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.