Reviewer 1

In their revision the authors have addressed the majority of my prior concerns. The claims that in my view were not well justified in the initial submission have mostly been softened or better supported. While uncertainties tailored to ambient measurements would have been ideal, I appreciate the challenge in doing so and approximate error estimates have now been included based on the Bazo et al. (2024) paper studying PIN100 performance in the lab which is satisfactory. There are however several points below for which I am still not following and/or believe need correction. In my opinion, the manuscript would be suitable for publication once these have been addressed.

We thank the reviewer for acknowledging the first review process by the authors, it has indeed helped to improve the quality of the manuscript. We proceed to answer each comment raised by the reviewer below.

GENERAL COMMENTS

<Regarding response to my initial comment on sources of key optical quantities>

Horvath (2015) proposes a stepwise extrapolation technique that relies on fits of log(P11) vs scattering angle and shows, that when this particular technique is used, errors in integrals over the full range of scattering angles (0-180°) are less than 5% for size parameters x≤15. However, the present manuscript simply says that regular linear extrapolation is used to fill in the truncated regions and also does not discuss the possibility of particles with size parameters greater than 15. This should be addressed.

- The reviewer is right, the extrapolation technique used in this work is not exactly the same as the one described in Horvath, (2015). Indeed, we have also done the stepwise extrapolation and compared it to our initial extrapolations. We have obtained differences in total scattering coefficients below 1% between both approaches, so we believe that we can assume that the errors of both extrapolation techniques are equivalent. According to Horvath, (2015): "*Using the stepwise extrapolation, the total scattering coefficient, the asymmetry parameter and the fraction of backscattered light can be attained with an accuracy far below 1%.*". Thus, for integrated values such as extinction coefficient, BS, SSA and asymmetry parameter errors are below 5%.

Horvath, (2015) also states "*for a particle sizes above 4 µm the deviations increase (although still in a range of <5%).*" These particles of around 4 µm in diameter would have size parameters between 19 and 31 in our wavelengths. According to Horvath, (2015), the extrapolation technique when these particles are present would still have errors below 5%.

To clarify these points, modifications in the text have been made (Lines 235-237)

"Stepwise extrapolations might be more consistent (i.e. Horvath, 2015), but our additional computations remarked that differences between linear and stepwise extrapolations were below 1% for σ_{sca}, g and B_s. For particle sizes above 4 µm the uncertainties can yield 5% (Horvath, 2015)."

I'm also wondering if an "illustration" of lidar ratio values is needed here. In my view, it would be much better to leave lidar ratios entirely out of the manuscript since they are determined primarily by extrapolation over an extremely variable region of F11 and are thus potentially subject to large very errors.

We appreciate referee suggestions. We think that our current analysis serves to illustrate the variability of lidar ratio, and also to highlight the difficulty of obtaining this parameter from polar nephelometry measurements. We have highlighted these issues in the revised manuscript

<Regarding response to my initial comment on unrealistic/nonphysical -F12/F11>

If I understand correctly, the authors are saying that the unrealistically low values of P12 are due to variability in the sample. I'm not completely following this explanation though. Each individual sample should still have physically plausible values of F12, even if the underlying aerosol was changing from one measurement sample to the next. And averaging many samples inside the physically plausible range could never produces means falling outside of that range.

We understand referee concerns and to clarify these issues we need to discuss details of how the instrument performs during dust measurements: When dealing with big particles forward scattering strongly predominates. That implies a challenge for PI-Neph measurements because it might saturate some angles at forward scattering and undersaturated others in the backward scattering. During the extreme events in March 2022, we had even to low camera gain to avoid supersaturations – which later implied specific calibrations. For the rest of the data series it didn´t happen and the camera gain was higher and kept constant. Calibration was specifically checked – see Bazo et al., (2023) for details. These specific modifications in the gain´s camera help to get appropriate measurements in all angles. Nevertheless, during the measurement process some saturations/undersaturation still happened but the data quality process that we followed was able to reject all saturated/undersaturated data. For the specific phase function measurements where many data points are rejected, all angles are eliminated because phase matrix measurements were considered non-reliable.

The determination of F_{12} is done by subtracting parallel and perpendicular phase function measurements and later dividing by F_{11}, which is computed as the sum of both components. The backward scattering regions are more sensitive to variations in the phase functions, and typically the lower signals are registered (although still they are not undersaturated). Also F_{11} minimums are founds in these regions. This can create a mathematical artifact that enhances variations and some outliers that even do not have physical meaning.

Nevertheless, we do not show hourly averages that fall outside of the physical range between 1 and -1. It is true that blue channel shows a non-typical pattern with low values of -F_{12}/ F_{11} in the range 100-150º for 405 nm, but studying the evolution of the extreme dust event a smooth variation is observed. In fact, the non-typical pattern in -F_{12}/ F_{11} at 405 nm is observed just before the event arrives and when it is withdrawing, while during the peaks of the events -F_{12}/ F_{11} at 405 nm follows the classical pattern measured in laboratory with pure dust samples. The PI-Neph is validated versus laboratory measurements using PSL (see Bazo et al., 2023 for details) plus the fact that the pattern at 405 nm is not always present make us trust our measurements, despite the noise.

To clarify all these points, we have modified the manuscript (Lines 413-428):

"Phase matrix elements were exhaustively monitored with the use of the PI-Neph during both extreme events. Given the high concentrations of large particles, the usual configuration of the measurements could lead to saturation of many angles in the forward scattering. Therefore, it was necessary to reduce the gain of the PI-Neph's camera, changing the dynamic range of the camera for obtaining non-saturated measurements at such high concentrations. But these changes were made to also guarantee enough signals in the backward region where the minimums are found. Nevertheless, sporadic pixels might present saturation/low SNR at some angles, but they were filtered out by the data quality criterion for the instrument (Section 2.2.1). For the specific phase function measurements where many data points are rejected, all angles are eliminated because phase matrix measurements were considered non-reliable. Moreover, -F_{12}/F_{11} is computed by subtracting first parallel and horizontal phase function, that in many angles are very similar. Dividing this small number by F_{11} can enhance the differences, being particularly critical in the angular regions where the minimum values of scattering are found (typically between 90-150º for dust particles). All these effects, although they are always present, imply larger random noise in the measurements. Actually, noisier patterns in measured F_{11} and -F_{12}/F_{11} affected by large particles when compared with measurements of anthropogenic origin has been already reported by the first versions of PI-Neph in the United States (Espinosa et al., 2018)."

I'd also be curious to hear the authors thoughts on why the aerosol was varying so drastically from one measurement to the next. I would expect the properties of an aerosol that has traveled many 100's of km (Sahara to Grenada) to be relatively homogenous.

The supplementary material shows the hourly evolution of the F_{11} and -F_{12}/F_{11} during the two extreme Saharan dust events in March 2022. Data temporal resolution is of 1 hour. In the supplementary material we observe a smooth variation of all phase matrix elements. The main changes are observed in -F_{12}/F_{11} at 405 nm but in our opinion the figure reveals a smooth change. We associated the variations in -F_{12}/F_{11} at 405 nm with the possible influence of other anthropogenic particles. During the peaks of extreme events such possible background anthropogenic particles seem to not have an impact on the mixture, while that does not happen before and after the peaks of the event.

For the analyses of the entire dataset where more common Saharan dust events reach the stations – none of them is classified as extreme – the non-typical pattern in -F_{12}/F_{11} at 405 nm is mostly present. This is supported by measurements of additional instrumentation that classified the aerosol sample as a mixture of dust with anthropogenic particles (Figure 10). Also, from the seasonal analyses the pattern observed for -F_{12}/F_{11} at 405 nm presents the lowest standard deviations (Figure 11). Simulations with GRASP help to support how those mixtures of dust with anthropogenic particles can alter the shape and spectral dependences of -F_{12}/F_{11}.

We believe that further studies are needed. On the one hand, more measurements at other locations will help to understand the different patterns of -F_{12}/F_{11} when influenced by Saharan dust particles. Also, further developments of GRASP will help to understand aerosol optical and microphysical properties differentiating between fine and coarse mode. Further studies with other modelling approaches for the scattering of big and non-spherical particles are also required (i.e. spheres, spheroids and irregular-hexahedral). We have slightly modified the text accordingly (Lines 795-797):

"Another issue to study is the use of irregular-hexahedral for modeling the scattering of large and non-spherical particles that might reproduce better polarization signals (Saito and Yang, 2021, 2023; Saito et al., 2021)."

And in the Conclusions section (Lines 883-884)

The possibility of implementing the irregular-hexahedral model would be also ideal to better understand polarization patterns.

The following references have been added:

Saito, M., and Yang, P. (2021): Advanced bulk optical models linking the backscattering and microphysical properties of mineral dust, *Geophysical Research Letters*, *48*, e2021GL095121. https://doi.org/10.1029/2021GL095121

Saito, M., Yang, P., Ding, J., and Liu, X. (2021): A comprehensive database of the optical properties of irregular aerosol particles for radiative transfer simulations. Journal of the Atmospheric Sciences, 78, 2089-2111.

Saito, M., & Yang, P. (2023). Quantifying the impact of the surface roughness of hexagonal ice crystals on backscattering properties for lidar-based remote sensing applications. *Geophysical Research Letters*, *50*, e2023GL104175. https://doi. org/10.1029/2023GL104175.

SPECIFIC COMMENTS (LN # links to the first submission of the manuscript)

LN 284: While Nousiainen & Kandler (2015) state several times that P11(180°) is lower in nonspherical particles, I could not find anything in that review supporting the idea that the total integral of P11 over the range 90° to 180° is lower in nonspherical particles. P11 generally increases with nonsphericity in the angular range of 90° to ~150° and this can cancel out the higher P11 values found in spheres at the more extreme backscattering angles (θ>150°). See for example the P11 panel of Figure 3 in Zhou et al. (2020).

- According to Horvath et al., (2018): "*The scattering function of the desert aerosol has a low back scattering, which is typical for non-spherical particles. Therefore it is to be expected that a characterization of desert aerosol particles could be achieved by considering the fraction of backscattered light. It is defined as the ratio of the integral of the volume scattering function between 90 and 180◦ divided by the integral over the full angle and is readily available once the volume scattering function is known*. […] *the backscattered fraction obtained from polar nephelometer measurements is lower for the aerosol dominated by desert particles, as expected.*"

Therefore, we believe that our statement is supported. We have changed the reference Nousiainen and Kandler, (2015) for the reference Horvath et al., (2018) in this statement.

Figure 5: To avoid confusion, if the Figure R1 version is not to be used in the final manuscript, I would at least suggest including a mention of the missing data to avoid confusion.

- The reviewer is right. We have added the following sentence between lines 360-361: "Note that Panel (a) in Figure 5 does not cover the beginning of the Saharan dust event due to lack of data related to supersaturation of the PI-Neph's measurements."

LN 504/707: I'm still not following what is meant by "large variability". Is this referring to the changes in P12/P11 on the order of ±0.2 that vary randomly and show little correlation over lengths greater than a few degrees in the red and green channels? Those variations are clearly nonphysical instrument artifact, likely stemming from lower aerosol concentration which result in low signal–to-noise. The dust measurements shown in Espinosa et al. (2018) – for which a citation was added to the text here – are also the result of instrument artifacts and do not represent a physical characteristic of the aerosol. This should be made clear.

We understand referee concerns. Apart from the natural variability of the sampled aerosol, measurements of phase matrix for large and non-spherical particles face additional challenges due to instrumental and physical limitations that ultimately might imply larger variability. On the one hand, these large and non-spherical particles present maximum scattering in the forward regions and minimum in the backward region, particularly in the region 90-150º. Thus, there are strong differences in the measured signals between both regions. As the referee comments, these issues were already raised in Espinosa et al., (2018). In our study, for the extreme Saharan dust events we had even to change the camera gain to avoid saturations in the forward scattering region. Nevertheless, we guarantee that the measurements presented are in the lineal dynamic range of the detector, filtering out saturations and under-saturations. On the other hand, we face limitations that the computation of $-F_{12}/F_{11}$ imposes. First, F_{12} is computed by subtracting horizontal and vertical phase function measurements that can be very similar for the case of dust. For the scattering region where minimums are found (90-150º), signals are weaker and thus are even more sensitive to errors in the measurements. This can be even enhanced when divided by F_{11}. Although these difficulties in the measurement of phase matrix for dust particles are systematic, they are ultimately traduced in random errors. In the revised manuscript, we have made all these points clearer (Lines 413-428):

"Phase matrix elements were exhaustively monitored with the use of the PI-Neph during both extreme events. Given the high concentrations of large particles, the usual configuration of the measurements could lead to saturation of many angles in the forward scattering. Therefore, it was necessary to reduce the gain of the PI-Neph's camera, changing the dynamic range of the camera for obtaining non-saturated measurements at such high concentrations. But these changes were made to also guarantee enough signals in the backward region where the minimums are found. Nevertheless, sporadic pixels might present saturation/low SNR at some angles, but they were filtered out by the data quality criterion for the instrument (Section 2.2.1). For the specific phase function measurements where many data points are rejected, all angles are eliminated because phase matrix measurements were considered non-reliable. Moreover, $-F_{12}/F_{11}$ is computed by subtracting first parallel and perpendicular phase functions, that in many angles are very similar. Dividing this small number by F_{11} can enhance the differences, being particularly critical in the angular regions where the minimum values of scattering are found (typically between 90-150º for dust particles). All these effects, although they are always present, imply larger random noise in the measurements. Actually, noisier patterns in measured F_{11} and $-F_{12}/F_{11}$ affected by large particles when compared with measurements of anthropogenic origin have been already reported by the first versions of PI-Neph in the United States (Espinosa et al., 2018)."

And we have included slight modifications in the discussions of the uncertainties. Now, between Lines 432-434 in the discussion of Figure 6 uncertainties.

"Note that the large standard deviation can be explained by the specific issues for the measurements of dust particles commented above".

In the discussion of Figure 7 uncertainties (Lines 482-484)

"Again, the specific issues related to the measurements of large particles can explain the deviations. Nevertheless, the larger deviations when compared with the previous events make us think that during this event there were more aerosol variability that is critical for the regions of the minimums in scattering."

In the discussion of uncertainties for Figure 9 (Lines 590-592)

“The standard deviations are larger (30% in in F_{11} and around 0.2 in $-F_{12}/F_{11}$) when compared with the extreme events, and despite the inherent issues in the measurement of phase matrix for big particles, it seems that the sample presents a more complex mixture with more variability during the 1-hour average.”

And finally in the discussion of Figure 11 (Lines 706-713)

“Particularly, for 660 and 515 nm large standard deviations are found in the region between 50º-150º while for 405 nm the standard deviations are considerably lower. Apart of the inherent limitations in the measurements in this scattering range, the results suggests that these $-F_{12}/F_{11}$ values at 660 and 515 nm are very sensitive to changing conditions in the aerosol that is sampled. Moreover, the other region that presents remarkable standard deviations for all wavelengths is the region of scattering angles above 170º. Those regions with large standard deviations are very sensitive to any change in particle type and size, which was demonstrated both from theoretical computations (Mischenko et al., 2002) and in laboratory measurements (Gomez-Martin et al., 2021)”.

REFERENCES

Espinosa, W. Reed, et al. "Retrievals of aerosol optical and microphysical properties from Imaging Polar Nephelometer scattering measurements." Atmospheric Measurement Techniques 10.3 (2017): 811-824.

Espinosa, W. Reed, et al. "Retrievals of aerosol size distribution, spherical fraction, and complex refractive index from airborne in situ angular light scattering and absorption measurements." Journal of Geophysical Research: Atmospheres 124.14 (2019): 7997-8024.

Zhou, Y., Levy, R. C., Remer, L. A., Mattoo, S., & Espinosa, W. R. (2020). Dust aerosol retrieval over the oceans with the MODIS/VIIRS dark target algorithm: 2. Nonspherical dust model. Earth and Space Science, 7, e2020EA001222. https://doi.org/10.1029/2020EA001222

Reviewer 2

Thank you for your detailed response to my comments. You could convince me that the lidar data for the specific dust outbreaks do not add further information to your study. And I hope that in future you'll find some nice comparisons of the lidar ratio at ground and in the lowest altitudes of the lidar observations. It is good that you have the lidar experts at the same university. Just a side note: Please take care with the altitude above sea level (SSC output) and the lidar data which are usually reported above ground level.

- We thank the referee for the comments and understandings, they have served to improve our knowledge and the manuscript. As stated in the first review process, there have been some intensive field campaigns to achieve this comparison between polar nephelometry and lidar. We hope to obtain results soon.

It is also sad that no size distributions were available for your observations. For an ACP publication I would have loved to see them, because the scope of the journal aims for a more holistic view on the single observations. That is why I just could mark "fair" for the scientific significance. Nevertheless, the article should be published. I strongly support your request to get bimodal size distribution which differentiate the optical properties (especially refractive index) between fine and coarse modes. Ideally, you would wait with the publication until this issue is solved to have a more complete modeling part in your study. However, I don't want to further postpone the publication and agree to do it without.

- We thank the referee for supporting the publication of our manuscript. We are working hard in collaboration with GRASP developers to optimize the retrievals. We are also working in evaluating the irregular-hexahedral model in the retrievals, which is time consuming

Concerning the uncertainties in Fig. 6, 7 & 9: Even if it just represents the atmospheric variability within the one hour of measurements, these error bars are important. In your Fig. R5, one sees that the results in the bottom left figure are clearly separated, whereas in the bottom center figure the variability is larger and the results are more similar. I would recommend to show these uncertainties in the paper as well and not only in the response to the reviewers. Please add these uncertainties to your 3 figures.

- We thank referee suggestions. In the revised manuscript we have added the error bars representing the standard deviation to Figures 6, 7 and 9.

Technical corrections:

Some new references in the introduction contain some letters of the surname, e.g., J. B. Renard or Reed Espinoza. It should not be the case.

EARLINET/ACTRIS – abbreviations should be written out once.

We thank the referee for pointing out these issues. Corrections have been added to the manuscript. Between Lines 239-240:

"Aerosols, Clouds and Trace gases Research Infrastructure (ACTRIS; https://www.actris.eu/)."

And between Lines 338-339:

"lidar measurements in AGORA in the framework of the European Aerosol Research Lidar Network (EARLINET; https://www.earlinet.org/)"