Dear anonymous reviewer,

Thank you very much for reviewing our paper and for your positive comments. We are extremely grateful for your kind words regarding the quality of our manuscript. We hope you find our answers and improvement satisfactory.

Kind regards,

Marcin Witek

Page 35 line 9 : "...nonspherical fraction artifacts are further enhanced, which could be due to the generally lower AODs than in V22 and thus more frequent misclassification between dust/cirrus particles." The atmospheric AODs have not become lower: the retrieved AODs are lower in V23 because of improved correction of certain artifacts. Why would an erroneously high AOD retrieval lead to a less erroneous retrieval of higher-order properties?

Re: Yes, we agree. We rephrased this sentence to read:

"In V23 retrievals these nonspherical fraction artifacts are further enhanced, which could be due to more frequent misclassification between dust/cirrus particles."

Page 32 line 10: "DW AOD retrievals in V23 are considerably smaller than in V22, thereby negatively impacting MISR's ability to differentiate between aerosol types and their AEs." I believe there is an error in this logic. Retrieved AOD over water was higher in V22 due to uncorrected stray light and possibly also underlight; these factors would be expected to manifest in (erroneously) higher retrieved AODs and (erroneously) more variable AEs. I believe this same logical error (lower retrieved AODs => lower S/N for higher-order properties) is repeated elsewhere in the manuscript, on page 33 line 9.

Re: We removed the last overly speculative sentence in this paragraph. We also rephrased wording on page 33 line 9 to read:

"This could be partially due to decreased ability of the MISR algorithm to distinguish between particle properties in low-AOD conditions."

Page 19 Line 26: "... only the red and NIR bands are used in the MISR dark water retrievals if the AOD is less than 0.50" This does not seem to be mentioned at all in Section 4.2.1, and seems pretty important! You should modify section 4.2.1 to more completely describe how the data from different bands are used differently.

Re: We added further explanation in section 4.2.1:

"The parameters v(l,j) and  $w_l$  are weights that depend on the wavelength and the availability of data (for details see e.g., Kalashnikova et al., 2013). For example,  $w_l$  are always equal to 1 for the MISR near-infrared (NIR) and red spectral channels, but vary between 0 and 1 for the green and blue bands depending on AOD. Different AOD-dependent  $w_l$  weights allow to mitigate the non-negligible contribution from the water leaving radiance to the TOA signal at shorter wavelengths."

Page 16 line 24: "The width of the combined distribution is proportional..." The term "proportional" implies scaling. What is the actual calculation used? Re: We added the following clarification: "The width of the combined distribution is proportional to the reported AOD uncertainty and calculated as the full width at half maximum divided by a scaling factor of  $2\sqrt{2 \ln 2}$ , assuming a normal distribution of  $f(\tau)$  (Witek et al., 2018b)."

Figures 13-14 If you use AOD weighting on the size mode fractions, this means that you are presenting an effective bulk modal fraction dMODE/dAOD, rather than e.g. a daily mean modal fraction such as would be used in a first-order climatology. You might consider adding a sentence describing this consequence of your weighting scheme for interpretation of your figures.

Re: We added the following clarification in section 6.2.3 where we first mention data averaging: "The differences in AE between V22 and V23 are visualized in **Error! Reference source not found.** The gridded AE values shown in **Error! Reference source not found.** are obtained by AOD-weighted averaging of individual AEs acquired over a year. Note that this averaging procedure is different than for example using daily values obtained from the L3 product and averaging them over the same period of time."

Page 19 line 12: "the minimum reflectance between model and observation constrains the wind speed in nearly all cases." This was not clear to me: what model? Is this a finding from Fox et al. (2007)? Re: We clarified and expanded the description of the wind speed retrieval used in V23. The new text reads:

"As described by Cox and Munk (1954), the peak surface reflectivity decreases, and the angular width of the glitter pattern increases systematically with wind speed. Given the range of view angles observed by the MISR instrument, it is possible to constrain the wind speed from the MISR data itself in some situations. For example, Fox et al. (2007) investigated in detail the degree to which the nine MISR view angles capture the wind-speed-related glitter pattern under a range of wind speeds and observing geometries. This work formed the basis of a simplified wind speed retrieval introduced in the new version of the aerosol product.

In V23, the modeled TOA reflectance,  $\rho_m$ , used for  $\chi^2_{abs}$  (Eq. 1) is pre-calculated at 3 discrete wind speeds: ws = 2, 5, and 7.5 m/s. At locations where no MISR channels are within the 40° glitter angle range, the algorithm selects the wind speed nearest to the monthly wind speed climatology from the TASC dataset. Otherwise, where at least one MISR channel is within the 40° glitter angle range, the algorithm selects the wind speed with the minimum equivalent reflectance difference,  $\Delta(ws, \tau)$ , between the model ( $\rho_m$ ) and the observation ( $\rho_{MISR}$ ), expressed as

$$\Delta(ws,\tau) = \sum_{l,j}^{4,9} v(l,j) \cdot \left(\frac{\rho_m(ws,l,j)}{\rho_{MISR}(l,j)} - 1\right)^2$$
(4)

where v(l, j) is 1 for channels within 40° glitter angle range; and 0 elsewhere. Note that the wind speed selected may be different for each aerosol optical model. The best estimate of wind speed reported in the V23 product is the wind speed selected for the aerosol optical model with the best fit according to  $\chi^2_{abs}$  and  $\chi^2_{maxdev}$  metrics."

Page 27 Line 19: "... setting the upper limit on the ratio to 16. The same constraint does not apply to retrievals over land." I think you mean that the success criteria for DW implies that any successful V23 retrieval within a 17.6km footprint implies a successful V22 retrieval for that footprint, while V23 land can have one or more successful 4.4km retrievals in a footprint with zero valid V22 retrievals. This could be more clearly explained.

## Re: We rephrased this description to make it clearer:

"Setting radiometric and algorithmic differences between V22 and V23 aside, if there are 16 valid V23 retrievals in a 17.6 km region, the DW success criterion implies that the same region also has to have a

valid V22 retrieval, setting the upper limit on the ratio to 16. The same constraint does not apply to retrievals over land—there could be multiple V23 but zero successful V22 retrievals within the same region."

Page 34 line 9: "...it is recommended to use SSA when AOD is above about 0.15..." Figures 9-13 are weighted by AOD, do they include retrieved properties for all AOD values or only for retrieved AOD above a threshold?

Re: All valid retrievals are included in the presented figures. Our goal in this introductory paper was to show all results. We debated adding additional figures where low-AOD retrievals are screened out, but this would increase the length of the paper considerably.

Page 44 Line 12: "There is, however, a substantial decrease..." This can be explained more clearly. Re: We rephrased this sentence to read:

"The standard deviations of the AE errors are in general quite similar for V22 and V23, although they are on average about 7% smaller in V23. Furthermore, the standard deviations decrease substantially, almost by 50%, between the smallest AOD bin and the bins with AODs about 0.15-0.2. This suggests increased uncertainty in derived AE values when AOD levels are low, which applies to both surface-based observations (Wagner and Silva, 2008) and MISR retrievals alike."

Page 12, line 9 and elsewhere (e.g. page 14, line 29): The NetCDF/HDF terminology for hierarchical levels used to separate different sets of data and metadata is "group." I think it's more appropriate to use this term rather than "subdirectory." <u>https://support.hdfgroup.org/HDF5/doc/Glossary.html</u> Re: Agreed. We changed "subdirectory" to group in two places in the text.

Page 24 Line 7: should be "northern hemisphere winter" Re: We don't distinguish between northern and southern hemisphere in the results shown in this section and in Figure 3.

Page 24 Line 23: "AOD peaks and valleys" => "AOD gradients" avoid using topography as metaphor Re: Corrected.