Dear Reviewer #1,
We thank you very much for carefully reading our manuscript and for your numerous comments and suggestions! In this reply letter, the comments of Reviewer #1 are given in **black** and our answers are written in **green**. The given line numbers refer to the revised version of the manuscript (without markup).

General comments

1. As mentioned above, the English needs to be improved prior to publication

—> Thank you again at this point for the many comments and suggestions concerning the linguistics of the draft. We implemented all of your suggestions and corrected a few more mistakes that we or Reviewer #2 found.

2. I found myself asking the question “why not estimate the maximum diameter of the 3D volume reconstructed following the method 2DVD uses (where each layer is assumed to be an ellipse) instead of using the Feret diameter approach?” This is not to say that I don’t think the Feret diameter approach demonstrated here is without merit, just that the authors may wish to address the choice of using the Feret diameters rather than the 3D reconstruction. I wouldn’t be surprised if a 3D reconstruction of snowflakes and aggregates is prone to overestimating the volume (and presumably the maximum dimension) and referencing literature discussing the issues with the 2DVD’s 3D reconstruction for snowflakes may be a helpful way to support the value of using the Feret diameter approach rather than a 3D reconstruction approach.

—> The 3D reconstruction based on stacking the ellipses on each other is a good approach for liquid droplets. However, in case of ice crystals, the volume is, as you supposed, largely overestimated. This is shown, for example, by the studies from Brandes et al. (2007) or Zhang et al. (2021) who experimentally investigated the 2DVD bulk density of ice crystals and found strong dependencies on the particle size, relative humidity, and further variables. Bulk densities down to 0.1 g/cm^3 were found under certain conditions. This is roughly 1/10 of the density of ice which means that those 2DVD particle volumes are differing up to one order of magnitude from the true particle volume. As Reviewer #2 pointed out that our precipitation rate in Fig. 9 is affected by that, we implemented the bulk density — size relation found by Zhang et al. (2021) into our precipitation rate calculation (Fig. 9 and Eq. 4). Following your suggestion, we added some sentences at the beginning of Section 3.1 in which we explain why we refrained from a 3D reconstruction (lines 238-242).

3. The manuscript would benefit from a discussion of how the maximum diameter estimate will be biased by horizontal particle motion. The bias due to horizontal motion should be fairly easy to estimate if the horizontal motion of the particle is known (obviously not possible with the actual observations, unfortunately). I feel like it should be fairly straight forward to derive an equation to relate the horizontal

particle motion to the resulting bias in the Feret diameter for an individual camera and such an equation may prove useful to users looking to understand the error characteristics of this new measurement.

—> Indeed, the maximum diameter retrieval would strongly benefit from such a correction. Helms et al. (2022) have applied an unskewing algorithm to 2DVD data and found a slightly better agreement to the simultaneously acquired PIP (Precipitation Imaging Package) data. Their method works in a way that the upper and lower most detected pixel of the 2DVD image should be vertically aligned. To our opinion, this makes much sense in case of droplets. In case of ice crystals, however, uncertainties should remain high. Helms et al. (2022) accordingly also conclude in their "Conclusions" Section, that "the corrected bounding-box-width measurements are still prone to error due to the motion skewing effects" after the application of the unskewing algorithm. For this reason, we decided to not deal with skewness correction at this point, while we added the potential of a skewness correction to the Conclusions section in line 496.

Specific comments

Title: The title would benefit from some mention of the maximum diameter estimate since introducing and evaluating this estimate is the core goal of the paper

—> We also thought about including the maximum diameter in the title. However, as the content also heavily deals with cloud microphysical studies in which also remote sensing instruments as well as HOLIMO are playing a crucial role, we would like to stay with the current title.

Line 112: Was AMF-1 deployed on Polarstern? If so, it may help to remind the reader of this by adding something like "aboard Polarstern" to the end of this sentence.

—> Yes, this paragraph is only about MOSAiC and Polarstern. Nevertheless, we added "aboard Polarstern" for clarification, as suggested. (line 117)

Line 123: I like that you mention the wind speeds as this was something I was wondering about. As a side note, removing "the conduction of" might improve the clarity of this sentence.

—> "the conduction of" is now removed.

Line 159: I'm guessing "dropped" would probably be a more accurate word than "thrown" here (as well as on line 163), but I'm not familiar with the specifics of the 2DVD calibration (beyond what has been mentioned here, of course). Also, are the spheres dropped one at a time or in batches? Later in the text, it is mentioned that the spheres might stick together, which makes me think they might be dropped in batches. It's probably worth clarifying this point in this section.

—> "thrown" was replaced by "dropped" (line 163) and by "dropped by hand" (line 167).

Lines 232 – 233: Given the importance of the Feret diameter to this paper, I think it may help the readers to contrast the Feret diameter with a traditional diameter (I had to search online to

understand the difference). Specifically, assuming my understanding of Feret diameter is correct, that the Feret diameter is the diameter of a 3D object from a single point a view.

—> Thanks for this important hint. We decided to include the circumscribing sphere diameter (dc) in this study. To summarize the results on that: We found that the maximum Feret diameter is very similar to dc. We included dc in Fig. 5 which is showing the similarity of dmax and dc for the calibration sphere data set. Our motivation for introducing the maximum Feret diameter was that, from a physical perspective, the definition of a maximum diameter should be the distance of its two outermost points. For that reason, we think that it is sufficient to briefly show the similarity between dmax and dc and to proceed with dmax in the paper. We also briefly investigated the difference of dmax and dc for both case studies and found that the two diameter types are almost identical for the snow measurements as well (see Fig. R1).

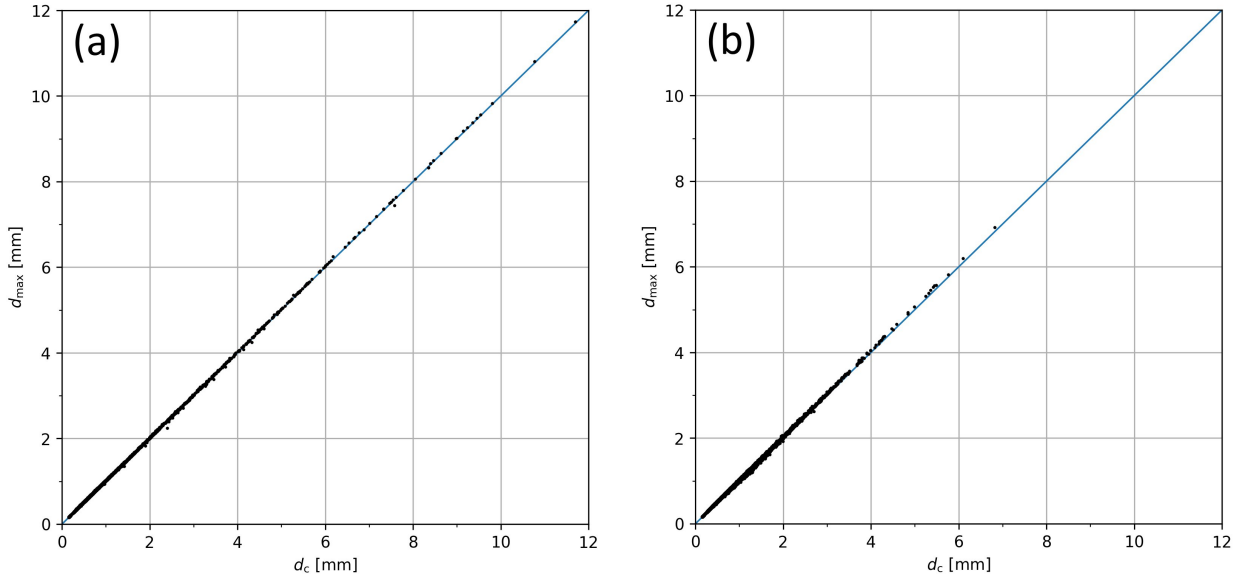

Figure R1: 2DVD dmax versus dc (a) on 17 January 2023 during PolarCAP/Cloudlab in Eriswil, Switzerland and (b) on 10 November 2019 from 11 to 15 UTC during MOSAiC.

Section 3.3: If you can find the information in the literature, it might be helpful to briefly mention what data was used to develop the LIRAS-ice retrieval. Given that this manuscript is introducing the Feret-diameter-based maximum diameter estimate, the typical concern about using the training data to validate an algorithm isn't an issue. That said, explicitly stating what instrument was used as the ground truth for the LIRAS-ice algorithm development would still be nice. Additionally, it would be helpful mention the vertical and temporal sampling intervals for the LIRAS-ice data in this section (alternatively, the LIRAS-ice data set could potentially be added to Table 1 by changing the "Instrument" column heading to "Data Source" or something similar). Finally, please also mention which instruments are used here to provide the input data for the LIRAS-ice retrieval (my guess would be the PollyXT lidar and the 94 GHz radar).

—> Thank you for pointing out that this Section needed some additional explanations. To clarify which instruments were used for the development of LIRAS-ice, the following

sentence was added in lines 294 to 295: “For the development of their retrieval, Bühl et al. (2019) used data of a PollyXT Raman lidar, a 35GHz cloud radar, and a radar wind profiler (RWP), albeit the usage of the RWP is optional.”
—> For the development of LIRAS-ice, clouds were investigated which did not precipitate down to the ground. Therefore, no ground-based in-situ instrument was used. This was also a particular motivation for us to compare the 2DVD data to the LIRAS-ice results.
—> To inform about the instruments and the sampling intervals in this study, the following sentences were added in lines 314 to 317: “For this study, the Raman lidar PollyXT and of the KAZR cloud radar which were operated during the MOSAiC campaign provide the observational basis (see Table 1). The vertical and temporal sampling intervals of the retrieved variables equal those of the processed cloud radar data which are approximately 30 m and 30 s, respectively.”

Line 310: Regarding “spheres that either cling together”, this is where I was unsure how the 2DVD calibration is performed. If the spheres are dropped one at a time, then spheres clinging together either shouldn’t be an issue or would be a bad calibration that should be excluded from the calibration data. Please clarify this is section 2.3.2.

—> The calibration was done again on 11 November 2025 under calm conditions and with particular care. This improved the results a lot by means of a much smaller distribution for each single sphere size and also the number of outliers got reduced a lot. Nevertheless, the smallest sphere sometimes stick to other spheres due to electrostatic charge or magnetism and can easily be overseen. Following your suggestion, they were manually filtered from the new dataset.

Line 324: Regarding “assumed to represent columnar crystals”, the wording here makes it unclear how thorough the authors were in determining what type of particles correspond to O <= 0.6. It might be helpful to expand on this slightly, even if just by mentioning that the authors visually inspected a sizable subset of the particles during this period and found columnar crystals to be the dominant particle type. The current wording also suggests an absoluteness that should probably be softened given that (presumably) the authors did not exhaustively examine all the particles with O<=0.6 that were observed by the 2DVD during this period. Also, is there a reason that these could not be needles?

—> We changed this sentence to “Characteristic examples from the two clusters show that many particles with O ≤ 0.6 have vertical extents of only a few pixels while reaching horizontal lengths of up to several millimetres (Fig. 8d). Therefore, their majority can be assumed to represent columnar crystals or needles.” (lines 351-353).

Figure 8a: Consider adding a horizontal reference line indicating O = 0.6 (I leave this up to author preference)

—> We added the horizontal dashed line.

Lines 341 – 343: I feel like there’s some ambiguity in the sentence that is leaving space for incorrect interpretations. In my experience, ice clouds will generally have a lower reflectivity than

liquid water clouds and snow will generally have a much lower reflectivity than liquid precipitation. I suspect what the authors are trying to say here is that the reflectivity increase was due to the ice crystals at the top of the cloud growing to precipitation-sized ice particles at the expense of liquid cloud droplets. It might also help with clarity to specify that these were supercooled liquid water droplets (assuming that was the case).

—> This is true. This one and the following sentence were modified so that the statement became more clear.

Line 378: It would be helpful to mention how these temperature measurements were obtained (e.g., balloon soundings?). Also, I suggest replacing “Within” with “Below”.

—> It is now mentioned that radiosonde data were used and a citation (Maturilli et al., 2021) was added. “Within” was replaced with “Below”. (lines 410-411)

Line 384: Please mention the height of the lowest height bin used in the average.

—> “starting at around 180 m,” was inserted (lines 416-417)

Lines 384 – 385: In the discussion, the authors mention that vertical wind shear was a concern. If this is the case, have you considered using only the lowest few height bins in the average and averaging over time instead to reduce the noise?

—> Yes, different options regarding the height bins were tested. One height bin is around 30m deep. So our range from around 180m to 400m equals the lowest seven height bins. We found that using only one height bin was disadvantageous because data gaps were too numerous and, especially in the case of the lowest height, bins with unrealistic results (‘outliers’) occurred too often. Further, we found that averaging to 300m, 400m, or 500m did not change the results much so that we chose 400m.

Lines 391 – 392 regarding the 60 s lag correlation working best: Based on Fig 10a, the typical fall speeds of the ice particles were around 1 m/s at the surface while, according to the discussion section, the LIRAS-ice data is averaged between 180 and 400 m heights. In a theoretical situation with zero vertical wind shear, my expectation would be that the shortest physically meaningful lag time would need to be 180 seconds (and somewhere around 290 seconds would probably be more reasonable) as that is how long it takes the particles to reach the 2DVD after falling from the lowest point in the LIRAS-ice data used in the correlation. Adding in wind shear complicates this, of course. A brief search suggests there were regular radiosonde launches from Polarstern, so it should be possible to for the authors to examine the wind shear vectors (in a du/dz and dv/dz sense) and compare them to the ship’s direction of travel to determine how the wind shear profile would affect the particles as they fall (possibly by assuming a 1 m/s fall speed and integrating the wind shear vertically). Even if the authors decide not to go so far as to estimate the expected lag based on shear and fall speed, I feel the result warrants further discussion that includes evidence to support the relatively short lag time.

—> Thank you for pointing out this interesting aspect. There was a balloon launch at 10:34 UTC and another one at 16:35 UTC. The radio sonde from the 10:34 UTC launch is showing wind speeds from 2.8 m/s (at 22m) to 5.7 m/s (150 - 200m) and directions from 164° (at 10m) to 133° (at around 180m). This means that there is both a directional as well as a

velocity wind shear between the ground and the lowest LIRAS-ice height bin. In addition to that, the wind direction has changed during the investigated time period: The 16:35 UTC launch shows directions from 170° at 20m to 182° at 180m. Between the two launches, the direction changed from 133° to 182° at 180m but remained approximately constant close to the ground which is likely reflected in the different slopes of the maximum radar reflectivities in Fig. 11 a. At 11 UTC, the slope of the fall streak is significantly different from one of the fall streaks at 13 UTC, for example. We conclude that the 60s can still considered to be plausible. This is mainly because it is an average value for which the correlation along the whole investigated time period is the best. One would, for example, only need a short time period in between, where 2DVD and LIRAS-ice investigated different particles which could affect the average correlation. Moreover, the correlation of the neighbouring time shifts (30s, 90s, etc.) was not differing too much from the one for 60s. We added the sentence “This value can be considered plausible given the presence of vertical wind shear (Maturilli et al., 2021)” in line 424.

Lines 405 – 407: Unless the calibration occurred under particularly windy conditions, shouldn’t the lateral movement be fairly small? I know the spheres are very small, but the distance they need to fall also looks fairly small based on Fig 2c. If you know when the calibration occurred, it might be worth checking the corresponding wind data. Also, the authors should mention the errors introduced when a small sphere is pixelized (I’m pretty sure this was mentioned earlier in the paper, but it is worth repeating here)

> —> We repeated the calibration under zero-wind conditions. Nevertheless, a few spheres appeared skewed in the data. We think that this can be explained by the fact that they were dropped by hand through the instrument. The calibration pattern from the one figure can only be used for the 10mm spheres but the others must be dropped by hand as the manufacturer does not provide a separate calibration pattern or similar. Therefore, we assume that using an adapted calibration pattern for all spheres would improve the calibration results even further. We modified this part of the discussion and present the improved results here (lines 436-441)).

Lines 419 – 422: This paragraph strikes me as circular reasoning: the 2DVD Feret-based maximum diameter estimate is good because it matches LIRAS-ice and the LIRAS-ice data is good because it matches the 2DVD Feret-based maximum diameter estimate. While I’d agree that the agreement of the two gives greater credibility to the individual data sets, presumably the LIRAS-ice data has been validated elsewhere in the literature and the agreement of the two is simply showing that the 2DVD Feret-based maximum diameter estimate is reasonable.

> —> Indeed, the LIRAS-ice retrieval has not yet been properly evaluated against any in-situ sensors. Only Ansmann et al. (2025) compared the ice crystal number concentration in cirrus clouds with the 2DVD particle number concentration on the ground and found that they agree within one order of magnitude or better. This is the reason, why we decided to perform this comparison between LIRAS-ice for the lowest height bins and 2DVD.

Lines 423 – 429: The authors might also want to mention the benefits of averaging in reducing noise, which was listed as the main motivation when talking about the averaging in the main body of the manuscript.

—> This is an important aspect, of course. We added “considering only one height bin would result in too strong temporal fluctuation of ICNC and maximum diameter.” (line 459).

Line 427 – 428 regarding “decreasing vertical resolution”: If this is a vertically pointing radar, I don’t think the vertical resolution would change with range. The across-beam resolution would change (i.e., horizontal resolution for a vertically pointing radar), perhaps this is what the authors meant?

—> You are right, the vertical resolution itself does likely not change in that height range. To generalise and simplify the statement, we changed “signal saturation near the instrument, ground clutter, the radar’s blind zone, and decreasing vertical resolution” to “near-field effects and beam geometry”. (line 463)

Technical Corrections / Suggestions

Line 15: “instruments” should be “instrument’s”

—> We followed the suggestion.

Table 1 Caption: Should “developed” be “deployed separately”? The first and second halves of the second sentence don’t seem to match up with one another if the word “developed” is the correct word. Also, mention in the caption that the superscripts indicate the instrument location.

—> We followed the suggestion, changed “developed” to “deployed separately” and added the sentence “The superscripts indicate the platform of each instrument and therefore its measurement location.”.

Table 1: “Resolution” isn’t really an accurate heading for this column. I’m not entirely sure what the best heading would be, however. The values appear to be a mixture of gate spacing, sampling frequency, uncertainties, and limitations. Maybe “Measurement Parameters”? Not sure.

—> We changed “Resolution” to “Specifications”.

Line 130: Suggest changing “the instrument is designed advantageous” to “the instrument design is advantageous”

—> We followed the suggestion.

Line 132: Suggest replacing “would not be processed” with “will have minimal impact on the measurements.”

—> We followed the suggestion.

Lines 132 – 133: Suggest replacing “that snow accumulates” with “snow accumulation” and “and gets blown into the virtual measuring area by wind gusts” with “that could otherwise interfere with measurements”

—> We followed the suggestion.

Line 137: Wouldn’t the constructed particle shape be two-dimensional rather than one-dimensional?

—> This is of course true, we changed “one” to “two”.

Line 140: Suggest “there are” be added before “several criteria”

—> We followed the suggestion.

Line 140: Suggest “needs to fulfil are considered” be replaced with “must fulfil”

—> We followed the suggestion.

Line 142: Suggest changing “In case” to “In the case”

—> We followed the suggestion.

Line 185 – 186: Regarding “horizontal winds that tilt falling hydrometeors”, if you are referring to the effects of horizontal motion on the reconstructed 2DVD image, “skew” is probably a better term than “tilt” just to clarify that you are not referring to the physical snowflake being rotated.

—> We followed the suggestion and reworded the sentence to “However, for solid particles, the applied method to calculate O can be highly error-prone, especially at significant horizontal winds which may let falling hydrometeors appear skewed in the camera images.” (lines 188-190).

Line 187: I think this is a stray fragment that didn’t get deleted during a previous edit of the manuscript.

—> This was shifted to the caption of Fig. 4 and the code is now published on Zenodo.

Line 219: Suggest replacing “2 mm of a well defined” with “2 mm within a well defined”

—> We followed the suggestion.

Line 232 – 233: Suggest adding “the” before “Feret diameter”

—> We followed the suggestion.

Line 241: Suggest removing “by” from “width is usually by many times larger”

—> We followed the suggestion.

Line 252: Suggest changing “height and particle” to “height while particle” and adding “than this” after “typically higher”

—> We followed the suggestion.

Line 265 (and elsewhere): I suspect that “dominant” is the word the authors are looking for rather than “dominating”

—> We replaced “dominating” with “dominant” here and at nine further positions in the manuscript.

Line 280: Suggest replacing “resembles best” with “best resembles”

—> We followed the suggestion.

Line 285: Suggest removing “further” (or replacing it with “hereafter”)

—> We followed the suggestion and replaced it with “hereafter”.

Lines 286 – 300: These paragraphs feel like they belong in a case selection section rather than as part of the current section

—> Due to a comment of Reviewer #2, this Section was revised. However, as we systematically applied this method throughout all case studies, we would like to keep it in the “Methods” Section. To formulate it more as a method, we changed the Section name from “Dominating particle shapes in 2DVD data” to “Identification of dominant particle shapes in 2DVD data”.

Line 292: Suggest removing “yet”

—> We followed the suggestion.

Line 299: Suggest changing “met the requirements most” to “best met the requirements”

—> We followed the suggestion.

Line 306: Suggest removing “well” from “can well be identified”

—> We followed the suggestion.

Line 346: Suggest replacing “according” with “corresponding”

—> We followed the suggestion.

Lines 348 – 367: I feel like these paragraphs are broken up in the wrong locations. Perhaps including the first sentence of the paragraph starting on Line 353 with the previous paragraph and moving the rest of this paragraph (i.e., “The relation of…” and onwards) to the start of the following paragraph?

—> We followed the suggestion.

Line 353: Suggest removing “the three-minute” and then adding “(three minutes each)” after “periods two and three”. Currently it sounds like there’s an extra three-minute period before the second and third periods.

—> We followed the suggestion.

Line 357: Suggest adding “the” before “major”

—> We followed the suggestion.

Line 366: Replace “constrain” with “constraint”

—> We followed the suggestion.

Line 373: Add “a” before “ground-based”

—> We followed the suggestion.

Line 380: Replace “on the ground” with “at the surface”

—> We followed the suggestion.

Figure 11: Suggest using the same color bar for both Fig 7 and Fig 11

—> We followed the suggestion.

Figure 11 caption: I suspect the word “example” is a better fit for the authors intent than “exemplary”

—> We followed the suggestion.

Line 390: Suggest replacing “which” with “required for” and removing “need”

—> We followed the suggestion.

Line 421: Suggest replacing “needful” with “useful”

—> We followed the suggestion.

Line 429: This sentence is a bit awkward and could use rewording

—> The sentence was reworded to “Therefore, the choice of a height range of 180\,m to 400\,m for averaging the LIRAS-ice results is well justified.” (lines 463-464).

Line 431: This sentence lists a 2 minute lag, but the main body of the text said a 60 s lag was used.

—> 60 s is the right value, so “2 minutes” was replaced by “60 s”

Line 437: Should the letter “V” instead be a lower-case nu?

—> Here, the volume is meant as deq is calculated from it. Some words were added for clarification.

Lines 453 – 455: This sentence is a bit awkward and could use rewording

—> The sentence was reworded to “The strong agreement with remote sensing observations supports including the instrument in future studies to enhance information on precipitation particle shape, size and number.” (lines 488-490)