We would like to thank the Reviewer for his/her thorough and detailed review as well as for the suggested papers. Our replies (regular font) for each comment (bold font) are provided below.

Reviewer #1

The authors assess the particle backscatter coefficient profiles from Aeolus/ ALADIN using co-located ground-based lidars at three locations in Greece, together with auxiliary model and satellite datasets. They attempt to attribute discrepancies between space borne and ground based lidars to (i) the natural variability of aerosols, (ii) instrument or retrieval limitations and/or (iii) spatial temporal co-location issues. This paper needs substantial improvements before it is worthy of publication in AMT.

Major comments

In a very general sense, numerous sentences throughout the paper are too wordy and their structure just too complicated. For example, "are utilized towards an optimum characterization of the probed atmospheric conditions under the absence of a classification scheme in Aeolus profiles" could be replaced by something along the lines of "are used to characterize the atmosphere as Aeolus/ALADIN does not provide an atmospheric classification product". This makes for a lengthy paper. Other examples are "obtained results" instead of simply writing "results" or "probed Atmosphere" instead of "atmosphere" throughout the paper.

We would like to thank the reviewer for his/her suggestions. In the revised manuscript, we have tried to reduce lengthy sentences and to "simplify" them thus improving the readability of the manuscript.

We recommend that the authors:

Use ALADIN in all caps or Aeolus/ALADIN consistently (instead of small caps and/or interchangeably using both names). Also, they should spell it out once when first introduced (i.e., "Atmospheric LAser Doppler Instrument").

Done.

Add a table describing the lidar(s) at each of the three locations. This table could contain, for example, the lidar's name, a small description, its limitations and uncertainties, its products. It could also contain information on how cloud screening is performed and possibly a dominant type of aerosols at each location (together with references related to these stations).

We believe that it is already provided the information requested by the reviewer. We would like to remind that in the submitted manuscript, there is a dedicated section discussing all the necessary technical aspects of each lidar system, the possible instrument deficiencies, the SCC automatic processing chain for lidar data (according to EARLINET) as well as the inversion methods applied during daytime and nighttime.

Add a brief description of the two models used in this study and, especially, their limitations. The way the analysis is written might sometimes give the impression that model results are considered as accurate as observations.

We would like to avoid extending the length of the manuscript by adding information for the CAMS and MERRA-2 since are the two most well-known reanalysis datasets widely used in aerosol (and atmospheric in general) studies. Nevertheless, we agree with the reviewer that our description in the submitted manuscript was not pretty clear and maybe there was a confusion regarding the way the reanalysis products are used in our study. In the revised text, we are clarifying that the aforementioned data are used just as an indicator of the aerosol load in the surrounding area of the station synergistically with the AERONET and Polly^{XT} observations, which always serve as the "truth". Below we are providing the relevant part in Section 7 in the revised manuscript.

"*Numerical outputs from reanalysis datasets (e.g., MERRA-2, CAMS) can be utilized as an indicator of aerosols' burden horizontal variation, taking advantage of their complete spatial coverage, their availability at high temporal frequency and their reliability in terms of total AOD (Innes et al., 2019; Gueymard and Yang, 2020). Nevertheless, such data are better to be utilized in a qualitative rather, particularly in terms of aerosol species, than a quantitative way since they cannot be superior of actual aerosol observations.*"

Add more text comparing their results to those in studies such as Baars et al. (2021) and Abril-Gago et al. (2022). In a more general sense, it would be helpful to open the paper by describing in more detail what their study brings to the table/how it complements other studies.

We have modified this part of the manuscript as suggested by the reviewer. Below is given the revised text

"*Abril-Gago et al. (2022) performed a statistical validation versus ground-based observations from three Iberian ACTRIS/EARLINET lidar stations affected mainly by dust and continental/anthropogenic aerosols. In their Cal/Val study, they processed AERONET optical properties related to particles' size and nature along with HYSPLIT air-mass backtrajectories towards characterizing the prevailing aerosol conditions.*"

Specify the way they average the MODIS AOD and add more analysis to their spatial characterization of aerosols at the three different stations. We believe that there is more to the characterization of aerosol spatial variation than simply averaging the AOD in different boxes. The authors should refer to previous studies such as Anderson et al., (2003), Sayer and Knobelpiesse (2019) or Shinozuka and Redemann (2011).

In the revised manuscript, we have clarified better the MODIS AOD averaging and have extended the analysis discussing the coefficient of variation and the spatial autocorrelation.

Discuss the limitations of their cloud screening of the Aeolus/ALADIN aerosol backscatter profiles using SEVIRI.

According to the product user guide, the quality of the CLM (Cloud Mask) is impacted by the following limitations:

1. The ECMWF temperature and humidity fields are not interpolated in time and space. This means that for all pixels within each segment the same temperature and humidity profile for the period for which the forecast has same validity time. This may lead to artificial straight lines in the display of the product
2. Within the MTP MPEF (Meteorological Product Extraction Facility) algorithms, a 10 level ECMWF temperature profile is used for the determination of e.g. the atmospheric correction, impacting the cloud detection especially in the lower troposphere

We have added a short description in the revised manuscript regarding the limitations of the SEVIRI CLM product. Please note that we are not mentioning the Meteosat-7 calibration bias with respect to MSG or IASI calibrations since these data are not relevant for our study period.

Discuss whether the Aeolus/ALADIN, with its 87 km horizontal resolution, is in fact able to characterize the aerosol natural variability at the three locations.

For the AOD spatial variability in the surrounding area of Antikythera, we are using the outputs from the MERRA-2 and CAMS reanalyses. Both products are derived at coarse spatial resolution (please see relevant plots in the supplementary material). This makes quite difficult to investigate the variability within the Aeolus BRC. An alternative approach would be to exploit MODIS L2 AODs, which are derived at fine spatial resolution

(10 km x 10 km). However, MODIS-Aqua crosses the region of interest once per day around 11-12 UTC and there is a non-negligible temporal departure with the Aeolus overpass time (both for dawn and dusk orbits). Therefore, since we don't have sufficient information (data available at fine spatial and temporal resolution would be suitable) we are not able to address adequately the comment raised by the reviewer.

Avoid strong statements such as "very good performance" when it pertains to a specific altitude, one case study and no quantification of the differences between space borne and ground based lidar in that case.

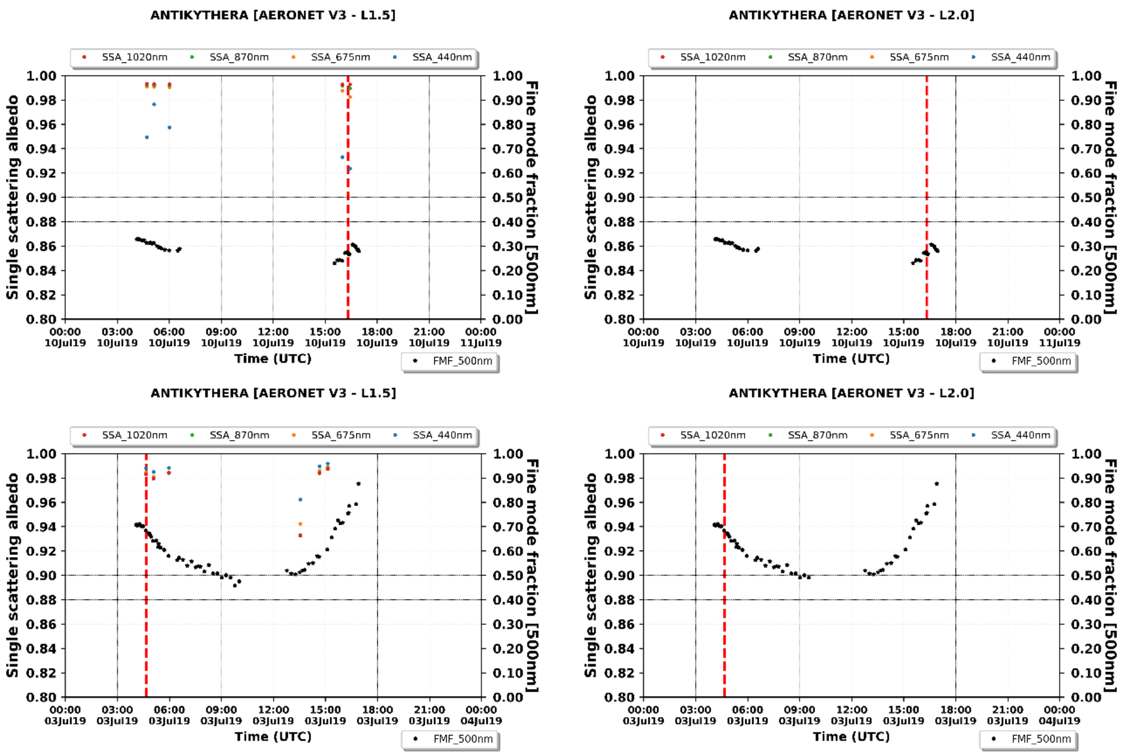
We have modified accordingly the text.

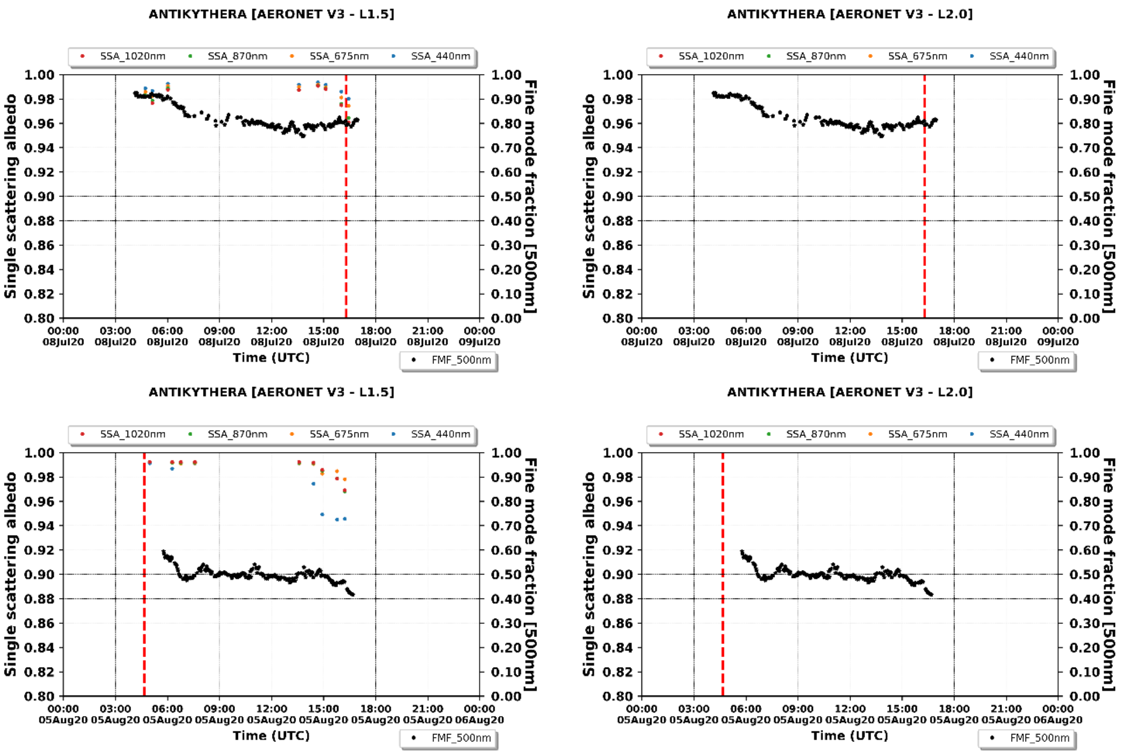
Use additional satellite derived aerosol information (e.g., CALIOP, TropOMI) to further characterize the aerosol during their case studies.

We haven't found coincident Aeolus-CALIPSO overpasses for the selected study cases. It is in our plans to incorporate in the future Aerosol Index and Aerosol Layer Height observations from TROPOMI.

Specify why they use AERONET level 1.5 instead of more accurate level 2 data (quality assured). Also, they should show AERONET-derived FMF and SSA at different wavelengths throughout the day (in addition to what they already do -- spectral AOD and angstrom exponent). This would be like Figure 8 in Abril-Gago et al. (2022). Let us remind the authors that, in addition to a size difference, a difference in SSA at two wavelengths from AERONET could point to the presence of dust versus smoke (e.g., Russel et al., 2014; Kacenelenbogen et al., 2022).

We would like to thank the reviewer for his/her comment. In our analysis we are using AERONET Level 2.0 data but there was a typo both in the manuscript as well as in the relevant plots in the supplement. This has been corrected in the revised manuscript and in the supplement. Regarding the SSA we are aware for its spectral dependence under dust/smoke conditions as it is shown in the studies mentioned by the reviewer as well as in Gilles et al. (2012). However, this was not feasible in our analysis due to the very limited availability of SSA retrievals both for Level 1.5 and 2.0 as it shown in plots below.





Therefore, in the revised supplement we are including only the FMF diurnal variation for each case and we have added the text accordingly.

Specify why they show Aeolus/ALADIN profiles that are cloud contaminated (or "unfiltered") in their analysis. It is not obvious why there would not be more disagreement between cloud-unfiltered space-borne and cloud-filtered ground-based profiles compared to cloud-filtered space and ground profiles. Or is this a way to test their SEVIRI-based cloud filtering method?

Under the presence of clouds, the backscatter coefficient is significantly higher than those when aerosols are recorded. Moreover, there are differences between the cloud and aerosol layers' structure. Based on these facts, we are expecting and we see differences between the cloud unfiltered and filtered profiles. It is well-known that in the raw L2A profiles there is not a discrimination between aerosols and clouds. Therefore, presenting the results from the comparison of the unfiltered Aeolus profiles against those acquired by the ground-based lidars, we are highlighting that this deficiency is critical and can lead in erroneous conclusions. This is further stretched when we are contrasting the results shown in Figures 4 (unfiltered) and 5 (filtered). At a first level, our intention is to highlight the necessity of including ancillary cloud information (SEVIRI in our case) for an appropriate assessment of Aeolus profiles. However, the most important is to highlight the imperative need to deploy a cross-channel, in a possible Aeolus follow on mission, that will facilitate a feature classification scheme, similarly done in the CALIPSO observations. We believe, that the above-mentioned aspects are well stated in our manuscript.

Figures should be called in the order they appear.

Done.

Shorten the conclusion.

We have reduced the conclusions.

<u>Detailed comments:</u>

Title: The authors might want to add "ALADIN" and "aerosol" in the title for increased searchability

We prefer to keep the original title

Line 29: Why not give examples after “a variety of aerosol species.”

We think that it is better to focus on the main aspects of the study. Moreover, in the introduction there is paragraph discussing all the aerosol types encountered over the broader Greek area.

Line 32: Why is PANACEA spelled out but not AERONET, CAMS, MERRA-2 etc...? We recommend either spelling none or all of them.

Because PANACEA, which is the Greek National RI, is not well-known to the community in contrast to AERONET, CAMS and MERRA-2.

Line 33: we recommend writing “sunphotometry observations ”... “model reanalysis” ...” modeled air mass back trajectories”

We think that it is already well stated in the submitted manuscript.

Line 36: Again, multiple sentences throughout the paper are too wordy. For example, “are utilized towards an optimum characterization of the probed atmospheric conditions under the absence of a classification scheme in Aeolus profiles” could be replaced by something along the lines of “are used to characterize the atmosphere as Aeolus/ALADIN does not provide an atmospheric classification product”.

We believe that the sentence is well written and its meaning is pretty clear.

Line 40: “very good” is too strong a statement here.

We have replaced with “good”.

Line 44: We recommend writing “46 identified cases when using [this time frame] at all three stations...”.

We prefer to keep our initial version.

Line 47: “positive tendency” could be replaced by “improvement” and “both Aeolus vertical scales” by “multiple Aeolus vertical scales”

We agree with the first part of the reviewer’s comment. The word “multiple” is misleading since there are only two Aeolus vertical scales.

Line 48: we recommend to replace “justified” by “explained” + “in the vertical the Aeolus performance”

We believe that the use of the word “justified” is appropriate and fits well in the sentence.

Line 49: “performance decreases” followed by the explanation for that decrease is not clear.

We have modified the sentence as follows:

“*In vertical, Aeolus performance downgrades at the lowermost bins due to either the contamination from surface signals or the increased noise levels for the Aeolus retrievals and to the overlap issues for the ground-based profiles.*”

Line 83: We recommend “Such observations are provided by networks... or by dedicated experimental airborne (Ansmann et al., 2011; Weinzierl et al., 2016) or shipborne campaigns (Bohlmann et al., 2018)”.

We would like to thank the reviewer for his/her suggestion but we prefer to keep the sentence in its current form.

Line 87: We recommend “characterization of aerosol vertical structure at global (e.g., Liu et al., 2008) ... was performed using CALIOP ... and CATS... respectively on the CALIPSO and the ISS...”. CATS could use other references such as Lee et al., (2019).

We have added the reference as suggested by the reviewer! Thanks a lot!

Line 108: “good quality” needs a reference.

Done

Line 116: ALADIN needs to be in all caps throughout.

Done.

Line 117: We find the description in Flament et al., (2021), a little clearer (e.g., “The UV laser beam is linearly polarized at the laser output. It goes through a quarter-wave plate before being routed towards the telescope and is thus transmitted towards the atmosphere with a circular polarization...”).

We have rephrased accordingly.

Line 126-128: This is important and should be explained in more detail. This paper is about validating Aeolus/ ALADIN. The limitations of the lidar should be clearly explained and other papers should be referenced.

There is a thorough discussion throughout the text regarding the limitations of the lidar whereas all the related papers are cited.

Line 130: We recommend “ALADIN” and why are “continuous” calibration and validation needed? Please explain.

We have changed “Aladin” to “ALADIN” and we have removed the word continuous.

Line 136: We recommend “L2A aerosol optical properties”.

We do not agree with the reviewer’s comment. Aeolus L2A optical properties refer both to aerosols and clouds (grouped as particulates) since there is not a separation between them.

Line 138: Regarding the “excellent agreement” here, we recommend adding some nuance. These results for a case study with a strong non-depolarizing aerosol, were ~satisfying only between ~4 and 8km.

We are clarifying better this sentence in the revised manuscript.

Line 142 – 156: If the type of aerosols over the three regions is discussed here, then you might consider not repeating it elsewhere (e.g., section 4). In general, we recommend adding a table describing the lidar(s) at each of the three locations. This table could contain, for example, the lidar’s name, a small description, its limitations and uncertainties, its products. It could also contain information on how cloud screening is performed from these ground-based lidars and possibly the dominant type of aerosols at each location (together with references related to these stations).

In this paragraph we are providing a summary of the aerosol types encountered within the broader Greek area. Actually, our intension is to note the coexistence of various aerosol species thus highlighting that the region of interest is ideal for the purposes of the study. On the contrary, in Section 4 we are specifying the aerosol type(s), per case, by exploiting the ancillary datasets (i.e., models and observations). Therefore, we are discussing about two different things and there is not overlap between these two parts of the manuscript. Regarding the second part of this comment please see our reply above.

Line 184: HSRL was already introduced on line 116.

Thanks!

Line 187: We recommend "are backscattered".

We agree. Thanks!

Line 206-215: What is the purpose of describing algorithms that are not used in the study (e.g., ICA and MCA)?

Actually, it is not a description but a short mention of the ICA and MCA algorithms just to inform a reader, who is not so much familiar with Aeolus, for their existence. However, we have removed this part in the revised manuscript.

Line 216: We recommend "the primary and most reliable".

We have modified this part in the revised text.

Line 217: We recommend "measured signals in the Mie channel".

Done.

Line 224: We recommend "signals in each channel"; also, the sentence is not clear.

This paragraph explains the cross-talk correction which is required due to the "cross-contamination" between the Rayleigh and Mie channels. The first sentence serves as a short statement whereas the following sentences describe explicitly the cross-talk issue attributed to the ALADIN HSRL instrument design.

Line 260 and section 4: Again, the three stations, type of lidar(s), products, uncertainties, limitations (e.g., overlap), etc. could really use a table. That table could also show a predominant aerosol type over the region and a median and standard deviation AOD from satellite(s).

We think that we have already provided a sufficient answer to this comment raised by the reviewer. Please see our previous replies.

Line 263: We recommend using "different" instead of "adverse".

Done.

Figure 1-i: It would be helpful to write "all three stations are within Xkm of each other".

We are providing only here the distances between the stations, as requested by the reviewer, and not in the caption of Figure 1 since we don't see why such information is helpful. The distances between ANT-ATH, ANT-THE and ATH-THE are equal to 273 km, 531 km and 305 km, respectively.

Line 272 to 275: The authors must mean "to ensure the consistency of all lidar-derived observations"?

Exactly! We have slightly rephrased this sentence.

Line 279: We recommend deleting "measurements" here.

We think that the word "measurements" fits well in the sentence.

Line 280-281: Why is this assumption plausible? Does it remain to be tested?

A well-mixed planetary boundary layer (PBL) is a common assumption, especially during afternoon hours above land when as mentioned in Stull, (1988; page 450, 11.2.1 The Mixed Layer Profile Shapes): *"…intense vertical mixing tends to leave conserved variables such as potential temperature and humidity nearly constant with height (see Fig 11.1). Even wind speed and direction are nearly constant over the bulk of the mixed layer."* Hence, the concentration of pollutants (similar to humidity) can be assumed to be nearly constant with height.

Under this assumption and in certain cases, lidar profiles can be linearly extended to the ground below the full overlap region when no information is available (i.e. in cases when we need to compare aerosol optical depth (AOD) from a sun-photometer, to the lidar derived AOD). Nevertheless, the altitudes below the full-overlap region (~800m for the ground-based lidar systems used in our study) were *not* accounted for herein, since i) we did not want to introduce any bias in our results caused by linearly interpolating ground-based profiles to the ground, ii) the Aeolus performance downgrades near the surface either due to surface reflectance or increased noise levels at the lowermost bins.

Since these altitudes were not accounted for, we removed this sentence from the manuscript, as it might be confusing for the reader.

Reference: *Stull, R.B.* An Introduction to Boundary Layer Meteorology*; Springer: Berlin/Heidelberg, Germany, 1988*.

Line 286-288: Doesn't this apply to all three stations? Also why not add biomass burning aerosols here?

It was not well stated in the submitted manuscript and we have remove it in the revised text.

Line 338: We recommend "Aerosol spatial variability in the vicinity of the PANACEA sites". A description of the dominant aerosol type at each station would fit well here but then the authors would have to delete it from the introduction to avoid repetition.

We would like to thank the reviewer for his/her suggestion but we prefer to keep the initial title.

Section 4.4: The purpose behind studying the spatial variability could be explained more clearly. Our understanding is that the authors are attempting to characterize spatial variability to explain a potential disagreement between Aeolus/ ALADIN and ground-based lidars. A disagreement could be due to imperfect spatial co-location and/or simply ALADIN's 87 km horizontal resolution. The authors are studying horizontal variability by using total column integrated AOD and that should be mentioned as well. There could be minimal horizontal variability but a strong vertical variability. It is also not clear how the authors have computed the mean AOD from MODIS. Is it a arithmetic or geometric mean? It does make a difference -- see e.g., Sayer and Knobelspiesse, (2019). Spatial characterization analysis usually uses mean and standard deviations within each satellite grid cell and/or the variation between consecutive satellite pixels or airborne measurements within a region (e.g., Anderson et al., 2003 or Shinozuka and Redemann et al, 2011).

We think that it is pretty clear the purpose of this short analysis and it is already explicitly stated in the submitted manuscript. We are copying below the relevant part from the text.

"*The aim of this introductory analysis is to investigate the horizontal homogeneity of the aerosol optical depth (AOD) in the respective broader areas, playing a key role in the comparison of ground-based and spaceborne profiles, which are not spatially coincident as it will be shown in a following section (i.e., collocation method)*"

As correctly mentioned by the reviewer, the disagreement between spaceborne and ground-based profiles can be due to the imperfect spatial collocation and/or the coarse spatial resolution of Aeolus BRC. However, we would like to remind that both aspects are already discussed in the submitted manuscript. For instance, we are providing below a part from the submitted text related to the reviewer's comment.

"*Aeolus retrievals are available at coarse along-track resolution (~90 km). This imposes limitations on their evaluation against point measurements, which are further exacerbated at sites where the heterogeneity of aerosol loads in the surrounding area of the station is pronounced, taking into account that the spatial collocation between spaceborne and ground-based retrievals is not exact.*"

In the revised text we are clarifying that the MODIS AODs correspond to the entire atmospheric column as well as that we are calculating the arithmetic mean. Regarding the vertical variability this is exactly what is examined in the current study either for specific cases or for the entire collocated sample. Please note that in the revised manuscript we have extended our analysis including the coefficient of variation and the spatial autocorrelation.

Fig. 2: The orange arrow is hard to see; It is also not clear if the analysis involving the 46 cases considers the closeness of the actual track to the station (e.g., better spatial colocation on July 1st).

We have changed the color of the arrow denoting Aeolus' flight direction. For each case, we are taking into account all BRCs residing within a circle area (of 120 km radius) centered at the station coordinates. Each case corresponds to a specific day and for this day the number of BRCs can be one, two or three (maximum). In the revised manuscript, we are clarifying better this point.

Line 381: The authors should discuss this "temporal window extension" in more detail and attempt to explain its consequences.

We have revised this part of the manuscript providing more details. Below is given the modified text.

"*For the ground-based observations, the aerosol backscatter profiles are derived considering a time window of ± 1 hour around the satellite overpass. Nevertheless, this temporal collocation criterion has been relaxed or shifted in few cases to improve the quality of the ground-based retrievals (i.e., by increasing the signal-to-noise ratio) as well as to increase the matched pairs with Aeolus L2A profiles. Both compromises are applied since the weather conditions favoring the development of persistent clouds may eliminate the number of simultaneous cases. It is noted, however, when the temporal window is shifted or relaxed we are taking into account the homogeneity of the atmospheric scene (probed by the ground lidar). For the Antikythera station we did not deviate from the pre-defined temporal criterion apart from one case study. In Thessaloniki and Athens, the time departure between Aeolus and ground-based profiles can vary from 1.5 to 2.5 hours. Overall, 43 cases are analyzed out of which 15 have been identified over Antikythera, 12 in Athens and the remaining 16 in Thessaloniki.*"

Line 385-397: We recommend "derived from radiances measured by SEVIRI", "indication of cloud presence"; the limitations of using this SEVIRI cloud mask should be discussed. For example, could SEVIRI be missing small broken water clouds? What about cirrus clouds? and what would be the consequences on the Aeolus/ ALADIN aerosol profiles?

We have modified the "*derived from radiances measured by SEVIRI*" part of the sentence in the manuscript, as suggested by the reviewer. Regarding the limitations of the SEVIRI CLM product, please see our reply in a relevant comment above.

Line 403: Regarding (ii), how will the authors differentiate the effects of natural variability, the imperfect co-location and the errors in the Aeolus/ ALADIN instrument? See e.g., section "nature versus noise" in Anderson et al., 2003. Regarding (iii), this was already demonstrated in numerous studies.

Please read our previous replies in relevant comments. We are not claiming that point III is a novelty of our study.

Line 406: We recommend "(...) Cal/Val study to facilitate the interpretation of our findings and to identify possible upgrades in the Aeolus/ALADIN retrievals."

We agree with the reviewer's suggestion and we have modified the text accordingly.

Line 409: We recommend "the results are depicted in Figure 3".

We think that we do not have to change something here.

Line 411: We recommend "... Aeolus retrievals are provided at a coarse horizontal and vertical resolution ..."

Done.

Line 420: We recommend "To depict the spatial patterns (...)".

Done.

Line 423: The fact that MERRA-2 and CAMS provide "aerosol products of high quality" is a strong statement and should be explained. The explanation should include model evaluation results from previous studies. Model aerosol optical properties and model aerosol speciation have serious limitations, which should absolutely be mentioned in the text.

The two reference studies in the submitted manuscript show that the AOD from both reanalysis datasets is a product of high quality. For example, please see the CAMS evaluation metrics given in Tables 3.2.1 (North Africa, Middle East and Europe) and 3.3.1 (Mediterranean) in Errera et al. (2021). In Gueymard and Yang (2020), the AOD evaluation metrics at station level, both for MERRA-2 and CAMS, depicted in Figure 6 reveal a very good agreement with observations. We fully agree with the reviewer that there are limitations on the modelled aerosol optical properties and aerosol speciation. Nevertheless, we are not claiming that the aerosol speciation is accurate and we are clarifying that the aerosol outputs from CAMS and MERRA-2 reanalyses are used just as an indicator (this is explicitly stated in the manuscript) to support a better characterization of the probed atmospheric scene. Even though the evaluation of CAMS and MERRA-2 is beyond the scope of this paper, we would like to mention that for the selected cases (Section 6.1) it seems that their performance is quite good in terms of capturing the aerosol load and the presence of coarse/fine particles.

Line 425: Why not use AERONET Level 2 (quality assured) instead of Level 1.5?

Please see our reply in a relevant comment above.

Line 427: We recommend "characterization of the aerosol load and size over the station"

Done.

Line 432: Figures should be called in the order they appear. Figure S4 is introduced before S1, S2 or S3.

Done.

Line 435: We recommend adding "at 550 nm".

Done.

Line 437: The truth should be in the sunphotometer direct measurements. This sentence, as written, could be interpreted as things being the over way around.

We think that it is pretty clear the meaning of this sentence. We are saying that the MERRA-2/CAMS findings are confirmed by the AERONET observations and the PollyXT retrievals (the following sentence).

Line 438: The Angstrom exponent should be briefly explained here (i.e., difference of AOD at two (or more?) wavelengths that informs on the particle size) and references for typical dust angstrom exponent should be added to the text (e.g., Dubovik et al., 2002).

We have added the wavelength pairs (440-870nm) as well as the reference suggested by the reviewer.

Figure S2: This10 min-worth of high VLDR content looks suspiciously high compared to the consecutive profiles in the curtain plot. How do the authors explain that dust was present for only 10min and then suddenly disappears?

We thank the reviewer for this comment that we believe is referring to the area indicated by the red box in the figure. As explained in the legend of Fig. S2, the 10-minute interval that seems to present higher VLDR values, corresponds to the calibration of depolarization measurements automatically performed by all Polly^{XT} lidar systems (Engelmann et al., 2016). The procedure followed is the ±45◦ -calibration method described in Freudenthaler et al. (2009) and Freudenthaler, (2016). As described in Engelmann et al. (2016), a polarizer placed in front of the detectors of the Polly^{XT} lidar system, rotates automatically three times per day, first at -45 (for 5 minutes) and then at +45 (for 5 minutes) degrees with respect to the polarization plane of the laser beam. When operating in normal mode, measurements are performed without the polarizer into the light path. Furthermore, in order for the aerosol optical properties to be retrieved, this 10-minute time interval is excluded from the data.

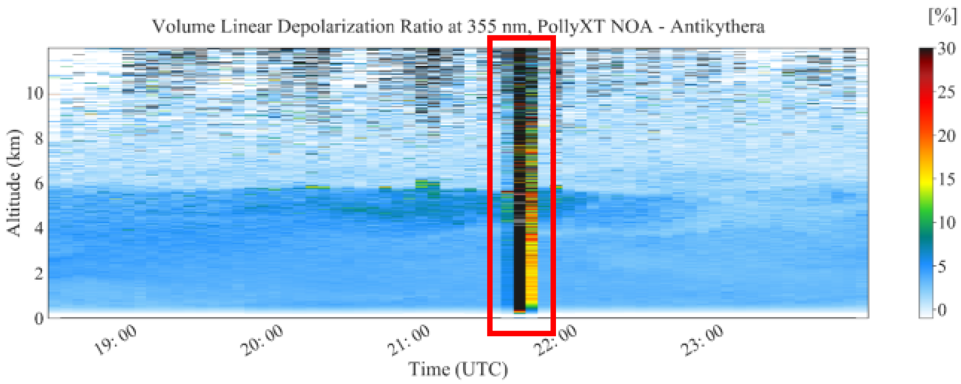


Figure S2: Time-height plot of the Volume Linear Depolarization Ratio (VLDR) at 355 nm at PANGEA station during 10 July 2019, 18:00 - 23:59 UTC. Station elevation is at 193 m a.s.l. The time period between 21:30 and 21:40 UTC corresponds to routine depolarization calibration measurements of the Polly^{XT} lidar system and is indicated by a thick black vertical line on the plot. Below 6 km, VLDR values (5 - 10%) indicate the presence of non-spherical, depolarizing particles of dust nature.

References:

Engelmann, R., Kanitz, T., Baars, H., Heese, B., Althausen, D., Skupin, A., Wandinger, U., Komppula, M., Stachlewska, I. S., Amiridis, V., Marinou, E., Mattis, I., Linné, H., and Ansmann, A.: The automated multiwavelength Raman polarization and water-vapor lidar PollyXT: the neXT generation, Atmos. Meas. Tech., 9, 1767–1784, https://doi.org/10.5194/amt-9-1767-2016, 2016.

Freudenthaler, V., Esselborn, M., Wiegner, M., Heese, B., Tesche, M., Ansmann, A., MüLLER, D., Althausen, D., Wirth, M., Fix, A., Ehret, G., Knippertz, P., Toledano, C., Gasteiger, J., Garhammer, M., and Seefeldner, M.: Depolarization ratio profiling at several wavelengths in pure Saharan dust during SAMUM 2006, Tellus B, 61, 165–179, https://doi.org/10.1111/j.1600-0889.2008.00396.x, 2009.

Freudenthaler, V.: About the effects of polarising optics on lidar signals and the Δ90 calibration, Atmos. Meas. Tech., 9, 4181–4255, https://doi.org/10.5194/amt-9-4181-2016, 2016.

Line 442: Why would this case be ideal for evaluating Aeolus/ ALADIN as we know Aeolus cannot measure non-spherical dust properly?

In a complete Cal/Val study, they must be addressed all the factors determining the agreement between spaceborne and ground-based retrievals. It is well-known that the misdetection of the cross-polar return signals by ALADIN results in a degradation of its performance for depolarizing targets (i.e., dust, volcanic ash, cirrus crystals). This is clearly stated between lines 473 and 477 in the submitted text. Therefore, in the test case of 10th July 2019 (Section 6.1.1) we want to demonstrate how much this deficiency affects the comparison between Aeolus and Polly^{XT} retrievals.

Line 449: What is meant by “statistical uncertainty margin”?

Below we are providing a short description of the backscatter error in Aeolus and ground-based profiles.

Aeolus

We are copying below from the last paragraph in Section 2.3.1 in Flament et al. (2021).

“*Equations have been derived to estimate the impact of the detection noise on measured signals S_{Ray} and S_{Mie} on retrieved β_p and α_p values. The derivation of these error estimates is fully explained in Flament et al. (2021) but is too cumbersome to be reported here. It is based on the assumption that the uncertainty of S_{Ray} and S_{Mie} is purely due to the Poisson counting noise and uses second-order developments. As a consequence, error estimates are valid as long as the level of noise is not too high; otherwise, the approximation introduced by the second-order developments becomes too coarse. The errors estimates do not take into account the impact of atmospheric heterogeneity within the BRC that increases the random noise on the BRC accumulation of observation level S_{Ray} and S_{Mie}. It nevertheless remains that they are useful to identify the β_p and α_p estimations that are reliable and then give a good idea of their accuracy.*”

Ground-based lidar

As explained in Mattis et al. (2016), statistical uncertainty in SCC Level 2 products is calculated either by Monte Carlo method or traditional error propagation. The former was selected in our study for all the considered profiles. In case of elastic lidar signals (as those used herein), the calculation of the uncertainty of particle backscatter coefficient profiles with the Monte Carlo method entails:

i) The overall statistical error of pre-processed signals. In case of photon-counting systems this can be evaluated for each photon-counting raw signal range bin as the square root of the corresponding count (D’Amico et al., 2016)
ii) the assumed particle lidar ratio value and uncertainty

iii) the statistical error of the signal within the selected calibration range

References:

D'Amico, G., Amodeo, A., Mattis, I., Freudenthaler, V., and Pappalardo, G.: EARLINET Single Calculus Chain – technical – Part 1: Pre-processing of raw lidar data, Atmos. Meas. Tech., 9, 491–507, https://doi.org/10.5194/amt-9-491-2016, 2016.

Mattis, I., D'Amico, G., Baars, H., Amodeo, A., Madonna, F., and Iarlori, M.: EARLINET Single Calculus Chain – technical – Part 2: Calculation of optical products, Atmos. Meas. Tech., 9, 3009–3029, https://doi.org/10.5194/amt-9-3009-2016, 2016.

We have modified the initial text as follows.

"*The colored dashed lines (Aeolus) and the pink shaded area (Polly^{XT}) correspond to the statistical uncertainty margins of the spaceborne (see Section 2.3.1 in Flament et al., (2021)) and the ground-based (D'Amico et al., 2016) retrievals, respectively. Both refer to the photocounting noise following a Poisson distribution.*"

Line 473: The authors mention "fine particles" but an explanation is missing here; we recommend "until their arrival over...".

In the previous sentence we are mentioning the "type" (i.e., anthropogenic, biomass) of the fine particles. Both "*until*" and "*till*" have the same meaning here.

Line 480: We recommend "AOS are mainly attributed ..."; the models seem to be taken, once again at face value here (i.e., they seem to be treated the same as observations, but model species (and their spatial variation) are sometimes not reliable).

We think that we have already provided sufficient answers of how the models are treated in our analysis. This is clearly stated in our replies as well as in the manuscript.

Line 484: We recommend "ALADIN reproduces the layer's structure well".

We would like to thank the reviewer for his/her suggestion but we prefer to keep the sentence as is.

Line 493-496: This sentence is not clear. The fact that MERRA-2 and CAMS aerosol optical properties and speciation disagree cannot be directly connected to a good or bad performance between models and AERONET. One would need to directly compare aerosol optical properties from MERRA-2 or CAMS to aerosol optical properties from AERONET. Also, the way this is written could make it sound like AERONET provides aerosol species, which it does not. Instead, it measures aerosol optical properties, which can be used to define aerosol types and that can be indirectly translated into aerosol chemical species in certain cases (Kacenelenbogen et al. 2022).

In our study, we don't give emphasis on the evaluation of the performance of the CAMS and MERRA-2 reanalysis products. We think that this is clear and it is beyond the scope of this paper. We assume that the reviewer is referring to the aerosol speciation and not in aerosol optical depth. For the latter one, the reanalysis and the observed AODs can be compared directly (depending whether AOD observations are assimilated or are independent in the assimilation scheme). The discussion in the manuscript is made in a more generic sense. We agree with the reviewer and we are aware that AERONET observations do not provide information for aerosol species. However, the combination of AERONET intensive and extensive aerosol properties, along with ancillary information (e.g., FLEXPART), can provide useful information (even though it is not complete). For this case (8th July 2020), MERRA-2 obviously fails to reproduce the AOD levels while the strong contribution from sea-salt particles is not justified by the high AERONET Ångström (1.5-1.6) and FMF (0.75-0.95) values, indicating the predominance of fine particles. Please note that we are referring to a specific case and we are not generalizing our results. A better assessment would require LUT tables of the extinction coefficient (at different wavelengths and RH levels, see Section 3 in Randles et al., (2017)) in order to compare the reanalysis optical properties against those given by AERONET.

In the revised manuscript we have modified this part which reads as follows:

"*On 8th July 2020, the broader area of the Antikythera island was under the impact of moderate-to-high aerosol loads, mainly consisting of organic and sulphate particles, in the western and southern sector of the station, based on CAMS simulated AODs (up to 0.5) (Fig. S12-ii). AERONET measurements, yield UV AODs up to 0.5 and Ångström exponent higher than 1.5 during early afternoon (Fig. S13) whereas the FMF is higher than 0.75 throughout the day (Fig. S14). MERRA-2 AOD patterns (Fig. S12-i) and speciation (strong contribution from marine and sulphate aerosols to the total aerosol load) are different from those of CAMS, without being very consistent with respect to the ground-based sunphotometer observations (Fig. S13, Fig. S14).*"

Line 502: We recommend "Aeolus performance depends on altitude according to Polly...".

We believe that our version is better stated.

Line 554: The reader needs to be reminded which case studies are included here – are those the 46 case studies of line 383?

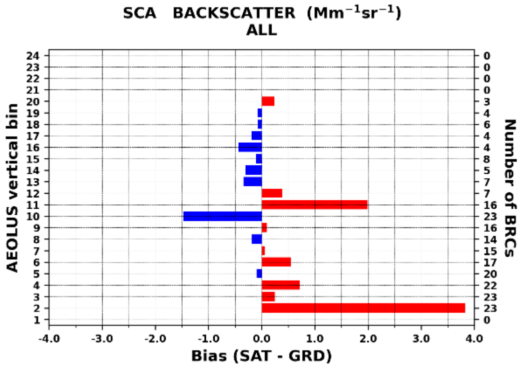
Thanks for the comment! In the revised manuscript we are clarifying better this point.

Line 563: The statement referring to "the contribution of depolarizing particles is quite low based on the ancillary dataset" needs more explanation and needs to be supported by some results.

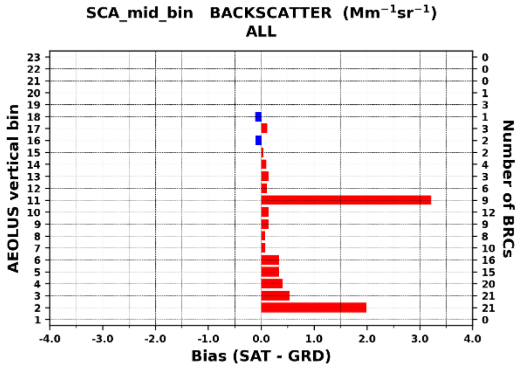
Please note that in the previous sentences we are mentioning that it is not possible to apply the conversion from linear to circular optical products for the retrievals acquired at Thessaloniki due to technical issues (related to the polarization purity of the emitted laser beam and the performance of the telescope lenses). Therefore, if we consider this conversion, a significant part of the collocated sample cannot be used in the statistical analysis thus reducing the robustness of our results. In the profiles derived in Athens and Antikythera, after applying the conversion in the ground-based profiles, we see that the differences between total (cross plus co) and Aeolus-like (co) backscatter are negligible due to the presence of spherical particles (the ancillary datasets have been used as indicator for further confirmation).

Here we are presenting the vertically resolved metrics (bias, RMSE) for the unfiltered (as in Fig. 4) and filtered (as in Fig. 5) Aeolus profiles. As reference, we are using the ground-based Aeolus-like backscatter retrievals after applying the conversions (from linear to circular optical products) presented in Paschou et al. (2021). Since depolarization measurements are not available at Thessaloniki, the sample for the statistical analysis contains profiles only from Athens and Antikythera. For the unfiltered Aeolus profiles, there are many similarities (except bins 13-19 for SCA) between the results presented here and those given in the manuscript. On the contrary, there are significant differences for the Aeolus cloud-filtered profiles. However, the number of BRCs considered for the metrics calculation is significantly reduced thus making questionable the robustness of the obtained findings. Drastic reductions appear also for the unfiltered Aeolus profiles. Summarizing, we believe that our decision to present in the manuscript the comparison of Aeolus backscatter retrievals against the total backscatter from ground-based lidars is well supported.

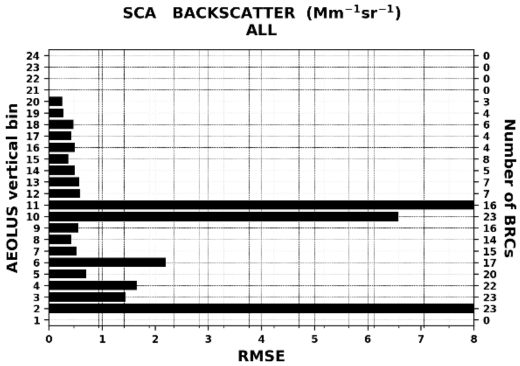
Unfiltered Aeolus profiles



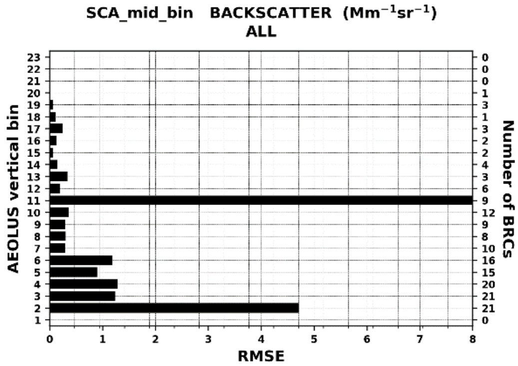
(i-a)



(i-b)

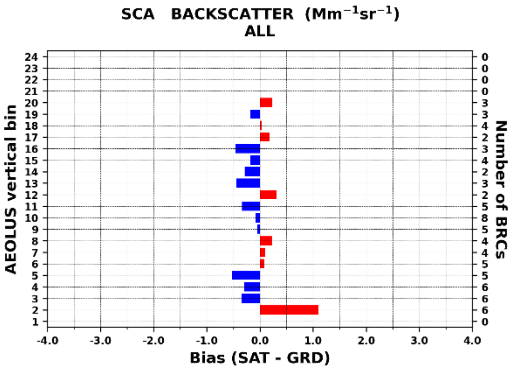


(ii-a)

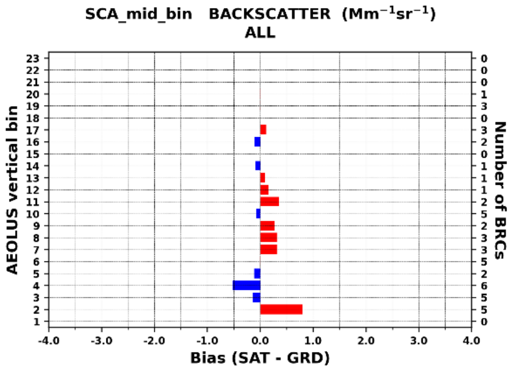


(ii-b)

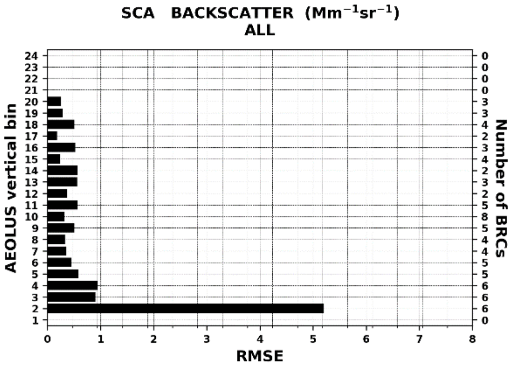
Cloud filtered Aeolus profiles



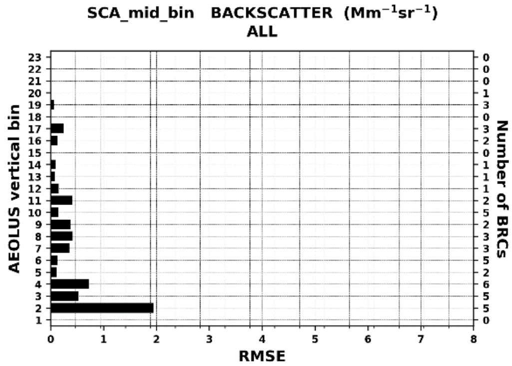
(i-a)



(i-b)



(ii-a)



(ii-b)

Figure 4: The metrics need to be described, like in Abril-Gago et al. (2022). Authors should specify why they show Aeolus/ALADIN profiles that are cloud contaminated (or "unfiltered") in their analysis. Is is not obvious why there would not be more disagreement between cloud-unfiltered space-borne and cloud-filtered ground-based profiles compared to cloud-filtered space and ground profiles. Or is this a way to test their SEVIRI-based cloud filtering method?

We don't think that we have to provide the formulas for the bias and the root mean square error. These two metrics are two of the most well-known and they have been applied in numerous studies in atmospheric sciences. In the revised manuscript we have added the following reference:

"*Wilks, D.S. Statistical Methods in the Atmospheric Sciences, 4th ed.; Elsevier: Cambridge, MA, USA, 2019.*"

Regarding the second part of the reviewer's comment, please see our reply in a previous similar comment raised by the reviewer.

Line 586: We recommend "Fig. 4" instead of "Fig. 5".

Done. Thanks for the correction!

Line 591-592: The authors should mention that "SCA mid-bin" is expected to perform better than SCA.

It is already mentioned in the text.

Line 597-599: This is a repeat from line 137.

In lines 136-138 we are mentioning the work of Abril-Gago et al. (2022) whereas in lines 597-599 we are stating that the findings between these two works are in a very good agreement.

Line 624: Do the authors mean low SNR instead of high SNR?

Thanks a lot for noticing our mistake! It has been corrected in the revised manuscript.

Figure 6-7: Again, why show cloud contaminated Aeolus/ALADIN profiles? Also, why show SCA instead of SCA_bin as the latter is expected to lead to better results (already shown in Fig. 5)?

Because we want to contrast the results between the two Aeolus vertical scales and between cloud filtered and unfiltered profiles.

Line 641: Again, can Aeolus/ALADIN profiles still be cloud contaminated after applying the SEVIRI cloud mask?

This might be possible, but we cannot be sure due to the lack of a classification scheme in Aeolus retrievals.

Line 641-643: This statement about not using QA flags appears too late in the text. It should be in the method or the Aeolus/ALADIN section.

We believe that it is already well placed.

Line 654: “many similarities” needs to be described in more detail.

We believe that it is pretty clear the meaning of this sentence. The scatterplots in Abril-Gago et al. (2022) and in the current study look very similar.

Line 682-688: The authors should explain why they expect a difference in performance between the ascending and descending orbital data. Grouping the data per orbit direction seems inconclusive and we question the usefulness of mentioning the results.

Considering the strong temporal variation of aerosol loads (even at short temporal scales) there could be differences between dawn (early morning) and dusk (early afternoon) orbits. Please note that we are briefly discussing this part and we are mentioning that we need more data (e.g., EARLINET study) in order to derive robust results.

Line 701-705: This appears too late in the text. It should be part of the comparison method between Aeolus/ALADIN and ground-based lidars.

We believe that this sentence fits better in Section 7 where we are discussing the Cal/Val aspects and the recommendations.

Line 702: It is not clear what the authors mean by “the theoretical assumptions”.

The formulas for the conversion from linear to circular optical products are valid under the absence of orientation of the suspended particles and multiple scattering effects.

Line 719-722: Again, the authors should add some nuance to the discussion on model performance.

We have rephrased the sentence.

<u>References:</u>

Abril-Gago, Jesús, et al. "Statistical validation of Aeolus L2A particle backscatter coefficient retrievals over ACTRIS/EARLINET stations on the Iberian Peninsula." Atmospheric Chemistry and Physics 22.2 (2022): 1425-1451.

Anderson, Theodore L., et al. "Mesoscale variations of tropospheric aerosols." Journal of the Atmospheric Sciences 60.1 (2003): 119-136.

Baars, H., Radenz, M., Floutsi, A. A., Engelmann, R., Althausen, D., Heese, B., et al. (2021). Californian wildfire smoke over Europe: A first example of the aerosol observing capabilities of Aeolus compared to ground-based lidar. Geophysical Research Letters, 48, e2020GL092194. https://doi.org/10.1029/2020GL092194

Dubovik, Oleg, et al. "Variability of absorption and optical properties of key aerosol types observed in worldwide locations." Journal of the atmospheric sciences 59.3 (2002): 590-608.

Flament, T., Trapon, D., Lacour, A., Dabas, A., Ehlers, F., and Huber, D.: Aeolus L2A aerosol optical properties product: standard correct algorithm and Mie correct algorithm, Atmos. Meas. Tech., 14, 7851–7871, https://doi.org/10.5194/amt-14-7851-2021, 2021.

Kacenelenbogen, Meloë SF, et al. "Identifying chemical aerosol signatures using optical suborbital observations: how much can optical properties tell us about aerosol composition?." Atmospheric Chemistry and Physics 22.6 (2022): 3713-3742.

Lee, Logan, et al. "Investigation of CATS aerosol products and application toward global diurnal variation of aerosols." Atmospheric Chemistry and Physics 19.19 (2019): 12687-12707.

Russell, Philip B., et al. "A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry." Journal of Geophysical Research: Atmospheres 119.16 (2014): 9838-9863.

Sayer, Andrew M., and Kirk D. Knobelspiesse. "How should we aggregate data? Methods accounting for the numerical distributions, with an assessment of aerosol optical depth." Atmospheric Chemistry and Physics 19.23 (2019): 15023-15048.

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.