Dear AMT Editors and Reviewers,

Thanks for starting the review process and for the precious comments and suggestions provided. In what follows we list the comments expressed by the Anonymous Referee #2 (paragraphs written in blue color), provide our answers (paragraphs written in black color) to the raised points and explain how we addressed each point in the revised manuscript.

Comment on amt-2022-61 from Anonymous Referee #2

This paper describes a method to estimate the 3D cloud envelope and development velocity using simulated images of a triplet of small satellites in a sun-synchronous orbit at 600 km height. The focus lies on trade wind cumulus and deep convection, while the methodology relies on stereo analysis and tracking to compute the 3D points of the cloud envelope and subsequent cloud motion estimation. The study assesses the feasibility and accuracy of the proposed mission design.

The paper is well written and structured and the study is thoroughly conducted and the topic is scientifically relevant. I suggest publication after the following points have been addressed:

1. Fig. 1:
Could be a bit more detailed. Just a few more numbers and lines if possible. For example the height of and the distance between the individual satellites might be good for directly understanding the geometric setup, which is important for stereo analysis and triangulation.

Authors' reply:
We edit the figure by adding a line with arrows that specifies the height of the satellites from ground and additional lines that indicate the baseline in case of two and three satellites.

Comment on amt-2022-61 from Anonymous Referee #2
2. Line 106:
"consequently simultaneously". Sounds a bit strange. Maybe just leave the "consequently" out.

Authors' reply:
Line 106: We remove "consequently" so that the phrase (lines 105-107, pp. 4) (lines 122-124 – pp. 5 in the revised manuscript) now becomes:
"The measurements of these space-borne sensors will simultaneously document the vertical cloud development retrieved by a stereoscopic method, the lightning activity and the distribution of water vapor at a high resolution by exploiting the multi-angle acquisitions."

Comment on amt-2022-61 from Anonymous Referee #2
3. Was it described somewhere what axis the along-track direction was (x or y)? Maybe add it to one of the initial figures so that analysis later is easier.

Authors' reply:
We now add x (along track) and y (across track) arrows in both Figure 2 and Figure 6 (that becomes Figure 5 in the revised manuscript).
In this respect we correct the phrase, lines 114,115 (pp. 7) (lines 132, 133 – pp. 7 in the revised manuscript):

"It should be noticed that satellite is moving from North to South and while the across track resolution remains almost constant the along track resolution is increasing for tilted views."

and rewrite as follows:

"It should be noticed that the satellites are moving from North to South (see x arrow in Figure 2) and while the across track resolution remains almost constant the along track resolution is decreasing for tilted views."

We also add in the caption of Figure 2 and Figure 6 (that becomes Figure 5 in the revised manuscript) the following sentence:

"The along and across track directions are identified by the x and y arrows, respectively."

Comment on amt-2022-61 from Anonymous Referee #2
4. Line 115 + 116:
For a tilted view shouldn't the track resolution be decreasing as a larger area is projected to a smaller image area? The same with the Ground Sampling Distance. Shouldn't it increase for a tilted view? Maybe just a misunderstanding.
Authors' reply:
That is correct. Thanks for pointing this out.
We correct the following sentence (lines 114-116, pp. 7) (lines 132-134 – pp. 7 in the revised manuscript):

"It should be noticed that satellite is moving from North to South and while the across track resolution remains almost constant the along track resolution is increasing for tilted views. This leads to an increase of the footprint in the along track direction and a reduction of the ground sampling distance (GSD)."

and rewrite as follows:

"It should be noticed that the satellites are moving from North to South (see x arrow in Figure 2) and while the across track resolution remains almost constant the along track resolution is decreasing for tilted views. This is due to the increase of the footprint in the along track direction and the consequent increase of the ground sampling distance (GSD)."

Comment on amt-2022-61 from Anonymous Referee #2
5. Line 185:
I think the field of view is always constant as it depends on the camera. If the angle representing the image projection of the ground area is meant then it should be smaller with tilted view, shouldn't it?

Authors' reply:
Lines 184,185 (pp. 12) (lines 266, 267 – pp. 16 in the revised manuscript), we correct the following phrase:
"Furthermore, pixel size increases with the distance from the image center and so does the field of view for tilted views."

by re-writing as follows:

"Furthermore, pixel size increases with the distance from the image center and so does the footprint for tilted views."

Comment on amt-2022-61 from Anonymous Referee #2
6. Line 278:
You write that the cameras are affine for a small tile of an image. Considering the field of view in this simulation of 1 degree, the cameras already are very weakly perspective. Does an additional tiling matter?

Authors' reply:
The s2p pipeline was conceived for 3D stereo processing of pushbroom camera images that are not rectifiable as opposed to images taken from pin-hole cameras. Errors in the rectification may result in a vertical disparity (epipolar error) between corresponding points in the rectified images, which may significantly affect the performance of the stereo matching. By approximating the sensor by an affine camera model on small image tiles leads in practice to an almost perfect rectification, with a very small epipolar error.
However, in our case, although simulations are weakly perspective, especially true when the relative distance (scene depth) between two points of a 3D object along the optical axis is much smaller than the average distance to the camera, we do not need further tiling as rectification works well (with no epipolar error) by setting the tile size equal to the ROI (865px x 865px).

For more clarity, after the following sentence (lines 277-278, pp. 18) (lines 314, 315 – pp. 18 in the revised manuscript):

"The s2p process pipeline can be summarized as follows: input images are first cut into tiles, where the cameras are assumed to be affine (i.e. perspective projection)."

we add:

"With respect to the calculations presented in this work, we use a tile size equal to the ROI (865px x 865px) as we achieve satisfactory rectification with no need of further reduction of the tiling."

Comment on amt-2022-61 from Anonymous Referee #2
7. Line 279:
Maybe shortly describe what an epipolar line is and why it's useful.
Authors' reply:
Lines 278, 279 (pp 18) (lines 316-318 – pp. 18 in the revised manuscript), we edit the following sentence:
"With regard to Fig. 8, the input reference (ref) and secondary images (sec) are first rectified (rec ref, rec sec) to simplify the search of matching features (stereo matching) along the epipolar lines."

by deleting "to simplify the search of matching features (stereo matching) along the epipolar line"

and then by adding the following:

"Stereo-rectification allows to restrict the search of corresponding features from the entire image to a single line. For any point p in the reference view, the corresponding point p' in the secondary image, provided that it exists, lies on the epipolar line of p, that is EL p. Analogously, p lies on the epipolar line of p', EL p'. There is a correspondence between the epipolar lines of the two views for images taken with pinhole cameras. In this case the epipolar lines are said to be conjugate. The purpose of rectification is that of resampling the images in such a way that matching points are located on the same row (epipolar lines become horizontal), thus simplifying the search of matching features and allowing to use conventional stereo matching algorithms."

Comment on amt-2022-61 from Anonymous Referee #2

8. Line 279 / Fig. 8:

You describe that you conduct a stereo image rectification in order to make the stereo analysis easier, aren't you? In that case the y-component is usually zero (which you write in line 285). Is that correct? Also the disparity (parallax) should have values between 0 and infinity (or negative). But in Fig. disparities are both negative and positive. A negative disparity would mean that the observed point is behind the cameras. Or do I miss something?

Authors' reply:

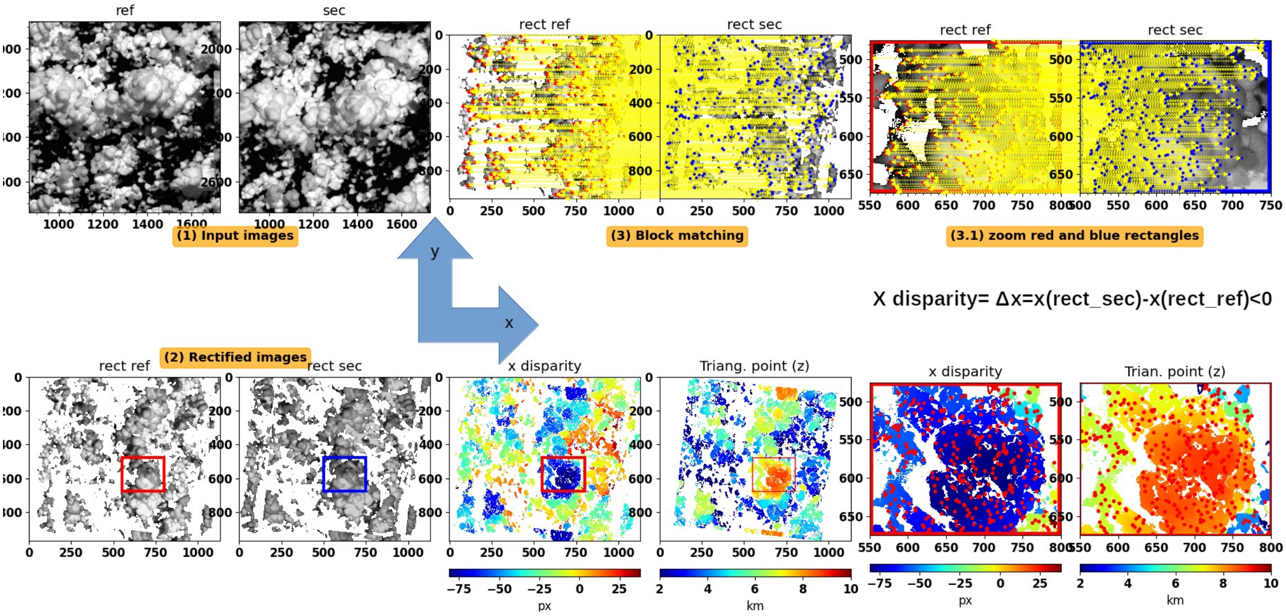
Image rectification is indeed done to make the search of matching features easier. In this way matching cloud pixels are found on the same horizontal lines (Δy=vertical disparity=0). Vertical disparity is 0 as this is the main purpose of rectification. If this was not the case this would introduce an "epipolar" error that would affect the performance of the stereo matching. Once the images are rectified the only disparity that we are left with is the disparity along the x axis. This is measured, for each pair of matching features with image coordinates (h, k), in the rectified reference image, and (h+Δx, k), in the rectified secondary image, as X disparity= Δx=x(rect_sec)-x(rect_ref).

Following are two examples of negative and positive disparity, respectively.

Case 1: Negative disparity - Points closer to the cameras

In this case disparity is negative as for any given pair of matching features it turns out that

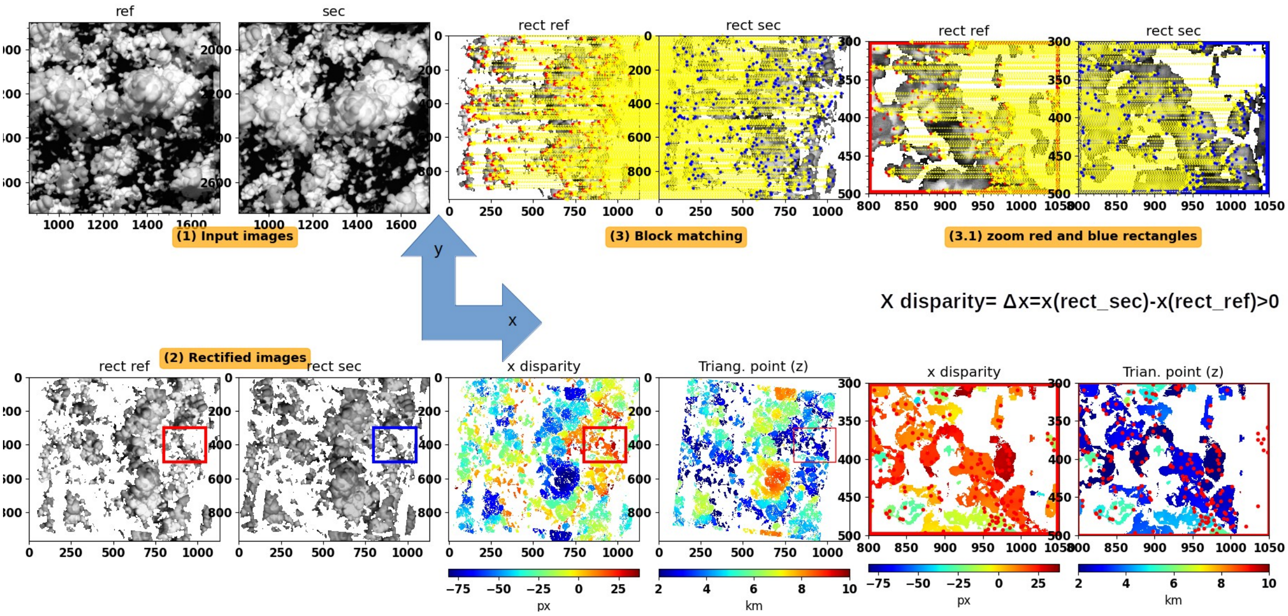
x rec sec < x rec ref (see x values for matching cloud structures).



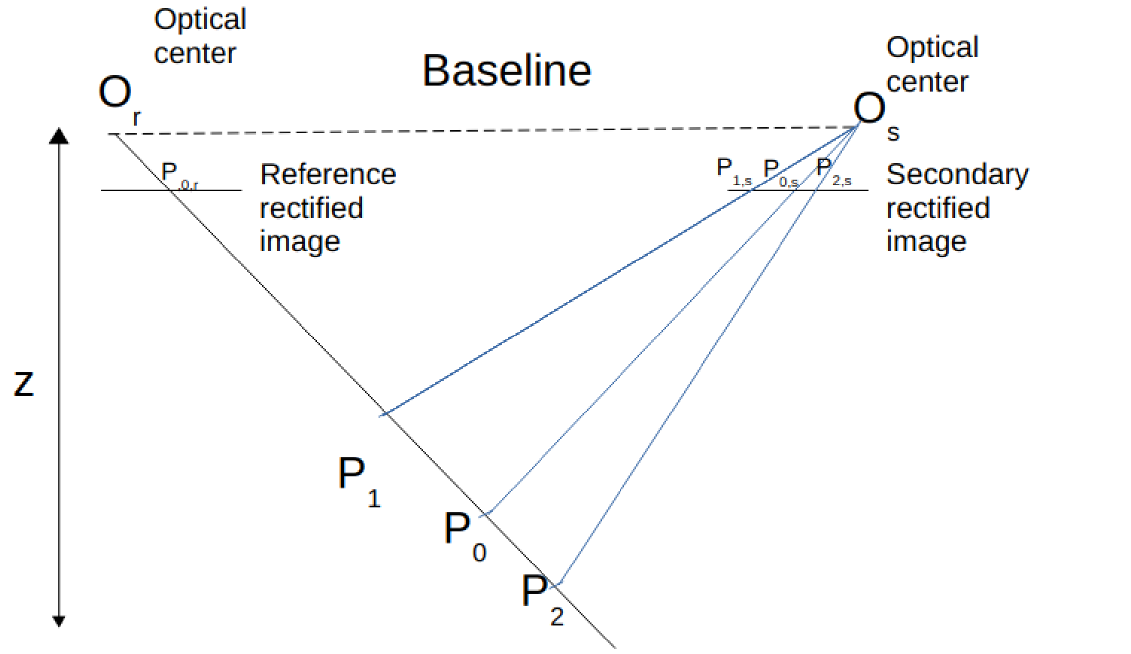
Case 2: Positive disparity: Points farther away from cameras

In this case disparity is positive as for any given pair of matching features it turns out that

x rec sec > x rec ref (see x values for matching cloud structures).



For further clarity we can refer to the following drawing, which is an alternative way of looking at things, meant to show, from geometric principles, the fact that a negative disparity is actually expected for any given point P_1 at a closer distance to the reference camera than to the secondary one and vice versa (positive disparity), for any given point P_2 closer to the secondary camera than to the reference one.



P_0 is equidistant from the two cameras, $P_{0,l}$ and $P_{0,r}$ have both the same coordinates (h,k) in the images anddisparity along x is nul

P_1 is closer to the left camera (reference), $P_{1,l}$ coordinates in the reference image are still (h,k) and $P_{1,r}$ coordinates in the secondary image is (h+dx) with the disparity dx < 0

P_2 is further away from the left camera (reference), $P_{2,l}$ coordinates in the reference image are still (h,k) and $P_{2,r}$ coordinates in the secondary image is (h+dx) with the disparity dx > 0

For further clarity we redo figure 8 that now includes the x and y axis coordinates for the two zoomed image insets, which correspond to the red and blue rectangles drawn in the rectified images, containing the uppermost part of the cloud top. This should highlight the fact that, for any given pair of matching features, disparity is negative if rectified reference x > rectified secondary x. We also rotate the input images 90 degrees clockwise for consistency with the along track direction x and add axis grids in order to emphasize the fact that after rectification the y disparity becomes 0.

Moreover, we change color scale to improve on data visibility. For consistency we change color scale throughout the manuscript (see Fig. 11a, 12a and 12b).

Lines 289-293 (pp. 19) (lines 333-337, pp. 19 in the revised manuscript) we rephrase the following sentence:

“The disparity dx associated to each tie point is the distance between two corresponding points in the rectified images (see Fig. 8, step 3). Cloud pixels closer to the camera (i.e. cloud top) having larger difference in relative shift along x in the two rectified images are associated to larger disparity values (deep red points), whereas points farther away from the camera (associated to lower difference in relative shift) are associated to lower disparities (light blue/brown pixels).”

as follows:

“The disparity dx associated to each pair of matching features is the distance between two corresponding points in the rectified images (see Fig. 8, step 3, dx=rect sec x – rect ref x). Cloud pixels closer to the cameras (i.e. cloud top) are associated to negative disparity values (deep blue points) as, for any given pair of matching features, rect sec x < rect ref x, whereas for points farther away from the cameras, associated to positive disparity, rect sec x > rect ref x (orange/red pixels).”

We also add in the caption of figure 8:

“Notice that Ref and Sec have been rotated 90 degrees clockwise for consistency with the orientation of the along track direction x”

Comment on amt-2022-61 from Anonymous Referee #2
9. Fig. 10:
What explains the large differences in Fig. 10F for A10 and A11? Shouldn’t the differences at least be symmetrical / similar to A1/A2? Similarly Fig. 10b. What could be a reason?
Authors’ reply:
As we have already pointed out in reply to similar remarks from reviewer 1, with respect to Figure 10, we have repeated the calculations and accordingly updated the figure that now includes calculations for all scenarios.
In doing so, and in particular with respect to Fig. 10f, we notice that the mean Y error for configurations 1-3 now becomes 5 m/s whereas previously it was about 15 m/s and differences become symmetrical.
We then correct the sentence, lines 357, 358 (pp. 24) (lines 406, 407 – pp. 25 in the revised manuscript):
«The mean difference (its absolute value) along z is less than 25 m while it is less than about 5 m and 15 m along x and y, respectively.»
as follows:
«The mean difference (its absolute value) along z is less than 25 m while it is less than about 5 m along x and y.»

In the abstract we also correct, lines 9-11 (pp. 1) (lines 12-14 – pp. 1 in the revised manuscript):
“The accuracy of the retrieval of cloud topography is quantified in terms of RMSE and bias that are respectively, less than 25 m and 15 m for the horizontal components and less than 40 m and 25 m for the vertical component.”
as follows:
“The accuracy of the retrieval of cloud topography is quantified in terms of RMSE and bias that are respectively, less than 25 m and 5 m for the horizontal components and less than 40 m and 25 m for the vertical component.”

With respect to Figure 10b, concerning the skewed distribution of the error along z for the A9-A11 views, we add the following sentence, after (lines 357, 358 – pp. 24) (lines 406, 407 - pp. 25 in the revised manuscript) “The mean difference (its absolute value) along z is less than 25 m while it is less than about 5 m along x and y.”:

“The skewed distribution of the error in Figure 10b, for the views A9-A11, may be due to the fact that fewer cloud features are visible as the clouds are less illuminated by the sun, with larger portions of the cloud field shaded, as it can be seen from Figure 5g, 5h and 5i.”

Comment on amt-2022-61 from Anonymous Referee #2

10. Sec. 5.3, comment: It is good that you mention possible differences due to the different distance estimation methods.

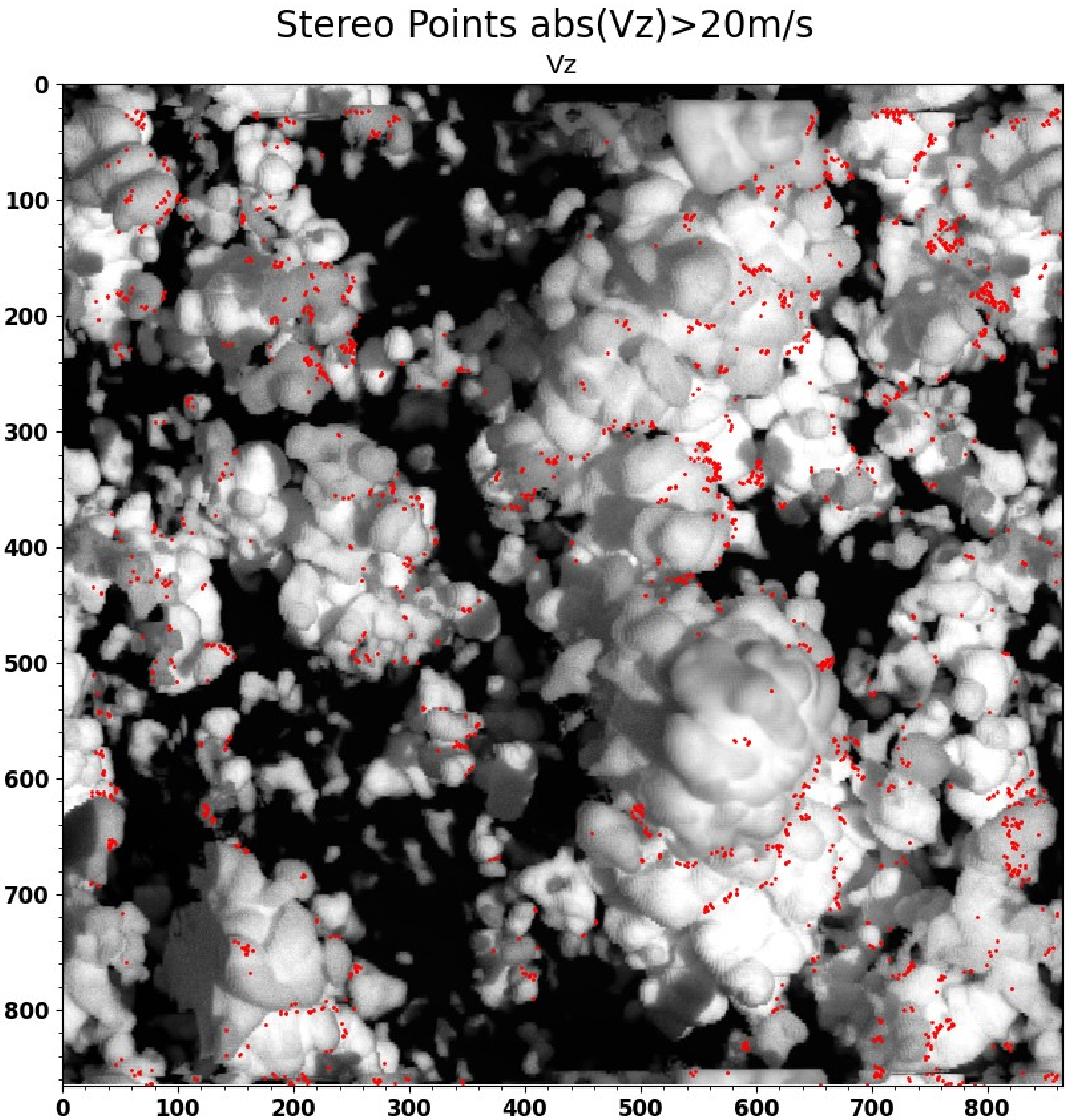
Comment on amt-2022-61 from Anonymous Referee #2

11. Fig. 12:

A sigma of 22.85 m/s in Fig. 12a for Vz seems a bit large considering the histogram.

Authors’ reply:

Thanks for this remark. That is correct. By redoing the calculations we found out that about 2% of the total stereo retrieved points, those shown in red in the following figure,



are associated to unrealistically high values of Vz with abs(Vz)>20m/s. As it can be seen these points are mostly located over the cloud edges or in shaded regions where stereo reconstruction is expected to be less accurate. We therefore screen out these points and by redoing the calculations we now find a sigma of about 5.1 m/s , 1.1 m/s and 6.8 m/s for z, x and y, respectively and mean values of about 1.6 m/s, 6.4 m/s and 5.9 m/s for z, x and y, respectively.

For consistency with these changes, we redo Fig. 12a, extend the range of values, for the x axis of the Vz histogram, up to 20 m/s and also update (line 408-410, pp.28) (lines 461, 462 – pp. 30 in the revised manuscript) the following phrase:

"The horizontal velocity components show that the cloud is moving along the diagonal direction with a mean velocity of (6.5, 6.0) m/s."

by rewriting:

“The horizontal velocity components show that the cloud is moving along the diagonal direction with a mean velocity of (6.4, 5.9) m/s.”

and the following phrase (line 410, pp. 28) (lines 463 – pp. 30 in the revised manuscript) as well:

“in the y direction the distribution is wider (σ = 6.5 m/s)”,

by rewriting:

“in the y direction the distribution is wider (σ = 6.8 m/s)”

Moreover, before the following sentence (line 411,412, pp.28) (lines 466-468 – pp. 30 in the revised manuscript):

“With regard to the GE velocity derived from the GT envelopes, the mean velocities of 0.6 m/s, 6.5 m/s and 6.1 m/s are consistent with the ST mean velocities.”

we add:

“It should be noticed that about 2% of the total number of stereo retrieved points, mostly located over the cloud edges or in shaded regions where stereo reconstruction is expected to be less accurate, are associated to unrealistically high values of Vz with abs(Vz)>20 m/s. Such points were filtered out.”

Comment on amt-2022-61 from Anonymous Referee #2
12. Line 443:
“attitude” → “altitude”?

Authors’ reply:
Line 443 (pp. 30) (line 501 – pp. 32 in the revised the manuscript), we mean indeed “attitude” in that by estimating a mean velocity over 100s we expect to smooth out the contribution to the error associated with satellite orientation.