Response to Cynthia Twohy.

We thank the reviewer for his thorough examination of our manuscript. The reviewer's suggestions have led to a significant revision and expansion of our work.

In the revision of our manuscript we have explored the possibility of deriving the AF that best matches computation and experiment, on a case by case basis. This is the main upgrade of our work. In this new version, we have acknowledged that the presented β or δ measured-computed comparison with respect to (R_{th}, AR) can only constrain a range of R_{th}. In fact, from figures 2 and 3, you can see that R_{th} must be about 0.5-0.8 μm, while the compatible ARs are all those between 0.3-0.5 and 1.5-3. Hence, at that stage, a particular choice of AR is somewhat arbitrary.
In the revision of the paper, we have performed additional simulations, and for each of the experimental (PSD, β, δ) datum we have seeked the AR and R_{th} that best matches the measured δ with the T-matrix calculations. AR and R_{th} were allowed to vary respectively in the ranges (0.3-0.6)U(1.5-3) and 0.5-0.8 μm.
Once these AR and R_{th} have been found, the same were used to calculate β and compare it with the experimental, on a case-by-case basis.
Of course, this procedure leads to an improvement in the agreement between experimental and calculated δs, while it does not change appreciably the agreement between the βs.

We have described this new procedure in the Abstract:

> *The parameters R_{th} and AR of our model have been varied between 0.1 and 2 μm and between 0.3 and 3, respectively, and the calculated backscattering coefficient and depolarization were compared with the observed ones.*
> *The best agreement was found for R_{th} between 0.5 and 0.8 μm, and for AR less than 0.55 and greater than 1.5.*
> *To further constrain the variability of AR within the identified intervals we have sought an agreement with the experimental data by varying AR on a case-by-case basis, and further optimizing the agreement by a proper choice of AR smaller than 0.55 and greater than 1.5, and R_{th} within the interval 0.5 and 0.8 μm. The ARs identified in this way cluster around the values 0.5 and 2.5.*

In paragraph 2.3 Variability with the threshold radius R_{th} and Aspect Ratio AR:

> *The result of this study allows to identify only the best R_{th}, resulting around 0.5-0.8 μm, while the ARs compatible with the measurements are all those between 0.3-0.55 and 1.5-3.*
> *To further constrain AR we have kept R_{th} at a fixed value, chosen between 0.5 and 0.8 μm and changed this value with a 0.1 μm step. For each of these fixed R_{th}, and separately for each PSD, we identified in the intervals (0.3-0.55), (1.5,3) the value of AR which best matched the observed δ with its computed value. Finally, for each PSD we selected the pair R_{th} and AR which provided the best match. Once the ARs and R_{th} have been selected by forcing the agreement between the δ_A, the same ones have been used for the calculation of the β_A.*

Figures 4 and 5 have been upgraded with the result of this new approach:

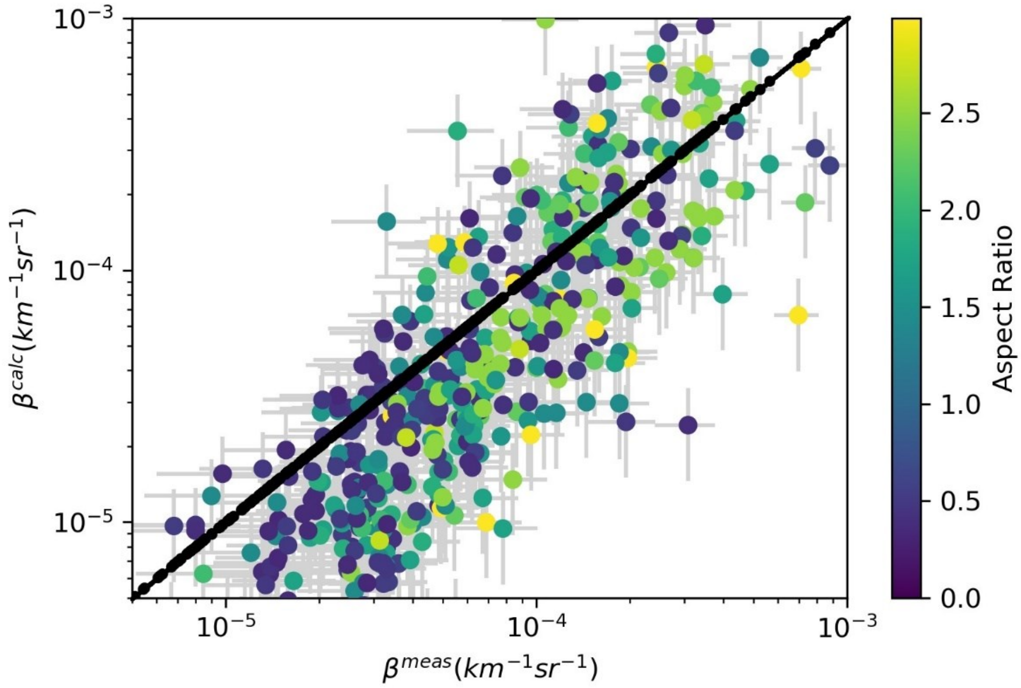

Figure 4. Scatterplot of computed vs measured particle backscattering coefficients β_A. The ARs used for the computations have been selected, case by case, to produce the best agreement between the β computed and measured, and are here represented in color coding. Only ARs in the intervals between 0.3 and 0.55, and between 1.5 and 3, have been considered. R_{th} was also selected within the interval 0.5-0.8 μm to provide the best match. We report data points with BR greater than 1.2, $\beta_{cross\ A}$ greater than 5 $10^{-6}km^{-1}sr^{-1}$ and temperature at the observation below 200 K.

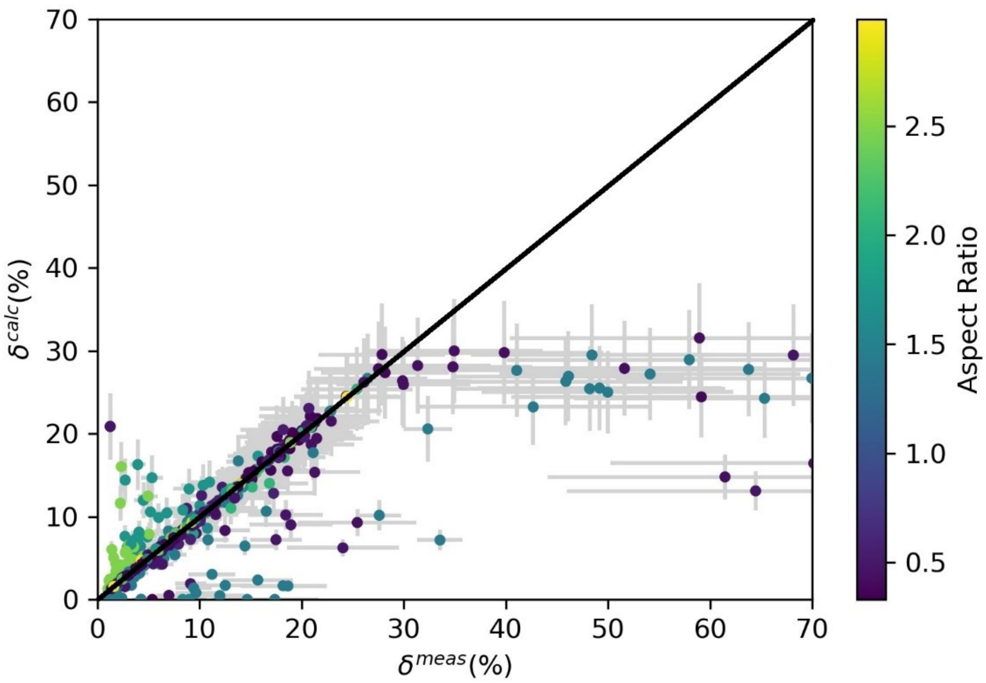

Figure 5. Scatterplot of computed vs measured particle depolarization δ_A. The ARs used for the computations are those that provided the best match between the δ_A computed and measured, and are here represented in color coding. Only ARs in the intervals between 0.3 and 0.55, and between 1.5 and 3, have been considered. R_{th} was also selected within the interval 0.5-0.8 μm to provide the best match. We report data points with BR greater than 1.2, β_{cross} A greater than 5 $10^{-6}km^{-1}sr^{-1}$ and temperature at the observation below 200 K.

We have also discussed the physical relevance of our results, in the par. <u>5. Discussion</u>:

The identification of the best R_{th} in the range 0.5 -0.8 μm supports what we already know from the theoretical understanding of NAT particle formation in PSC and from measurements (Deshler et al., 2003b). Concerning particle shape, in our model all solid particles in a single PSD share the same AR, but different PSDs can have different ARs. This approach could suggest that the choice of the AR which, case by case, optimizes the agreement between calculations and measurements, may be the result of chance rather than physics. There are two facts that counter this criticism.

First, it appears that the selected ARs may be related to the shape of the PSD. Figure 6 shows the 2D-histogram by occurrence of ARs and of $N(r > 0.7\mu m)/N_{tot}$, the ratio between particles with radius greater than 0.7 μm and total particles, which is a parameter related to the PSD shape. In Figure 6 the AR are not distributed randomly. Conversely, there is a tendency for the AR to grow as the percentage of large particles increases. In fact AR values tend to peak around 0.5 in the lower $N(r > 0.7\mu m)=N_{tot}$ range, while tend to cluster around 2.5 when $N(r > 0.7\mu m)=N_{tot}$ is higher. The shape of the PSD mirrors particle formation conditions and history, is linked to the presence of solid particles, as already highlighted in the discussion of Figure 1, and is likely linked to the average particle shape as well.

Second, if we consider the sequences of measurements acquired in individual balloon flights, the corresponding sequences of selected ARs do not evolve randomly but, conversely, are auto-correlated. An example of this behavior is shown in Figure 7, where the time series of β and δ are reported respectively with red and blue dots. The ARs that provide the best agreement between experiment and simulation are shown with black dots. It can be seen that temporally contiguous observations often result in the selection of the same AR. Contiguous observations of PSD are likely to have similar characteristics in terms of microphysics, and this seems to be correctly reflected in the constancy of AR. We are therefore confident that our method produces results with a physics-based content.

We report here as well, for completeness, new Figures 6 and 7:

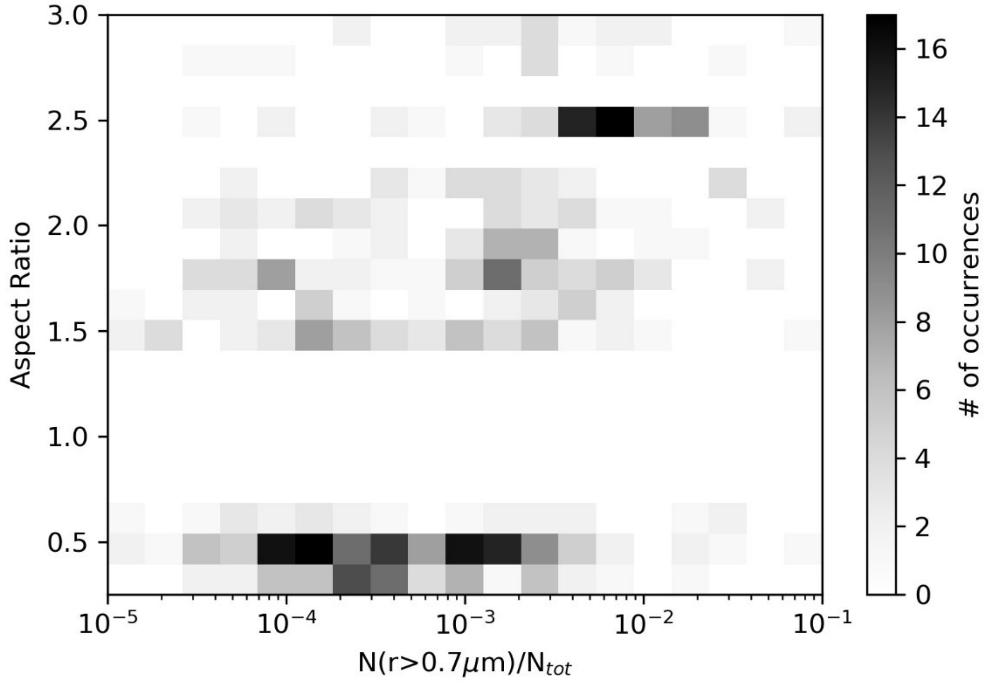

Figure 6. 2D-histogram of occurrence of ARs and of N(r < 0.7μm)=N$_{tot}$, the ratio between particles with radius greater than 0.7 μm and total particles. Only ARs in the intervals between 0.3 and 0.55, and between 1.5 and 3, have been considered.

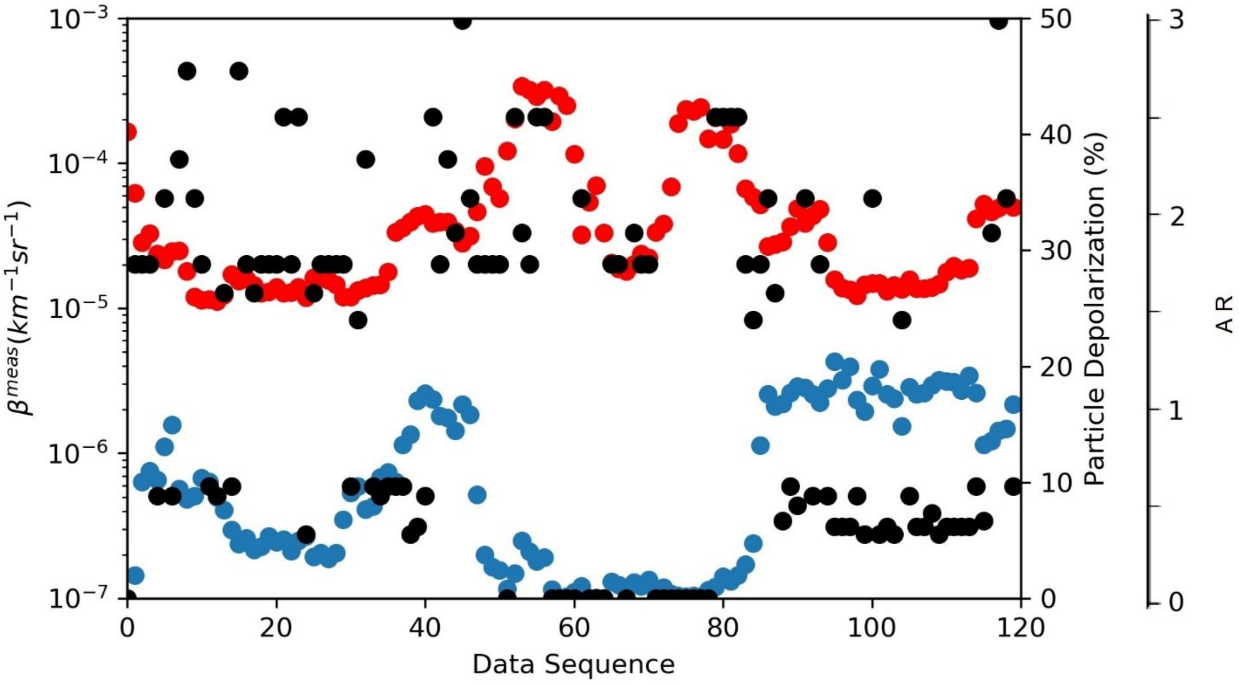

Figure 7: Sequences of b (red dots) and d (blue dots) measured on a balloon flight on December 9th 2001, from Kiruna, Sweden. Each data point represents an average over 60s.

Black dots represents the ARs providing the best match between the d and those computed from concomitant measurements of PSD.

the depolarization plots (Fig 5) do sometime show good agreement at low backscatter and high depolarization—it would be useful to examine these particular cases and see what is different. Under these conditions, what is really happening physically? Are these primarily large or small particles and what is the predominant phase/composition/ temperature that is contributing to agreement with the model? Likewise for the cases with poor agreement. For those cases, could a change in phase/shape/AR for some particles bring things into better accord? Plots of backscatter and depolarization for varying cases with associated size distributions (see below) could be helpful in determining what is causing the variations in model success.

The 4. Discussion paragraph has been extensively rewritten and is now addressing more closely the reviewers remarks.

In general, our model leads to good correlations between measured and modeled βs. For the δs the measurements are well reproduced by the calculations in many instances, as is the case for many of the selected ARs in the range 0.3-0.55. However, there are other cases in which the agreement is worse (when the best ARs have been selected in the range 1.5-3), or does not occur at all, as in the cases of observed depolarizations greater than 30%. In these latter cases, the impossibility of reproducing the observed values even under the hypothesis of a completely solid particles implies that, for those PSDs, our model is not able to produce the observed depolarizations. In these particular cases in which the model performs particularly badly, there may be problems of inhomogeneities of the cloud. These cases come from Antarctic observations, for which the microphysical observations from the balloon and the optical ones from ground-based lidar are separated geometrically, so that the two instruments sample air masses separated by several tens of kilometres, and it may be the case that some clouds were not homogeneous on such spatial scales.

For what concerns the possibility of implementing a size-dependent AR approach, we write:

Different shapes produce different polarization, according to T-Matrix. This has also been proven experimentally since the early work of Sassen and Hsueh (1998) and Freudenthaler et al. (1996) that showed how lidar depolarization ratios in persisting contrails ranged from 10% to 70%, depending on the stage of their growth and on temperature. In the T-matrix theory, for fixed AR, the depolarization depends on the particle size and maximizes for particular sizes. There is certainly a way to assume a particle size-dependent AR in our PSDs so as to reconcile the computations with the observed values. However, such an approach would have little physical basis and could only be justified to maximize the agreement of calculations. Therefore, we have not explored this possibility further, although it is possible that our simplified hypothesis of a common AR for every particle may be the cause of the bad agreement between data and calculations in some case.

I also suggest that the authors show some particle size distributions to give an idea of the actual distribution of particles within the sampled PSCs.

Concerning the shape of the PSDs which is suggested to display, our study is based on 473 data points (i.e. 473 triplets of PSD, backscattering coefficient, and depolarization). A three panel plot showing: i. the time series of PSD (color plot), ii. the corresponding backscatter coefficient (line plot); iii. the corresponding depolarization (line plot) could be produced, but given the range of the observations such representation may be difficult to interpret. However, in the revision of the manuscript, when discussing the AR distribution in new Figure 6, we have correlated it with a parameter characterizing the shape of the PSD.

> **How well are they represented with a unimodal or bimodal lognormal distribution? Could differences in the fit contribute to some of the errors in calculated results?**
> **What is the uncertainty in the size distributions themselves (as well as in the other measurements)? These are likely to have significant error as well, particularly for larger, aspherical particles. By including the measurement uncertainties and combining them with the model uncertainties such as in refractive index, you could add error bars to Figures 4 and 5 and have a better idea how the model is really performing.**

The treatment of uncertainty has been elaborated, and error bars have been used in the new figures 4 and 5. Errors have been discussed in <u>2.1 Dataset:</u>

> *Experimental errors in the particle Backscatter Ratio (R-1) are estimated to be 5%, but not less than 0.05 in absolute value, while the error in volume depolarization is about 10%–15%. Additional uncertainty comes from the determination of pressure and temperature from radiosoundings, needed to compute β_A and δ_A (Adriani et al., 2004).*
> *[…]*
> *Particle size histograms are fitted to unimodal or bimodal lognormal size distributions, which are the representation of size distribution used in this work. The uncertainties on the determination of the parameters of the mono/bimodal lognormals were determined by Deshler et al. (2003b) with Monte Carlo simulations. These were 20% for distribution width, 30% for median radii and 10% for modal particle concentrations.*

And in <u>3. Results</u>:

The uncertainties associated with the measured β_A and δ_A derive from the error analysis for the single lidar data, which can be found in Adriani et al. (2004) or from the standard deviation for the averaged data, depending on which is greater. The uncertainties on the calculated β_A and δ_A, are 40% as determined by Deshler et al. (2003a) for any moment of a PSD derived from the OPC measurements. Deshler et al. determined this through a Monte Carlo simulation which used the uncertainties of the OPC size and concentration measurements to quantify the uncertainties in the PSD parameters and their subsequent moments.

> **Figs 4 and 5: Calculated vs measured backscatter coefficients are shown but there are no regressions to evaluate the goodness of fit.**

We have now provided the Pearson correlation coefficient (resulting to be 0.56) for the goodness of the 1:1 fit for the β comparison and added 1:1 lines to the data in Figs. 4 and 5.
We did not perform goodness-of-fit tests for the comparison of δs. In this case it is clear that there is a set of well-aligned points along the 1:1 line, and sets of points that deviate from it in a non-

random way. We have discussed the different characteristics of these sets in the 3 Result paragraph, totally rewritten, which we report here in its entirety:

3 Results

Figure 4 reports the scatterplot of measured vs computed β_A, colour coded in terms of AR. The figure represents the analogue of Figure 4 in Snels et al. (2021), where in the present case we have used a larger dataset, including now four Arctic balloon flights, and used T-Matrix instead of a factor 0.5 reduction in the Mie backscattering. Figure 5 reports the scatterplot of measured vs computed δ_A similarly color coded in terms of AR. The uncertainties associated with the measured β_A and δ_A derive from the error analysis for the single lidar data, which can be found in Adriani et al. (2004) or from the standard deviation for the averaged data, depending on which is greater. The uncertainties associated with the measured β_A and δ_A derive from the error analysis for the single lidar data, which can be found in Adriani et al. (2004) or from the standard deviation for the averaged data, depending on which is greater. The uncertainties on the calculated β_A and δ_A, are 40% as determined by Deshler et al. (2003a) for any moment of a PSD derived from the OPC measurements. Deshler et al. determined this through a Monte Carlo simulation which used the uncertainties of the OPC size and concentration measurements to quantify the uncertainties in the PSD parameters and their subsequent moments.

Despite the dispersion in Figure 4 the points cluster around the straight line $\beta_{calc}=\beta_{meas}$, indicating the agreement between computation and measurements can be considered fine for β_A with the exception of β values below 4 $10^{-5}km^{-1}sr^{-1}$ where the β_{calc} underestimate the measurements. Such underestimation seems to be of the order of $10^{-5}km^{-1}sr^{-1}$, a magnitude compatible with possible inaccuracies in the calibration of the lidar data. The Pearson correlation coefficient for the entire dataset is 0.56, and increases if the lower values of β are neglected.

The δ_A scatterplot shows the presence of a good number of points that align along the $\delta_{calc}=\delta_{meas}$ correlation line, with AR selected mainly around the value 0.5. However, for depolarization values greater than 30% there is no AR that will reproduce the measurements. These points correspond to those presented in Figure 1, with low values of BR and high values of the concentration ratio of large to total particles. They mainly come from three single observational periods of about one minute each, characterized by air temperatures between 184-188 K. Given the magnitude of the depolarization, it is possible that those observations are not referable to clouds in mixed phase, but rather to clouds of predominantly solid particles. For that particular set of points, we also explored the possibility that all particles were solid, but even under this assumption the comparison with the experimental data did not improve appreciably.

In Figure 5 for depolarizations lower than 15%, the points which deviate, by excess or defect, from the 1:1 straight line have predominantly AR greater than 1.5. So it seems that selected ARs greater than 1.5 generally produce a worse correlation. From Figure 4 we observe that AR values in the range (0.3-0.55) tend to be associated with medium-low β values, while AR values in the range (1.5-3) are mainly associated with medium-high β.

To conclude, the choice of R_{th} in a range between 0.5 and 0.8 μm leads to a reasonably good agreement between the β's, but there seems to be a discrepancy between the calculated value and the measurements in their lower range of variability.

From Figure 4 such mismatch, which makes the measurements larger than the calculations, seems to be of the order of $10^{-5}km^{-1}sr^{-1}$. The selection of the AR that produces the best agreement with the observed δ's leads to three results: i. The ARs in the range 0.3-0.55 tend to be selected in correspondence with medium-low β's, the ARs in the range 1.5-3 in *correspondence with medium-high β's. ii. ARs in the 0.3-0-5 range reproduce the measurements well, except for some observations where the depolarizations are greater than 30%; iii. the ARs in the 1.5-3 range reproduce the measurements less well; iv. There is no AR that will allow the calculations to reproduce the measurements for depolarizations greater than 30%.*

The plots at top and bottom of Fig 4 look identical except at low backscatter—is this correct?

Yes, it was. However, we have updated figure 4 with the new results.

the caption for Fig 5 says backscatter coefficient, when what is plotted is depolarization.

We are sorry for this. We have corrected our mistake.

Finally, we wish to thank the reviewer for her patience in spotting out all our typographical and grammatical errors. We apologize for the poor quality of the written English, responsibility of the first author only. This has been corrected in the revised manuscript.