An introduction of 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.~~

Our responses are in regular font.

Thank you very much for your thorough review and insightful comments on our manuscript. We appreciate the time and effort you have dedicated to evaluating our work and your constructive feedback. Your suggestions have been invaluable in helping us improve the quality and clarity of our manuscript. Below, you will find our point-by-point responses to your comments, along with the revisions made to the manuscript.

Easily fixed issues:

Pu et al 2021

The new reference to Pu et al. 2021 (L29) is about snow darkening and melting as a consequence of it. As such it is not suitable as reference for "importance of snowflake shapes for our understanding of atmospheric science.

Thank you for your advice. We have removed this reference in the revised manuscript.

Fig 2

The beams extending from the four cameras in Fig 2 are confusing. They are not mentioned neither in the sentence (L104) referring to Fig 2, nor in the caption of Fig 2. Fig 2 shows also the high-speed camera Cam3, apparently with a conical beam extending from it. That camera is not mentioned in L104. For clarity, I would add labels in Fig 2 pointing to the three telecentric lenses, the high-speed camera or lens, and the LED light beams. If you still want to keep the additional beams extending from the lenses, then I would mention them somehow in the caption.

Thank you for your advice. We have added labels in Fig 2 and also referred to them in the caption and in L104.

The three telecentric lenses and LED lighting beams of 3D-PPI are illustrated in Fig. 2.

The four cameras, lenses, and LED lights, including the additional beams of 3D-PPI, are illustrated in Fig. 2.

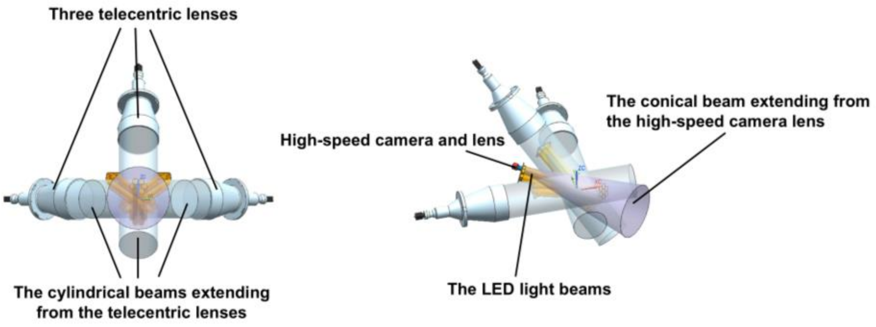

Figure 2. The two views of three telecentric lenses and LED light beams.

Figure 2. The two views of four cameras and lenses, including the additional beams extending from the lenses and the LED light beams.

Three dimensions of OV

The sentence in L104-106 has several issues and should be rephrased. The "three dimensions … is a x b x d" is grammatically and mathematically wrong. The three dimensions are a, b, and d, not a x b x d. The previous issue is better fixed with using the appropriate term instead of "interior rectangle". I think "cuboid" would be correct here (OV of one camera is a cuboid defined by the three dimensions …)

Thank you for your advice. We have revised this part:

To clarify, the three dimensions of observation volume (OV) of one high-resolution camera is a × b × d (170mm × 125mm × 88mm), which represent the length, width, and depth of field of view respectively.

To clarify, the observation volume (OV) of one high-resolution camera is a cuboid defined by the three dimensions: a, b, and d (170mm, 125mm and 88mm), which represent the length, width, and depth of field of view respectively.

Da vs Dp

Thank you for adopting this way of showing results of ceramic spheres. Remove sentence L209-210 which doesn't apply anymore.

Thank you for your advice. We have removed this sentence in the revised manuscript.

Size-dependent OV

*The new Eq (10) to calculate the effective OV as a function of snowflake size seems wrong to me. At each border or edge only D/2 needs to be removed. If a particle of size D is at least D/2 away from the border (distance between particle centre and border), then it will not touch it. So: Vi = (a-Di)(b-Di)*d*

Thank you for your comments, we have modified the formula as you suggested.

Further discussions

REVISED Sec. 3.2 Estimation of pixel resolution (previously Calibration of image binarization) and NEW Sec. 3.3 Calibration of image binarization

For me the estimation of pixel resolution comes too early (before image processing) and requires image binarization (so that that is done twice). Image binarization is optimized, not calibrated. In your response you try to justify why spherical targets are better than planar micrometer scales, where there would be an issue with orienting them perfectly. The disadvantage of the spherical targets is the necessity to do a manual binarization in addition to the adaptive thresholding. The apparent size of spherical objects depends on any binarization. I believe that this is introduces a larger uncertainty than what would be related to imperfect alignment of a micrometer slide. Some more detailed comments below:

Image processing is described in separate sections (3 and 4): The image binarization (now described in the new Sect 3.3) is for me not a "calibration" as the title of the subsection suggests (but the determination of optimal binarization). I would consider image binarization as one of the steps in image processing. Subsection 4.1 is now called "Image processing" and describes the steps of noise removal and segmentation. It would be a clearer structure to first describe the complete image processing and only then results (pixel resolution) and algorithms that use this image processing. Sect 3.3 should be part of image processing (Sect 4.1).

For doing the above, you would need to do the binarization only once (with adaptive thresholding). Now it is done twice: In Sect. 3.2, Dp is "counted manually" to determine pixel resolution, which means that a manual binarization is done. Then, in Sect 3.3, binarization is done again with adaptive thresholding. The "optimal" sensitivity coefficient c is effectively the coefficient where adaptive thresholding approximates closest the manual binarization from Sect 3.2. (It is understandable and

fine that you will not change from your spheres to anything else now. But consider the following:
It would be more transparent to find "manually" the best sensitivity coefficient c (i.e. decide visually what looks to be a good binarization) and make this part of the description of image processing. Then the estimation of pixel resolution could be done with images processed according to that image processing. That would also avoid having twice similar analyses (related to 6b and 7c/d).

Thank you for your comments and advice.

Firstly, we have removed the section "3.2 pixel resolution estimation " for the following reasons. We have completed an accurate estimation of the pixel resolution by shooting standard calibration plates including micrometer slide (shown in Figure 1), which yields a value of 41.6μm·px^{-1}. As you say, estimating camera resolution using spheres relies on image binarization and manual processing, and the estimated pixel resolution will not necessarily be more accurate than the 41.6 μm·px^{-1} given by the manufacturer.

Secondly, we removed the section "3.3 calibration of image binarization " and the content about image binarization were added to section "4.1 Image Processing".

Thirdly, we photographed the ceramic spheres after image processing. Further image processing such as binarization was performed and the errors was calculated.

Fourthly, the previous method of determining the sensitivity coefficient *C* seems to be a bit complicated and not very meaningful. It is indeed a simple and effective method to find the optimal *C* through visual assessment. We have revised this part as your advice:

> The sensitivity coefficient *C* is crucial; it adjusts how the local mean is used to set the threshold. A smaller C favors classifying pixels as foreground, while a larger *C* favors background classification. We have manually adjusted the sensitivity coefficient *C* to determine the optimal value of 0.4. This process involved visually assessing the binarization outcomes for various *C* values to identify which value best distinguishes between foreground objects and background.

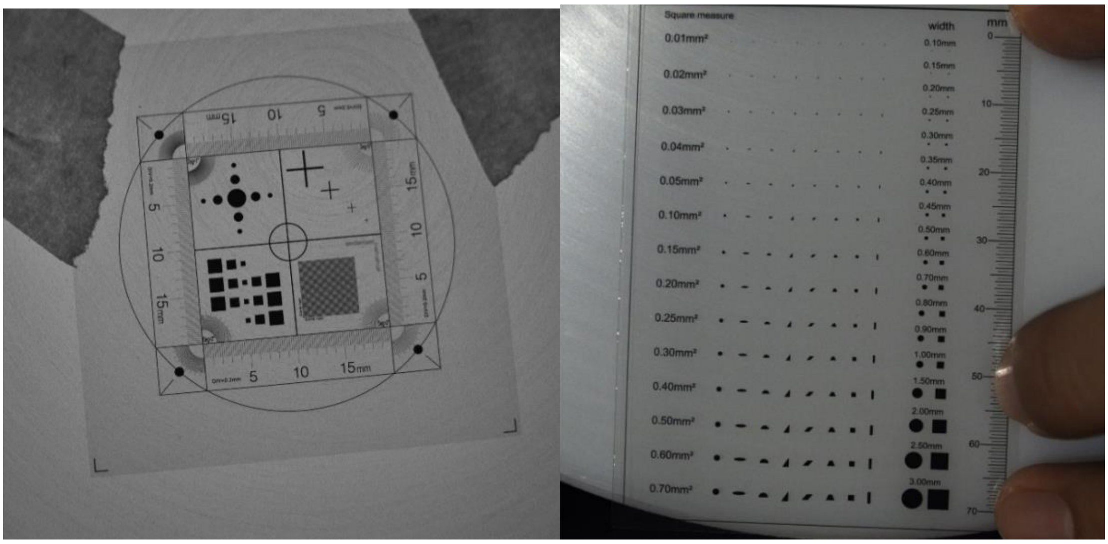

Figure 1. The images of standard calibration plates.

MTF method to determine the depth of field

I appreciate the details on how you determined the depth of field and that you checked and corrected it.

There are two things I didn't understand though:

• *What are "different spatial frequencies"? This is probably a minor thing or me not understanding.*

Thank you for your comments. To clarify, spatial frequency describes the rate that a stimulus changes across space. For images recorded in this manuscript, high - frequency components in an image refers to the sharp edges, fine textures and other detailed information, while low - frequency components refer to large, smooth areas and overall contours.

When using the Modulation Transfer Function (MTF) to determine the depth of field (DOF), we consider different spatial frequencies because a lens's imaging ability for different levels of detail varies with the object distance. Higher spatial frequencies reflect the lens's ability to resolve fine details, and lower spatial frequencies represent its performance in depicting large - scale features.

Reference:

(https://evidentscientific.com/en/microscope-resource/knowledge-hub/anatomy/mtfintro)

Ashoor, M. and Khorshidi, A.: Modeling modulation transfer function based on analytical functions in imaging systems, The European Physical Journal Plus, 138, https://doi.org/10.1140/epjp/s13360-023-03884-8, 2023.

• *How are snowflakes below the MTF threshold "deemed fuzzy and considered outside the depth of field"? You claim that, consequently, your "algorithm effectively excludes these particles from identification". How does this work? Snowflakes are* *detected/identified if they are binarized as a connected region of more than 20 pixels. What is the relation to the MTF threshold?*

Thank you for your comments. In our previous response, we mentioned that the depth of field of the lens was determined by calculating the MTF function. Due to its high accuracy, this MTF method is commonly employed by lens manufacturers to determine the depth of field. MTF is time-consuming and not suitable for rapidly processing a large number of snowflake images to exclude those deemed fuzzy that are out of focus. In the revised manuscript, we have explained how we exclude deemed fuzzy snowflakes outside the depth of field by calculating the variance of the Laplacian:

> Secondly, combine regions into a single region of interest (ROI) when the distance between the closest points of connected regions in a single image is detected to be less than 0.5 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, discard the blurred particles outside the depth of field. To avoid detecting the particles completely out of focus, in the greyscale image before binarization, the mean grey value of the ROI region must be at least 20 greater than the mean grey value of the image and the variance of the Laplacian of the ROI grey value must be at least 10. Fourthly, discard the particles at the edge of the image. If the connected region of a particle contains points located at the edges of the image, the particle is considered not to be fully captured, and it should be discarded.

Description of image binarization Sec. 3.3

I appreciate the added details about the binarization method (adaptive thresholding). However, I have some questions and see a few issues:

• Could you cite some description of this method in the literature? Without that I think that I need some more information.

Thank you for your advice. We apologize for our oversight. We have added literature in the revised manuscript.

> Bataineh, B., Abdullah, S. N. H. S., and Omar, K.: An adaptive local binarization method for document images based on a novel thresholding method and dynamic windows, Pattern Recognit. Lett., 32, 1805-1813, https://doi.org/10.1016/j.patrec.2011.08.001, 2011.

• L221 and Eq (6): what is the "local mean $\mu(u,v)$"? Mean of what? How is it calculated? What is the specified neighborhood? How is that adjusted by the sensitivity coefficient C?

Thank you for your comments. The local mean $\mu(u,v)$ refers to the average brightness of pixels within a specified neighborhood around the pixel (u,v). This neighborhood can be defined in various ways, such as a fixed-size window or a dynamically sized region based on the local image characteristics. The specific method of calculation depends on the implementation, but typically involves averaging the brightness values of the pixels within the chosen neighborhood.

• *MRE defined by Eq (7): this seems to be the average of the two means of the absolute relative errors. Being based on absolute values it is always positive.*

Later you refer to the "MRE of Dmax" (L243) and "MRE of Deq" (L244). According to Eq (7) there is only one MRE, which is based on both Dmax and Deq.

Then, you also refer to "relative errors" of Dmax and Deq. These are not absolute values but positive or negative. From Fig 7 I assume that they are determined as (Dmaxi-Dai)/Dai, which is different from what Eq (7) would suggest ((Dai-Dmaxi)/Dai).

If the worst relative error is -7% then it is strictly speaking wrong to call that the "maximum relative error", which would be +2% in case of Deq (Fig 7f).

Thank you for your comments. We apologize for the confusion caused by the addition of Eq (7) in the last response. We have removed the Eq (7) in the revised manuscript. Therefore, in the revised manuscript there is no MRE, only "relative errors", which are not absolute values but positive or negative.

Thank you for your advice. The "maximum error" is indeed inappropriate and we have changed it to "worst error".

• *The definition of Dmax (L231) is different from the definition of Dmax later used in Sect. 5.2 (L397). The definition in L397 is the one I would expect here. I would call Dmax "maximum dimension" not "maximum size" as doen in Sect 5.2.*

Thank you for your comments. We apologize for our mistakes. To avoid redundancy, we have revised the notation in L397 by renaming D_{max} to DV_{max}. The corresponding content in Fig. 14 has also been updated accordingly. Additionally, we have revised the definitions of both variables, as detailed below:

L231: the D_{max} is ~~the distance between the two farthest points of the particle profile~~

the diameter of the smallest enclosing circle

L397: ~~maximum size Dmax (diameter of the smallest enclosing circle)~~

maximum dimension DV_{max} (distance between the two farthest points on the surface of the particle)

Pixel noise (L274-275)

Referring to detected regions with less than 20 connected pixels as "pixel noise (no larger than 20 pixels)" is now clearer than "small noises" previously. It implies that all regions with less than 20 connected pixels are indeed noise, i.e. not related to actual snowflakes. I am not sure this is true in general. Could these "noise" features be caused *by snowflakes that are outside the depth of field, or by small snowflakes that are too small to be detected by 3D-PPI? So, rather than and/or in addition to "prevent these noises from being mistakenly detected as small snowflakes" it should say "exclude features of small snowflakes that cannot be detected from analysis"?*

Thank you for your comments. As you pointed out, these so - called "pixel noises" could indeed be small snowflakes or snowflakes outside the depth of field (We have already mentioned the snowflakes outside the depth of field in the revised manuscript.) Given their small size, it is extremely difficult to extract the features of these tiny regions. Thus, we excluded them from the analysis. We have revised the sentence as your advice:

> Thirdly, discard the particles with an area smaller than 20 pixels (Equivalent to 0.035 mm^2, D_{max} is about 0.2mm), which enables the removal of pixel noise or small snowflakes (no larger than 20 pixels) from the image, to ~~prevent these noises from being mistakenly detected as small snowflakes~~ exclude features of small snowflakes that cannot be detected from analysis.

2-mm gap criterion

Your new criterion for joining regions is better than the previous one. It, however, still allows that small regions would be joined across a gap that can be larger than these regions. A 20-pixel region has about 14 to 5 pixels across, and a 2-mm gap corresponds to almost 50 pixels. I.e., two 5-pixel regions could be joined even if they are separated by a about ten times larger gap. I think two such regions should rather be excluded. Would they indeed be belonging to the same snowflake, then that would mean that a large part of this snowflake would have been missed (not been detected by binarization) likely due to being out of depth of field.

Thank you for your comments. We fully recognize that our previous criterion for joining regions might not be reasonable, which could lead to the situation that you described. After careful consideration, we have raised the standard for combining. Now, only connected regions with an area larger than 20 pixels and a distance of less than 0.5 mm between their closest points will be combined. As a result, we have reordered the sentences in the manuscript by first discarding connected regions with an area larger than 20 pixels, and then combine regions into a single particle when the distance between the closest points of connected regions in a single image is detected to be less than 0.5 mm apart. This adjustment allows us to preserve only the shape features of the main part of the particles and effectively avoids the problem that you mentioned. We have revised the sentence in the manuscript:

> ~~(ii) Particle detection. Firstly, detect the connected regions in binarized images. Secondly, combine regions into a single particle when the distance between the closest points of connected regions with an area larger than 20 pixels in a single image is detected to be less than 0.5 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, discard the particles with an area smaller than 20 pixels (Equivalent to 0.035 mm2, Dmax is about 0.2mm), which enables the removal of pixel noise or small snowflakes (no larger than 20 pixels) from the image, to exclude features of small snowflakes that cannot be detected from analysis.

(ii) Particle detection. Firstly, in binarized images, detect the connected regions with an area larger than 20 pixels (Equivalent to 0.035 mm^2, Dmax is about 0.2mm), which enables the removal of pixel noise or small snowflakes (no larger than 20 pixels) from the image, to exclude features of small snowflakes that cannot be detected from the analysis. Secondly, combine regions into a single region of interest (ROI) when the distance between the closest points of connected regions in a single image is detected to be less than 0.5 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. This method enhances the accuracy of foreground particle detection, particularly in images with complex backgrounds and uneven illumination. Thirdly, discard the blurred particles outside the depth of field. To avoid detecting the particles completely out of focus, in the greyscale image before binarization, the mean grey value of the ROI region must be at least 20 greater than the mean grey value of the image and the variance of the Laplacian of the ROI grey value must be at least 10.