We thank the reviewers for the careful reading and the suggestions that helped us improve our manuscript. The answers to the reviewers will be given after each comment in bold font.

Reply to Referee #2

general comments

The manuscript entitled „Characterization of aerosol over the Eastern Mediterranean by polarization sensitive Raman lidar measurements during A-LIFE – aerosol type classification and type separation“ by Groß et al. promises an important contribution in the field of aerosol typing from different measurement techniques. Nevertheless, I feel slightly disappointed after reading the manuscript for two main reasons.

1. It seems like parts of the manuscript have been written by different people. Some sections are well written, others are quite sloppy in usage of language and presentation of figures.

We carefully revised the manuscript to hopefully achieved to better homogenize the different parts of the manuscript and improve insufficient usage of language and presentation.

2. The abstract describes the goal of the A-LIFE campaign to “characterize dust in complex mixtures”. But a detailed discussion of aerosol mixtures is missing in the paper. Instead the case studies focus -again - on situations with pure aerosol types and the discussion of derived aerosol types and dust fractions is not comprehensive enough.

We believe that showing this pure aerosol cases is important for the significance of the investigation of mixtures later in this paper. We want to make sure – and show – that optical properties different from the know values for pure cases are really due to mixture and not due to transport or aging. Thus, we prefer to keep the case studies as they are. However, we agree, that the discussion of the derived aerosol types and the dust fraction could be extended and improved. We revised these parts of the manuscript.

Some important points for improvement

Section 2.2:

It is unclear how the extinction coefficients from day-time measurements (presented in Fig. 7) were derived. Please, elaborate in more detail. Furthermore, concerning the retrieved uncertainties, it is insufficient to just provide a reference to previous work. Especially since the captions of Figures 4-6 claim that all error bars only show systematic uncertainties, however the references clearly describe the handling of statistical errors as well. Please, add some sentences like “systematic errors of lidar ratio include the uncertainties of backscatter calibration, …, ”

For the day-time measurements the extinction coefficient is retrieved with the Fernald-Klett algorithm, as described in the Methodology section. As described, this algorithm uses the lidar ratio as an input.

We did not include the statistical errors as those are only available for the Raman measurements and were not included in the references for the depolarization measurements. We would have to come up with a new method of error propagation for the statistical error of the depolarization measurements. Furthermore, the statistical uncertainty strongly depends on the averaging length. However, we included a short description of the what is included in the systematic errors.

Sections 2.3 – 2.5:

The reader may get the impression that data of both lidar systems have been analyzed with the same software tools, and/or by the same experts. Please, elaborate.

We modified the description of the systems to make clear that this is not the case.

Section 2.6:

Please, provide technical details (e.g. thresholds) of the aerosol typing here, instead of in section 3.5.1.

We moved the technical details from section 3.5.1 to the Methodology section and extended the description of the aerosol typing (including thresholds).

Section 2.7:

Since the publication by Weinzierl et al. is not yet published, the reader needs more details here, not just a reference. Such details should entail: what exactly are the derived aerosol type "mixtures?" What components do those mixtures contain? It is also entirely unclear, what a (moderately) polluted mixture with or without coarse mode might be. Please, provide descriptive examples like "sea salt mixed with smoke". To that point, it is unclear whether the in-situ classification scheme depends on FLEXPART input or not. If it is not independent, why provide the comparison in Table 2?

We expected the publication by Weinzierl et al. to be published yet. As this is not the case, we removed the reference to this publication. However, a description of the method and also of the complex mixtures of dust and pollution is given by Teri et al., 2022 and 2024. Both publications are cited in our study. A detailed evaluation of mixtures and how that effects the in-situ measurements and optical properties is given in Teri et al., 2024 (doi.org/10.5194/egusphere-2024-701). It is beyond the scope of this study to repeat it here. However, we included some further information of the in-situ measurements for days classified as mixtures from the in-situ measurements and as pure from the lidar measurements in a table in the supplement.

Section 2.8:

Is it possible to provide thresholds for the aerosol typing here?

We just used the pre-classification. For details on how the typing is actually done we have to refer to the publications describing FLEXPART.

Section 2.9:

If the highly sophisticated FLEXPART simulations were available for the campaign, why is it necessary to work with basic HYSPLIT trajectories, too? Please, elaborate why FLEXPART simulations were not used for analysis of the lidar measurements.

The FLEXPART simulations were run for the aircraft measurements (flight track and altitude). Thus, they were not available for the full lidar profile. As it was not the intention of this study to include a full investigation of the source region and transport of the aerosols we used HYSPLIT to highlight, that the airmasses for the different case studies were indeed advected from different regions.

Section 3.1:

It is mentioned here that smoke aerosols have been present during the campaign. But section 2.4 does not describe the handling of smoke in the aerosol typing procedures, please elaborate.

Smoke and pollution cannot be distinguished from the remote sensing instrumentation as they are quite similar in the measured optical properties. That is why we referring to pollution/smoke throughout the text. As that, it is included in Section 2.4 and after a revision of the aerosol typing from AERONET also in section 2.7

Figure 1:

Change minor ticks in a way that each interval corresponds to one day. Indicate Falcon overpass times and altitudes by symbols and times of lidar case studies by vertical lines

We changed that.

Figure 2:

It is difficult to visually separate the different symbols and colors. In addition, grid lines would be appreciated. Further, the time series of AOD at 340 and 1020 nm could be omitted as the wavelength dependency is already illustrated by angstrom exponents.

The reviewer is right, that the wavelength dependency is already illustrated by the angstrom exponent. Still, we want to keep the information in this plot. For better visualization we changed the color of the symbols.

Section 3.2:

The introduction of this section claims that the presentation of three case studies with pure aerosol conditions is important for later studies of aerosol mixtures. However, section 1 mentions the availability of many previous field campaigns for studying pure aerosol conditions and the importance of the A-LIFE campaign for investigating mixed conditions. It is very important to add at least one extended case study with mixed conditions. The extended presentation of the mixed case should also include profiles of aerosol type and dust fraction as well as data of non-lidar measurements (AERONET, FALCON). In order to keep the manuscript short, case studies of 9 and 21 April are unnecessary.

We do not want to include other case studies or remove case studies presented in the current version of the manuscript. We have chosen these case studies to ensure, that optical properties

different from the know values for pure cases are really due to mixture and not due to transport or aging. A more detailed evaluation of the mixtures would be beyond the scope of this study. A detailed analysis of the in-situ measurements can be found by Teri et al., 2024.

Figures 4-6:

Adding a label (a), b), c), d)) to the different panels would be more reader-friendly. Additional panels with profiles of angstrom exponents and dust fraction need to be added. These will allow for a better comparison with AERONET data and provide a better connection with Table 2. The Quicklook could be as a top row panel above the profile panels.

We believe that the figures are quite clear to understand and thus do not need any labeling. We also don't want to include more information like the dust fraction, as we investigate rather pure cases here. The profile of the dust fraction would not provide additional benefit. As the case studies are done for night-time measurements they are not directly comparable with the daytime AERONET measurements. Thus, giving the Angström Exponent would not allow for a better comparison. It furthermore can be derived from the extinction coefficient of both wavelengths.

It is very advanced that the authors present systematic uncertainties of all profiles. Nevertheless, there must be statistical uncertainties, too. As such they need to be included in the plot.

We did not include the statistical errors as those are only available for the Raman measurements and were not included in the references for the depolarization measurements. We would have to come up with a new method of error propagation for the statistical error of the depolarization measurements.

Section 3.4:

Please reconsider whether it is really the altitude /signal-to-noise ratio which limits the retrieval of PLDR. In most cases, it is a too low aerosol amount making the retrieval of PLDR mathematically unstable. Thusly, it depends on atmospheric situation, not measurement setup whether PLDR can be retrieved or not.

Actually, it is the small signal to noise ratio that limits the retrieval of the PLDR from the measured VLDR, but it is also correct that the smaller the backscatter ratio is, the higher the SNR must be to retrieve the PLDR with a given relative uncertainty. We include this in the sentence as:

"*As the signal to noise ratio was too small to retrieve the PLDR with sufficient accuracy for overpasses at flight altitudes >7 km with low backscatter ratios, we restrict our evaluation to the extinction coefficient in those cases.*"

Further, it is unclear from which of the two lidar systems the reported values and findings originate. It can be assumed that this data come from POLIS system. Why are no findings from PollyXT reported in this section?

We included in the text, that the lidar values were measured by POLIS. We also investigated if there are differences between the different lidar systems but found good agreement within the uncertainty ranges. For a better visualization of the comparison we decided to not include

the POLLY^{XT} data in Figure 7. But we extended the text to give the information about POLIS and POLLY^{XT} agreement.

Again, there is the need for further explanation on how the extinction coefficients have been obtained from daytime measurements.

For the day-time measurements the extinction coefficient is retrieved with the Fernald-Klett algorithm, as described in the Methodology section.

Section 3.5:

The introduction paragraph of this section is only a replication of previous statements.

We removed part of this introduction paragraph.

Section 3.5.1:

The description of the typing algorithm should go to section 2. Thresholds used for the typing should be printed as lines in Figure 8. It would also be helpful for the reader if the individual points are color-coded by type. The highlighted daily mean values have no real benefit because the change between atmospheric situations does not follow calendar days. It would be better to highlight the measurements during FALCON flights which were used for the typing in Table 2.

We moved the description of the typing to section 2 and extended the description. We do not want to include lines in Figure 8 to highlight the thresholds but color-coded the different types. We kept the daily mean values for the measurements used in Table 2.

Section 3.5.2:

The description of aerosol typing from lidar data is in conflict with the description in section 2.4. It rather seems that the typing has been obtained from ancillary data (trajectories) and not from optical properties. How else can it be explained that the new type “Arabian dust” was identified in this study, but was not known in previous studies?

For the identification of the Arabian dust cases ancillary data (trajectories, source investigation, satellite observations) were used. The optical properties were than investigated in detail, as also presented in a case study and in the following included into the aerosol typing using the lidar ratio difference to distinguish between Saharan dust and Arabian dust. Furthermore, we included also the investigations of former studies to come up with a threshold to distinguish between Saharan and Arabian dust. We included the following text to make that clearer.

‘*Floutsi et al. (2023) already included the separation between Saharan and Arabian dust. But they rather did a data collection than a classification. HETEAC (Wandinger et al., 2023) was made more flexible to be applied to multiwavelength observations. The resulting HETEAC-Flex (Floutsi et al., 2024) includes optical properties for Saharan dust separate from Arabian dust. We could check how well their optical properties are in line with the A-LIFE observations.*’

Table 1 and Figure 9 needs a much more detailed discussion. Why are there missing values in Table 1? What are the measurement times? It would also be helpful for the comparison, if aerosol types and dust fractions derived from both lidars would be included in the table and their differences discussed.

Missing values occur when no measurements were available at the specific times. We added this clarification in the figure caption. We also added the measurement time for which the values have been retrieved. But we do not want to include dust fractions in this table, as for our understanding it would not provide any benefit for further aerosol classification.

Figure 9:

The difference between “dust mixture” and “polluted dust” needs to be explained in this paper. References to previous works is not sufficient here.

To be more precise here, we added that ‘dust mixtures’ refer to mixtures of dust with marine.

Some individual data points require more discussion. There is one “polluted marine” (25 April?) which is clearly located within the pollution cluster and one “dust mixture” (20 April?) between Saharan and Arabian dust clusters. How can those data points be so far outside their clusters if the optical properties have been used for typing?

We had a typo in running our classification. We corrected for that.

Section 4.1 and Table 2:

First of all, a translation table explaining all the different terms for aerosol types is dearly missing in this paper. Terms are different from instrument to instrument, but also from section to section (e.g. lidar “mixed pollution” and AERONET “polluted mixture” were not mentioned anywhere before Table 2). Those inaccuracies can be solved by a careful editing of the text.

The reviewer is right, there are inaccuracies in the description of the AERONET based classification; especially with respect to the categories. We revised the AERONET classification part to be closer to the method proposed by Toledano et al. With that we reduced the number of categories to marine (not available in this study), dust, dust mixture, polluted mixture and pollution/smoke. We revised the description of the AERONET classification to be clearer.

‘For the AERONET based aerosol typing we use the scatter plot of Ångström exponent (440-870 nm) vs. AOD at 500 nm as proposed by Toledano et al. (2009, 2011). Values of the Ångström exponent (AE) larger than ~1.2 serve as indication for smoke/pollution, independent of the AOD. Ångström exponents of <0.5 serve as indication for dust (Toledano et al., 2009, 2011, 2019) or marine aerosols. Following Toledano et al., 2011, a threshold of AOD=0.15 is used to separate dust and marine aerosols. Measurement points with AOD<0.15 and AE<0.5 are classified as marine, while measurement points with AOD>0.15 and AE<0.5 are classified as dust. Ångström exponents between 0.5 and 1.2 serve as indication for mixtures. We further subdivide this value range in dust mixtures for AE larger 0.5 a value of the AERONET derived Fine Mode Fraction larger 0.5 (AE values of ~0.8). Values with Fine Mode Fraction < 0.5 (AE values >~0.8) and AE values smaller 1.2 are classified as polluted mixture.’

The quality and usefulness of this paper could significantly be improved by a discussion on how typing from one method (e.g. lidar) could be compared to typing from another method (e.g. in-situ). Obviously, such comparisons are difficult because instruments measure different quantities (optical properties vs. size distribution). Nevertheless, if a direct “translation” is impossible, the authors should at least discuss the difficulties.

Thank you for pointing this out. It is indeed difficult to directly translate the typing of from one method to another method. We mention this in the text including the following paragraph:

‘The comparison of the different aerosol typing schemes highlight that, although the dominating aerosol type is captured quite well, it is hard to directly compare the outcome in detail. The different schemes rely on different measured quantities (e.g. optical properties vs. size distribution and microphysical properties). Thus, also the results of the aerosol classification schemes can provide a different degree of detail. It is important to carefully investigate if the chosen method provides sufficient information for specific studies they are used.’

In general, the discussion of Table 2 is way too short. This table contains the main findings of the paper (which has “aerosol type classification” in the title) and deserves more attention. Especially since it would be interesting to learn more about the days with complex situations (5,6,11 April as mentioned in the text). Why not present one of those complex cases in the case study section?

We extended the discussion of Table 2, as the reviewer is right, this is one of the main findings of this paper. Except for the 5 April we did not include these days as a case study, but we included a table in the supplement presenting information on the in-situ measured data of these complex days. For a more detailed analysis we refer to the publication by Teri et al. 2024 (doi.org/10.5194/egusphere-2024-701) investigating in depth the in-situ measurements and the contribution of pollution in dust mixtures.

Section 4.2:

The introduction of in-situ total mass concentrations as a “referee” after discussing the differences between lidar and FLEXPART retrievals (lines 450-452) seems odd.

We agree and changed that.

It would be helpful to add lidar derived total mass concentrations (not only dust component) to Figure 10 for a better comparison with in-situ total mass concentration. Uncertainties due to omitting the non-dust fraction in mass estimates should at least be discussed.

We looked at the comparison of the lidar derived total mass concentration and the in-situ derived total mass concentration. However, we found not a significant improvement of the agreement. This is due to the different volume sampled by lidar and in-situ but also due to the different assumptions considered in calculating the total mass concentration. Another reason is, that the optical properties during the dust dominated events are mainly determined by dust, while the in-situ measurements also have a good characterization of the minor contributing aerosols to be included in the calculation. We expanded like the following:

‘The total mass concentration even exceeds the lidar derived dust mass concentration for most of those days, especially when the dust mass concentration was large. Considering also the non-dust contribution in the comparison does not result in a significant improvement of the comparison. The lidar derived optical properties during dust dominating day are mainly

determined by dust aerosol, while the in-situ measurements better characterize the minor contributing aerosol components which are included in the in-situ derived total mass concentration. Further differences occur due to the different averaging time of the lidar and the in-situ measurements and thus the sampled volume.'

Would it be possible to estimate a dust fraction from in-situ data? The subclasses “pure”, “moderately-polluted” and “polluted” should at least allow for a rough estimate.

It might be possible to do so, but it is not in the scope of this study. It might be done in a follow on in-situ based study.

Why are no Polly^{XT} results included in Figure 10? Please elaborate.

POLIS and POLLYXT data were analyzed and provided by different groups. The main lidar system used in this study is the POLIS lidar. PollyXT is used to cross-check the values of POLIS. The measurements and analysis of the two systems do not show significant differences. As we would use the same method to retrieve the dust fraction and dust mass concentration we do not expect a benefit of including the PollyXT data in Figure 10.

Minor points:

L41: add comma after properties

Done

L53: add comma after step. A next step -> the next step

Done

L 179: the sentence has 2 times “obtained”

Changed

Figure 3: the 24h symbols are not visible.

We revised the Figure to make the 24h symbols more visible.

L243-244: strange wording to start a section.

Changed

L 246: lidar measurement show -> the lidar measurement shows

Corrected

Figure 7: the lower panel is too busy. It would be better to present the dust/ non-dust extinction coefficients in a separate panel. Again, statistical uncertainties are missing.

We followed the advice of Reviewer 1 and removed the dust/non-dust extinction coefficients and just kept the dust backscatter and volume fraction.

Table 2: again, what are statistical errors?

The statistical error strongly depends on the averaging length. Thus, we do not think, that the values would benefit from giving them here. However, we choose the averaging length in a way that the standard deviation was small and well below the statistical uncertainties.

Figure 9: What are the error bars? The symbols in the legend are very small and hard to see. The legend should list all aerosol types. It is confusing to have some of them listed in the legend, others in the caption. The plots would be easier to read without aerosol types which are not used in this study (like volcanic ash). This would allow to plot the existing pollution points in gray, allowing the pollution data points of this study to be plotted in black for better contrast, especially in the left panel. Another option for clarification would be not to show all individual previous data points, but only cluster boundaries as polygon lines.

The error bars show the mean systematic error. We added that in the figure caption. We furthermore increased the legend for better visibility. However, we do not want to change the figure in a way to remove former data points.