An introduction of the Three-Dimensional Precipitation Particles Imager (3D-PPI)

Response to the reviewers

Jiayi Shi, Xichuan Liu, Lei Liu, Liying Liu, Peng Wang

Original Referee comments are in italic

manuscript text is indented, with added text underlined and ~~removed text crossed out.~~

We would like to thank the reviewers for their very helpful comments. We revised the manuscript thoroughly and responded to all of the reviewers' comments.

Important points in the order they appear in the manuscript:

1) Previous instruments are presented such as the VISSS and the MASC. When describing the MASC the following statement is done (73-75): "Nevertheless, only 10^2–10^4 particles were observed during a typical snowfall event (Gergely and Garrett, 2016), which is insufficient to permit the reliable estimation of the particle size distribution (PSD) (Gergely and Garrett, 2016)." I don't see why this sample size would be insufficient and I don't see Gergely and Garrett, 2016 saying that. It is a bit misleading to look at sampled number of particles during specific storms rather than snowfall rates or snow number concentrations. The authors should instead look at the observation volume or something similar if they want to put the MASC in relation to 3D-PPI. The MASC has three cameras, as the 3D-PPI, but they are all located in one plane. However, the MASC was extended to five cameras with the additional two not being in the same horizontal plane as the original three cameras (Notaros et al. 2016, Kleinkort et al. 2017).

Thank you for your comments. We apologize for our mistakes. The reference we cited does not mention the relevant information that we stated. We did intend to compare 3D-PPI to MASC in terms of the particle capture efficiency. The sampling area of MASC is 2.5 cm^2 with a frame rate of 2 fps, while the 3D-PPI has a larger sampling (observation) volume of 1464 cm^3 (due to the use of a telecentric lens) and a higher frame rate of 5 fps.

The MASC captured 10,000 images during nearly 2 months, excluding out-of-focus particles, which were captured at very few particles per minute. On the contrary, one of the advantages of 3D-PPI is that in just 13 hours we recorded over 880,000 snowflakes. Therefore, for 3D-PPI, it is sufficient to reliably estimate a PSD on the minute temporal scales needed to capture changes in precipitation particle properties. Additionally, we

are not comparing the two instruments in terms of camera position design here. We have modified the sentences in the revised manuscript (72-77).

~~Nevertheless, only 102–104 particles were observed during a typical snowfall event (Gergely and Garrett, 2016), which is insufficient to permit the reliable estimation of the particle size distribution (PSD) (Gergely and Garrett, 2016).~~ Incorporating the rich dataset from the Multi-Angle Snowflake Camera (MASC), the 3D-GAN model is adeptly trained to reconstruct the intricate three-dimensional architecture of snowflakes, thereby unlocking new dimensions in the study of snowfall microphysics (Leinonen et al.,2021). Furthermore, the MASCDB, is a comprehensive database of images, descriptors, and microphysical properties of individual snowflakes in free fall, as presented by, showcases the MASC's exceptional potential for contributing to the field of atmospheric science by providing an extensive and detailed resource for studying the microphysical properties of snowflakes (Grazioli et al.,2022).

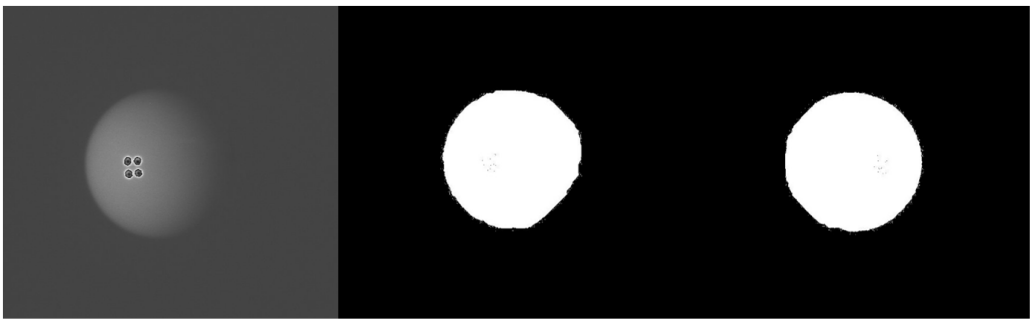
2) a) What is the actual resolution (combination of optical resolution and illumination that allows details of a certain size to be resolved)? What are the smallest details that can be resolved? This is not discussed but an important detail of a new instrument. I can only guess from the images (mostly Fig. 1c). The images look similar to MASC images of snow, likely due to a similar illumination scheme. The smallest details are very faintly grey and seem to disappear in the black background. Due to this and due to the missing description of the detection algorithm and thresholds (see further comment on image processing (comment 8) below), it is not possible to do a fair judgment of the actual resolution for detection of small details (e.g. thin branches).

Thank you for your comments. The pixel resolution of 3D-PPI is 41.6 $\mu m \cdot px^{-1}$, which theoretically means that the smallest recognizable details (e.g. thin branches) of a large snowflake is 0.0416mm. The actual resolution is influenced by both the optical resolution and the illumination scheme. The illumination provided by the LED array is designed to enhance the contrast and visibility of snowflakes, but it can also affect the perception of smaller details, particularly those that are faint against a dark background. The faint gray details you observed in Fig. 1c may indeed be challenging to discern, which can complicate the assessment of small features like thin branches. We are currently refining our detection algorithm and thresholds of image processing algorithms to improve the identification of these subtle details.

b) Judging from Fig 6b, the sizes of the smallest ceramic spheres are underestimated. For each camera, the linear fit has a negative offset (see comment below) with the observed size being on or below that fit. So, there seems to be a systematic bias for sizing of the smallest details. Please comment on this.

Thank you for your comments. We acknowledge that the reflectivity of ceramic spheres differs from that of snowflakes. The smooth surface of the ceramic spheres can lead to variations in brightness across different areas in the ceramic spheres' image. Some regions appear very bright, while others may appear quite dark. The darker areas are more susceptible to being misidentified as background, which can contribute to an underestimation of their sizes. Before the linear fitting in Fig. 6b, we performed manual

adjustments to mitigate the software's misidentifications (illustrated in the following figure). However, this systematic bias in the sizing of the smallest details remains a challenge.



The image of ceramic spheres **Detect by software** **Detect manually**

In contrast, when capturing real snowflakes with rough surfaces, their diffuse reflection may result in a more uniform brightness distribution on a single snowflake image, Therefore, the systematic bias of the real snowflakes is expected to be much lower than that of the ceramic spheres.

c) Related to this is in 207 "Snowflakes that are too small in diameter are ignored". What is this diameter, how did you decide on its value? Also, in the next sentence you state that you connect apparently separate detected regions if they are up to 4mm apart. This indicates that you expect that smaller details that may be connecting these regions are not being detected.

Thank you for your comments. Diameter means D_{max}, the detected connected regions in binarized images with the D_{max} less than 20 pixels (equivalent to approximately 0.035 mm², D_{max} is about 0.2mm) are ignored. This threshold is the optimal value chosen through testing. The value cannot be too large, or more snowflakes will be ignored; nor can it be too small, or some small noise spots will be recognized as snowflakes mistakenly. For the second problem, the apparently separate detected regions should be connected firstly if they are up to 4mm apart, and then discard particles that are too small. We have modified the corresponding text in the revised manuscript (221-227).

> (ii) Particles detection. Firstly, detect the connected regions in binarized images. Secondly, combine regions into a single particle when the centers of connected regions in a single image are detected to be less than 4 mm apart. This step is necessary because a single particle may sometimes be perceived as two separate particles due to its position near the edge of the image processing threshold. Thirdly, discarding the particles with an area greater than 20 pixels (Equivalent to 0.035 mm^2, D_{max} is about 0.2mm) enables the removal of small noises from the image, to prevent these noises from being mistakenly identified as small snowflakes.

3) Inconsistent and wrong or confusing use of "resolution", "pixel resolution", "pixel size", and "magnification". Revise and use better and consistent terminology throughout the paper.

Thank you for your advice. We have modified the terminology in the revised manuscript as follows: resolution: 4096×3000 or 720×540; pixel resolution: 41.6 or 265.4 μm·px−1; pixel size: 3.45μm × 3.45μm or 6.9μm × 6.9μm; magnification: 0.083 or 0.026. These have been clearly specified in Table 1 of the revised manuscript.

4) 102, "observation volume size of 1505.327cm3": Part of the observation volume is out of focus (as depth of focus with 104mm is smaller than the FOV dimensions). Three digits after the decimal point are not needed.

Thank you for your advice. The effective observation volume (OV) of 3D-PPI should be the intersection of the OVs of three high-resolution cameras, which is smaller than the OV of a single camera (170mm ×125mm × 104mm). The observation volume has been recalculated, which is 1464 cm^3. We have modified the corresponding descriptions in the revised manuscript (104-108).

> To clarify, the observation volume (OV) of one high-resolution camera is the interior rectangle of the observation volume of one telecentric lens with three dimensions a × b × d (170mm × 125mm × 104mm), which represents the length, width, and depth of field of view respectively. The intersection of three OVs of three high-resolution cameras forms the effective OV of 3D-PPI, the volume size is 1464 cm^3. Only the particles falling within the effective OV of 3D-PPI can be simultaneously captured by the three high-resolution cameras.

5) 156 "super-determined": Does this refer to an overdetermined system with more equations than unknowns? What are the unknowns here in that case? Do "super-determined" and "super-deterministic" refer to the same thing?

Sorry for our mistakes. Both the "super-determined" and "super-deterministic" refer to the overdetermined linear system, where the number of equations exceeds the number of unknowns. In Eq. (2), the other two matrices are known except KM_i, which is an unknown 2 × 4 matrix. The number of columns of the two known matrices is the number of chosen checkerboard corner points and is much more than 4. Therefore, Eq. (2) solving KM_i is equivalent to solving two overdetermined linear systems with the number of equations larger than the number of unknowns. The unknowns of the two equations are the first and second rows of KM_i respectively. We have modified the corresponding sentences in the revised manuscript (175-176).

> As can be seen from Fig. 5b, the value of J is much larger than 4, so Eq. (2) is equivalent to overdetermined linear systems. Further, the least squares method is used to optimally estimate the projection matrices (KM) for each high-resolution camera.

6) 163-169, explanations why the planar two-dimensional checkerboard grid cannot be used for calibration:

a) 164-165, "but then the values of all three-dimensional world coordinate points are the same": this sounds very general; explain better what you mean, which points/coordinates are the same?

What we mean is that there is only one WCS. We have deleted this sentence to avoid confusion.

b) The sentence continues that with that A equals a matrix, which is the same matrix as the matrix KMi is multiplied with in Eq (2).

The matrix *A* has been deleted.

c) The matrix A has not been defined before. Please do that.

The matrix *A* has been deleted.

d) 165-167, "The third and fourth-row values are the same …" This sentence is not correct. Please revise. The fourth-row values are all 1, did you want to say that all values in the third row are 1, or are the same but not necessarily 1 but any other value?

Sorry for our mistakes. We intended to state that the third-row values are the same. Furthermore, since the fourth row consists entirely of 1s, if the values in the third row are all the same, then the third and fourth rows are linearly dependent, which means the matrix *A* is impossible to invert. We have replaced it with a simpler statement, shown in the next reply.

e) 167, "determinant of A is 0": as far as I know, the determinant is defined for square matrices, but A is not (unless j=4). f) 168, "impossible to inverse" should be "impossible to invert"

Thank you for your advice. Actually, *A* is not a square matrix and has no determinant. Since the third and fourth rows are linearly related, the rank of *A* is 3. Changed as suggested. To be understood clearly, we have simplified the descriptions in the revised manuscript (176-178), which can also answer the above comments a) to e).

> It is worth noting that during the solution (optimal estimation) it is important to make sure that no row of W_J (except the row with value 1) can be the same value. In other words, the selected points cannot be in the same plane in the WCS. That is why we select a 3D checkerboard rather than a 2D checkerboard.

7) 171-172, in this sentence at the end of Sect 3.1, you present an "average reprojection error". Nowhere in Sect 3.1 you present any calibration results related to the theoretical treatment presented. The theory presented seems needed to determine KMi, which are needed for the matching algorithm. Define what the average reprojection error is and how you have determined it. Here you say 0.32 pixels, in the Conclusion 0.4 pixels.

Thank you for your comments and sorry for our mistakes. The average reprojection error is a key metric used to evaluate the accuracy of our estimated projection matrix. It is calculated as the mean distance between the observed 2D image points and their corresponding projected points obtained from the estimated projection matrix. This error indicates how well our model captures the actual image data. The calibration result referred to here is the estimated projection matrix *KM*. We have modified and improved this section in the revised manuscript (179-181).

To assess the accuracy of the estimated projection matrices (*KM*) of three cameras, we calculated the average reprojection error, which is calculated as the mean distance between the identified 2D image points (uj, vj) and their corresponding projected points $KM \cdot (X_{wj}, Y_{wj}, Z_{wj})^T$. The ~~calibration~~ results show that the average reprojection error for three high-resolution cameras is 0.32 pixels.

Besides, we have changed to 0.32 pixels in the conclusion of the revised manuscript.

8) Determination of pixel resolution (41.5um/px) in Sect 3.2 (Calibration of image binarization) and image processing in Sect. 4.1:

a) 181, "is optimally binarized manually": How has this manual binarization been performed. Is each image treated differently? Is this using the image processing algorithm described in Sect. 4 (where the details of the detection are missing, see comment about "adaptive thresholding below")?

Thanks for your comments. The meaning of manual binarization has been illustrated in the previous figure in 2b). We performed manual adjustments to mitigate the obvious software misidentifications. Each image was treated differently. The image processing algorithm mentioned in Sect. 4 is not used here. In this part, accurate pixel resolution is obtained by capturing ceramic spheres and linear fitting, so we need to treat each image differently and perform perfect processing. In other words, since the perfect binarization result of each ceramic sphere in the image is circular theoretically, optimal manual binarization is to make the ceramic sphere in the image more circular, to avoid missing detection details.

b) You determine the pixel resolution from the reciprocals of the slopes in Fig 6b (Dmax/px vs Dmax/mm). You are not discussing the role of the offset of between 0.5px to 4px in Equations (3)-(5).

Thank you for your comments. The non-zero intercepts observed in the linear fits (0.5 px to 4 px) might be attributed to several factors: systematic errors introduced during the calibration process due to lens distortion or misalignment; image processing artifacts from the binarization method that may affect edge detection. Additionally, the pixel resolution of the cameras and the physical characteristics of the ceramic spheres, such as surface texture, can also cause these offsets. We have added these discussions in the revised manuscript (211-214).

c) The pixel resolution should also result from the calibrations using chessboards. Has this been attempted and the values compared to the pixel resolution reported in Sect 3.2?

Thank you for your comments. To clarify, the word "calibration" in Sec. 3.1 and Sec. 3.2 has different purposes. Sec.3.1 is only for estimating the projection matrix, which contains the internal and external parameters of the camera, and we are not estimating the full range of internal and external parameters, as having the projection matrix is sufficient for the 3D reconstruction later on. Sec.3.2 is for estimating more accurately the pixel resolution, which is a crucial internal parameter.

The reasons for not using the checkerboard grid to obtain image resolution are as follows: pixel resolution can only be estimated when the checkerboard plane is parallel to the camera image plane, which is difficult to guarantee. The advantage of the ceramic sphere in comparison is that its image taken at any angle is round.

d) The pixel resolution using chessboards would be similar to using micrometer or millimeter scales. The advantage over ceramic spheres would be that it would not depend on image processing (selection of specific grey level threshold to detect the contours of the ceramic spheres), which may result in under- or over-sizing of the spheres. The details of this image processing are, however, not disclosed. You only state (Sect. 4.1, 204) that the images are "binarized through adaptive thresholding" without giving any reference or explaining what adaptive thresholding is in your specific case. Please explain this method.

Thank you for your comments. According to your comments 8) a) and our response, the checkerboard was used to estimate the projection matrix, rather than using micrometer or millimeter scales. Regarding the adaptive threshold method, we have added the following description in the revised manuscript (219-221).

> Adaptive thresholding dynamically calculates the threshold for smaller regions of the image, allowing for better handling of varying lighting conditions. This method enhances the accuracy of foreground particle detection, particularly in images with complex backgrounds and uneven illumination.

e) In Sect. 4.1 describing the image processing you show results from imaging the ceramic spheres (Fig. 8a and 8b) in terms of measured size vs real size (both in mm). You must have used the pixel resolution resulting from imaging the ceramic spheres (Sect. 3.2, Fig 6b). For the "calibration of image binarization" you must have used the image processing algorithms described in Sect. 4.1. That means that Fig 8 does not show anything new or independent from previously reported Fig 6b. You are presenting the same ceramic sphere measurements in two related ways. Consequently, the error analysis related to Fig 8 is only a re-interpretation of the error analysis of Fig 6b.

Thank you for your comments. As explained above, in Sec.3.2 we did not use the image processing algorithm in Sec. 4.1, the measurements were processed by optimal manual binarization, and the section was only intended to get a more accurate pixel resolution. Instead, the measurements in Fig.8 were obtained by batch processing using the image processing algorithm mentioned in Sec.4.1. The purpose of Fig. 8 is to evaluate the reliability of the image processing algorithm as its error is mainly caused by the image processing algorithm.

f) In any case, the presented error analyses are unclear and confusing: L196-197, "The estimated random errors from the normalized root square errors, derived from the observed and true size difference …" Unclear what "observed and true size difference" is. Unclear what "random" refers to? Does it refer to variations in 20 observations of the size of the same ceramic sphere?

Thank you for your comments and sorry for our unclear descriptions. Actually, no error

analysis is needed for this section, we just want to estimate the pixel resolution of each high-resolution camera by the linear fitting.

We removed the text:

> ~~The estimated random errors from the normalized root square errors, derived from the observed and true predicted size difference, were 1.8% (Cam0), 2.6% (Cam1), and 1.5% (Cam2) respectively, indicating that random errors in fitting straight lines is are negligible.~~

g) Due to the offset in your linear least squares fits, the reciprocals of the coefficient px/mm (24px/mm here) are not equal to the coefficients that would result from fitting true size vs observed size (mm/px). Redo that fit or comment.

Thank you for your comments. Actually, the inverse of the slope of the fitted straight line is very close to the pixel resolution after unit conversion, as follows:

$$\frac{1}{24\,\mathrm{px/mm}} = \frac{1000\,\mu\mathrm{m}}{24\,\mathrm{px}} \approx 41.67\,\mu\mathrm{m}\cdot\mathrm{px}^{-1}$$

h) 227, "average absolute error of Dmax measurements for all diameters of small spheres is -0.048mm," An absolute error is positive, but you state that it is -0.048mm. Same for Deq with - 0.33mm. Then it is unclear what the "average relative error" (229, 230) is.

Thank you for your comments and sorry for our mistakes. The "absolute error" actually refers to the "error". The error is defined as: the measured value - the true value, where a positive value indicates that the size of the particle tends to be overestimated, and a negative value indicates that it tends to be underestimated. Similarly, the relative error is calculated as: (measured value - true value) / true value.

9) 206-207, "Detecting connected regions … enables the removal of small noise …" Missing/wrong logic: What is enabling the removal of noise from image? What is small noise on the image?

Thank you for your comments. Actually, there might be individual noise points or artifacts in the binarized image that were misidentified as particles. By connecting the regions within an area greater than 20 pixels (equivalent to approximately 0.035 mm², D_{max} is about 0.2mm), we can effectively filter out these small noises and the particles smaller than 0.2 mm. Considering that the 3D shape reconstruction of snowflakes mainly focuses on larger particles, the above noise removal process does not have a significant impact on the observations.

10) 207, "Snowflakes that are too small in diameter are ignored" What is this diameter, how did you decide on its value? This is related to the comment about actual resolution. Here you seem to have defined a smallest particle that is accepted.

Thank you for your comments. The diameter refers to D_{max}, the detected connected regions in binarized images with the D_{max} less than 20 pixels (equivalent to

approximately 0.035 mm²) are ignored. This value is determined through pre-testing. It cannot be too large, or more snowflakes will be ignored; nor can it be too small, or some small noise spots will be mistakenly recognized as snowflakes.

Actually, the smallest acceptable particle D_{max} is 0.2mm does not contradict that the pixel resolution of 3D-PPI is 41.6 μm·px^{-1}. The pixel resolution means that the smallest recognizable detail of a large snowflake is 0.0416mm, rather than the smallest recognizable snowflake.

11) 243-252, explanation of matching algorithm: This is a clever way to use the matrices KMi (i=1,2,3) from camera calibration for matching. The following issues/unclarities result from wrong language or unclear formulation.

a) I would not re-use the index i, that previously was used to refer to the three cameras, to refer now to a particle number. Use another index or no index.

Thank you for your advice. The index *i* was removed in the revised manuscript, because we then modified the manuscript to describe the matching process for only one particle.

b) 245, "i underdetermined linear equations": Is there one underdetermined linear equation for each particle? Refer to an equation number in the paper or explain better what these equations are (and why/how underdetermined).

Thank you for your comments. Actually, there is one underdetermined linear equation for each particle centroid. In Eq. (2), when $J = 1$, KM and C_1 are known and W_1 is solved for, i.e., 3D points are solved from 2D points, this is equivalent to solving the underdetermined equation, which does not have unique solutions.

c) Would be good to add equation (if appropriate in Appendix), the solution of which is Li. Then also another equation showing the multiplication of KM1 and KM2 to produce the projection of Li onto Cam1 and Cam2 image planes.

Thank you for your comments. We have supplemented Appendix B with a more detailed description of the equations:

Appendix B: Additional equation solving process

To solve for (X_w,Y_w,Z_w) based on the known points P(u,v) and KM_0, simplify Eq. (2) by removing 1 from the second term:

$$KM_0 \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \xrightarrow{\text{Simplify}} A \cdot \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = B \quad \text{(B.1)}$$

Where A is a known 2 × 3 matrix and B is a known 2 × 1 matrix, so it is equivalent to underdetermined linear equations. And the solution is not unique and shown in Eq. (C.2):

$$\begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix} = U t + V = \begin{bmatrix} U_1 t + V_1 \\ U_2 t + V_2 \\ U_3 t + V_3 \end{bmatrix} \quad \text{(B.2)}$$

Where, both U and V are 3 × 1 matrixes, t is all real number. Therefore, all solutions form a straight line L in 3D space WCS. In other words, this process implements the back-projection of P onto the line L.

Furtherly, project the L onto the planes of Cam1 and Cam2 by multiplying the projection matrices KM_1 and KM_2, respectively, shown in Eq. (C.3):

$$L_{p1} = KM_1 \cdot \begin{bmatrix} U_1 t + V_1 \\ U_2 t + V_2 \\ U_3 t + V_3 \\ 1 \end{bmatrix} = \begin{bmatrix} f_1(t) \\ g_1(t) \end{bmatrix} \qquad L_{p2} = KM_2 \cdot \begin{bmatrix} U_1 t + V_1 \\ U_2 t + V_2 \\ U_3 t + V_3 \\ 1 \end{bmatrix} = \begin{bmatrix} f_2(t) \\ g_2(t) \end{bmatrix} \quad \text{(B.3)}$$

Where, L_{p1} and L_{p2} denote the point sets of projections of L onto the Cam1 and Cam2 planes respectively. The functions $f_1(t)$, $g_1(t)$, $f_2(t)$, and $g_2(t)$ are all linear functions of t. L_{p1} and L_{p2}. Therefore, L_{p1} and L_{p2} represent straight lines in the plane. Determine the range of t to ensure that the line is within the image range (4096×3000) to get the corresponding line segments.

d) 246, "ith straight lines" should be "i straight lines"?

Thank you for your comments. The index *"i"* has been removed in the revised manuscript.

e) 247, "the lines Li is the back-projection" should be "the lines Li are the back-projections".

Thank you for your advice. We have modified as suggested in the revised manuscript.

f) 248, "multiplying the projection matrices KM1 and KM2," should be "multiplying the projection matrices KM2 and KM3, respectively, with Li,"

Thank you for your advice. We have modified as suggested in the revised manuscript.

g) 249, "resulting in ith line segments" should be "resulting in i line segments on each of the image planes of Cam1 and Cam2"

Thank you for your advice. We have modified as suggested in the revised manuscript.

h) 251-252: from the text it is unclear if each particle has to be on all three lines, or each particle on its respective line in each of the images. Describe for one particle (to avoid confusion between THREE cameras, and THREE particles/lines).

Thank you for your advice. It refers to each particle on the respective line in each of the images. We have described one particle matching process to avoid confusion.

To be understood clearly, we have revised the descriptions in the manuscript, which can also answer the above comments a) to g).

The revised text is as follows (259-273):

For one particle, the matching algorithm is implemented in detail as follows (Fig. 9):

(i) Detect the centroid coordinates of the particle P (u, v) in the PCS from Cam0 (Fig. 10a).

(ii) Using the projection matrix KM_0 of Cam0, the underdetermined linear equations corresponding to P are solved to obtain a straight line L in WCS. Specifically, it is equivalent to solving W_1 in Eq. (2) with KM and C_1 known, where the solution to W_1 is not unique and all solutions are the L in the WCS. The L represents all points in 3D space that can be projected onto P by KM_0, in other words, the line L is the back-projection of the point P in WCS.

(iii) Project the L onto the planes of Cam1 and Cam2 by multiplying the projection matrices KM_1 and KM_2, respectively, with L, resulting in the line segments on each of the image planes of Cam1 and Cam2 (Fig. 10b, c). The exact derivation of the formulae in (ii) and (iii) are described in detail in Appendix B.

(iv) Detect the particles that the line segments pass through on Cam1 and Cam2 respectively. If the line segments do not pass through any particles in Cam1 or Cam2, it is a failed matching, meaning that the particle does not appear in the effective OV.

By performing the above matching for each particle detected by cam0, the location of this particle in Cam1 and Cam2 can be found. Fig. 10 shows the three particles detected in Cam0 and the matching of each particle in Cam1 and Cam2.

12) 259-294, 3D reconstruction: The description of how particle is located to enable/simplify 3D reconstruction is unclear. Here a few things that make it difficult to follow:

a) Add new sub-section (now it is part of 4.2 Particle matching and localization).

Thank you for your advice. We have modified it as suggested in the revised manuscript.

b) 259, Not sure what voxels are and why/how they are used traditionally. Provide a reference.

Thank you for your comments and sorry for our mistakes. Our method is not comparable to traditional methods, so we decided to remove this sentence.

c) 260-261, Explain or discuss how are "traditional methods" "computationally inefficient".

Sorry for our mistake. Our method is not comparable to traditional methods, so we decided to remove this sentence.

e) 264, "particles' localizations" should be "particles' locations"?

Thank you for your advice. We have modified it as suggested in the revised manuscript.

f) 266-271, Unclear why irregular particles pose a problem. You probably need to be

more specific in explaining the issue.

Thank you for your comments. We were originally trying to state that the localization is relatively straightforward for particles with regular geometries, on the contrary, it is somewhat more complex and requires the methods described below for complex snowflakes. However, we later found out that it is not true. Therefore, this sentence was removed in the revised manuscript.

We removed the text:

> ~~For particles with regular geometries, back-projection of the contour centroids suffices to determine the centroid of the 3D reconstructed object. However, snowflakes often exhibit highly irregular shapes. The center of the three-dimensional body projected onto the contour may not coincide with the center of the two-dimensional contour, rendering the straightforward intersection of back-projected lines inapplicable and complicating the reconstruction process.~~

g) 277-278, confusing use of indices: "Lines L1", I thought L1 is one line, the back-projected line from P1 on image of Cam0. "L2" according to earlier explanations (247), this is the back-projected line from P2 (particle 2) on image of Cam0, but here is redefined.

Thank you for your comments and sorry for our mistakes. We have modified the previous text, L_1, L_2, P_1, P_2 are no longer existed. So, in this section, we still use these symbols.

h) How can L2, a line, be "represented as a 2-row by 1-column matrix"? 279-280, "P2" and "P2'", what is P2 (it is not the P2 on image of Cam0, 243)? It becomes more and more tedious to follow the remaining explanations. This whole section should be revised for consistency and clarity.

Thank you for your advice. We have modified the section in the revised manuscript (280-303):

> After completing camera calibration and particle matching, 3D reconstruction for each particle needs to be performed. However, since each particle only occupies a small space in the effective OV, we propose a method that involves preliminarily locating particles in WCS before proceeding with subsequent 3D reconstruction. This method leverages the positions of single particles in three images to identify the minimal cuboid capable of containing the particles, thereby accurately pinpointing the particles' localizations.
>
> For a single particle, the pixel coordinates of the centroid point of the particle in Cam0, Cam1, and Cam2, respectively, have been detected and the subsequent 3D spatial localization steps in the WCS are as follows (Fig. 11):
>
> (i) Find the back-projection line L_1 of point P_1 by KM_0. The underdetermined linear equation corresponding to P_1 is solved to obtain a straight line L_1 in WCS. This implementation principle is similar to the second step of particle matching

mentioned above. Eq. (B.1) and Eq. (B.2) in Appendix B explain this process.

(ii) The lines L_1 are projected onto the planes of Cam1 by multiplying the projection matrices KM1, resulting in line segment L_2, which is represented as a 2-row by 1-column matrix. This corresponds to the segment of the Lp1 within the image boundaries as described in Eq. (B.3).

(iii) Find the point P_2' on L_2 that is closest to P_2. Due to the irregular shape of the particle, P_2' does not necessarily coincide with P_2.

(iv) Following the same approach as in step 1, determine the back-projection line L3 of point P_2' by KM_1.

(v) Localize the 3D coordinates of the intersection of L_1 and L_3 in the WCS, that is Pc, which is the centroid of the target cuboid, and further determine the side lengths of the cuboid. From the previous steps, L_1 and L_3 are destined to intersect in the WCS, and the intersection point is regarded as the centroid of the rectangle, whose side lengths can be determined by converting from the pixel dimensions in the particle image to the actual physical dimensions in the WCS.

(vi) Finally, verify that the projection of the Pc point through KM_2 in Cam2 is near the P3 point and within the particle contour, otherwise, it is a failed localization.

The particle's position in the WCS should be inside the region of the cuboid determined by localization, which will next be discretized into numerous smaller voxel grids to finely perform 3D reconstruction.

13) Sect 4.3: It is not immediately clear if and how this connects to explanations in 259-294. 296-297: "silhouettes that have been serially calibrated using multiple viewpoints around the target" The section starts with unclear formulations like this one (are silhouettes contours, multiple viewpoints refer to Cam0,1,2, what does "serially calibrated mean"?). I did not review the remaining of this section. After revising the previous section, this section should be revised too.

Thank you for your constructive comments. The Visual Hull (VH) method is employed for reconstructing the 3D shape of snowflakes by utilizing silhouettes—defined here as the contour outlines of the snowflakes as viewed from various perspectives. Multiple viewpoints, such as cameras positioned at angles Cam0, Cam1, and Cam2, are used to capture the silhouettes. Each silhouette is carefully calibrated to ensure accuracy in the reconstruction process. The term "serially calibrated" refers to the process of calibrating the camera parameters and positions for each viewpoint.

We have revised this part (307-311):

~~The Visual Hull (VH) method is used to reconstruct the 3D shape of snowflakes. This approach utilizes silhouettes that have been serially calibrated using multiple viewpoints around the target, thereby enabling the reconstruction of its three-dimensional shape.~~ The Visual Hull (VH) method is a technique used to reconstruct the three-dimensional shape of snowflakes by utilizing silhouettes, which are the

outlines of the snowflakes as seen from different camera angles. In this process, multiple cameras are positioned around the snowflake at various viewpoints (Cam0, Cam1, and Cam2 in 3D-PPI). Each camera captures the silhouette of the snowflake, and these silhouettes are carefully calibrated to ensure that they accurately represent the snowflake's geometry.

14)

a) Explain better Eq (7).

b) "Nima is the number of particles" should be "Nima is the number of acquired images"?

Thank you for your advice. We have modified it as suggested in the revised manuscript.

c) You state that (339) "the probability of capturing the same snowflake in two consecutive frames is very low." How often can the same snowflake be captured twice? Discuss if anything is done to account for this or, if not, how big of an error this may cause.

Thank you for your advice. Since the high-resolution camera samples at a frame rate of 5fps (with a time interval Δt=0.2s), particles cannot be captured twice when the horizontal velocity of the particle v_h > 0.85m/s (0.17m / 0.2s) or the vertical velocity v_v > 0.625m/s (0.125m / 0.2s).

When v_h < 0.85m/s and v_v < 0.625m/s, the probability that the particle is caught twice in the horizontal and vertical directions are respectively:

$$\begin{cases} \mathrm{Prob}_h = 1 - \dfrac{\triangle t \cdot v_h}{0.17} \\ \mathrm{Prob}_v = 1 - \dfrac{\triangle t \cdot v_v}{0.125} \end{cases}$$

Horizontal and vertical capture probabilities are 50% when v_h and v_v of the particles are approximately 0.4 m/s and 0.3 m/s respectively. Particles with small velocities have a higher probability of being captured more than twice, which is indeed an important factor contributing to the inaccuracy of PSD measurements. Subsequently, we will consider increasing Δt (decreasing the frame rate of the high-resolution camera) to alleviate this problem.

d) You are getting the PSDs only from Cam0 here. That means $V_{observation}$ is different from the observation volume reported in 102 (1505cm3). $V_{observation}$ is calculated using the depth of focus (104mm). How well-defined is the depth of focus? Are particles outside depth of focus detected but then rejected reliably?

Thank you for your comments. In the revised manuscript, $V_{observation}$ refers to the volume of OV, which is the interior rectangle of the observed volume of the telecentric lens, with three dimensions a × b × d (170mm × 125mm × 104mm), representing the length, width, and depth of field of view respectively. To clarify, the "104mm" refers to the

depth of field rather than the depth of focus, referring to the range that can be clearly imaged perpendicular to the direction of the lens. It has been tested that particles outside the depth of field cannot be clearly imaged and therefore cannot be detected.

15) From Fig16b I would guess that the PSD peaks at Deq between 1 and 2 mm, whereas in Fig 13 the PSD peaks at Deq of about 0.4mm. Is everything consistent here and I am only being confused and misled by something?

Thank you for your comments. We would like to clarify that Fig.16 represents the particle velocity distribution measured on April 6, while Fig.13 shows measurements taken on March 28 and 29. The differing periods account for the variations in the observed PSD peak values.

We apologize for any confusion caused by this discrepancy. Unfortunately, due to data storage limitations during the field experiments, we were unable to retain high-resolution camera images and high-speed camera images from the same period for direct comparison. We will continue to conduct the field experiments in the next winter and analyze the observation data, thank you for your understanding, .

16) Negative velocities:

a) L416-417 "The average value of the horizontal velocity component measured by 3D-PPI is +0.05m/s (positive and negative values indicate westward and eastward velocities, respectively), and the standard deviation is 2.56m/s (Fig. 16a)." Unclear if your average is an absolute average (or do you consider positive and negative values when averaging)? Consequently, the meaning of the standard deviation is unclear.

Thank you for your comments. In calculating these statistics, both positive and negative values were taken into account, so they are not absolute averages. The average velocity and standard deviation values have been updated to +0.41 m/s and 0.73 m/s in the revised manuscript (432-433).

b) I would expect horizontal speed to be characterized by speed (absolute value) and direction (0 – 360degrees). Instead, you use positive and negative values to indicate westward and eastward. Why choosing to give info on west-east, and not north-south or the actual direction in degrees?

Thank you for your comments. The 3D-PPI instrument was installed with its camera facing south, side facing the prevailing west wind direction. As a result, the measurements from this orientation primarily capture the east-west horizontal velocity component, represented by positive values for westward velocity and negative values for eastward velocity. We acknowledge that a single high-speed camera cannot measure speeds in all directions from 0 to 360 degrees.

c) If you choose to include directional information, why are you not analyzing the direction. I would expect the horizontal speed direction to correlate with local wind direction. Was the local wind measured and compared to horizontal speed? This could be part of a discussion how wind affects measurements.

Thank you for your comments. Actually, the horizontal speed direction has a high correlation with the local wind direction. Unfortunately, we have not collected the local wind observation data during this study, we acknowledge the significance of wind effects on measurements and will consider this in future observations.

Minor issues

1) The three cameras are numbered 0,1,2 (Cam0, ...), whereas you use the indices i=1,2,3 to indicate the three cameras (e.g. KM1, KM2, KM3). Can you use the same indices to reduce confusion?

Thank you for your advice. We have mapped the three camera projection matrices KM_0, KM_1 and KM_2 to the camera numbers (Cam0, Cam1 and Cam2) to reduce confusion.

2) "and" in wrong place in a list (85, 185).

Thank you for your advice. Changed as suggested.

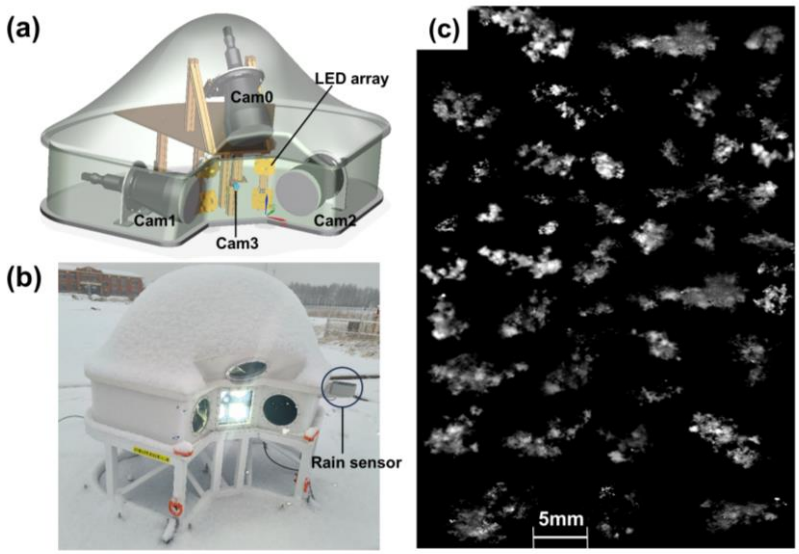
3) 86-87, "capacitive rain sensor is adopted as a trigger, the cameras only work when the precipitation occurs." Rain sensors detect rain. When you say the cameras only work when precipitation occurs, do you mean when it rains? I would expect you want to measure with snow but perhaps not with rain? Please clarify.

Thank you for your comments. To clarify the point in Lines 86-87, we would like to emphasize that the capacitive rain sensor is designed to trigger the operation of the 3D-PPI instrument whenever precipitation occurs, regardless of whether it is rain or snow. The sensor detects any moisture on the instrument's surface, and its heating element ensures that snow is melted and then sensed. We have added the following description in the revised manuscript (88-90).

> To improve the instrument's working efficiency, a capacitive rain sensor is adopted as a trigger, the cameras only work when the precipitation occurs. The sensor detects any moisture on the instrument's surface, and its heating element ensures that snow is melted and then sensed.

4) Fig 1a: It would be useful to label the different parts. Without labels it is, for example, not clear which/where is the fourth non-telecentric camera. Labels are a complement to more clarity in the text.

Thank you for your advice, we have added the labels in Figure 1.



5) 98-99, "45° angle relative to the optical axis of the high-speed camera" Be more specific: two cameras are positioned, at 45 degrees, on either side of the high-speed camera in the same horizontal plane, the third .... at 45 degrees vertically above

Thank you for your advice, we have revised the corresponding sentences in 100-102:

> The high-speed camera is positioned horizontally at the center, while ~~the three high-resolution cameras are oriented at a 45° angle relative to the optical axis of the high-speed camera.~~ the Cam1 and Cam2 are positioned, at 45 degrees, on either side of the high-speed camera in the same horizontal plane, and the Cam0 is positioned at 45 degrees vertically above (Fig. 1a).

6) 100-101, "overlapping region of the LED lighting beams" and intersection of the "three rectangular light columns"??? I don't think you are talking about the LED lightning?! Do you mean the intersection of the FOVs of the three cameras (which may be approximated by rectangular light columns)? But I assume the LED lightning beams are larger than the FOVs (and not exactly rectangular and change in cross section).

Sorry for our unclear descriptions. To clarify, the observation volume (OV) of one high-resolution camera is the interior rectangle of the observation volume of one telecentric lens with three dimensions a × b × d (170mm × 125mm × 104mm), which represent the length, width, and depth of field of view respectively. The intersection of three OVs of three high-resolution cameras forms the effective OV of 3D-PPI, the volume size is 1464 cm^3. The high-brightness LED array light sources are situated on the same side as the cameras to illuminate the observation volume. We have revised the corresponding sentences in revised manuscript (102-107):

> The high-brightness LED array light sources are situated on the same side as the cameras to illuminate the observation volume. The cylindrical observation volume of the three telecentric lenses and LED lighting beams of 3D-PPI is illustrated in Fig. 2. To clarify, the observation volume (OV) of one high-resolution camera is the interior rectangle of the observation volume of one telecentric lens with three

dimensions a × b × d (170mm × 125mm × 104mm), which represent the length, width, and depth of field of view respectively. The intersection of three OVs of three high-resolution cameras forms the effective OV of 3D-PPI, the volume size is 1464 cm^3.

7) Be consistent with units. 102 you state the FOV as 17cm and 12.5cm, later you write 170mm and 125mm.

Thank you for your advice. Changed as suggested.

8) 101-103, the text referring to Fig 2 describes light reflected and scattered by snow particles, which is not shown in Fig 2.

Thank you for your advice. Changed as suggested in revised manuscript (103-104):

The cylindrical observation volume of the three telecentric lenses and LED lighting beams of 3D-PPI is illustrated in Fig. 2.

9) Fig 2: You have talked about rectangular columns earlier; you should now show these (rather than the circular columns you do show). "Optical structure" doesn't seem to be the appropriate expression. Some more info in the caption would be useful.

Thank you for your comments. We have modified the sentences in revised manuscript (110):

Figure 2. The observation volume of the three telecentric lenses and LED lighting beams. ~~Optical structure of 3D-PPI~~

10) 114-115, "which leads to a difference in the method of performing 3D reconstruction later in Sec. 4". Are you referring to a difference of the method". How is the method different from what?

Thank you for your comments. We would like to clarify that the different imaging schemes will result in slight variations in the methods used for 3D reconstruction. We have modified the sentences in revised manuscript (120-123):

Unlike non-telecentric lenses, which produce larger images of nearby objects and smaller images of distant objects (Fig. 3a), telecentric lenses are based on the principle of parallel light imaging, resulting in identical objects at different distances from the lens having the same size in the image (Fig. 3c). This difference in imaging scheme will lead to distinct methods for three-dimensional reconstruction (Fig. 3b, d), which will be discussed in detail in Sec. 4. ~~which leads to a difference in the method of performing 3D reconstruction later in Sec. 4 (Fig. 3b, d).~~

11) 127 "The LED light sources are arranged in a parallel configuration, leading to a unidirectional power supply interface." Are you talking about electrical set-up or spatial placement of LEDs? Are you talking about LEDs or LED arrays? What do you mean with "unidirectional power supply interface"?

Thank you for your comments.

Electrical Setup vs. Spatial Placement: We are referring to the electrical setup of the

LED light sources, which are arranged in a parallel configuration. This configuration allows for a single power supply interface to power the LEDs.

LEDs vs. LED Arrays: We are specifically discussing LED arrays, where multiple LEDs are grouped together for improved illumination.

Unidirectional Power Supply Interface: The term "unidirectional power supply interface" refers to a single power connection that supplies power to all LEDs in the parallel configuration, ensuring that the power is distributed uniformly across the array.

12) 128: specify (or re-word) "consistent light output" as it is unclear.

Thank you for your comments. To clarify, we have rephrased this to specify that it refers to a stable and uniform light intensity produced by the LEDs. We have modified the sentences as follows in response to 11) and 12) in revised manuscript (135-139):

> ~~The LED light sources are arranged in a parallel configuration, leading to a unidirectional power supply interface. The LED operates in either constant current mode or trigger mode, with the former ensuring a consistent light output, thereby enhancing the uniformity of illumination within the observation volume.~~
>
> The LED light sources are configured in parallel, allowing for a single power supply connection that distributes power to the entire array. Each LED can operate in either constant current mode or trigger mode. In constant current mode, the LEDs provide stable and uniform light intensity, which enhances the uniformity of illumination within the OVs.

13) Fig 4a is not needed, you have the same information as Fig 4b.

Thank you for your advice. Changed as suggested.

14) 139-140: "projection matrix KMi of the transformation relationship between the 3D spatial points and each pixel plane pixel point in the world coordinate system"

a) Re-formulate this for clarity. What is a "pixel plane pixel point" if it is not a mistake? A pixel point is not "in the world coordinate system".

Thank you for your comments. We have rephrased in revised manuscript (147-151):

> The geometric model for telecentric lens imaging is described in detail in Appendix A, where four coordinate systems (WCS, CCS, ICS, and PCS) are defined and their transformations are derived. The purpose of camera calibration in this section is to estimate the projection matrix ~~KMi~~ ,KM_0, KM_1, and KM_2 for three high-resolution cameras, which enables the coordinates transformation of 3D spatial points in the WCS to their corresponding projection points in the PCS. ~~of the transformation relationship between the 3D spatial points and each pixel plane pixel point in the world coordinate system. The geometric model for telecentric lens imaging is described in detail in Appendix A.~~

b) Also, define the World Coordinate System when you first use it.

Thank you for your advice. In the previous sentence, we have already mentioned that

WCS has been defined in Appendix A.

15) 142-144: Something is missing or wrong in this sentence (in particular "and the apparent 3D …").

Thank you for your comments. Considering that the 3D checkerboard has been introduced later, this sentence has been deleted

> ~~Since the chessboard grid (Fig. 5b) has an ideal regularity and is easy to be recognized, it is used here as a calibration plate for camera calibration, and the apparent 3D coordinate point to the image pixel coordinate point is a linear change without considering the camera/lens distortion.~~

16) a) 153-154: Define/describe the 3D checkerboard.

Thank you for your advice. The 3D checkerboard has been defined in the revised manuscript.

b) I would be consistently using only "checkerboard" or only "chessboard".

Thank you for your comments. It should be "checkerboard" rather than "chessboard". It has been changed to 'checkerboard' throughout the manuscript.

c) Reconsider sentence: "from the same localization using three cameras" is wrong. The three cameras image from three different views/locations.

Thank you for your advice. We have rephrased as follows in the revised manuscript (161-167):

> To obtain the projection matrices, the following steps are executed: ~~Firstly, establish the position of the WCS by capturing images of a 3D checkerboard grid from the same localization using three cameras.~~ Firstly, make a 3D checkerboard and establish the WCS. Attach three high-precision 2D checkerboards to three mutually perpendicular flat boards to form a 3D checkerboard. The three plane intersections are used as the WCS origin O, and the two plane intersections are used as the X, Y, and Z axes respectively (Fig. 5b). The 3D checkerboard is placed in a position that defines the WCS. This position ensures that three high-resolution cameras can capture checkerboard corner points. Secondly, physically measure the precise coordinates of all checkerboard corner points in the WCS (X_{wj}, Y_{wj}, Z_{wj}) (j denotes the number of corner points), and identify the pixel coordinates (u_j, v_j) of ~~these~~ j^{th} corner points in the PCS of each camera image.

17) 157 "such as Eq. (2)": should that be "shown in Eq. (2)"? (Eq (2) is not an example but shows exactly the equations you want)

Thank you for your comments. Changed as suggested.

18) 170-171, "The two … define a common …WCS" Two planes do not define a WCS. Should it be "The three…"?

Thank you for your comments. This sentence has been removed in the revised

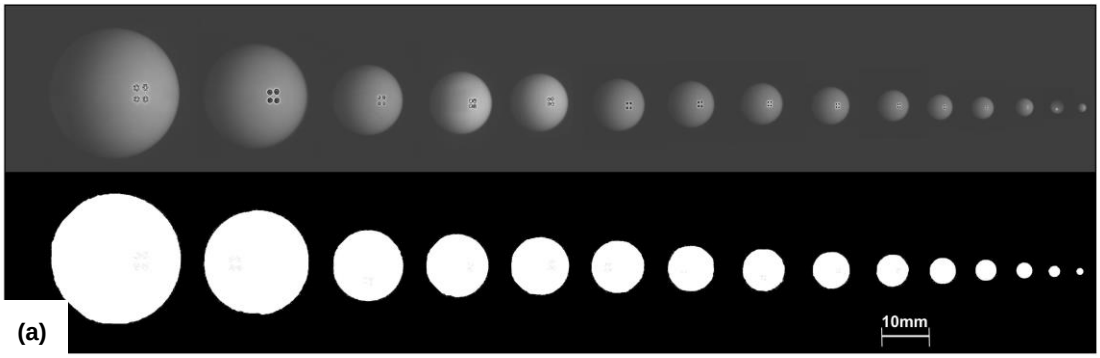
manuscript.

19) Eq.s (3)-(5): "Dmax [um]" should be "Dmax [mm]"

Thank you for your comments. Changed as suggested.

20) Fig 6a and 6b: You show 13 spheres in Fig 6a and 15 points on the plot Fig 6b. How many spheres did you use?

Sorry for our mistakes. We used a total of 15 spheres, we have made the necessary corrections in Fig. 6a.



21) 209, "is necessary is an essential step" Check and correct.

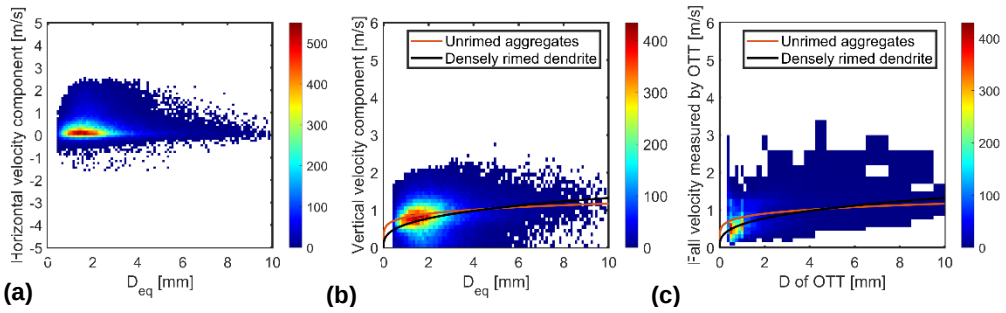
Thank you for your comments. Changed as suggested.

22) 239, "which poses a challenge for particle identification from the images captured by three cameras" Did you mean a "challenge for particle matching"?

Yes, we mean "challenge for particle matching" rather than identification. We have modified it in the revised manuscript.

23) Comparison between 3D-PPI and OTT measurements Fig 16, What are the numbers on the colour scale? Would be interesting to see speed distribution for a few particle sizes (you mentioned such a distribution in 402 "snowflake velocity distribution with diameter was calculated").

Thank you for your comments. The color scale denotes the number of snowflakes measured in the corresponding bins. (Sorry for the error in the previous Fig. 13, please refer to the new one.)



We have revised the corresponding description and added some necessary information

in revised manuscript (418-421):

> From 0800 UTC to 0830 UTC on 6 April 2024, the Cam3 of 3D-PPI recorded ~~322,267~~ 77042 valid snowflakes, and the snowflake velocity distribution with diameter was calculated. for each snowflake, horizontal and vertical velocities were calculated. The distributions of horizontal ~~velocity component~~ and vertical velocity components as a function of D_{eq} are further plotted as a scatter density plot and compared to the results measured by OTT at the same period, which is shown in Fig.16. The color scale denotes the number of snowflakes measured in the corresponding bins.

24) 343-344, "PSDs are described …" a) What is the meaning of this sentence? b) "across a larger range of sizes"?

Thank you for your comments. We try to convey that the PSD described by D_{max} covers a broader range of particle sizes compared to that described by D_{eq}. However, using D_{eq} to describe the PSD appears to be more valuable in terms of analysis. b) "across a larger range of sizes" means covering a broader range of particle sizes in the PSD plot for the horizontal coordinates. To avoid misunderstanding, we have removed this sentence in the revised manuscript.

> ~~PSDs are described by Dmax as opposed to Deq across a larger range of sizes, and it may be more valuable to describe them with Deq~~

25) 344, "The peaks of Deq…" Deq does not peak. Be more precise and correct with your formulations.

Thank you for your comments. To clarify, the highest number density of particles is found at D_{eq} 0.4 mm for both chosen days. In the revised manuscript (359-360):

> ~~The peaks of Deq were all near 0.4 mm and varied very little over the days.~~ The PSDs of particles with a D_{eq} of about 0.4 mm were highest for both chosen days.

26) a) 348-352, "Comparison of temporal plots…" Long sentence with several statements that are unclear. b) Same next sentence (meaning of "which means the aggregation of snowflakes was weakened"). Please revise.

Thank you for your advice. We have revised this part in the revised manuscript (362-367).

> ~~Comparison of temporal plots, some periods (19:00 to 19:50, 20:50 to 22:00, and after 23:30 UTC on March 28; 01:30 to 04:30, and 06:00 to 07:59 on March 29) have fewer snowflakes per unit volume, while the average size is larger and the deviation of Deq and Dmax is larger, which may be a period when the snowflakes are sparsely distributed in space, with a high degree of aggregation of individual snowflakes and a more complex shape. On the contrary, in other periods (19:50 to 20:50 and 22:00 to 23:30 UTC on March 28; 00:00 to 01:30, and 04:30 to 06:00 on March 29), the particle counts per unit volume were smaller, while the average size and the deviation of Deq and Dmax was larger, which means the aggregation of snowflakes was weakened.~~

In comparing the temporal plots (Fig. 13c, d, e, f), certain periods (19:00 to 19:50, 20:50 to 22:00, and after 23:30 UTC on March 28; 01:30 to 04:30, and 06:00 to 07:59 on March 29) exhibited a smaller number of particles per unit volume, with larger average sizes and greater difference between D_{eq} and D_{max}. This indicates a higher degree of aggregation and potentially more complex shapes of individual snowflakes during these times. Conversely, other periods (19:50 to 20:50 and 22:00 to 23:30 UTC on March 28; 00:00 to 01:30, and 04:30 to 06:00 on March 29) showed a larger number of particles per unit volume, smaller average sizes, and reduced difference between D_{eq} and D_{max}.

27) Eq. (8): What is surface area S? Surface area of the 3D reconstructed hull?

Thank you for your comments. The surface area *S* indeed refers to the surface area of the 3D reconstructed hull of the observed particles. This value is calculated based on the 3D model generated during the reconstruction process. In the revised manuscript (377):

Sp is derived from the *V* and *S* (surface area of 3D reconstructed hull) and characterizes the to which 3D particles approach the sphere:

28) 381, "so blurring that particle motion is insignificant" Check sentence for correct English.

Sorry for our unclear description. We have revised the sentence in the revised manuscript (394-396):

~~The single exposure time of a high-speed camera is 20 μs so blurring that particle motion is insignificant,~~ The exposure time for a single frame is 20 μs, which renders the motion blur of the particles negligible. Additionally, the time interval between two consecutive frames is 5 ms, allowing the same particle to be captured multiple times, thus enabling accurate velocity calculations.

29) 383, "and the same particle is merged into a single image in Fig. 15a" Describe better what was merged into a single image (and that it is done only to better visualize something in the paper/Fig).

Thank you for your comments. For better visualization of the particle movement, we have merged several consecutive frames of the same particle into a single image in Fig. 15a. This approach is only for improved clarity in illustrating the particle's motion. We have revised the corresponding sentence in the revised manuscript (396-397):

~~The same particle may appear in several consecutive images two or more times, and the same particle is merged into a single image in Fig. 15a,~~ The same particle from consecutive frames is merged into a single image in Fig. 15a to enhance the visualization of its movement. The speed calculation schematic is shown in Fig. 15b.

30) In Sect. 6 Conclusion: 446, "pre-calibration" is used only once so that it is unclear what it refers to (add reference to section in paper and use consistent names). Revise the whole Section after revising the manuscript.

Thank you for your comments. It refers to the camera calibration mentioned in section 3.1 rather than pre-calibration. Changed as suggested.

31) Comparisons without clear reference: 342, "across a larger range of sizes" 345, "more and more concentrated small particles" 345-346, "average particle size was consistently smaller" smaller than? (also unclear what "consistently" means) 391, "generally not more than 20%." What is 20%? 435-436, "estimate PSD more accurately" 436, "calculation of velocity more accurately"

Thank you for your comments.

342, "across a larger range of sizes" means PSDs described by D_{max} cover a broader range of particle sizes compared to those described by Deq, which has been mentioned in 25). We have removed it in the revised manuscript.

345-346, The sentence "There were more and more concentrated small particles on March 28 compared to March 29 (Fig. 13a), and the average particle size was consistently smaller (Fig. 13e)." is not sufficiently clear and does not add substantial meaning to the discussion. Therefore, we have removed it in the revised manuscript.

391, "generally not more than 20%." We added some necessary explanations in the revised manuscript (405-407):

> Given that the size of the particles captured in two consecutive frames does not change significantly, the D_{max} and D_{eq} of the particles are similar, generally not more than 20% (the D_{max} or D_{eq} value of the particle in the next frame deviates within ±20% of the previous particle).

435-436, "estimate PSD more accurately". There is no comparison here, so we removed "more" in the revised manuscript.

436, "calculation of velocity more accurately". There is no comparison here, so we removed "more" in the revised manuscript.