Author Responses

We thank the reviewers for their valuable comments and helpful suggestions, which we think helped us to significantly improve the quality and clarity of the manuscript. Our responses and revisions are enumerated below. Beyond changes in response to reviewer comments, we made minor copy edits.

Reviewer #1:

This paper introduces an approach to performing cloud cover analysis by means of image segmentation, designed specifically to work with a new type of all-sky camera operated by the Atmospheric Radiation Measurement (ARM) program, which replaces an older camera model used for multiple decades. The objective of the algorithm is to provide continuity to cloud cover estimates from all-sky camera images by the ARM program, and perform well at multiple sites of deployment.

The paper is well-written, and the visualizations are helpful (but image resolution needs to be improved for some of the figures, see line-by-line comments). The algorithm's performance is assessed both quantitatively (using a test dataset) and more qualitatively (by comparison with other collocated measurement data), and this performance seems fit for purpose.

We thank the reviewer for their feedback. We note that the apparent image resolution is the result of the PDF rendering and not the actual resolution of the images, which are at 300 dpi or higher without compression (higher than EGU's minimum requirements).

My main critique of the current paper is that it reads more as an algorithm description document than a scientific paper. The scientific value of the algorithm and its output is clear to me, but I think the paper could improve in placing the algorithm in context of other approaches discussed in the literature and how the performance compares to those. Specifically, the method to classify cloudiness operates on a per-pixel basis, which is likely inferior to segmentation algorithms that consider the entire image at once (e.g. a convolutional neural network). Although it is mentioned that computational efficiency is required for generating the algorithm's output in near real-time, it is not exactly clear what the limiting factors are and whether the simplified approach is justified: especially when little to no successfully implemented alternative approaches are mentioned. It is thus also difficult for the reader to assess the performance of the algorithm. As an example, the paper states that "Clear sky pixels are properly classified at an impressive rate of 91%", but how can we possibly say that this is impressive if no reference performance is given?

Thank you for raising these concerns. Indeed, this manuscript serves more as an algorithm description with evaluation and demonstration, which we think aligns with AMT's scope:

"…The main subject areas comprise … techniques of data processing and information retrieval for gases, aerosols, and clouds. Papers submitted to AMT must contain atmospheric measurements, laboratory measurements relevant for atmospheric science, and/or theoretical calculations of measurements simulations with detailed error analysis including instrument simulations."

We agree that some context was missing as to why we used pixel segmentation and not more sophisticated algorithms such as CNN or transformer networks, for example. We now discuss that in the text at the beginning of Section 2.2:

"An operational product such as ASISKYCOVER requires a computationally-efficient engine that will render the data product suitable for near real-time production. The estimation of cloud cover in given all-sky images can be performed using a per-pixel or an all-at-once segmentation applied to image sectors or the full image. Examples of the all-at-once approach include the application of advanced ML algorithms, such as convolutional neural networks (CNNs) or transformer networks, to all-sky images (e.g., Dematties et al., 2023; Xie et al., 2020). To reconcile the geometric inconsistency between the rectangular filter shape on which these methods operate and the effective circular shape of all-sky images, these methods are commonly applied to full images. This means that:

1. *Changes in the relative position of the effective image (e.g., different housing positioning in new deployments) might require training of site-specific models.*
2. *More critically, such all-at-once approaches present a task with significantly greater degrees of freedom relative to the single pixel level segmentation, which necessitates extensive training datasets (e.g., Dematties et al., 2023). This requirement is especially crucial if one is interested at mitigating the influence of naturally-occurring artifacts (insects or dirt in FOV, rain events, etc.), and the production of a physically-consistent product (e.g., with minimal cloud cover time series "jumps" artifacts stemming from the highly non-linear decision weighting).*

Moreover, the training and, from an operational perspective, the application of these methods are generally much more computationally demanding, both in terms of memory and processor requirements, than the application of per-pixel segmentation to full images. For these reasons, pixel segmentation is the most appropriate approach for this product.

After testing the output of multiple classification packages for the pixel segmentation task, …"

Beyond that, we now note in the final paragraph of Section 2.2 that our per-pixel segmentation is equivalent or superior to other CNN and transformer networks from the literature:

*"When assessing the testing dataset using such an approach, the algorithm's robustness is further demonstrated (**Error! Reference source not found.**b)… When we exclude from the evaluated pixels of the testing dataset scenes with SZA > 80°, which are challenging to all algorithms we are aware of (for reference, the TSI product does not provide output beyond an SZA of 87°), we see even more robust results with merely 2.7% of the clear pixels and 3.6% of the cloudy pixels being classified incorrectly (**Error! Reference source not found.**c). These results are equivalent or superior to other cloud cover segmentation studies implementing ML algorithms such as CNN and transformers (cf. Dematties et al., 2023; Fa et al., 2019; Xie et al., 2020), especially when considering the evaluation of an extensive and fully independent testing data that includes sites not used in the training process. "*

We also removed the *"impressive"* adjective from the sentence mentioned by the reviewer to further address the point concerning lack of references.

Some other broader comments/questions I have:

The discussion of the RSS approach for the uncertainty quantification needs more detail, or at the very least a reference to a resource that further explains it.

We added a reference to one of R.A. Fisher's work on a similar topic and added some text to emphasize that we are not making typical assumptions in RSS-related analyses but provide purely empirical justification. (Note that in all correlation coefficient statistical significance calculations, or RMS errors for that matter, in which the RSS part is the denominator, we implicitly assume fully independent normally distributed samples):

"The RSS approach for the cloudy pixels is used to balance the inherent tendency of the trained model to have the sum of cloudy class probabilities much greater than the probability of the clear class, yet to account for their unknown correlation patterns. Note that, unlike the traditional RSS approach for uncertainty estimation or error analysis (e.g., Fisher, 1915), the cloudy classes here are not assumed to be normally distributed nor are they treated as fully independent because autocorrelations necessarily exist during model operation."

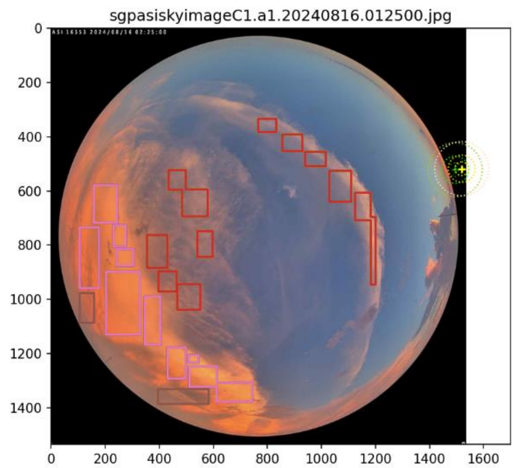
The calibration of the uncertainty estimates for the cloud coverage is not assessed. Firstly, when the algorithm says that the cloud cover is 50% +/- 10%, what does this mean? If it means something like "there's a X% chance that the actual cloud cover is between 40% and 60%", then the correctness of this uncertainty band could be assessed quantitatively.

We understand that the reviewer is interested in a more precise definition of the cloud cover uncertainty. We think of the uncertainty band as the range at which we are comfortable with the algorithm's output based on our extensive experience working with this instrument. In some of the tens of tests/tuning exercises we performed trying different algorithms (e.g., RF and boosting packages, etc.) and/or thresholding, workflows, and hyperparameter, or their combinations thereof, the uncertainty band was much smaller or larger than it currently is typically, and the product behavior was somewhat different. We acknowledge that this is likely not the definition the reviewer was looking for. However, we feel that stating that "the uncertainty represents an X% chance …" would be misleading because this presumably accurate definition heavily relies on the evaluation dataset, which necessarily does not account for the infinite scene possibilities (i.e., how do we define "representativeness"?). Related to the previous comment, our definition of an uncertain pixel was based on an arbitrary threshold, which was evaluated using expert judgment against a small library of ~60 images with different, both simple and challenging, segmentation scenes. We can treat that part of our development as the ASISKYCOVER calibration, which followed by the validation against other products and instruments presented in Section 3.1. To our knowledge, this ostensibly subjective tuning workflow is no different from the development workflow of similar or vastly different cloud cover assessment algorithms, which all (hopefully) have gone through. Note that this subjectivity does not contradict our statement of the all-sky imagers providing an "objective assessment of cloud cover", since that only refers to algorithms vs. the historic observers' okta approach. We added some text to the sentence describing the uncertain pixel definition to reflect its arbitrary nature:

"For a given cloudy class pixel (one of thin, intermediate, or thick cloud), we set the pixel as "uncertain" if the RSS of all cloudy class decision probabilities is below an arbitrarily selected threshold of 0.5, which we evaluated using expert judgment of a set of ~60 test cases (not shown). Similarly, for a given clear class pixel, we set the pixel as "uncertain" if the clear class decision probability is below the same threshold of 0.5. "

Given that bounding boxes are used for labeling, I imagine that very few areas near cloud boundaries (which are typically not rectangular) have been incorporated into the labeling process. How does this affect the algorithm's performance? Shouldn't you create some pixel-by-pixel annotations to really understand the performance of the algorithm?

We think that this is an excellent point. Indeed, the rectangular shape is somewhat limiting, which necessitated annotating a large number of rectangles in certain scenes such as in the example below (included in the training dataset), particularly because we were inclined to incorporate regions around cloud boundaries, as the reviewer has suggested. We agree that pixel-by-pixel annotation is possible; however, we suspect this might have become a more tedious annotation task. We also note that because our objective was to represent cloud edges and the special geometry created by the fisheye configuration of all-sky imagers, both in training and testing, we initially had a pixel class representing "cloud (thick) side", which represented the (side) edge of thick clouds (apparent purple rectangles in the example below), most commonly visible at higher pixel zenith angles. However, we eventually merged that class into the thick cloud class because this merging resulted in better algorithm performance in both cloud detection and classification.



One of the objectives of the algorithm is to perform well at different deployment sites. But this is not explicitly assessed in the discussion of the pixel segmentation algorithm, correct?

Indeed, we did not explicitly separate the data from the site not included in the training dataset (CoURAGE) from independent samples from sites from which some data were used for training. We now explicitly inform the reader about the percentage of samples of the testing dataset that were taken from that fully independent site:

"To test the output of the segmentation algorithm, we also constructed a testing dataset fully independent from the training dataset … The testing samples were gathered from … the ARM Coast-Urban-Rural Atmospheric Gradient Experiment in Maryland (CuORAGE; Davis et al., 2024) (3 days); BNF (2 days); and CAPE-K (1 day). Note that none of the CoURAGE ASI data (~74% of the test samples) were included in the training dataset, which increases the representativeness of the testing dataset, given that observations collected by that specific ASI instrument were not included in the training dataset."

To further address the reviewer's comment, we generated the confusion matrix using only the CoURAGE samples. As shown below and indicated by the precision, recall, and F1-score, the ASISKYCOVER algorithm performance is actually better than suggested from the entire testing dataset. Note that the number of testing samples for the mask pixels is quite limited (hundreds relative to millions of clear and cloudy samples) and are all associated with periods when SZA > 90° (Solar disk below the horizon).

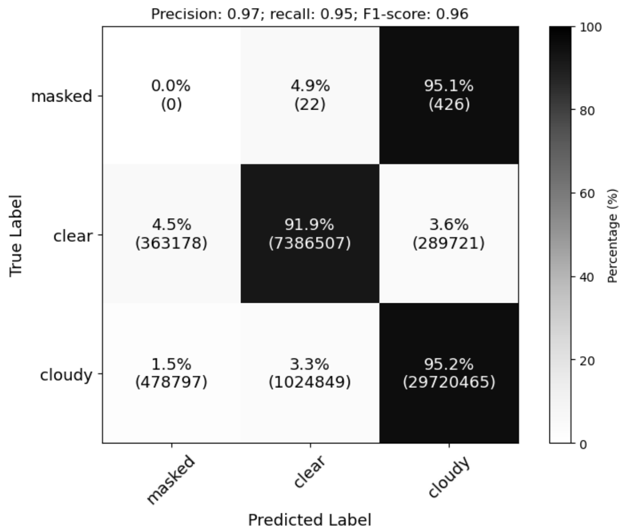


Figure 1: Missing horizontal axis label for histogram. Different colors for the class bounding boxes are occasionally hard to discern from each other. Perhaps add text labels to them within the images? Generally, the figure resolution is a bit too low. It would be good to add a title to each image that indicates clearly what deployment it is from and what the date + (local) time is. The information is there already in the filename, but it's not very clear.

We think those are excellent suggestions. We added the horizontal axis label for the histogram, a legend for the bounding boxes, site name at the bottom right each panel, and a few text labels within a selected set of these boxes. Regarding the figure resolution, as noted in our first response, this is a PDF rendering issue and therefore is irrelevant for publication. With all our enhancements to this figure, its caption now reads (note that with the addition of Figure 1, this is now Figure 2):

"Figure 1: (a-i) Examples of images from different ARM deployments with manually labeled sectors of clear (orange), thin clouds (red), intermediate-thick clouds (purple), thick clouds (brown), and artifacts (masked; yellow), which are included in the training dataset (see legend and a selected set of in-box labels). In the depicted images, the x- and y-axis values represent pixel indices. The compass rose at the top right of each panel denotes the location of the north, south, east, and west directions, with compass tilting reflecting the site-specific solar misalignment. (Note that the ASI images appear mirrored in the east-west direction; see Section 2.1 for details). The concentric green and yellow circles designate the analytically calculated and misalignment-corrected solar disk positions, respectively (outer circle radius of 100 pixels). Solar zenith and azimuth angles are shown at the bottom left corner of each image. The corresponding ARM deployments are shown at the bottom-right of each panel (ENA — Eastern North Atlantic; BNF — Bankhead National Forest; SGP — Southern Great Plains; KCG — Cloud And Precipitation Experiment at Kennaook; NSA — North Slope of Alaska). (j) Distribution of training (black) and testing (red) pixel samples as a function of solar zenith angle (SZA; bin size of 2°)."

Line 102: should "exists" not be "exist" ?

Fixed. Thank you.

Figure 2: Would also benefit from a higher resolution. Some of the uncertainty estimates would lead to physically infeasible cloud cover values (e.g. fig b has a lower uncertainty bound below 0%). It may be worth commenting on this, at the very least.

Now noted in the text (Sect. 2.2):

"(Note that cloud cover is not suggested to exceed 100% or fall below 0% when considering the resolved value and uncertainty)."

Line 162: "bogus cloud cover" is more exact terminology available to describe this?

Changed to *"spurious cloud cover"*.

Line 178: "mis-alignment" spelled as "misalignment" elsewhere in the paper

Now using misalignment everywhere.

Equation 2: Might be helpful to state the units of the angles here?

Radians. Now stated.

Line 244: one symbol is italic, the other isn't. And is "solar zenith angles" here correct?

Good catch. It is the pixel zenith angle. Fixed. Thank you!

Line 260: Shouldn't all the numbers here be written as 75th, 95th etc?

Right. Fixed.

Figure 5: Quite hard to distinguish between the KAZR and CEIL curves.

We changed the CEIL curve color to be more distinct from the KAZR.

<u>Reviewer #2:</u>

This manuscript presents a significantly updated and improved image segmentation approach developed for the All Sky Imager (ASI) deployed at Atmospheric Research Measurement (ARM) user facilities after the retirement of the long-running TSI instrument. The paper is written with clear prose and thorough analysis, and promises to be a highly cited paper, as every future investigation that uses the sky cover or cloud mask output will cite this paper. So it is a timely, relevant, and highly useful publication for the community. I provide ample comments below in my efforts to improve the clarity and relevance of the message in this paper, even though the paper is already of a very high quality and relevance. The comment platform restarts all numbering - see the attached PDF for sequential numbering.

We appreciate the comprehensive and thoughtful recommendations as well as kind words.

Specific Comments:

I would like the paper to explicitly provide guidance on how future users of this dataset should use the "uncertainty" information. This comment is motivated by my observation that researchers may tend to ignore or incorrectly use uncertainty metrics. For example, in the discussion of Figure 3c (page 10), it was indicated that the 23.7% confusion of "masked" pixels being classified as "cloudy" pixels was attributed to the "halo" found around artifacts such as the bird, and these pixels are treated as "uncertain." So … is there a way to exclude "uncertain" pixels from the final product (sky cover and cloud mask images)? What would be the benefits and drawbacks from simply excluding all “uncertain” pixels from the final products? (For example, in Figure 2, the cloud masks would convert all of the hatched areas into masked areas.)

We think that it is an excellent question. We often observe similar behavior in that uncertainty is overlooked by users. As we also responded to the first reviewer on a similar question, we view the uncertainty band as the range within which, based on our extensive experience working with this instrument, we are generally comfortable with the algorithm’s output (i.e., a measure of confidence in the algorithm output). Our definition of an uncertain pixel was based on an arbitrary threshold, which was evaluated using expert judgment against a small library of ~60 images with different, both simple and challenging, segmentation scenes. We can treat that part of our development as the ASISKYCOVER calibration, which is followed by the validation against other products and instruments presented in Sect. 3.1. To our knowledge, this ostensibly subjective tuning workflow is no different from the development workflow of similar or vastly different cloud cover assessment algorithms, which all (hopefully) have gone through. We added text to the sentence describing the uncertain pixel definition to reflect its arbitrary nature, and also added a new paragraph to discuss the uncertainty per the reviewer’s recommendation. This paragraph is now also referred to when discussing the bird scene classification (Sect. 2.2):

“For a given cloudy class pixel (one of thin, intermediate, or thick cloud), we set the pixel as “uncertain” if the RSS of all cloudy class decision probabilities is below an arbitrarily selected threshold of 0.5, which we evaluated using expert judgment of a set of ~60 test cases (not shown). Similarly, for a given clear class pixel, we set the pixel as “uncertain” if the clear class decision probability is below the same threshold of 0.5. “Mask” class pixels are uncertain by definition for the cloud fraction calculation as discussed below and in the following sub-section, and are therefore excluded from these calculations… We recommend that the total uncertainty (the union of “mask” and “uncertain” pixels) associated with the resolved cloud cover quantities for given sky images be integrated and considered in analyses involving the ASISKYCOVER product. From a qualitative standpoint, we think that the total uncertainty can be used as a general measure of confidence in the segmentation of a given scene (e.g., low uncertainty — higher confidence; high uncertainty — lower confidence, with a higher likelihood of a challenging or artifact-contaminated scene). From a quantitative standpoint, the uncertain pixels are treated no differently than masked pixels, meaning that masked and uncertain pixels effectively represent image regions where we have no reliable information, either due to artifacts or segmentation model limitations. Therefore, the resolved cloud cover, serving as our best estimate, should be considered together with the total uncertainty, as a measure of the lower and upper bound of our estimate. (Note that cloud cover is not suggested to exceed 100% or fall below 0% when considering the resolved value and uncertainty). We think that incorporation of the uncertainty in analyses is especially critical in cases where it is substantial (e.g., exceeding arbitrary values such as 30%, 40%, or 50%), in which case the resolved cloud cover should be taken with a grain of salt, as the convolution of the scene’s cloud configuration with the spatial configuration of the masked sectors has many more degrees of freedom…. In ***Error! Reference source not found.****b, we see that the bird “artifact” is effectively fully masked, given that in this case, the cloudy classes’ “halo” surrounding it are denoted as “uncertain” (hatched areas in the plot). However, as noted above, the results for such a highly uncertain scene should be taken with a grain of salt.”*

A sentence or two about the approach to the ASI near-zenith product would be helpful. I was surprised to learn in the caption to Figure 8 that it requires 99% or more of (~5400) pixels within the 5 degrees off-zenith. This didn’t make sense to me. How many pixels fall in this 5-deg circle around zenith? Is the ~5400 pixels the product of the pixels in this circle times the number of images? I would assume that a 5-minute running mean was also taken, consistent with the processing for KAZR and CEIL?

Good catch. We apparently kept our notes justifying that 99% threshold (instead of a typical baseline threshold of 50%) out of the manuscript. This is now clarified in the text (Sect 3.3):

"Figure 4 shows the 2024 SGP central site ASISKYCOVER-determined cloud cover occurrence histogram as a function of SZA normalized for each SZA bin, together with … its near-zenith (NZ) product, the TSI, and the KAZR-CEIL combination (see legend). Here, the NZ product is processed using a quasi-binary approach equivalent to the CEIL and KAZR point (i.e., non-spatial) measurement nature. To determine NZ cloud occurrence on a given scene (i.e., a value of 1), we require at least 99% of the ~5400 pixels within 5° off-zenith to be classified as one of the cloud classes (value set to 0 otherwise). The resultant binary time series is then used to calculate the 5-min running means. The 99% threshold mitigates the low SZA solar flaring effects discussed above, at the limited expense of some under-estimation in scenes with partial cloud cover within the small 5° off-zenith region."

Note that we removed the now-redundant explanation from the caption of Figure 9, which now reads:

"… The purple solid and dotted curves represent the 5-minute running average mean and mean ± sample uncertainty cloud cover as a function of SZA determined using the ASISKYCOVER near-zenith product (NZ; see text), rendering a more equivalent comparison to the KAZR-CEIL data…"

While working on the response to this comment, we also noticed that on Figure 9, we didn't clip the mean ± uncertainty values to the range of [0, 100] prior to the averaging, which resulted in slightly overestimated uncertainty range for ASI (now corrected). Regarding the number of pixels within the near-zenith region, that is indeed the number of pixels within that area per image. Consider that 180° are spread over an effective image diameter of ~1500 pixel lengths. This translates to a 5° radius of ~41 pixel lengths → ~5400 pixels within the near-zenith region.

Since this will probably be the canonical paper for utilization of the ASI data stream, have you considered including a photograph of the instrument? I noticed that one is not included on the ARM ASI instrument page, and it could be nice to see what it looks like! (just a suggestion – your choice.)

Great idea. A photograph of the ASI is now given in the new Figure 1.

I would like to know more about how the manually determined training and testing pixels were created. I see now, after carefully reading the paper, that the labeling process is illustrated in Figure 1, but that was not apparent upon my first reading.

Indeed, and we hope that our additions and edits to the text given above further address this point.

A) This could perhaps be improved by a first sentence in the caption to Figure 1 indicating the main purpose of the figure, before highlighting the distribution of training & testing pixels relative to SZA? Moreover, the histogram could come at the bottom of the figure, because it seems secondary to the labeling boxes in panels b-j. Or the histogram could get moved to a separate figure altogether, perhaps that includes other qualities of the sampled pixels.

We switch the order of the sample images and histogram. Incorporating other suggestions for this figure (see comment below), the caption now includes explicit reference to "manually labeled sectors" and reads:

"Figure 2: (a-i) Examples of images from different ARM deployments with manually labeled sectors of clear (orange), thin clouds (red), intermediate-thick clouds (purple), thick clouds (brown), and artifacts (masked; yellow), which are included in the training dataset (see legend and a selected set of in-box labels). In the depicted images, the x- and y-axis values represent pixel indices. The compass rose at the top right of each panel denotes the location of the north, south, east, and west directions, with compass tilting reflecting the site-specific solar misalignment. (Note that the ASI images appear mirrored in the east-west direction; see Section 2.1 for details). The concentric green and yellow circles designate the analytically calculated and misalignment-corrected solar disk positions, respectively (outer circle radius of 100 pixels). Solar zenith and azimuth angles are shown at the bottom left corner of each image. The corresponding ARM deployments are shown at the bottom-right of each panel (ENA — Eastern North Atlantic; BNF — Bankhead National Forest; SGP — Southern Great Plains; KCG — Cloud And Precipitation Experiment at Kennaook; NSA — North Slope of Alaska). (j) Distribution of training (black) and testing (red) pixel samples as a function of solar zenith angle (SZA; bin size of 2°)."

B) I also see that the discussion about manual labelling is located in the paragraph around line 175. It seems odd to me that if "… we attempted to sample the various features in a uniform matter …" then only the distribution of SZA is included in Figure 1. I would expect to also see the distribution of the different classes (clear, thin, etc.), and the distribution of pixel location (elevation and azimuth angle).

The distribution of classes would have been critical should we have not used a balanced weighting of classes during training, as already noted in the text (paragraph proceeding that noted by the reviewer):

"The LightGBM segmentation model is trained using weights that consider the number of training samples for each class,…"

When we started the development of ASISKYCOVER, we initially integrated dependence on SZA and pixel location, which holds great promise, and among other benefits, could mitigate the solar flaring mentioned above. We already refer to those first tests in the text (in the paragraph preceding the one noted by the reviewer):

"The SZA impacts image lighting, such as sky brightness and hue… We comprehensively tested and evaluated the option of integrating the SZA, as well as other angles, such as pixel azimuth angle, as features. While this exercise showed promising results in many given scenes, bulk analysis of algorithm output using those additional features indicated the occasional occurrence of spurious cloud cover "step-like" changes between consecutive images, corresponding to decision tree splits of those features. Since integration of such additional features significantly increases the training data parameter space, the training dataset size should be increased in size and carefully curated to mitigate such biases. For now, however, we keep this explicit angle-dependent segmentation for future software updates …"

We agree with the reviewer that once those features will be integrated in ASISKYCOVER as part of future updates, "representative" distributions and cross-distributions of the various angles will be crucial to ensure a robust product with smooth behavior.

C) (curiosity question) Does the output of this segmentation model appear to be sensitive or insensitive to the sampling and labeling of the manual dataset?

Generally not, given the balancing procedure mentioned above.

I appreciated the discussion about the calibration for the solar zenith and azimuth angle, but I feel that the prose should be more clear, and an equation might facilitate communication, particularly the paragraph around line 110. A) That is: The analytical SZA and azimuth is well known from astronomical factors. I understand that you are using a manual labeling approach to identify the center of the sun location on the image. What is a "denoted image"? When you mention "calculated angles," is this the true SZA? Expected image location of the SZA? Manually determined SZA? Or difference between the expected image location and manually determined location? The same four questions hold for the azimuth angle, of course! I recommend clear and consistent language. A simple equation might help to clarify this discussion significantly. B) Furthermore, it is stated that "the positioning biases characterized for different ARM deployments are robust (e.g. see concentric yellow circles in Figure 1)" and indicates that the "denoted versus calculated angles" have a standard deviation on the order of 1 degree. I'm not sure what "robust" means here – small? Consisitent over time for each deployment? – but examination of Figure 1 shows that the yellow and green circles show significant disagreement in 1b, c, i, and j (4 out of 9), and are noticeably different in all other thumbnail images – e.g. the green circle is not overlapping the yellow circle. I would expect a 1-degree difference to be indistinguishable by eye, so that suggests that I don't understand something fundamental about the "denoted versus calculated" angles. Please clarify!

Thanks for this suggestion. We revised the text and added relevant equations to help communicate our processing and evaluation. We think that the added text also helps readers to orient themselves with regard to the ASI images and their non-intuitive geometric configuration:

"the image center coordinates, x_c and y_c, are determined for individual images by finding the x-axis index corresponding to the maximum y-axis image diameter, and vice versa… To determine the SZA and azimuth angle biases, the center of the solar disk is manually identified in a set of images where it is visible, preferably from different times of day… we find that the manual labeling approach is best suited for this task… Following this manual identification of the solar disk center, we use astronomical factors and site coordinates to calculate the azimuth angle (ϕ) and SZA (θ) corresponding to the time of each of the annotated images. Under an equidistant projection assumption, we can analytically convert those angles (in units of radians) to pixel coordinates x and y using:

(1a) $x = x_c + r\,sin(\phi + \pi) = x_c - r\,sin(\phi)$

(1b) $y = y_c + r\,cos(\phi + \pi) = y_c - r\,cos(\phi)$,

where r is the radial distance of the calculated solar position from the effective image center, which is given by:

(2) $r = \frac{r_{eff}}{\pi/2} \cdot \theta$,

where $r_{eff} = d_{eff}/2$ is the effective image radius (d_{eff} being the effective image diameter). Pixel zenith and azimuth angles from the manually annotated solar disk center x and y coordinates can be calculated using the inverse forms of equations 1 and 2. The shift of ϕ by a factor of π in the trigonometric functions in equation 1 stems from the upward-looking ASI geometry, which, unlike maps or nadir-pointing satellite imagery, results in an apparent east-west mirroring and a north-south flipping. This different geometry means that when ASI images are shown with the north direction at the top, the east direction (and hence, sunrise, for example) will be located at their left part, unlike the common intuitive interpretation of such images, where the east direction is to the right of the north. We also note that while an equidistant projection is a specialized form of barrel distortion, we do not apply other barrel distortion corrections, given the ASI lens's relatively minor distortion values, smaller than 0.3%, 0.8%, 1.9%, and 3.8% at pixel elevation angles of 60°, 70°, 80°, and 90°, respectively, based on tables provided by the ASI vendor (cf. Long et al., 2006, their fig. 5). These small distortion values essentially translate to cloud cover errors of a fraction of a percentage point, much smaller than the typical scene uncertainty (see Sections 2.3 and 3.1).

Using the Pixel zenith and azimuth angles, corresponding to the manually annotated solar disk center, and their deviations from the calculated SZA and azimuth angle, the site-specific biases of those angles are determined as the average deviation in the annotated image subset. The standard deviation metric calculated from the annotated image subset serves as a measure of robustness in bias characterization. Based on that metric, the positioning biases observed across different ARM deployments are robust, as indicated by the standard deviation of the differences between the annotated and calculated angles, typically on the order of 1° or less (not shown). For the SGP central site analyzed here, these biases are relatively small at -0.23±1.51° and 2.69±1.34° for the solar azimuth and SZA, respectively (e.g., compare the positions of the yellow and green circles in Figure 2d to Figure 2g)."

Technical Corrections:

Line 51: There is a citation to (ASI-16) that I cannot find in the bibliography. Is there a publication that describes the fundamentals of the ASI deployed at the ARM sites, such as image size (pixel dimension), image capture frequency, dynamic range (presumably 256 bit), and anything else? If not, then this publication could serve this need!

Good point. We are not aware of such a publication other than the vendor's brochures. We added a new table (Table 1), which provides some ASI specs.

Line 85-90. This indicates that the image center is found by the black circle in the ASI images. Yet the black circle must be set by the imaging rig, right? Because the world doesn't become black below the horizon. So what makes us think that the center of this black circle represents true zenith?

We assume but cannot guarantee "true zenith". Our misalignment analysis suggests that errors are small, but the reviewer is correct that in reality, there could be multiple components that are misaligned, resulting in the calculated biases.

Lines 100-105. I found myself wondering about possible barrel distortion here, and you discuss it around lines 135-140. Consider moving the lens distortion discussion to this section. (optional – your choice!)

Done (see our response to the last major comment above).

Figure 1

The axis labels are too small to read, including the legend in 1a and the SZA & AZ annotations in each thumbnail.

The colors of the boxes are not at all clear from the text description in the caption (orange, yellow, purple and brown). Could you please include a legend with the colored lines, because I cannot tell them apart in the images!

See my comments above suggesting that the histogram comes at the bottom of the image, the caption begins with an orientation for the reader, and/or the histogram gets moved to a separate figure that contains histograms for other properties as well.

Even with a legend with line color, it is tremendously hard to see the boxes and colors. Could this image be enlarged to full-page size?

I am curious – where is "North" in these images? I would expect it towards the top of the image, with "East" towards the right, but the azimuth angles don't support that assumption. Could you please add a compass rose or North arrow for clarification?

We appreciate these suggestions. We added to each panel a compass, which also reflects the solar misalignment, as well as a legend to the figure and a few in-box label titles. We also increased the figure size to a full page and moved the sample histogram to the bottom of the figure in accordance with the reviewer's suggestion.

Regarding the confusing directions (where is the north, etc.), here's a simple experiment: tilt your head up such that your forehead points north and chin points south. Where is the east direction?

…

…

…

This apparent flipping of the E/W directions is what is happening in the ASI images. It is now discussed in the text (see discussion about equations 1 and 2 above) and also noted in the figure caption given above.

Figure 2

Again, the font size for the legends in the "cloud masks," the colorbar labels, and the image titles are too small to read!

An image such as this begs for a "truth" column to the right of the segmented image to enable a discussion of the spatial patterns that have been accurately or inaccurately labeled. Which leads me to wonder: what would be the most perfect, "true" segmentation result for these four images? (!)

Font sizes were increased per the reviewer's recommendations. We understand the interest in a "truth" column, but given that the shown scenes are not taken from the training dataset, we have no "truth" segmentation and constructing one would require a significant amount of work. We also understand the interest in the perfect segmentation (e.g., one can imagine more cirrus clouds classified in the peripheral regions of the example third from the top).

Table 1: is the third column needed?

No. We removed it and encapsulated the third column information in the caption:

"Table 1: ASISKYCOVER pixel segmentation algorithm classes. The different cloud classes qualitatively represent optical thickness. The masked pixels are excluded from the cloud cover calculation."

Para near line 175: Perhaps consider adding the geographic context (e.g. Tasmania, Alabama, US) for the field sites that don't include the geographic location, and also for the CuORAGE site around line 199. (suggestion)

Done. Thanks!

A citation or additional information about the root-sum-square analysis would be helpful

We revised the text per the reviewer's request to better reflect our approach and thought process:

"…For a given cloudy class pixel (one of thin, intermediate, or thick cloud), we set the pixel as "uncertain" if the RSS of all cloudy class decision probabilities is below an arbitrarily selected threshold of 0.5, which we evaluated using expert judgment of a set of ~60 test cases (not shown). Similarly, for a given clear class pixel, we set the pixel as "uncertain" if the clear class decision probability is below the same threshold of 0.5. "Mask" class pixels are uncertain by definition for the cloud fraction calculation as discussed below and in the following sub-section, and are therefore excluded from these calculations. The RSS approach for the cloudy pixels is used to balance the inherent tendency of the trained model to have the sum of cloudy class probabilities much greater than the probability of the clear class, yet to account for their unknown correlation patterns. Note that, unlike the traditional RSS approach for uncertainty estimation or error analysis (e.g., Fisher, 1915), the cloudy classes here are not assumed to be normally distributed nor are they treated as fully independent because autocorrelations necessarily exist during model operation."

Figure 3.

Avoid red font on a blue background! I suggest a grayscale background with black text on the non-diagonal cells, and white text on the diagonal cells.

Fix the massive size of (a), (b), and (c).

Done.

Line 244: theta_i1, and theta_i2 should not be the SOLAR zenith angles, but perhaps the pixel zenith angles? Also, the "vertical edges" of a pixel is not clear – do you really mean the y-axis limits (vertical)? Wouldn't a conversion to radial coordinates make more sense? Otherwise with "vertical" pixel zenith angle for a pixel at y ~ 800 would be identical. (using the pixel counts from Figures 1 & 2)

Good catch and observation. Now reads:

"… where θ_{i1} and θ_{i2} are the pixel zenith angles (in radians) at the edges of a given pixel i along the radial direction from the effective image center. "

Line 245. I am curious – if the N_valid does not include "uncertain" pixels, then presumably neither does N_cld. So I wonder then where the "uncertainty" comes from, for example in Figure 8, where it appears to be around 10-20%?

That is correct. The text now reads:

"... where f is the cloud fraction, N_i^{cld} *and* N_i^{valid} *denote the number of valid cloudy and all valid pixels, respectively, and ..."*

The solar disk mask (radius of 100 pixels) typically accounts for 2-4%, depending on its location in the radial direction (greater when closer to the center). The remainder of the uncertainty primarily emerges in cloud transition scenes (e.g., cloud edges as in Fig. 3a) and artifacts (e.g., Fig. 2f, Fig. 3b).

Line 282. Sentence beginning "In addition …" The previous sentence discussed the cloud cover around sunset, but this sentence appears to discuss the morning, Please provide a transition for the reader.

This sentence was reworded. The text now reads:

"... insinuating the possibility that the "intermediate" class fraction could be overestimated in this case to a certain extent, i.e., classified as "intermediate" instead of "thick". Similarly, right after sunrise, it is likely that the thin-dominated period is misclassified, given that the solar disk state is classified as "blocked by clouds", i.e., the low-light conditions render "thick" (or "intermediate") clouds appear as the "thin" class. "

Line 285 and the thumbnail sky image in Figure 5 at 14:00: You cite "the challenge in even manually interpreting …" It is entirely unclear from the print copy of this paper what is going on in the thumbnail at 14:00! It appears to be a clear (blue) sky? Yet the ceilometer reads 100% cloud cover (< 7 km), and the text indicates low clouds. Could the experts please clarify what the "true" sky cover is like in this case, for example?

It is indeed cloudy. We reworded the sentence:

"Examination of the inset ASI snapshots supports this general interpretation of the depicted time series although using other ARM datasets (as shown below) is needed to confirm the occurrence of intermediate and/or thick clouds around sunrise. (Note the challenge in even manually interpreting the 14:00 UTC snapshot taken shortly after sunrise; based on that snapshot, is it clear or cloudy?)"

Figure 5:

Again I find myself wondering about north, since the sun appears to set in the southeast.

The text discusses SZA = 80-deg. Could you indicate when that occurs in the figure? (e.g. civil twilight etc., or time of day?)

It might be nice to include lines that connect each image thumbnail to the time of its capture, since they don't all appear immediately under their image time. I think this would help you tell your story.

Any chance you could enlarge the font size for the axis tick labels and axis labels? (I understand that the legend size is limited!) That said – please reduce the font size for the thumbnail image times!

Consider adding the ASI-determined cloud cover (f) for each thumbnail.

We appreciate these suggestions. We now indicate the 80° crosspoints, use larger font size for axes, and smaller font size for the thumbnails. Regarding the other suggestions, we feel that this figure is already pretty condensed and therefore prefer not to add additional information (cloud cover numbers, compass rose, etc.). In fact, at some point in our internal iterations of the manuscript, we had connecting lines between the thumbnails and their associated times and cloud cover quantities (as the reviewer suggested) but were then recommended to remove them for these overcrowding reasons.

Full year analysis: I am curious about times when the thin clouds (presumably higher altitude) might be occluded by the lower, thicker clouds. I'm not sure if this is relevant or important, though, given the snapshot-nature of this data presentation.

We agree that deconfounding such potential effects on the cloud type partitioning will be an interesting task, but one that we think is beyond the scope of this demonstration as we also note in our response to the next comment.

Full year analysis: a number of cloud climatologies have been done at the SGP site. Is there any way to compare this one-year output to any of those published datasets?

There is a way for a qualitative comparison if we assume interchangeability between cloud thickness in this study and cloud altitude in other studies. We added some text with references to some related SGP studies:

"… ASISKYCOVER suggests a cloud cover of 56% over the course of 2024, in agreement with the climatological results of Wang and Zhao (2017) for the SGP site based on the ARM Active Remote Sensing of Clouds (ARSCL; Clothiaux et al., 2001) dataset. Forty-nine percent of this annual cloud cover is classified as thin clouds. Assuming interchangeability between cloud thickness in this study and cloud altitude in other studies, this result is in qualitative agreement with previous climatologies for the SGP site (cf. Dong et al., 2006, their fig. 1; Mace and Benson, 2008, their fig. 4).

A 31-day running mean cloud cover evolution throughout the year is depicted in ***Error! Reference source not found.****… Assuming interchangeability between cloud thickness and altitude (similar to the examples in Table 1), we observe qualitative agreement with climatological analyses of annual cloudiness and cloud-type variability over the SGP site (cf. Dong et al., 2006; Mace and Benson, 2008). However, additional analysis using other instrument datasets beyond the scope of this work is required to evaluate the influence of only using SGP daytime data (SZA < 90°) and its dependence on cloud type (e.g., Dong et al., 2005; Zhao et al., 2017), as well as to deconfound the influence of occluding cloud layers on the apparent occurrence of different cloud types."*

Figure 6 (all optional suggestions for your consideration)

You might consider a different color choice, so "thin" corresponds to a lighter color, and "thick" to a darker color for easier interpretation.

I would also recommend text labels under each shaded region reading, for example "thick," "thin," and "intermediate" for easier interpretation.

I'm sure you carefully designed the figure, but I would have expected the "thick" category to be on the bottom, partially because it is the primary ASI observation since it tends to occur due to lower clouds. Also, when "thick" is high, I would expect more "thin" clouds to be occluded, indicating that the shaded regions should be dependent upon the lower region.

We liked the idea of reversing the cloud type order such that "thick" is at the bottom and "thin" at the top, which we agree is more intuitive, especially when we discuss interchangeability. We also added labels over the shaded regions. The color scheme was not modified to retain the consistency with Figure 3 and Figure 6.

Figure 7

Again, please reduce the font size for (a), (b) etc.

I find the color contrast in Figure 7b challenging, between the 25 < CC < 50 and the 50 < CC < 75 conditioning. Perhaps a different hue?

In figure 7c, I'm curious about the mostly clear images. Are there any images with CC = 0%? If they have one pixel that is "thin," does it fall in the "thin dominated" category? I wonder if a lower limit on CC is appropriate here, to help account for clear sky images in which no cloud type is dominated?

Consider including in the caption the data analyzed: presumably the SGP images from 2024?

We agree that a lower CC limit makes a lot of sense. We applied a 5% CC threshold for this analysis and the numbers changed a bit, primarily in the case of thick-dominated clouds, which now exhibit better ASI-TSI agreement, as well as smaller deviations between thick and intermediately-thick clouds. Therefore, we removed the following sentence from the text, which we think is no longer justified:

"Those statistical differences between the thick and intermediate-thick cloud scenes tentatively suggest that there is value to the 3-class definition of cloud thickness in the ASISKYCOVER product. "

We note that even if the 3-class definition has less distinct difference from this examined aspect, we still think it is critical.

Regarding the other recommendations, we reduced the font size and increased the contrast on the CC panel. Site and year were added to the caption, which now reads:

"Figure 3: ASISKYCOVER deviations from the TSI cloud cover product kernel density estimations (KDEs) for SGP in 2024 while conditioning on different variables: … (c) dominating cloud type per scene with CC greater than 5% based on ASISKYCOVER, …"

Line 372: "(or clear sky pixels as clear for that matter, see Figure 3)" This is not apparent to me! The clear classification (horizontal row – true label) looks good in image 3a!

Good point. We removed that part of the sentence and retained the reference to Figure 3 (now Figure 4), which does exhibit classification challenge for thin clouds. The sentence now reads:

"Figure 3c provides a clear indication of the challenge in consistently classifying thin cloud pixels (see ***Error! Reference source not found.****)"*

Line 375: I feel that this final sentence of this paragraph is an important key conclusion of your paper! "the dominating cloud type per scene might be the best predictor for cloud detection algorithm consistency." Consider highlighting this in the abstract & conclusion? Unfortunately, the second clause to this sentence is not clear to me ("and emphasize the model performance …")

Sentence was reworded:

*"Those clear KDE patterns when partitioning based on dominating cloud type insinuate that the dominating cloud type per scene might be the best predictor for cloud detection algorithm consistency, and corroborate the class-dependent algorithm performance test results (**Error! Reference source not found.**)."*

We now highlight this conclusion in the abstract:

"A comparison with co-located TSI data tentatively suggests that the dominant cloud type per scene might be the best predictor of inter-consistency between cloud detection algorithms. This ASI-TSI comparison and evaluation against other ARM measurements, such as zenith-pointing radars and lidars underscores the ASISKYCOVER's potential to improve cloud cover analyses and data evaluation efforts, as well as to be integrated into higher-level data products that synergize instrument suites to generate new and insightful information."

We already refer to this cloud type-dependent argument (to some extent) in the Conclusions:

"… While still suffering some artifacts, especially when classifying scenes dominated by thin clouds and/or cases at high SZA, … "

Line 381: regarding ASI solar flaring, perhaps refer to Figure 1, panels c-i?

This sentence primarily refers to the TSI, and hence, we think that a reference to Figure 1 (now Figure 2) will be confusing.

Figure 8:

Does this figure again analyze the 2024 observations from SGP?

In the bar chart, does "Number of samples" correspond to images? Pixels? 5-minute observations?

I also want to make sure that the x-axis label for the bar chart should truly

Line 411 in the caption to Figure 8: close the parentheses after (see Section 3.1).

Thank you for these recommendations. The figure caption now reads:

"Figure 4: (Top) ASISKYCOVER-determined cloud cover occurrence frequency histogram as a function of SZA (bin sizes of 4% and 3°, respectively) normalized for each SZA bin (i.e., each column of histogram bins sums to 1) for SGP in 2024. … The pink and yellow curves designate the … the TSI and the 5-minute running average combination of ceilometer cloud base detections and KAZR moderate sensitivity (MD) mode echoes above 4 km, respectively (see Section 3.1). The purple solid and dotted curves represent the … near-zenith product (NZ; see text), … (Bottom) The total number of samples (images) per SZA bin range."

Line 447: "gamma configurations" is not clear.

Changed to: *"color saturation level and gain configurations"*