# Response to the Anonymous Referee #1 comments for the manuscript

*"Cloud height mapping using multi simultaneous sky images from an all-sky camera network"*

By Celia Herrero del Barrio et al. in AMT

First, we are grateful for the effort of referee #1 and his/her review in detail. Reviewer comments are in black font (RC), and author comments (AC) in red font.

## Author's answer to Anonymous Referee #1

**General comments:**

**RC:** This manuscript presents a comprehensive and well developed methodology for retrieving cloud base height (CBH) and cloud top height (CTH) using a network of 20 all-sky cameras deployed around Valladolid (Spain). The approach combines stereoscopic reconstruction, advanced filtering, and multi-camera aggregation to produce near-real-time cloud height maps with spatial resolution of 50 m.

The study is clearly motivated and addresses a relevant challenge in atmospheric remote sensing, as it fille the the gap between high-resolution but local ground-based instruments (e.g., ceilometers) and spatially extensive but temporally limited satellite measurements.

Validation against Sentinel-2 cloud masks and ceilometer observations spanning ∼2.5 years strengthens the proposed method. The authors report good CBH performance, with $R^2 \approx 0.93$ with low bias, demonstrating significant potential for operational applications such as solar nowcasting and cloud monitoring.

Overall, the manuscript represents a valuable contribution to atmospheric measurement techniques, but several aspects require clarification, and detailed discussion before publication.

1. **RC:** The main novelty of this work is to scale the existing stereographyc approach using all-sky cameras to a large network in automate and real time processing. However, more explicite explanation of this novelty to differenttiate respect to the key prior works would be desired. In particular, the authors should try to clarify wether the main innovations lies in network scale, agregation strategy or near real time implementations with improved accuracy.

   **AC**: We thank the reviewer for this comment and think that the innovation lies in combining the improvements in the three aforementioned aspects within an operational framework. We have tried to highlight it in the conclusions as follows:

   Pg 35, ln 759: *This work proposes a robust methodology for the identification and three dimensional localization of clouds using stereoscopic calculations, implemented through an unprecedented 20 all-sky camera dense network deployed around the city of Valladolid (Spain). It presents a novel aggregation strategy in which each camera is paired with every overlapping camera, and the resulting heights are combined onto a common spatial grid, together with a fully automated, near real time operational implementation, generating openly available cloud height products every 5 minutes during daytime and, as a further novelty, every 2 minutes at nighttime. The workflow also brings together for the first time a combination of different recommendations from previous studies, such as downsizing images, using rectified row-oriented pairs of simultaneous images to improve finding pixel correlations, using blocks of neighboring pixels in the correlation to avoid mismatches, using different baseline* *distances and applying smoothing filters that improve the segmentation of clouds. In addition, this implementation is also relevant due to the simultaneous application of all these novelties and aggregation strategies, while significantly optimizing the processes to achieve a high computational efficiency. For the case study shown in Section 3 (28 December 2025 at 11:20 UTC), the total processing time to retrieve cloud heights for the 16 available camera pairings was 102 s using our server infrastructure. However, this processing time is significantly optimized in the operational version through parallelization. By executing 40 processes simultaneously, a total of 120 cases (captured between 07:30 and 17:25 UTC) were processed in just 6 min 44 s. This results in an effective average processing time of 3.37 s per case. This allows for the generation of cloud position maps, including cloud base and top height (CBH and CTH), in near-real time, representing a substantial improvement over previous implementations.*

2. **RC:** The methodology is well detailed but frequently it is difficult to follow. Please improve the detail in the next aspects:

   - The image rectifications and projection stage (Section 3.1.1) would be benefit of a clearrer mahtematical formulation, maye including an schematic diagram summarizing transformations and coordinate system.

     **AC**: We have improved the formulation of the projection procedure, explicitly mentioned the equation for the coordinate transformation:

     Pg 10, ln 245: *Then the geographical coordinates of the cameras are used to compute the relative displacement between both cameras. With this distance between the cameras on a local cartesian system ($\Delta x$, $\Delta y$), it is possible to obtain the relative orientation angle between the reference and auxiliary camera as: $\beta = arctan\left(\frac{\Delta x}{\Delta y}\right)$. After that, both the reference and auxiliary fisheye images are transformed into planar projections. This is possible due to the geometrical calibration parameters of each camera. As mentioned in Section 2.1, the geometrical calibration provides the azimuth and zenith angles from which each pixel is viewed by the camera, but also the pixel coordinates of the zenith (camera center) and the relationship between the pixel radial distance (r) to this center position and the sky zenith angle ($\theta$); this relationship is given by the coefficients of a second order polynomial. In a first step, the azimuth angles are corrected by an azimuth offset ($\Delta\phi$) to align with north-up. To obtain the planar projections of the images, a Cartesian grid with uniform spacing and same resolution as the input image is generated at a fixed height. In this work, the grid is defined as a 4 km square at Z = 0.6 km above the camera. This grid is oriented according to the angle $\beta$ relative to the North, ensuring that both images are projected into the same reference system. This alignment is crucial as it enables a direct pixel by pixel comparison along the rows of both planar images. For each point in the grid $(X, Y)$, the corresponding zenith and azimuth angles are calculated knowing the height and distance from each point to the center of the grid $\left(\theta = arctan\left(\frac{\sqrt{X^2+Y^2}}{Z}\right), \phi = arctan\left(\frac{X}{Y}\right)\right)$. Then the polynomial $r(\theta)$ and the azimuth offset $\Delta\phi$ are used to get the pixel coordinates from the viewing angles ($u = u_0 - r(\theta)\sin(\phi + \Delta\phi)$, $v = v_0 - r(\theta)\cos(\phi + \Delta\phi)$, where $x_0$, $y_0$ are the optical center coordinates obtained from the calibration). Finally, bilinear interpolation is performed on these pixel values (u,v) for each channel of the RGB image, evaluating each pixel of the original fisheye image.*

   - The derivation of the uncertainties propagation of the stereoscopic height equation (eq. 1) would be intereting, instead of the reported ±1 pixel shifts.

     **AC**: We thank the reviewer for this suggestion and would try to clarify the approach used. The derivation of Eq1. shows that the uncertanties depend explicitly on the angular errors in zenith and azimuth and the precision with which the geographical coordinates and altitude of each camera and therefore the distance between them are known. Regarding the angular terms, their uncertainty arises from the geometrical calibration of the cameras, which according to Antuña-Sánchez et al., (2022) is on the order of 1-2 pixels. In the proposed methodology, the ±1 pixel shift used to estimate the retrieval uncertainty is applied on the downsampled 200x200 pixel image, which corresponds to approximately ±10 pixels in the original 2000x2000 pixel resolution image, corresponding to a bigger error than the caused by the uncertainties in zenith and azimuth values. This choice was intentional to ensure a conservative approach. It exceeds the geometric calibration uncertainty and so it absorbs residual pixel matching errors that may not be fully captured by the RGB-correlation threshold filter. Intentionally it can also account for the uncertainties in the precision of the geographical coordinates of the cameras and ensures a conservative error of ±1km for all the retrieved heights, not deppending in the differences between the cameras, smoothing the errors from the calibration and ensuring the computational cost is still efficient. This has been explicitly added in the new version of the manuscript.

Pg 13, ln 308: *The uncertainty in the retrieved cloud height values is obtained by slightly shifting ±1 pixels the column positions of the maximum RGB-correlation coefficient value in the auxiliary image. For each displaced position, the cloud height is recalculated using equation* **??**. *Finally, for each pixel, the absolute differences between the original cloud height and the ±1 recalculated values are computed and the maximum of these two is considered as the uncertainty on the retrieved pixel cloud height. This uncertainty is greater than the uncertainties derived from zenith and azimuth because the error from these angles is around 1-2 pixels (Antuña-Sánchez et al., 2022), whereas in this case, because of the image downsizing, the ±1 pixel shift corresponds to ±10 pixels in the original image.*

Pg 14, ln 325: *Also, cloud height values whose associated uncertainty exceed the chosen threshold of 1 km are also discarded. The error in the assumed height will be 1 km with this threshold; this is a conservative choice, however, in many cases the actual error should be smaller. The quality of the pixel matching is then considered: cloud heights derived from pixels with a RGB-correlation coefficient lower than an established threshold of 2.25, or from undefined values, are rejected, as these cases correspond to poor or ambiguous correlations. Finally, the calculated segmentation sky mask (see Section 2.1) of the reference image is used to restrict the analysis to pixels identified as cloudy (cloud and thin cloud classes), thereby avoiding spurious correlations in cloud-free regions.*

- How the thresholds criteria have been choosen? Please justify them quantitatively even from the empirical point of view. Are the ersults sensitive to this thresholds?

  **AC**: The selection of the different thresholds was based on an analysis using a set of representative reference cases. These cases included scenes with low clouds, high clouds, multilayers, and cloud fields with easily identifiable shapes and contours that allowed a visual assessment of the performance. For these reference cases, cloud heights were retrieved while systematically varying the threshold values and comparing the results with cloud base height measurements from the ceilometer, together with the cloud shape observed in satellite imagery. The correlation coefficient threshold was selected after adopting the image downsampling strategy. Reducing the image size to 200×200 pixels smooths part of the small scale image texture. Under these conditions, using a higher correlation coefficient threshold resulted in the rejection of an excessive number of valid matches and a significant reduction in spatial coverage. The chosen value was selected finding a qualitative compromise to discard most of cloud-free points without rejecting a huge amount of real cloud points in the images.

The minimum and maximum cloud height filters were defined according to the geometry and field of view when pairing the all-sky cameras. The selected height limits ensure that only cloud targets observed within a reliable angular range are considered. The uncertainty threshold, as well as the minimum number of cameras and matching points required for a valid retrieval, were selected by evaluating the compromise between removing outliers and retaining a sufficient number of observations. Relaxing these criteria increased the occurrence of unrealistic cloud heights caused by cloud-free correlations, image artifacts, or features such as dirt on the camera dome that could be mistakenly correlated between different cameras. On the other hand, stricter thresholds significantly reduced the number of valid retrievals. Overall, the cloud height estimations show a moderate sensitivity to the selected thresholds. Variations within reasonable ranges mainly affect the number of points and the occurrence of outliers and should be checked if the density of cameras in the network is different. This has been addressed in the new version of the manuscript.

Pg 37, ln 778: *A fundamental aspect of this methodology is its robust filtering process, which utilizes different criteria based on defining cloud levels and frequency analysis to validate the cloud height retrievals from the multiple camera pairs. These thresholds were selected empirically from a set of representative reference cases, balancing the rejection of outliers against the retention of a sufficient number of valid retrievals. This filtering stage is essential for integrating the vast amount of data generated by the network, as it successfully discards unrealistic values and allows for the determination of both CBH and CTH values.*

3. **RC:** The methodology strongly depends on a supervised segmentation model (U-Net), which is just trained for daytime conditions. Since nighttime and twilight performance are cleary degraded with respect to daytime performance, please discuss the potential bias induced by segmentation uncertainties on CBH and CTH retrieval. Could be the proposed method work without this segmentation process?

   **AC**:We acknowledge that the segmentation performance is indeed optimized for daytime conditions, as it is based on a U-Net trained on such data. However, this does not affect the stereoscopic methodology or its performance, as the correlation and cloud height estimation are applied to all pixels regardless of the segmentation mask. It is only applied after that to discard pixels that are not classified as clouds. The reason for this is that pixels which are not classified as clouds, but for which a height is retrieved, should be correlation outliers. Therefore, there is no need to include them in the aggregation of cloud height estimates among all the camera pairs. Consequently, the use of this U-Net mask only affects to the data points available after the cloud height estimation process. The methodology would work without it, however, that would require significantly stricter outlier and smoothing criteria. Employing the U-Net allows for much more efficient filtering and smoothing, as applying those processes to all the camera pair cloud height estimations would be computationally expensive. It would also be possible to use other segmentation methods not based on neural networks, such as those relying on the red–blue ratio, among others.

4. **RC:** The validation for CTH and CBH against ceilometers and Satelite could be problematic, especially in case of thick clouds. Can you comment on that?

   **AC**: Regarding the ceilometer comparison, we agree that thick clouds present difficulties for the validation because of signal saturation. Following the suggestion of reviewer 2 we have changed the discussion in section 4.2 in order to show the differences when filtering out the cases when the ceilometer could not penetrate the cloud, therefore the cloud penetration depth used to estimate the cloud top height was not reliable (See new Figure 13). With this change we want to enforce the idea that it is not a straightforward comparison and there are some difficulties in the retrieval of this parameter from ground based measurements. On the other hand, the comparison with Sentinel-2 (Section 4.1) concerns only the cloud mask (cloud/no-cloud classification), since Sentinel-2 does not provide cloud height information, then the difficulties in this comparison could come more from the thin clouds than for the thicker ones, that are easier to be detected by the satellite. For sure there would be differences when comparing the CTH from the satellite and from the cameras for thick clouds but this is the scope for future work.

5. **RC:** The authors identify some performance limitations (e.g., low clouds, high clouds at night, geometry constraints) but, did you quantify them more systematically? Are they in relation with uncertainties of the method or realted to previosu uncertainties (geometric calibration of the all-sky imager, segmentation, etc)

   **AC**: Although this comment is really relevant, we would like to clarify that the determination of the performance limitations is already provided throughout the manuscript, related to specific sources of uncertainty. The impact of illumination conditions, which involves nighttime, is quantified in Table 3, supported by the comparison metrics ($R^2$, MBE, RMSE, SD) presented in Figures 12 and 13. In the discussion of these figures, the limitations arising from the segmentation model were also addressed, ranging from dirt or reflections on the domes being misclassified as clouds to the lower performance observed at nighttime. The source of error, related to dome dirt and segmentation artifacts, is more difficult to quantify systematically; however, it has been qualitatively observed that spurious cloud detections increase during periods when the cameras are less frequently maintained, such as in August, when dome contamination tends to be higher. The cloud height retrieval limitations for specific cloud types (e.g., low-level clouds, thick or overcast layers where cloud top detection is challenging) are quantified in Table 4 and Figure 14, which displays the height difference boxplots as a function of altitude. Calibration related limitations are minimized by imposing the $\pm 1$ pixel uncertainty below 1 km. Constraints related to the camera distances and how this affects to the capability to see higher or lower clouds are addressed in Section 4.1, where it is mentioned that in the perimeter areas, the network lacks sensitivity for lower clouds, since the high density of cameras is in the center of the network, and it is needed short camera distances for resolving low altitude clouds. This is again addressed in section 4.2 when comparing with a ceilometer further away from the center of the network. In this section is quantified how for longer camera distances is only possible to retrieve higher clouds. In addition, Figures 8 and 9 provide further information to quantify the limitations with camera distances and camera availability in different parts of the domain. Figure 8 shows, for each grid bin, the number of contributing cameras (Panel A) and cloud points (Panel B), while Figure 9 (Panel C) shows the standard deviation of the retrieved heights within each bin. Other limitations, such as those arising from the geometric calibration, camera coordinates determination, and correlation errors, have been addressed in the Methodology (Section 3). As mentioned before, the $\pm 1$ km uncertainty, derived from the conservative $\pm 1$ pixel shift (equivalent to $\pm 10$ pixels in the original image resolution, see our response above), is designed to account for these combined sources of error while still maintaining a compromise with computational efficiency.

## Minor comments:

**RC:** The manuscript is generally well written, but some sentences are too long and could be simplified.

**AC**: We have tried to improve the readiness of the manuscript in its new version.

**RC:** Please correct minor grammatical persintent issues (e.g., change “All this cameras” by “All these cameras”).

**AC**: This has been corrected in the new version of the manuscript.

# Response to the Anonymous Referee #2 comments for the manuscript

*"Cloud height mapping using multi simultaneous sky images from an all-sky camera network"*

By Celia Herrero del Barrio et al. in AMT

First, we are grateful for the effort of referee #1 and his/her review in detail. Reviewer comments are in black font (RC), and author comments (AC) in red font.

## Author's answer to Anonymous Referee #2

**General comments:**

**RC:** In this study, Herrero del Barrio et al. present an improved methodology that uses a network of all-sky cameras to derive cloud base height (CBH) and cloud top height (CTH) maps at high spatial resolution (50 m) in near real time, every 5 min during daytime and every 2 min during nighttime. A thorough comparison with Sentinel-2 products and ceilometer observations highlights the strengths and limitations of the method. The objectives of the study are clearly defined and are addressed through a thorough analysis, and the proposed methodology for deriving cloud masks and cloud height maps is of significance for the remote sensing of clouds. I consider the topic and results of this manuscript to fit the scope of AMT. However, I have some general comments that should be addressed prior to publication.

- **RC:** I suggest stating at the very beginning of the Methodology section what is currently mentioned only in the Conclusions (lines 770–772): "It is important to remark that by using various camera pairs with different distances, the system captures parts of the vertical cloud structure, not only the cloud bases." This could also be reflected in Fig. 4. It would then be immediately clear to the reader why, for each grid bin, the minimum and maximum retrieved cloud heights are assigned to the CBH and CTH, respectively. In the same context, I would also suggest bringing forward what currently appears only in the Results section (lines 533–535): "cameras observe clouds from the surface, seeing mostly the cloud base and its sides; therefore, the CTH calculation is expected to be less accurate and precise, as the top part of the clouds is often obstructed by the cloud itself and only seen by a few viewing angles."

  **AC**: Following the sugestion we have reformulated the beginning of the methodology as follows, removing the redundant parts afterwards:

  Pg 9, ln 226: *This section describes the proposed methodology for computing the cloud heights maps (specifically CBH and CTH) from the simultaneous all-sky images, from the mentioned PRESENTE network, captured at a given time. The procedure for image processing, cloud height computation, and cloud height map generation is explained. This algorithm can be implemented continuously to process the pictures taken by all the all-sky cameras of the network. It is important to remark that by using various camera pairs with different baselines, the system captures parts of the vertical cloud structure, enabling the retrieval of both CBH and CTH. For each grid bin, the minimum and maximum retrieved cloud heights are assigned to CBH and CTH, respectively. It should be noted, however, that cameras observe clouds from the surface; consequently, CTH calculation is expected to be less accurate and precise than CBH, as the top part of the clouds is often obstructed by the cloud layer itself and is only visible from a few viewing angles.*

- **RC:** I find the evaluation of the camera-derived CTH (which is expected to be underestimated, given the methodological limitations discussed above) against ceilometer observations problematic, since the ceilometer also underestimates the CTH. I would suggest presenting first, in the main text, the cases currently shown in the supplement (Fig. S5), in which the ceilometer detects an additional cloud layer above the current one, as these provide a more reliable ground truth for the CTH and would strengthen the evaluation of the presented methodology.

  **AC**: Following the recommendation, we have moved the results from Figure S5 in the supplement to the main discussion in Section 4.2, creating a new Figure 13 that presents both the unfiltered CTH and the filtered CTH. We prefer to keep the unfiltered CTH in the discussion because applying the cloud penetration depth filter can be overly strict, limiting the dataset to multilayer cases and removing specifically most of the high-level clouds. Furthermore, by showing both datasets, we highlight that there are numerous cases where the ceilometer signal saturates when attempting to retrieve the CTH, meaning that comparisons and validations against this ground-based instrument must be interpreted with caution. This section has been rewritten for the analysis of these new figures.

- **RC:** I believe the study would benefit from a comparison of the results with those of previous studies mentioned in the introduction (where applicable), and from highlighting the novelty of the present study in the last paragraph of the introduction.

  **AC**: We thank the reviewer for pointing this out, and we agree that a comparison with other methods would be very valuable. However, in the case of cloud height retrieval from a single camera pair, the underlying geometrical principles are essentially the same in all works, so the comparison is less significant at that level. We believe such a comparison would be more meaningful against other networks composed of more than two cameras, since this is one of the main novelty of the present study lies: not in the geometry of individual camera pairs, but in how the methodology has been designed to exploit an entire network of 20 cameras. While previous studies have generally focused on individual camera pairs, our approach aggregates spatial information from multiple simultaneous viewing perspectives of the same cloud, allowing many more points to be correlated across camera pairs and yielding a substantially more precise and complete three-dimensional reconstruction. This is achieved while still ensuring robust filtering of the resulting cloud height estimates, and while keeping the whole methodology operational and computationally efficient, as demonstrated through its implementation in near real time over a larger spatial domain. However, to the best of our knowledge, no other network of comparable scale and density is currently available for such a comparison. We believe that one of the most interesting future developments will be to evaluate the scalability of the methodology by applying it directly to other camera networks, provided that the camera calibration parameters are available and that the filtering criteria are adapted to the number of available camera pairs. Extending the approach to additional locations and networks would provide valuable insight into its transferability and operational potential, and we would be very interested in pursuing this in future work. We have highlighted the novelty of this study in the conclusions.

  Pg 35, ln 759: *This work proposes a robust methodology for the identification and three dimensional localization of clouds using stereoscopic calculations, implemented through an unprecedented 20 all-sky camera dense network deployed around the city of Valladolid (Spain). It presents a novel aggregation strategy in which each camera is paired with every overlapping camera, and the resulting heights are combined onto a common spatial grid, together with a fully automated, near real time operational implementation, generating openly available cloud height products every 5 minutes during daytime and, as a further novelty, every 2 minutes at nighttime. The workflow also brings together for the first time a combination* *of different recommendations from previous studies, such as downsizing images, using rectified row-oriented pairs of simultaneous images to improve finding pixel correlations, using blocks of neighboring pixels in the correlation to avoid mismatches, using different baseline distances and applying smoothing filters that improve the segmentation of clouds. In addition, this implementation is also relevant due to the simultaneous application of all these novelties and aggregation strategies, while significantly optimizing the processes to achieve a high computational efficiency. For the case study shown in Section 3 (28 December 2025 at 11:20 UTC), the total processing time to retrieve cloud heights for the 16 available camera pairings was 102 s using our server infrastructure. However, this processing time is significantly optimized in the operational version through parallelization. By executing 40 processes simultaneously, a total of 120 cases (captured between 07:30 and 17:25 UTC) were processed in just 6 min 44 s. This results in an effective average processing time of 3.37 s per case. This allows for the generation of cloud position maps, including cloud base and top height (CBH and CTH), in near-real time, representing a substantial improvement over previous implementations.*

## Specific comments:

**RC:** Line 32: It is not clear which short-term forecasting applications the authors are referring to here. Lines 32–34 should be refined. I suggest the authors focus on the fact that the detailed information on the cloud field provided by all-sky cameras (as their method provides) is important for capturing the spatio-temporal variability of cloudiness (supported by relevant publications). In general, more references, and more recent ones, should be cited in this paragraph.

**AC**: It has been specified that we refer to cloudiness short term forecasting and added the apropiated references:

Pg 2, ln 28: *Cloud height mapping is interesting for all the aforementioned reasons, but it also has other applications. Knowing CBH with spatial resolution can also be very useful for cloud monitoring, as it helps to understand formation and dissipation processes, as well as the evolution of clouds under different regimes (Seiz and Davies, 2006). This is particularly important for climate and weather prediction models. Another application is solar nowcasting, which requires capturing the high spatio-temporal variability of cloudiness to predict the solar energy budget in near-real time (Cirés et al., 2019). Accurate short-term cloudiness forecasting combined with detailed information on the 3D cloud field have been proven essential for the integration of renewable energy into power grids, as it allows for the prediction of irradiance fluctuations and ramp-rate events, thereby facilitating more efficient plant management and grid stability (Gregor et al., 2023; Papachristopoulou et al., 2024; Papadimitriou et al., 2026). These estimations are also crucial for other issues related to visibility, particularly in aviation but also for monitoring local atmospheric conditions that affect road and maritime transport, as well as for civil protection in adverse weather situations (Isaac et al., 2014).*

**RC:** Line 51: This statement needs explanation. Please either remove it or elaborate further.

**AC**: We have elaborated on the statement by specifying that the cost-effectiveness refers to the more accessible technology and low maintenance requirements compared to active remote sensing instruments while providing valuable scientific data:

Pg 2, ln 51: *In this framework, the use of all-sky cameras has increased as these instruments are relatively simple and accessible technology, have low maintenance requirements, and ease of deployment compared to active remote sensing instruments such as lidars and cloud radars, while still providing valuable scientific information for monitoring cloud properties such as cloud cover, cloud type or cloud modification factor among others (Calbó and Sabburg, 2008; Li et al., 2022; González- Fernández et al., 2024). In addition, all-sky cameras have a wide field of view, providing sky images with information of the spatial cloud field with high temporal resolution.*

**RC:** Fig. 8 & 9: The white dots are difficult to see. Please consider using a different color or marker shape or increasing the figure size.

**AC**: In order to improve the figure, the dots have been changed to black color and their size has been increased.

**RC:** Line 482: Should this read "becomes more constrained" or "less constrained"? This part may need further elaboration.

**AC**: We have clarified in the manuscript that the stereoscopic configuration is more geometrically constrained toward the perimeter. This is because at the edges grid bins lack observation from opposing cameras, then is not possible to do the triangulation require, particularly for low clouds.

Pg 22, ln 500: *The field of view of the network can also be observed for the present cloud heights, as clouds above 41.9°N latitude are not visible. In the central part of the network, the higher camera density results in shorter camera distances, which ensures strong overlap and reliable detection of low to mid level clouds. Moreover, these central cameras can be combined with more distant peripheral stations, providing larger camera distances that improve the triangulation geometry for higher cloud layers within the same region of the map. In contrast, grid bins located near the outer boundary of the network do not benefit from this geometric complementarity. While clouds in these bins may be observed by nearby stations within the central part of the network, they lack additional outer stations that could provide overlapping views from the opposite side. Consequently, the stereoscopic configuration becomes geometrically constrained toward the perimeter, particularly affecting the reconstruction of lower cloud layers.*

**RC:** Lines 545-546: "the grid bins within a 150 m radius of the ceilometer location (31 bins in total) have been considered to be a more representative area." — Is there any justification for this choice? Physical reasoning, previous studies, or a sensitivity analysis?

**AC**: The 150 m radius was empirically selected to ensure robust validation after already showing the point-based comparison in Table 2 and Figure 11. This aggregated value acts as a collocation window that compensates for the temporal mismatch between the ceilometer's 15 s integration period and the cameras' instantaneous capture, accounting for cloud movement. Furthermore, it ensures a sufficient sample size (31 grid bins) to minimize the impact of outliers.

**RC:** Lines 555-559: Is text missing here, or should this part be included in the discussion of Table 2, since it refers to the cases where the cameras indicate cloud-free conditions while the ceilometer detects clouds?

**AC**: The percentage of clouds detected by the cameras and not by the ceilometer was the same for Table 2 and the "All SZA" conditions in Table 3, which was confusing. We have moved the suggested lines to the discussion of Table 2 to avoid confusion and improve readiness.

Pg 26, ln 579: *These results, presented in the first row of Table 3, show an overall agreement of 66% predicting clouds, which represents an improvement over the point to point comparison. The case when cameras detected clouds not identified by the ceilometer remains the same at only 3%.*

**RC:** Lines 565-566: "This is likely due to the strong presence of the sun in these images, which produces significant saturation if not obstructed by clouds." — Perhaps refer here to Fig. 7 (Cam025)? Is that the case?

**AC**: Thank you for the good appreciation. It is not entirely clear whether this issue occurs in the case mentioned, as in Figure 7 (Cam C025), clouds can be distinguished around the sun in the unmasked image. However, is possible to see it in Figure S1 in Cam C017, where some pixels around the sun appear really white in the image and are misclassified as clouds. This has been commented in the new version of the manuscript

Pg 26, ln 587: *This is likely due to the strong presence of the sun in these images, which* *produces significant saturation if not obstructed by clouds (one example of this can be seen in the supplementary material Fig S1 Cam C017).*

**RC:** Line 568: Do these correlations correspond to cases where dirt is present on one camera only?

**AC**: Not in general. As can be seen in Figure S4, particularly for cameras C014, C015, and C016, several white spots are visible on the camera domes that do not correspond to clouds. These artifacts are either misclassified as clouds by the segmentation model or generate false correlations between images, appearing as one of the sources of outliers mentioned in the manuscript. Moreover, cameras tend to become dirty simultaneously, since events such as rainfall or desert dust episodes typically affect the entire network rather than individual cameras.

## Technical corrections:

**RC:** Figure S2: White dots . . . instead of black. However, see me 3rd specific comment.

**AC**: Changed to black color in the new version of the manuscript.

**RC:** Figure 11. 1st line base instead of "bsae"

**AC**: Corrected in the new version of the manuscript.

**RC:** Figure S5: Section 4.2 instead of Section ??

**AC**: Corrected in the new version of the manuscript.

**RC:** Figure 14: two instead of three different days

**AC**: Corrected in the new version of the manuscript.