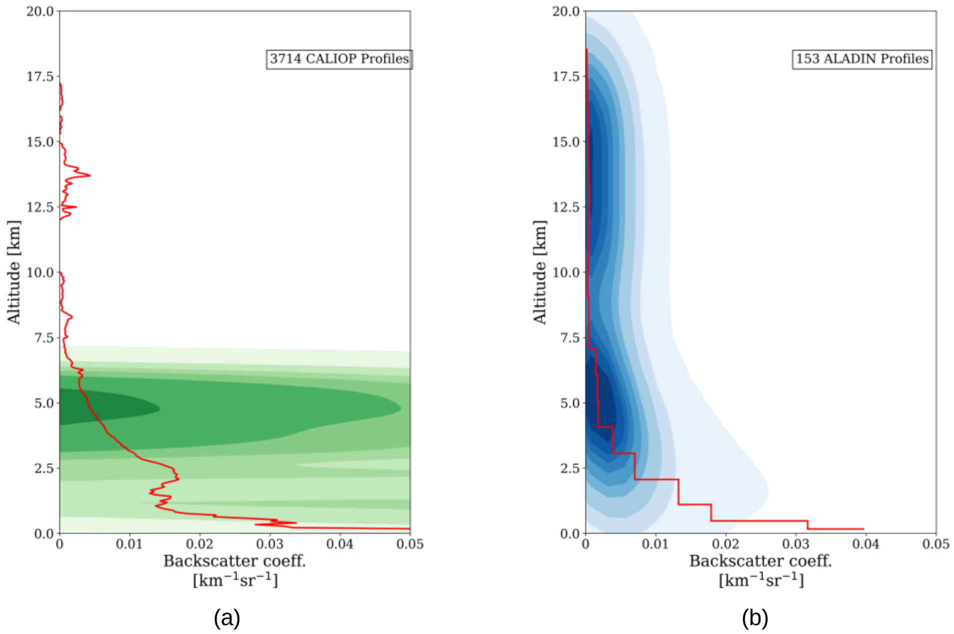

Figure S1. CALIOP (a) and ALADIN (b) backscatter coefficients in linear scale.

- Results presented in Figures 4a, 4b, 5b, and, 5d: you may provide statistical metrics reporting on the intercomparison of backscatter and extinction coefficient profiles (e.g., σ, r2, mean/relative biases, …). Prior to doing this analysis the authors should elaborate how they get CALIPSO to the same horizontal and vertical resolution to Aeolus.

The following statistical analysis has been added to support the discussion of Fig.4:

"...... In general, CALIOP and ALADIN show good consistency in detecting dust aerosols, with evidence of dust being uplifted to 7 km. Within the main aerosol layer from 1.5 to 7.5 km in altitude, the mean backscatter coefficients retrieved by CALIOP and ALADIN show a strong correlation, with an R-square (R2) of 0.967. At ∼3.5 km, the altitude with the most valid retrievals, ALADIN's retrieved backscatter coefficient averages 0.004 $km^{-1}sr^{-1}$. CALIOP, which offers a higher vertical resolution, has an average backscatter coefficient of 0.01 $km^{-1}sr^{-1}$ when adjusted to match ALADIN's vertical resolution."

The following statistical analysis has been added to support the discussion of Fig.5:

"The two instruments generally show a good agreement in their extinction coefficients within the dust layer, with an R2 value of 0.992 for mean extinction retrievals between 1.5 and 7.5 km altitude. However, some disparities are also apparent. For instance, at the altitude of ∼3.5 km, ALADIN has an extinction coefficient of 0.057 km^{-1} while CALIOP has an extinction coefficient of 0.046 km^{-1}."

- You can apply linear scales also to figures 6(b,c,e,f)and 7(b,c,e,f).

Thank you for your suggestion regarding Figures 6 and 7. As with our tests in Fig.4, we have similar concern in using linear scales for these figures. The extinction coefficients in the case studies of Figures 6 and 7 exhibit significant variations, due to the extensive spatial coverage (over 16 degrees in latitude) and the necessity of maintaining consistent scale ranges across different layers in the subplots. These factors make the logarithmic scale more suitable for effectively representing the data. We hope this explanation clarifies our decision to retain the logarithmic scale in these figures.

- Line 198 and Figure 3: According to the authors the method of Ashpole and Washington (2012) is applied. Since this is a crucial section, please provide discussion on the method, assumption, performance, and uncertainties. Since CALIPSO is used, which is the reason for not implementing CALIPSO aerosol subtype classification as dust identified aerosol layers?

We have expanded the following discussion on the dust flagging method: "This dust flagging method utilises the infrared channels of SEVIRI for the detection of dust events, proving to be effective in consistently identifying moderate to heavy dust outbreaks across the central and western Sahara". The performance and uncertainties of this method are comprehensively detailed in Ashpole and Washington (2012) and we note that this request contradicts that of Reviewer 2 to minimise the discussion of cloud flagging methods.

Regarding CALIPSO's aerosol classification, we refrained from using the dust subtype products due to the known issue of misclassification among various aerosol subtypes (Chen at al., 2010).

- Lines 242-248: This sentence actually is generic to a degree that it doesn't provide any information, since none of the aforementioned source of discrepancies is assessed and no effort in quantifying the effect of each factor is provided in the manuscript. Please elaborate more on this.

The following statistical analysis has been added to support the discussion of Fig.4:
"...... In general, CALIOP and ALADIN show good consistency in detecting dust aerosols, with evidence of dust being uplifted to 7 km. Within the main aerosol layer from 1.5 to 7.5 km in altitude, the mean backscatter coefficients retrieved by CALIOP and ALADIN show a strong correlation, with an R-square (R2) of 0.967. At ~3.5 km, the altitude with the most valid retrievals, ALADIN's retrieved backscatter coefficient averages 0.004 km−1sr−1. CALIOP, which offers a higher vertical resolution, has an average backscatter coefficient of 0.01 km−1sr−1 when adjusted to match ALADIN's vertical resolution."

- Please describe clearly in Section 6 how you apply corrections to CALIPSO and provide the formulas.

We have updated the relevant paragraph to articulate the correction process more clearly:
"The extinction coefficient $\alpha_{532(corr)}$ is then corrected by multiplying it with $LR_{updated}/LR_{CALIOP}$, where $LR_{updated}$ is set to 63.5 sr and LR_{CALIOP} is derived from each individual CALIOP profile. This scaling method is an approximation, as varying the lidar ratio can influence the lidar profile by impacting the backscatter retrieval during the Klett inversion process. This alteration in backscatter retrieval, in turn, affects the subsequent extinction retrieval.

- Please take care of the units in the manuscript, in some places they are missing.

We have checked the units throughout the manuscript.