A Novel Segmentation Algorithm for the ARM User Facility All-Sky Imagers Using Machine Learning Applications

Israel Silber[1], Donna M. Flynn[1], Jennifer M. Comstock[2], Erol L. Cromwell[2], Brian D. Ermold[2]

[1] Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, Richland, WA, 99352, USA
[2] Advanced Computing, Mathematics, and Data Division, Pacific Northwest National Laboratory, Richland, WA, 99352, USA

Correspondence to: Israel Silber (israel.silber@pnnl.gov)

Abstract. Clouds play a pivotal role in modulating the Earth's energy budget through the reflection of incoming solar radiation and the trapping of outgoing longwave radiation. Ground-based all-sky imagers offer an objective assessment of cloud cover and can be used to estimate solar irradiance, classify cloud types, track cloud movement, and serve as a benchmark for the evaluation of satellite and reanalysis data products. The Atmospheric Radiation Measurement (ARM) user facility has utilized all-sky imagers for more than 25 years to monitor cloud cover and augment its comprehensive suite of atmospheric measurements. Following the retirement of its Total Sky Imager (TSI), ARM recently deployed the TSI's successor, the All-Sky Imager (ASI-16 camera systems). To provide a smooth transition and continuity to the vast amount of knowledge gathered by the TSI over the years, while addressing typical deployment issues, we developed a novel pixel segmentation algorithm, the ASI Sky Cover (ASISKYCOVER). ASISKYCOVER builds on the different strengths and properties of the TSI processing algorithm while integrating machine learning techniques, ensuring data validity and accuracy across diverse atmospheric conditions. It enhances cloud cover characterization with new features such as artifact detection and uncertainty quantification. ASISKYCOVER also includes cloud cover estimates for near-zenith (narrow field-of-view) and reduces susceptibility to false detections. This study introduces ASISKYCOVER, details its algorithm framework, and demonstrates its capabilities using a year-long dataset from the ARM Southern Great Plains site. A Ccomparisons 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, are presented, underscoresing 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.

1 Introduction

Clouds modulate the atmospheric energy budget by reflecting part of the incoming solar shortwave radiation back to space and partially emitting outgoing longwave radiation back towards the surface (greenhouse effect). The exact contribution of clouds to the energy budget depends on their microphysical properties and their coverage, the characterization of which is still deficient (e.g., Grosvenor et al., 2018; Loeb et al., 2018).

Satellite data products provide a comprehensive long-term cloud cover record but suffer from biases in characterizing near-surface and optically thin high-level clouds, and are generally challenged above ice surfaces (e.g., Kato et al., 2012; Maahn et al., 2014; Minnis et al., 2008). Moreover, while geostationary satellites provide samples at a high temporal resolution at small spatial scales, they are confined to low- to mid-latitudes (typically below ~50°), rendering a paucity of high-latitude cloud cover data.

Ground-based measurements complement, and in some cases, replace satellite data for certain purposes, such as multi-scale process understanding. These measurements provide a wealth of information, which, although local, is typically of high accuracy and value, especially for validating satellite products. For example, ground-based all-sky imagers can offer an objective assessment of cloud cover surrounding deployment sites using various algorithms (e.g., Esteves et al., 2021; Li et al., 2022; Sarangi et al., 2025). Imager-based data products can be used to estimate solar irradiance (e.g., Logothetis et al., 2022; Manandhar et al., 2023; Park et al., 2021; Pfister et al., 2003; Song et al., 2022), classify cloud types (e.g., Buch et al., 1995; Calbó and Sabburg, 2008; Wang et al., 2020), track cloud movement (e.g., Raut et al., 2023), and serve as a benchmark for the evaluation of satellite and reanalysis data products (e.g., Ackerman and Cox, 1981; Werkmeister et al., 2015), as well as other ground-based cloud cover estimates (e.g., Kassianov et al., 2005; Riley et al., 2020). Moreover, when considerations of the camera frame of reference equivalent to surface visual observers are made (see Long et al., 2006), all-sky image records can be combined with historical records of fractional sky coverage made by human observers (e.g., Free et al., 2016) to generate long-term climatology extending back to well before the satellite era (Free and Sun, 2013).

The ARM user facility has been operating all-sky imagers for a quarter of a century to characterize cloud coverage and supplement its other measurements. Specifically, the total sky imager (TSI; Morris, 2005) was used throughout this period until the recent retirement of all TSI instruments. Two TSI instruments at the ARM Southern Great Plains site in Oklahoma (SGP; Sisterson et al., 2016) and the Eastern North Atlantic central site at Graciosa Island (ENA; see Mather, 2024; Wood et al., 2015) continued to be operated alongside the TSI successor, the All-Sky-Imager (ASI-16; Schreder-CMS, 2017), for evaluation and data continuity purposes, as discussed below, until their decommission in August and July of 2025, respectively.

The TSI's advanced cloud cover algorithm (Long et al., 2006) provides a reliable cloud cover data product in an extra wide field-of-view (FOV) covering the sky up to 80° off-zenith. The data product includes an error mitigation routine and a unique feature of correcting the estimated cloud cover from the projected image sensor to an observer's frame of reference, facilitating robust comparisons with historical records. This analytical correction renders the cloud cover algorithm significantly more computationally efficient than other methods that perform this correction via various image distortion corrections.

Unlike the TSI, its successor, the ASI (Figure 1), has no moving parts and operates a Complementary Metal-Oxide-Semiconductor (CMOS) sensor, which is less prone to saturation and flaring effects than charge-coupled device (CCD) sensors, such as the one operated by the TSI, resulting in a higher dynamic range in pixels surrounding bright objects such as the solar disk, which renders a mechanically rotating shadowband obsolete (see Table 1 for ASI specifications).

Figure 1: **(Left) The ARM All-Sky-Imager (ASI) mounted at the Southern Great Plains (SGP) central site and (right) zoom on the ASI housing and fish-eye lens.**

The ASI vendor-provided image processing software allows for some control over the partitioning of cloudy and clear sky areas in the ASI image by setting specific thresholds. However, because different sky scenes (midday, sunset, partial solar disk coverage, polluted days, etc.) generate different variations and contrast between the RGB channels and channel ratios of clear and cloudy pixels, fixed thresholds are unsuitable for an operational product aimed at applying to the diversity of atmospheric scenes typical to different ARM deployments throughout the year. Moreover, the ASI vendor-provided software does not offer features critical to the dynamic nature of ARM sites and deployments, such as automatic determination of the effective image center and border, solar disk characterization, and evaluation of ASI camera misalignment and time-dependent offsets.

Table 1: The ARM ASI specifications.

Focal length	1.47 mm
Aperture	F2.8
Sensor	1/1.8 inch Progressive CMOS
Minimal illumination	0.18 Lux at F2.2 (Color)
Dynamic range	100 dB
Image resolution	1536 × 1536 pixels
Image capture rate	30 sec at all sites except for SGP (15 sec)

Here, we present an advanced machine-learning-powered cloud cover data product, the ASI Sky Cover (hereafter, ASISKYCOVER). The ASISKYCOVER product is consistent with the TSI processing algorithms in many aspects, such as the extensive solid angle coverage and processing of cloud cover to represent the camera (visual observer) position's frame of reference. However, this product provides essential upgrades to ensure more accurate diagnostics and reduced susceptibility to false detections. It offers new features such as artifact detection, uncertainty metrics, and the estimation of near-zenith cloud cover, facilitating direct comparisons with remote-sensing datasets and streamlining cloud cover evaluation, as well as the synergistic use of ASI data with other zenith-pointing ARM measurements, including radar, lidars, and radiometers.

In the following sections, we describe the ASISKYCOVER processing algorithm (Section 2) and present an analysis of the ASISKYCOVER output for the entire year of 2024 at the ARM SGP site, along with a comparison to the co-located TSI, radar, and ceilometer instruments. Conclusions and a short outlook are given in Ssection. 4.

2 Methodology

The ASISKYCOVER algorithm comprises multiple steps, including image preprocessing and site-dependent correction, a machine learning (ML)- powered pixel segmentation algorithm, segmentation post-processing, and an ML-powered solar disk state classification algorithm. Those components are described below.

2.1 Image Preprocessing and Deployment Site Configuration and Corrections

Before applying the ASISKYCOVER segmentation algorithm, ASI images require minimal preprocessing to determine the effective (circular) image center and edges. This processing step is essential since the effective position of the image can influence the cloud cover calculation as a result of inaccurate representation of the projected sky area represented by a given pixel (see Section 2.3). Slight deviations in image center and edge coordinates are expected between ASI units, given the exact positioning of the ASI housing relative to the ASI's CMOS sensor (see Figure 1). While such a property is essentially fixed for a given unit, slight deviations can occasionally occur (e.g., as a result of strong winds). Therefore, the image center coordinates, x_c and y_c, are~~is~~ determined for individual images by finding the x-axis index corresponding to the maximum y-axis image diameter, and vice versa. Edges are determined by finding "near-black" pixels, i.e., pixels with red-green-blue (RGB) brightness values near 0 (on an integer scale of 0 to 255).

With this method of determining the effective image edges, ASI site deployment misalignment can be characterized in preparation for the segmentation task. Assuming that the ASI position (in azimuth and zenith) is fixed throughout a given deployment, instrument misalignment needs to be characterized only once. In the event of a unit position change or in cases where there is an interest in reevaluating the unit's alignment, multiple characterization cycles must be applied. Note that, unlike sky imagers with a shadowband or arm, such as the TSI, solar zenith angle (SZA) and azimuth angle biases are generally more challenging to determine by visually inspecting a given image.

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. Automated algorithms to detect the solar position exist~~s~~ in the literature (e.g., Savoy et al., 2016). However, the errors associated with such algorithms are ostensibly greater than those with manually ~~denoted~~ annotated images, mainly because of the requirement for clear-sky images combined with field-deployment constraints (e.g., dirt on the camera, cases of which were tested but not shown). Combined with the fixed nature of ARM deployments, which eliminates the need for frequent re~~e~~valuation of solar position misalignment, 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~~Given the site coordinates and using geometrical considerations to analytically convert a given pixel to an expected~~ SZA and azimuth angle, the site-specific biases of those angles are determined as the~~-~~ average deviation in the annotated image subset~~ of the angles derived from the solar disk center pixel indices in the denoted images from the calculated angles~~. The standard deviation metric calculated from the annotated image subset serves as a measure of robustness in bias characterization. Based on that metric, ~~T~~the positioning biases observed across~~characterized for~~ different ARM deployments are robust~~ (e.g., see concentric yellow circles in Figure 1)~~, ~~suggested~~ as indicated by the standard deviation of the differences between the ~~denoted~~ annotated ~~versus~~ and calculated angle~~ differences~~s, 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~~ positions~~ in Figure 2~~Figure 1~~d~~e~~ to Figure 2~~Figure 1~~g~~h~~).

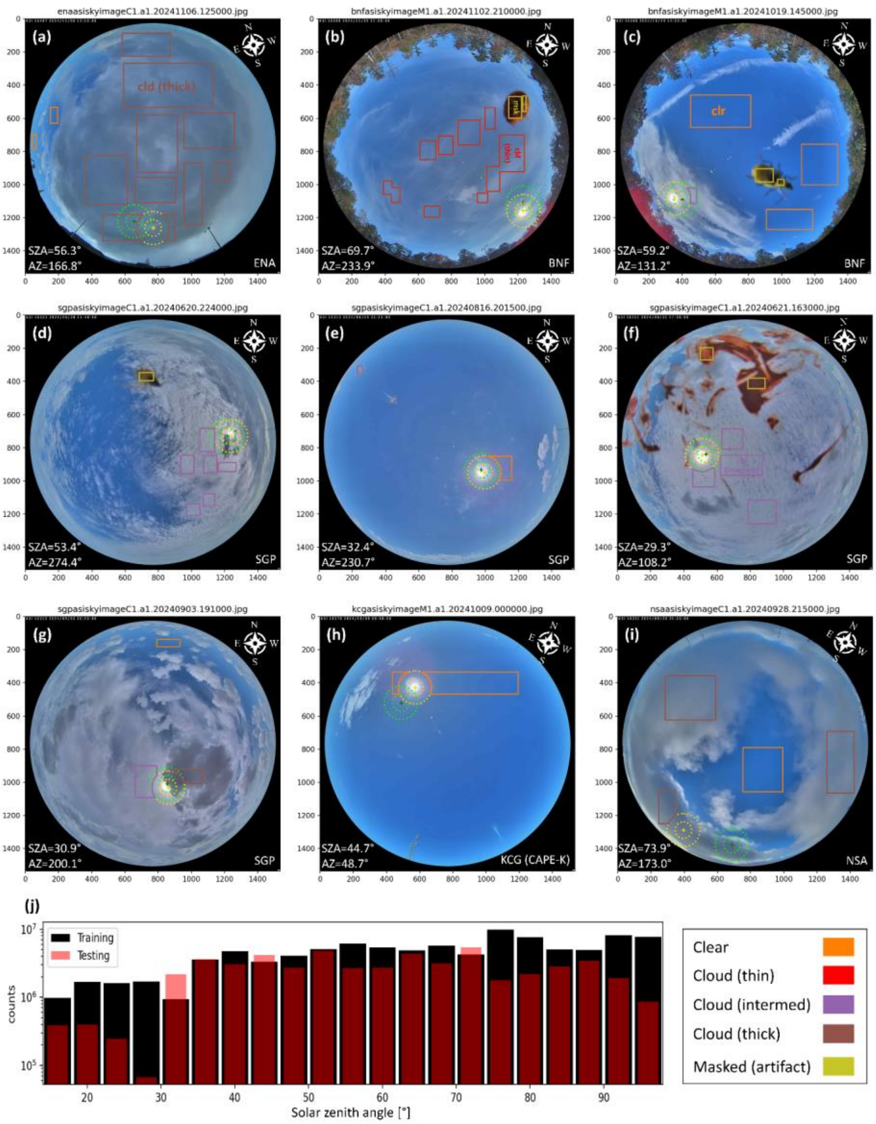
Finally, the ASISKYCOVER algorithm can utilize deployment masks that can be manually generated to exclude fixed objects visible in pixels within its effective FOV (80° off-zenith) used in the cloud cover calculations, if such objects exist in a given ASI deployment. The SGP central site deployment, however, is clear of such obstructing objects within the ASI FOV, and therefore, a mask is not applied in this case. In deployments with obstructing objects (e.g., the masked antenna at the top of Figure 3~~Figure 2~~a), the masked pixels are excluded from the cloud cover calculation but are considered in the cloud cover uncertainty calculation discussed below (see section 2.3).

2.2 Pixel Segmentation Algorithm

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).



enaasiskyimageC1.a1.20241106.125000.jpg
bnfasiskyimageM1.a1.20241102.210000.jpg
bnfasiskyimageM1.a1.20241019.145000.jpg
sgpasiskyimageC1.a1.20240620.224000.jpg
sgpasiskyimageC1.a1.20240816.201500.jpg
sgpasiskyimageC1.a1.20240621.163000.jpg
sgpasiskyimageC1.a1.20240903.191000.jpg
kcgasiskyimageM1.a1.20241009.000000.jpg
nsaasiskyimageC1.a1.20240928.215000.jpg
(a) cld (thick) SZA=56.3° AZ=166.8° ENA
(b) SZA=69.7° AZ=233.9° BNF
(c) clr SZA=59.2° AZ=131.2° BNF
(d) SZA=53.4° AZ=274.4° SGP
(e) SZA=32.4° AZ=230.7° SGP
(f) SZA=29.3° AZ=108.2° SGP
(g) SZA=30.9° AZ=200.1° SGP
(h) SZA=44.7° AZ=48.7° KCG (CAPE-K)
(i) SZA=73.9° AZ=173.0° NSA
(j)
Training
Testing
counts
Solar zenith angle [°]
Clear
Cloud (thin)
Cloud (intermed)
Cloud (thick)
Masked (artifact)

Figure 2~~1~~: ~~(a) Distribution of training (black) and testing (red) samples as a function of solar zenith angle (SZA; bin size of 2°).~~ (a~~b~~-i~~j~~) 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°).

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, we ultimately decided to utilize the gradient boosting classification algorithm in the LightGBM software package (Ke et al., 2017). Pixels are classified into one of five classes: clear, thin cloud, intermediate-thick cloud, thick cloud, and masked (Table 2~~Table 1~~). The cloud thickness qualitatively refers to the optical thickness of the observed clouds and provides a finer, much-needed differentiation relative to the "thin" and "opaque" classes of the TSI segmentation algorithm (Long et al., 2006). The "masked" class pixels are excluded from the cloud cover calculation but are accounted for as part of the uncertainty calculations in case they are within the 80° off-zenith FOV (see Section 2.3).

Table 2~~1~~: ASISKYCOVER pixel segmentation algorithm classes. The different cloud classes qualitatively represent optical thickness. The masked pixels are excluded from the cloud cover calculation.

Class	Example
Clear	
Thin cloud	Cirrus clouds
Intermediate-thick cloud	Altostratus clouds
Thick cloud	Cumulus clouds
Masked	Dirt on the ASI housing; bugs, birds, people, or flying objects within the image field-of-view

Formatted Table

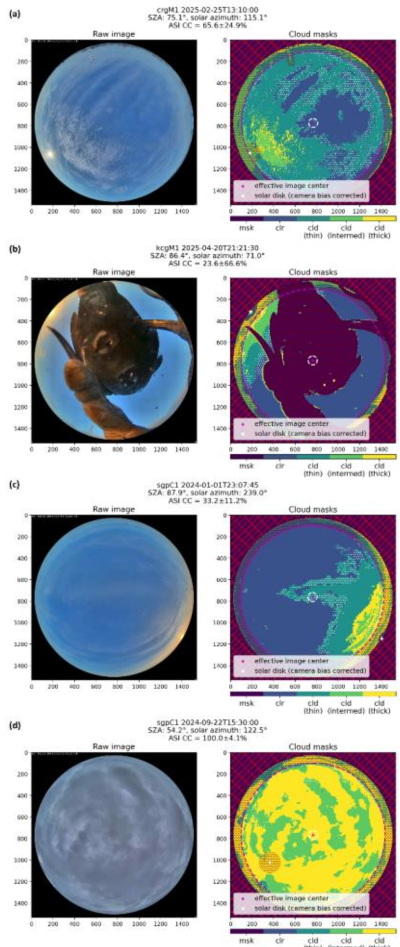


Figure 3~~2~~: ASISKYCOVER pixel segmentation algorithm output for selected scenes. (left) Raw ASI image (Note that the ASI images appear mirrored in the east-west direction; see Section 2.1 for details), (right) algorithm output. The resolved classes are given in the legend at the bottom of the classification masks. In each classification mask, the red '+' denotes the determined image center, the dashed magenta and white circles designate the image areas covering up to 80° and 5° off-zenith, respectively, the white star marker indicates the corrected position of the solar disk, dotted white hatched areas denote uncertain classification pixels, and the dotted magenta hatched areas mark masked pixels (site mask + solar disk mask). The ASI image site, facility and time, as well as the corresponding SZA, azimuth angle, and algorithm-resolved cloud cover (following section 2.3) are given in the title of each image pair.

~~We note that we do not correct for lens barrel distortion, given that the ASI lens distortion is relatively minor, with 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).~~

To classify image pixels, the ASISKYCOVER segmentation algorithm implements four features as input for the LightGBM algorithm: the pixel's red, green, and blue channels, and the red-to-blue channel ratio. Some features are more traditional, such as the blue channel and the red-to-blue channel ratio (e.g., Dev et al., 2017; Long et al., 2006), while the green channel is primarily used for artifact ("masked" class) detections, and is also impactful for classifying ASI scenes during twilight periods (not shown).

The SZA impacts image lighting, such as sky brightness and hue, especially at very high SZAs due to stronger Rayleigh scattering at shorter wavelengths. In those cases, the near-zenith clear-sky region is typically darker due to smaller scattering phase function values at angles of 90° and 270°, which present an additional challenge. However, these effects generally are not as impactful and quite gradual at lower SZAs. 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 ~~bogus~~ 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 and limit the SZA dependencies to two segmentation models: one for daytime periods (SZA < 90°) and another for twilight times (SZA ≥ 90°). This separation enables the daytime model to implicitly consider the influence of hue and brightness variations on the existing four features, without over-weighting features based on the twilight periods, which are characterized by rapidly changing conditions and unique red-blue-green channel combinations. The dedicated twilight segmentation model is trained to handle these conditions, which are considered challenging for segmentation algorithms in the first place.

The training data for the segmentation model consists of roughly 81 million samples (pixels) from 513 images collected during different days at multiple ARM sites: SGP central site (9 days); the north slope of Alaska central site at Utqiaġvik (NSA; Verlinde et al., 2016) (1 day); the ENA central site (2 days); the Cloud And Precipitation Experiment at Kennaook, Tasmania (CAPE-K; Mace et al., 2023) (4 days); and the ongoing deployment at the Bankhead National Forest, Alabama (BNF; Kuang et al., 2023) (4 days). Those pixels were chosen by manually labeling image sectors with uniform classes. In the labelling process, we attempted to sample the various features in a uniform matter, as demonstrated in the histogram shown in Figure 2~~Figure 1ja~~ for the SZA. In constructing the training dataset, we automatically exclude pixels within a 100-pixel radius from the mis~~-~~alignment-corrected solar disk center to prevent biases in the trained classification model as a result of commonly-saturated pixels over or in the immediate surrounding of the solar disk (e.g., panels b, d, f, g, h, and i in Figure 2~~Figure 1~~).

The LightGBM segmentation model is trained using weights that consider the number of training samples for each class, such that the effective class weight ratios are $1:\sqrt{1/3}:\sqrt{1/3}:\sqrt{1/3}:1$ for the clear, thin, intermediate, thick, and mask classes, respectively. The weighting of the cloudy classes is used to set the root-sum-square (RSS) weight of the cloudy classes equal to 1, thereby accounting to some extent for the higher number of cloudy vs. clear classes.

The segmentation model decision probabilities are used to determine uncertain classifications with the objective of being able to robustly indicate the cloudy-clear distinction, rather than the specific cloud class (in case of one of the "cloud" class pixels). 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 estimationes 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.

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.

To test the output of the segmentation algorithm, we also constructed a testing dataset fully independent from the training dataset (sample histograms as a function of SZA are shown in Figure 2Figure 1ja). The testing dataset consists of roughly 53 million manually-labeled samples (pixels), rendering an effective train-test split of approximately 60-40. The testing samples were gathered from 207 images collected at 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.

Figure 3~~Figure 2~~ demonstrates the output of the segmentation algorithm after it is applied to untrained images from CoURAGE (Figure 3~~Figure 2~~a; part of the testing dataset), CAPE-K (Figure 3~~Figure 2~~b; part of the testing dataset), and SGP (Figure 3~~Figure 2~~c and Figure 3~~Figure 2~~d). Figure 3~~Figure 2~~a illustrates the model performance in a highly challenging scene with fractal cloud cover, in which case the calculated cloud cover is 65.6±24.9% (note that this is the cloud cover value after pixel area weighting described in Section 2.3 below). In this case, cloud type and occurrence determination is robust, though a small fraction of the observed cirrus cloud pixels around the right side of the image are being misclassified as clear. In Figure 3~~Figure 2~~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. Figure 3~~Figure 2~~c demonstrates the highly challenging classification around sunrise and sunset times. In this case, cloud cover, dominated by high-level cirrus clouds, is underestimated, mainly due to pixel misclassifications of the left part of the image. In Figure 3~~Figure 2~~d, the classification output of a fully overcast scene dominated by thick clouds is displayed, in which case the resultant cloud cover is 100±4.1%.

By applying the segmentation algorithm on the full testing dataset, we can examine the algorithm's robustness by generating confusion matrices (Figure 4~~Figure 3~~). The confusion matrix for all pixel classes (Figure 4~~Figure 3~~a) suggests that artifacts are properly masked 98% of the time. Clear sky pixels are properly classified at a~~n impressive~~ rate of 91%, while the specific cloudy classes are properly classified in more than two-third of their relevant testing samples (pixels) (roughly 81% for the intermediate class). Misclassifications are typically of "close" classes. For example, intermediate-thick cloud classified as a thick cloud (and vice versa) or a thin cloud classified as clear. The latter distinction is especially challenging at cloud edges (see Koren et al., 2007), and in specific scenarios, it also challenges human observers. In general, cirrus clouds pose the greatest classification challenge, especially at larger SZA values, as demonstrated in Figure 3~~Figure 2~~c, for example (cf. Dev et al., 2017; Long et al., 2006; Pfister et al., 2003). Thus, deployments to polluted sites or cirrus-dominated sites could benefit from recalibration of the model weights and potentially revisiting the training dataset. However, for most applications, assuming the primary goal is separating clear and cloudy pixels, all cloudy classes can be combined (post-processing) to evaluate the general algorithm performance.

When assessing the testing~~raining~~ dataset using such an approach, the algorithm's robustness is further demonstrated (Figure 4~~Figure 3~~b). In this case, only 6.2% of the clear pixels and 5.5% of the cloudy pixels are classified incorrectly. 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 (Figure 4~~Figure 3~~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. Note that the smaller correctly "masked" pixels in this case are the result of the exclusion of the bird image (Figure 3~~Figure 2~~b), taken when the solar disk was just above the horizon. With only a few bugs and some dirt left on the residual testing dataset, and the tendency to suggest a cloud "halo" around artifacts, we received this lower percentage. However, most of those incorrectly classified pixels are treated as "uncertain" and therefore should primarily be excluded from the cloud fraction calculations.

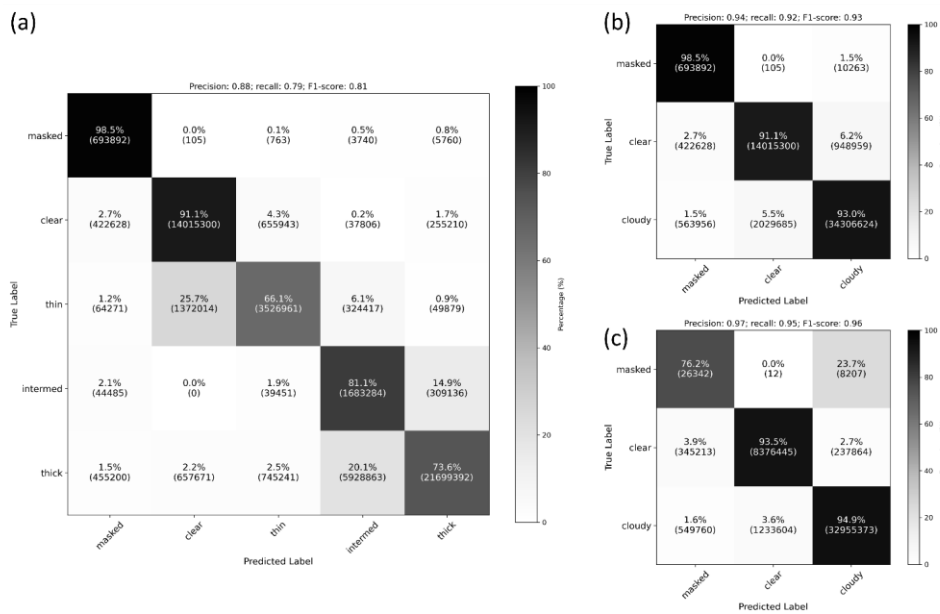


Figure ~~4~~3: Confusion matrices of the segmentation algorithm output against the testing dataset. (a) Using all segmentation algorithm classes; (b) merging all cloudy classes to focus on cloudy-clear separation; and (c) same, but excluding testing samples with SZA exceeding 80°. The weighted precision, recall, and F1-score metrics for each testing subset are given in panel titles.

2.3 Segmentation Post-Processing

Following the segmentation algorithm operation, cloud cover is calculated by weighting each pixel by the effective sky area it represents in the observer's (camera) frame of reference. Following Long et al. (2006), this calculation is done using:

~~(1)~~(3) $$f = \frac{\sum l_i N_i^{cld}}{\sum l_i N_i^{valid}},$$

where f is the cloud fraction, N_i^{cld} and N_i^{valid} denote ~~given~~ the number of valid cloudy and all valid pixels, respectively, and l_i is the pixel area correction factor calculated as:

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

~~(2)~~(4) $$l_i = \frac{2(cos\theta_{i1} - cos\theta_{i2})}{(\theta_{i1}^2 - \theta_{i2}^2)},$$

where θ_{i1} and θ_{i2} are the pixel~~solar~~ zenith angles (in radians) at the ~~vertical~~ edges of a given pixel *i* along the radial direction from the effective image center. Here, valid pixels are pixels within the 80° off-zenith effective FOV, not masked by site or solar disk masks (see Section 2.1), not classified as "masked" by the segmentation algorithm, and not considered uncertain. Finally, in addition to the cloud cover in the full 80° FOV, ASISKYCOVER provides cloud cover estimates for a narrow FOV, in a region covering up to 5° off-zenith (e.g., the white circles in the right panels of Figure 3~~Figure 2~~).

2.4 Solar Disk State Classification Algorithm

In addition to the segmentation algorithm, we have trained a simplified binary random forest (RF) model to determine the solar disk state, which supplements the ASISKYCOVER product, with the possible solar disk state classes being "sun blocked" (by clouds) and "solar disk seen". Note that we do not further partition the state classes to "solar disk seen through clear sky", "solar disk seen through thin or intermediate-thick clouds", and "partial solar disk blockage by clouds". Classification of these solar states would require the trained model to be spatially-informed, suggesting that a convolutional neural network architecture should be used for a task of this type. Future updates to the ASISKYCOVER product could include such a more comprehensive solar disk state classification component.

The current RF solar disk state classification model is trained using the Scikit-Learn package (Pedregosa et al., 2011) and implements a set of features characterizing the brightness of pixels surrounding the solar disk center using only the red and blue channels. Those features are the minimum, maximum, mean, standard deviation, interdecile range, and the 5^{th}, 10^{th}, 25^{th}, 50^{th}, 75^{th}, 90^{th}, and 95^{th} percentiles of pixels within circles with radii of 15, 25, and 50 pixels centered at the location of the solar disk center on training images (799 samples); the same manually labeled solar disk-designated images used to characterize the solar disk mis~~-~~alignment (Section 2.1).

Evaluation of the binary classification using an independent testing dataset consisting of 258 samples (roughly 75-25 train-test split) suggests robust performance of this classification model (Figure 5~~Figure 4~~) with all "solar disk seen" samples and 87% of "sun blocked" samples being classified correctly. Model robustness is further indicated by an F1-score, precision, and recall values of 0.973, 0.947, and 1.000, respectively.

Formatted: Outline numbered + Level: 1 + Numbering Style: 1, 2, 3, … + Start at: 1 + Alignment: Left + Aligned at: 0" + Indent at: 0.25"

Formatted: Font: Italic

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

Formatted: Superscript

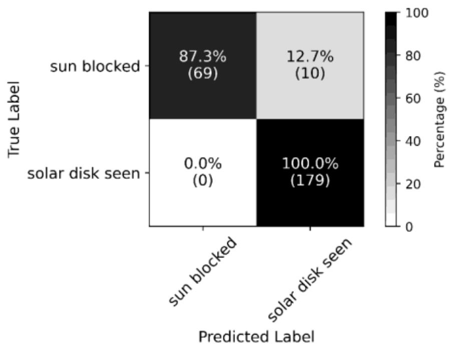


Figure 5~~4~~: Confusion matrix of the solar disk state classification algorithm output against the testing data.

3 Results

3.1 Single-day time series

Figure 6~~Figure 5~~ illustrates a time series of the ASISKICOVER cloud cover data for January 1, 2024, at the SGP central site, demonstrating some of the information provided by this data product. As indicated by the top panel, representing the full FOV, cloud coverage exhibits a highly variable pattern prior to sunrise, followed by a steady and sharp rise to overcast conditions commensurate with sunrise. During this sharp cloud cover increase, clouds are initially classified as "thin", but then quickly transition to thick-dominated clouds. The overcast conditions persist until after 19:00 UTC, when cloud type transitions back to thin-dominated, with several local bursts of cloud occurrence. The light pink curve, denoting the TSI sky cover product, describes a similar story with cloud cover largely staying within the ASISKYCOVER uncertainty range. However, the cloud onset occurs only after sunrise with a distinct offset relative to the ASISKYCOVER. In addition, the TSI cloud cover appears significantly underpredicted relative to ASISKYCOVER after 22:00 UTC, ~1 hour before sunset. Examination of the solar disk state classification output (pink markers) indicates that indeed the solar disk is visible after 19:00 UTC but is generally blocked by clouds before that, 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~~In addition, given that the solar disk state is classified as "blocked by clouds"~~, it is likely that the thin-dominated period ~~right after sunrise~~ 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. 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 ~~; note~~ 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?)~~.~~

To supplement the evaluation of the time series output, Ka-band ARM Zenith Radar (KAZR; Widener et al., 2012) and ARM ceilometer (CEIL; Morris, 2016) cloud occurrence time series data are illustrated in the bottom panel of Figure 6~~Figure 5~~ together with the near-zenith output of ASISKICOVER. The KAZR and CEIL cloud occurrence percentages are averaged in 5-min running mean windows to mitigate their temporal point measurement operation difference from the spatial nature of the ASI. CEIL cloud occurrence instances are determined when the instrument indicates the occurrence of at least one cloud base height. KAZR-determined cloud instances are sample times with one or more moderate sensitivity (MD) mode echoes with signal-to-noise ratio exceeding -16 dB (see Silber et al., 2018) at radar range gates greater than 4 km. This dual instrument depiction accounts for the maximum CEIL (Vaisala CL31 model) range of 7 km, the lack of its cloud base height sensitivity to ice clouds, and the commonly occurring SGP KAZR echoes generated by biota at ranges of up to a few kilometers (below the freezing level). This multi-instrument depiction also streamlines figure interpretation. As such, the continuing CEIL cloud occurrence of 100% preceding sunrise, commensurate with 0% KAZR occurrence, suggests the occurrence of low-level clouds, and that the ASISKYCOVER cloud cover is significantly underpredicted before sunrise. In comparison, the TSI's 0% cloud cover is a starker artifact. These challenging lighting conditions provide the main incentive for flagging ASISKYCOVER data at SZA values exceeding 80° (vertical dotted black lines in Figure 6). That said, Figure 6~~Figure 5~~ suggests that clouds are intermittently indicated by the classification algorithm during the civil twilight. As shown from times later on the depicted day (after 18:30 UTC), the continuous CEIL-based cloud cover diminishes while the KAZR-based cloud cover increases in transient bursts, suggesting high cloud (cirrus) occurrence, consistent with the ASISKYCOVER analysis. However, while the change in total cloud cover is gradual in the ASISKYCOVER output, changes are more abrupt in the CEIL and KAZR time series data. The near-zenith product output displayed in the bottom panel of Figure 6~~Figure 5~~ resembles this abrupt change in cloud occurrence and further shows a general agreement between ASISKYCOVER and the CEIL and KAZR instruments (e.g., between 18:00 and 22:00 UTC).

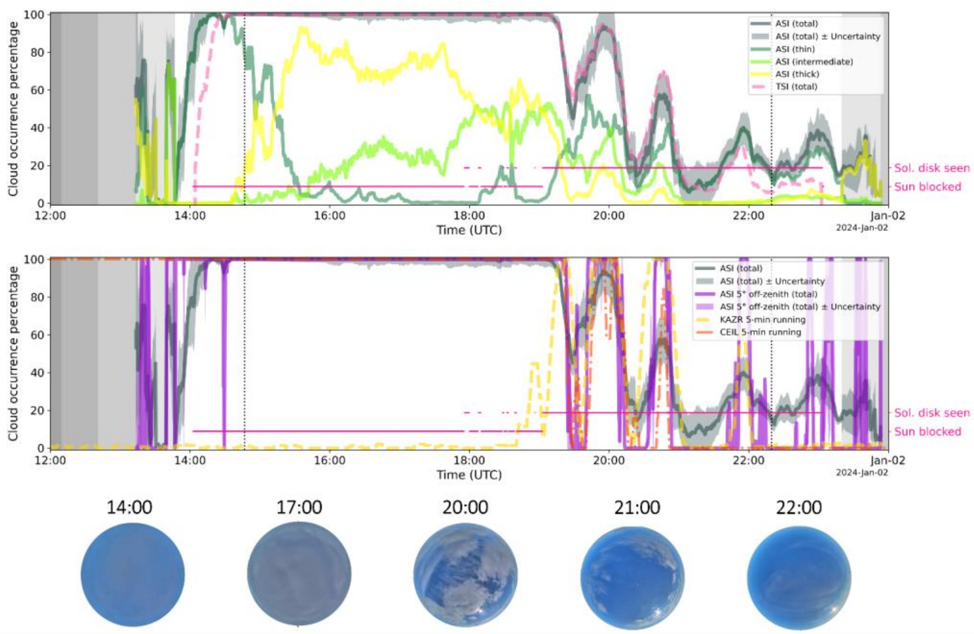


Figure 65: ASISKYCOVER cloud cover time series for the SGP central facility on January 1, 2024. (Top) total cloud cover time series using the full 80° off-zenith field-of-view together with time series for the three cloud classes (see legend). The light pink curve illustrates the TSI sky cover time series. The pink markers denote the solar disk state classification output (see right y-axis labels). The Grey-shaded regions extending until nearly 14:00 UTC and from roughly 23:00 UTC designate (in decreasing brightness) the civil twilight, nautical twilight, astronomical twilight, and nighttime. The vertical dotted black lines denote 80° SZA crosspoints. (Bottom) Similar, but showing the near-zenith (5° off-zenith) product time series together with the 5-min running mean Ka-band ARM Zenith Radar (KAZR; using moderate sensitivity mode echoes above 4 km range) and ceilometer cloud occurrence (see legend). A selected set of raw ASI images are shown at the bottom of the figure together with their corresponding times. (Note that the ASI images appear mirrored in the east-west direction; see Section 2.1 for details).

3.2 Full-Year Analysis

Using the full year of processed ASISKYCOVER data, cloud cover over the SGP central site can be characterized for daytime conditions. Here, we exclude samples corresponding to SZA values exceeding 90°, which are considered less accurate (see previous section). 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. 49%Forty-nine percent of which of this annual cloud cover isare 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).

~~We note that further analysis is required to examine the representativeness of these cloud cover and cloud type partitioning trends throughout the year over the SGP area, as well as the influence of only analyzing daytime data (SZA < 90°), which would necessitate the use of other instrument datasets.~~

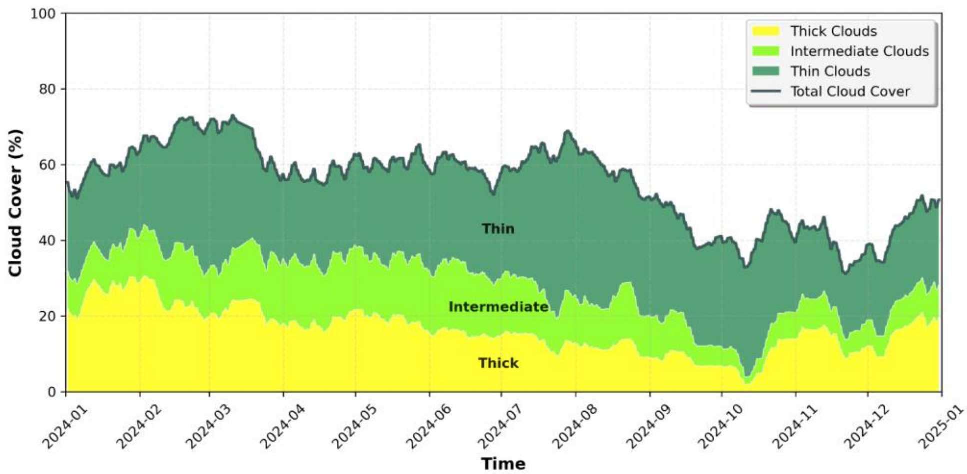


Figure 7~~6~~: Cloud cover 31-day running-mean time series for the SGP site as determined by ASISKYCOVER for 2024. The stacked colored areas represent the distribution of cloud types (see legend and labels).

~~-~~A 31-day running mean cloud cover evolution throughout the year is depicted in Figure 7~~Figure 6~~. Cloud cover at the beginning of the year fluctuates around 55% in early January, with roughly equal contributions of thin and thick clouds. It then gradually increases until March, with the most prevalent classified cloud type being thick clouds until early February. From here through the end of the year, the thin cloud class is the most prevalent, though for the most part, its contribution to the total cloud cover is still smaller than the combination of the intermediate-thick and thick cloud types. The apparent thin cloud cover first peaks in early March with a 31-day mean cover of ~38% around the time when the total cloud cover has a global maximum (31-day mean of 73% at a period centered on March 13). Total cloud cover trends downwards starting in mid-March and stabilizes around 60% between April and early July. It then gradually increases until late July, peaking at 68%, commensurate with the global thin cloud cover peak of 42%. From that point, total cloud cover generally trends downwards until mid-October, when thin clouds contribute close to 90% of cloud cover over 31 days centered on October 11, when the 31-day total cloud cover has a local minimum of 33%. The total cloud cover continues to fluctuate around 40% through early December, with the global 31-day mean total cloud cover minimum of 31% in a window centered on November 22. Finally, from early

December, total cloud cover trends towards 50% with an increasing prevalence of thick clouds, in what appears to be a setup for a repetition of early 2024 trends. Assuming interchangeability between cloud thickness and altitude (similar to the examples in Table 2), 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.

Formatted: Normal

3.3 Bulk Comparison to Other ARM Datasets

Without a clear cloud cover benchmark, we begin this bulk comparison to the TSI product, the output dataset of which is the only direct equivalent to the ASISKYCOVER dataset. As suggested by the single-day time series presented in Figure 6~~Figure 5~~, ASISKYCOVER is generally consistent with the TSI product. However, the two products tend to have larger deviations and inconsistencies under certain conditions than others.

Figure 8~~Figure 7~~ shows kernel density estimations (KDEs) of the ASISKYCOVER cloud cover deviations from the TSI cloud cover data product using the SGP central site 2024 data, conditioned on different variables. The interpretation of the depicted KDEs is based on numerous observed ASISKYCOVER and TSI scenes, which were qualitatively evaluated using expert judgment. On average, ASISKYCOVER exhibits a slightly higher tendency to detect more clouds than the TSI (by less than one percentage point) (Figure 8~~Figure 7~~a). However, the dataset difference variability is quite extensive, exceeding 14 percentage points.

Examination of deviations while conditioning on the ASISKYCOVER total cloud cover (Figure 8~~Figure 7~~b) suggests that the TSI tends to detect a cloud cover that is ~6% higher in cases of clear to low cloud cover (up to 25% based on ASISKYCOVER). In these cases, the TSI processing often incorrectly classifies clear pixels as cloudy, primarily due to solar flaring and resultant saturation of large image sectors. However, in broken cloud scenarios, there are cases in which the ASKISKYCOVER incorrectly classifies a cloudy pixel as clear. In those cases, the pixels are often flagged as uncertain, an ASISKYCOVER feature that is not reflected in this comparison. Scenes with extensive cloud cover (75% up to overcast) are characterized by better agreement between the two data products, with relatively minor variability shown by the narrow KDE and a small standard deviation (SD) of ~9%. Such scenes are relatively more straightforward to classify, given that the most challenging image regions to classify are typically located at cloud edges. Vice versa, in relatively complex scenes with a mixture of cloudy and clear pixels, the product disagreement reaches its maximum with SD at ~19%.

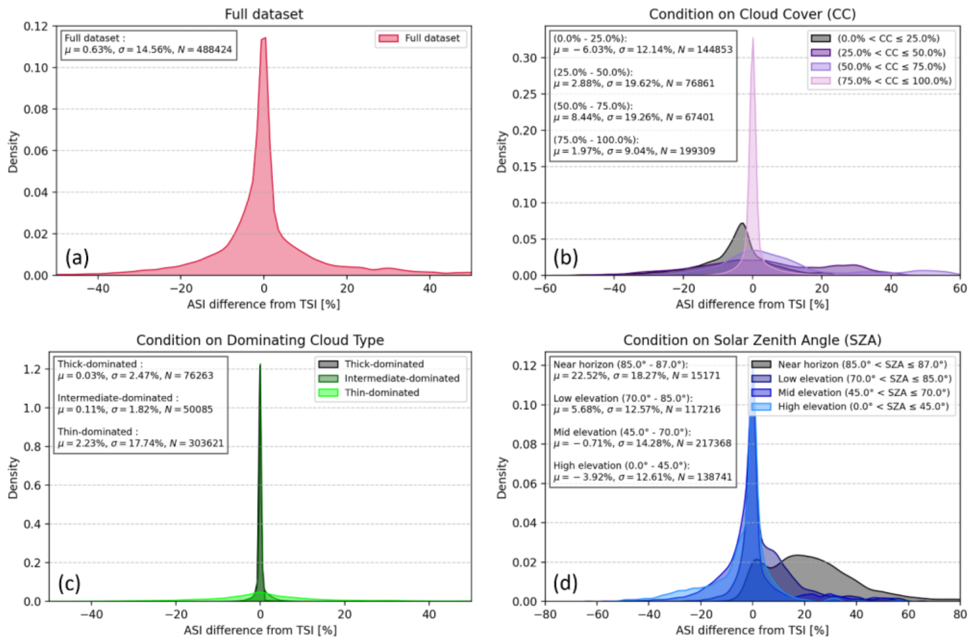


Figure 8~~7~~: ASISKYCOVER deviations from the TSI cloud cover product kernel density estimations (KDEs) for SGP in 2024 while conditioning on different variables: (a) Full dataset (no conditioning), (b) ASISKYCOVER cloud cover (CC), (c) dominating cloud type per scene with CC greater than 5% based on ASISKYCOVER, and (d) SZA (see legend). Mean deviation (μ), standard deviation (σ), and number of TSI-ASISKYCOVER sample pairs in the conditioned data subset (N) are given on the top left of each panel.

The occurrence of thick clouds often characterizes overcast conditions over the SGP. Conditioning on the dominating (most prevalent per scene) cloud class (Figure 8~~Figure 7~~c), it is indicated that both data products have good agreement, with thick-dominated scene deviations smaller than 0.05~~1~~% on average and an SD of merely a few percent. The instrument cross-agreement is even better for scenes with intermediately -thick cloud dominance, where the SD is smaller than 2%. ~~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 the TSI data product exhibits a very strong general tendency to classify clouds as "opaque" regardless of their true cloud thickness, and hence, to also underestimate the occurrence of thin clouds; not shown). Figure 8~~Figure 7~~c provides a clear indication of the challenge in consistently classifying thin cloud pixels (~~or clear sky pixels as clear for that matter;~~ see Figure 4~~Figure 3~~). While on average, ASISKYCOVER detects total cloud cover higher by 1 percentage point in thin-cloud dominated scenes, the variability is distinctively higher (see KDEs), with SD on the order of 18~~7~~%. 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 ~~emphasize~~ corroborate the class-dependent ~~model~~ algorithm performance test results ~~demonstrated above using the test dataset~~ (Figure 4~~Figure 3~~).

Conditioning on SZA (Figure 8~~Figure 7~~d), product deviation variability remains large in all of the four examined ranges, but the average deviation provides additional information on instrument tendencies (note that the highest SZA is set to 87°, the maximum SZA value for which the TSI classification algorithm is applied). At high solar elevation (SZAs smaller than 45°), the TSI tends to overestimate the total cloud cover by ~4 percentage points. This overestimation is primarily driven by very strong solar flaring saturating a large number of pixels, rendering them classified as cloudy. We note that the ASI at SGP suffers as well from solar flaring effects, though to a lesser extent than the TSI. However, our evaluation (not shown) suggests that ASI flaring effects are often starting to show during clear-sky scenes in the near-zenith pixels, and hence, the near-zenith product at SZA < 60°, which could be mitigated, to some extent, by using it in a quasi-binary approach as demonstrated below. At mid solar elevation (SZA between 45° and 70°), ASISKYCOVER and the TSI have reasonable agreement, with an average deviation of roughly half a percentage point. However, at higher SZAs, the ASI tends to suffer more from solar effects in the form of internal reflections, which are enhanced when the ASI has dirt on its housing, resulting in ASISKYCOVER overestimation (~6 percentage points more classified cloudy pixels relative to the TSI in SZA between 70° and 85°). When the solar disk is near the horizon, however, ASISKYCOVER indicates total cloud cover that is higher by more than 20 percentage points relative to the TSI, which we attribute primarily to a loss of TSI sensitivity rather than ASI-related artifacts.

While this analysis points to some instrument and algorithm weaknesses, the limited conditioned variable ranges and lack of consideration of classification uncertainty, which the ASISKYCOVER data product provides, result in an incomplete picture regarding the ASISKYCOVER product evaluation. To perform a robust comparison to other ARM datasets, we examine the total cloud cover as a function of SZA — a variable independent of ASISKYCOVER output. Unlike a simple long-term time series analysis, for example, this approach deconvolves the changing length of day influence on the determined cloud cover, and facilitates the identification of instrument biases and tendencies, which were suggested in Figure 8~~Figure 7~~. We compare the ASISKYCOVER total cloud cover to the TSI sky cover data and the combination of the 5-min running mean KAZR and CEIL data described above (see Section 3.1). We note that for all practical purposes, we treat the KAZR-CEIL combination as insensitive to SZA and therefore unbiased on a first order, noting that most solar effects are typically detected when compared against nighttime data, which are excluded from this analysis.

Figure 9~~Figure 8~~ shows the 2024 SGP central site ASISKYCOVER-determined cloud cover occurrence histogram as a function of SZA normalized for each SZA bin, together with mean cloud cover curves as a function of SZA for the ASISKYCOVER output, 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. The histogram at the bottm of Figure 9 indicates that the ASISKYCOVER total cloud cover distribution is generally bi-model regardless of SZA, with peaks at the distribution edges, i.e., clear (0-4% cover) and overcast (96-100% cover) conditions (see the darker histogram bins at the top and bottom of the depicted data). However, the clear conditions peak (histogram bottom) is much weaker and occasionally absent in specific SZA ranges.

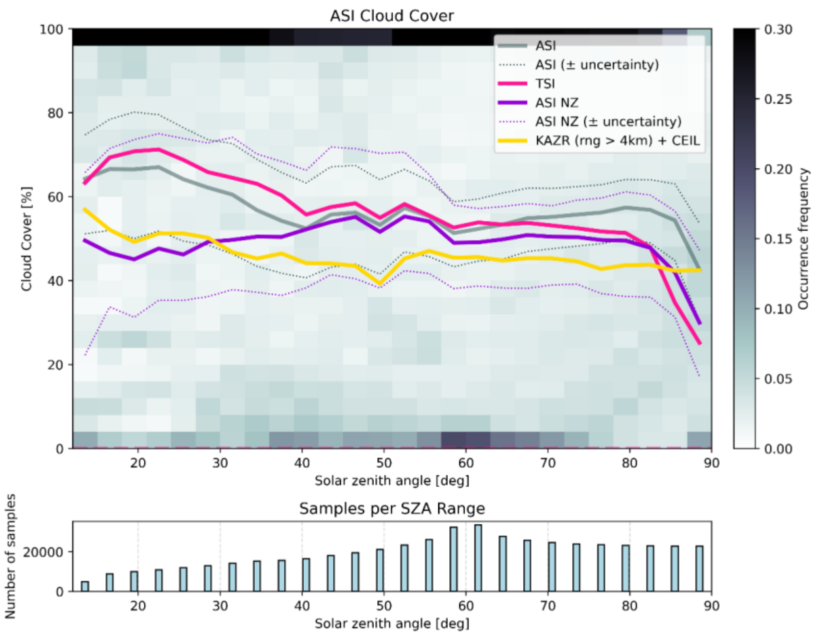


Figure 98: (Top) ASISKYCOVER-determined cloud cover occurrence frequency histogram as a function of SZA (bin sizes of 4% and 3°, respectively) normalized for each SZA bin (i.e., each column of histogram bins sums to 1) for SGP in 2024. The mean cloud cover as a function of SZA is illustrated by the grey curve, with the dotted grey curves designating the mean ± sample uncertainties. The pink and yellow curves designate the mean cloud cover as a function of SZA for 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 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) product, requiring 99% or more of (~5400) pixels within 5° off-zenith to indicate cloud occurrence, rendering a more equivalent comparison to the KAZR-CEIL data. (Bottom) The total number of samples (images) per SZA bin range.

Consistent with KDE analysis above, the reduced peak at SZAs smaller than 36° is the result of solar flaring effects, resulting in sporadic higher cloud cover counts. These sporadic counts contribute to the increasing total cloud cover with a decreasing SZA pattern delineated by the mean ASISKYCOVER curve (grey). The TSI-determined cloud cover (pink) exhibits a similar trend, though with more pronounced overestimation. While one could suggest that this gradual enhancement is a mere artifact, the KAZR-CEIL curve (yellow) shows a similar behavior, which suggests cloud cover enhancement around local noon. Moreover, the cloud cover anti-correlation with SZA was not detected in an equivalent 1-year bulk analysis we performed of ASISKYCOVER output for the ARM ENA site (not shown).

Here, the ASISKYCOVER and TSI mean-cloud cover patterns appear to follow each other up to 60°, also consistent with the KAZR-CEIL patterns. From this SZA to roughly 80°, the ASISKYCOVER total cloud cover gradually increases, exceeding the TSI value, yet with the TSI remaining within the ASISKYCOVER uncertainty range (dotted grey curve). As indicated in the analysis above, this gradual increase results from the incorrect classification of clear pixels as cloudy, principally due to the combination of ASI housing dirt and internal reflection. In those cases, pixels are mostly classified as thin clouds, which explains the ASKISKYCOVER overestimation in low cloud cover scenes (~~Figure 7~~Figure 8b) and part of the variability in thin-dominated scenes (Figure 8~~Figure 7~~c) While these optical effects are observed throughout given ASI images, they are most pronounced close to the solar disk; at these SZA ranges, around the side of the image. Evidently, the NZ product (purple curve in Figure 9~~Figure 8~~)~~, processed here in a near binary approach equivalent to the CEIL and KAZR,~~ does not display this bogus gradual increase. Moreover, the equivalent ENA analysis (not shown) did not indicate any such artificial patterns, suggesting that this artifact tendency could be driven by local conditions (pollution, etc.) and influenced by the specific ASI instrument deployed at SGP. Above 80°, both the ASISKYCOVER and TSI products lose their cloud detection capability, though the ASISKYCOVER does so in a much weaker fashion, while retaining quantitative bulk agreement with the KAZR-CEIL combination. This agreement with the KAZR-CEIL is largely seen throughout all SZA ranges when accounting for the uncertainty ranges, in both the full FOV and NZ products. Furthermore, the mean curve pattern similarities tentatively suggest that ASISKYCOVER could be combined with zenith-pointing instruments for certain applications. We note that this full-FOV correspondence with point-measurement datasets might be somewhat different, depending on the prevailing cloud types (see Riley et al., 2020; Wagner and Kleiss, 2016).

4 Conclusions and Outlook

The ASISKYCOVER data product provides a thorough and comprehensive solution to the cloud cover determination challenge using all-sky imagers. It mitigates the influence of various artifacts such as bugs and dirt, and handles a wide range of scenarios. While still suffering some artifacts, especially when classifying scenes dominated by thin clouds and/or cases at high SZA, those artifacts are ameliorated by the ML-driven algorithm and are generally accounted for in the data product's uncertainty estimates. The algorithm is theoretically transferable to other all-sky imager platforms, likely with a new set of training data to address the different sensitivities, ~~gamma~~ color saturation level and gain configurations, and resilience to pixel saturation and "solar flaring" in other image sensors.

ASISKYCOVER datasets ~~will soon be~~are already being routinely generated for active and past ARM deployments involving the ASI instrument, and ~~will be~~are available via the ARM Data Discovery. This data product will serve as the basis for ARM value-added data products (VAPs) that could include spatial classification and cloud height and movement tracking datasets handling multiple instruments and consecutive images, advanced artifact filtering by considerations of fixed objects in consecutive images, and more. Future development of the ASISKYCOVER algorithm will include spatially feature-based solar disk state classification (e.g., using convolutional neural networks), which will facilitate robust separation between conditions such as partial solar disk blocking by clouds and the distinction between the solar disk visibility through thin clouds or cloud-free air, as well as dedicated configurations for highly polluted sites and events. As demonstrated here, the introduction of ASISKYCOVER ensures the continuation of the long-term cloud cover records ARM collected using the TSI, while the various features of ASISKYCOVER open new avenues for future developments and dataset enhancements.

Data Availability

~~Releases of the~~ ASI sky cover (ASISKYCOVER) time series and cloud mask (ASICLDMASK) datasets (Silber et al., 2026a, b, c) ~~will be~~are available on the ARM Data Discovery (https://adc.arm.gov/discovery/). ASI sky images (Flynn, 2024), TSI sky images, ~~and~~ cloud mask, and cloud cover (Flynn and Morris, 2000a, b, c), KAZR moderate sensitivity (MD) mode~~-~~ (Feng et al., 2011), and ceilometer (Zhang et al., 1996) datasets used in this study are available on the ARM Data Discovery (https://adc.arm.gov/discovery/; last access: 21 February 2025).

Author Contribution

Conceptualization: IS and DMF
Formal analysis, investigation, methodology, visualization, and original draft preparation: IS
Project management: JMC
Data curation and validation: IS, ELC, and BDE
Manuscript review and editing: all authors

Competing Interests

The authors declare that they have no competing interests.

Acknowledgments

The authors thank Adam Theisen, Roj Marchand, Scott Collis, ~~and~~ Bhupendra Raut, and Evgueni Kassianov for helpful discussions and valuable feedback on ASISKYCOVER. Data were obtained from the ARM user facility, a U.S. Department of Energy (DOE) Office of Science user facility managed by the Biological and Environmental Research (BER) program.

Financial Support

This research was supported by the ARM user facility, a U.S. Department of Energy (DOE) Office of Science user facility managed by the Biological and Environmental Research (BER) program. Pacific Northwest National Laboratory is operated for the U.S. Department of Energy by Battelle under contract DE-AC05-76RL01830.

References

Ackerman, S. A. and Cox, S. K.: Comparison of Satellite and All-Sky Camera Estimates of Cloud Cover during GATE, J. Appl. Meteorol., 20, 581–587, https://doi.org/10.1175/1520-0450(1981)020%253C0581:COSAAS%253E2.0.CO;2, 1981.

Buch, J., K. A., Sun, C. H., and Thorne, L. R.: Cloud classification using whole-sky imager data, in: Report Number: CONF-9503140-, 1995.

Calbó, J. and Sabburg, J.: Feature Extraction from Whole-Sky Ground-Based Images for Cloud-Type Recognition, J. Atmospheric Ocean. Technol., 25, 3–14, https://doi.org/10.1175/2007JTECHA959.1, 2008.

Clothiaux, E. E., Miller, M. A., Perez, R. C., Turner, D. D., Moran, K. P., Martner, B. E., Ackerman, T. P., Mace, G. G., Marchand, R. T., Widener, K. B., and others: The ARM millimeter wave cloud radars (MMCRs) and the active remote sensing of clouds (ARSCL) value added product (VAP), DOE Tech Memo. ARM VAP-002.1, 2001.

Davis, K., Zaitchik, B., Asa-Awuku, A., Bou-Zeid, E., Baidar, S., Boxe, C., Brewer, W. A., Chiao, S., Damoah, R., DeCarlo, P., Demoz, B., Dickerson, R., Giometto, M., Gonzalez-Cruz, J., Jensen, M., Kuang, C., Lamer, K., Li, X., Lombardo, K., Miles, N., Niyogi, D., Pan, Y., Peters, J., Ramamurthy, P., Peng, W., Richardson, S., Sakai, R., Waugh, D., and Zhang, J.: Coastal-Urban-Rural Atmospheric Gradient Experiment (CoURAGE) Science Plan, United States, https://doi.org/10.2172/2429693, 2024.

Dematties, D., Raut, B. A., Park, S., Jackson, R. C., Shahkarami, S., Kim, Y., Sankaran, R., Beckman, P., Collis, S. M., and Ferrier, N.: Let's Unleash the Network Judgment: A Self-Supervised Approach for Cloud Image Analysis, Artif. Intell. Earth Syst., 1–34, https://doi.org/10.1175/AIES-D-22-0063.1, 2023.

Dev, S., Lee, Y. H., and Winkler, S.: Color-Based Segmentation of Sky/Cloud Images From Ground-Based Cameras, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., 10, 231–242, https://doi.org/10.1109/JSTARS.2016.2558474, 2017.

Dong, X., Minnis, P., and Xi, B.: A Climatology of Midlatitude Continental Clouds from the ARM SGP Central Facility: Part I: Low-Level Cloud Macrophysical, Microphysical, and Radiative Properties, J. Clim., 18, 1391–1410, https://doi.org/10.1175/JCLI3342.1, 2005.

Dong, X., Xi, B., and Minnis, P.: A Climatology of Midlatitude Continental Clouds from the ARM SGP Central Facility. Part II: Cloud Fraction and Surface Radiative Forcing, J. Clim., 19, 1765–1783, https://doi.org/10.1175/JCLI3710.1, 2006.

Esteves, J., Cao, Y., Da Silva, N. P., Pestana, R., and Wang, Z.: Identification of clouds using an all-sky imager, in: 2021 IEEE Madrid PowerTech, 2021 IEEE Madrid PowerTech, 1–5, https://doi.org/10.1109/PowerTech46648.2021.9494868, 2021.

Fa, T., Xie, W., Wang, Y., and Xia, Y.: Development of an all-sky imaging system for cloud cover assessment, Appl. Opt., 58, 5516, https://doi.org/10.1364/AO.58.005516, 2019.

Feng, Y.-C., Lindenmaier, I., Deng, M., Wendler, T., Johnson, K., Matthews, A., Castro, V., and Rocque, M.: kazrmd (a1), https://doi.org/10.5439/1976091, 2011.

Fisher, R. A.: Frequency Distribution of the Values of the Correlation Coefficient in Samples from an Indefinitely Large Population, Biometrika, 10, 507, https://doi.org/10.2307/2331838, 1915.

Flynn, D.: asiskyimage (a1), https://doi.org/10.5439/1890632, 2024.

Flynn, D. and Morris, V.: tsicldmask (a1), https://doi.org/10.5439/1992208, 2000a.

Flynn, D. and Morris, V.: tsiskycover (b1), https://doi.org/10.5439/1992207, 2000b.

Flynn, D. and Morris, V.: tsiskyimage (a1), https://doi.org/10.5439/1992206, 2000c.

Free, M. and Sun, B.: Time-Varying Biases in U.S. Total Cloud Cover Data, J. Atmospheric Ocean. Technol., 30, 2838–2849, https://doi.org/10.1175/JTECH-D-13-00026.1, 2013.

Free, M., Sun, B., and Yoo, H. L.: Comparison between Total Cloud Cover in Four Reanalysis Products and Cloud Measured by Visual Observations at U.S. Weather Stations, J. Clim., 29, 2015–2021, https://doi.org/10.1175/JCLI-D-15-0637.1, 2016.

Grosvenor, D. P., Sourdeval, O., Zuidema, P., Ackerman, A., Alexandrov, M. D., Bennartz, R., Boers, R., Cairns, B., Chiu, J. C., Christensen, M., Deneke, H., Diamond, M., Feingold, G., Fridlind, A., Hünerbein, A., Knist, C., Kollias, P., Marshak, A., McCoy, D., Merk, D., Painemal, D., Rausch, J., Rosenfeld, D., Russchenberg, H., Seifert, P., Sinclair, K., Stier, P., van Diedenhoven, B., Wendisch, M., Werner, F., Wood, R., Zhang, Z., and Quaas, J.: Remote Sensing of Droplet Number Concentration in Warm Clouds: A Review of the Current State of Knowledge and Perspectives, Rev. Geophys., 56, 409–453, https://doi.org/10.1029/2017RG000593, 2018.

Kassianov, E., Long, C. N., and Ovtchinnikov, M.: Cloud Sky Cover versus Cloud Fraction: Whole-Sky Simulations and Observations, J. Appl. Meteorol., 44, 86–98, https://doi.org/10.1175/JAM-2184.1, 2005.

Kato, S., Loeb, N. G., Rutan, D. A., Rose, F. G., Sun-Mack, S., Miller, W. F., and Chen, Y.: Uncertainty Estimate of Surface Irradiances Computed with MODIS-, CALIPSO-, and CloudSat-Derived Cloud and Aerosol Properties, Surv. Geophys., 33, 395–412, https://doi.org/10.1007/s10712-012-9179-x, 2012.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y.: LightGBM: a highly efficient gradient boosting decision tree, in: Proceedings of the 31st International Conference on Neural Information Processing Systems, 3149–3157, 2017.

Koren, I., Remer, L. A., Kaufman, Y. J., Rudich, Y., and Martins, J. V.: On the twilight zone between clouds and aerosols, Geophys. Res. Lett., 34, 2007GL029253, https://doi.org/10.1029/2007GL029253, 2007.

Kuang, C., Giangrande, S., Serbin, S., Elsaesser, G., Gentine, P., Heus, T., Oue, M., Peters, J., Smith, J., Steiner, A., McComiskey, A., Jensen, M., Sedlacek, A., Kollias, P., Vogelmann, A., Morrison, H., Petters, M., and Turner, D.: Science Plan for the Deployment of the Third ARM Mobile Facility to the Southeastern United States at the Bankhead National Forest, Alabama (AMF3 BNF), United States, https://doi.org/10.2172/2280575, 2023.

Li, X., Wang, B., Qiu, B., and Wu, C.: An all-sky camera image classification method using cloud cover features, Atmospheric Meas. Tech., 15, 3629–3639, https://doi.org/10.5194/amt-15-3629-2022, 2022.

Loeb, N. G., Yang, P., Rose, F. G., Hong, G., Sun-Mack, S., Minnis, P., Kato, S., Ham, S.-H., Smith, W. L., Hioki, S., and Tang, G.: Impact of Ice Cloud Microphysics on Satellite Cloud Retrievals and Broadband Flux Radiative Transfer Model Calculations, J. Clim., 31, 1851–1864, https://doi.org/10.1175/JCLI-D-17-0426.1, 2018.

Logothetis, S.-A., Salamalikis, V., Wilbert, S., Remund, J., Zarzalejo, L. F., Xie, Y., Nouri, B., Ntavelis, E., Nou, J., Hendrikx, N., Visser, L., Sengupta, M., Pó, M., Chauvin, R., Grieu, S., Blum, N., Van Sark, W., and Kazantzidis, A.: Benchmarking of solar irradiance nowcast performance derived from all-sky imagers, Renew. Energy, 199, 246–261, https://doi.org/10.1016/j.renene.2022.08.127, 2022.

Long, C. N., Sabburg, J. M., Calbó, J., and Pagès, D.: Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images, J. Atmospheric Ocean. Technol., 23, 633–652, https://doi.org/10.1175/JTECH1875.1, 2006.

Maahn, M., Burgard, C., Crewell, S., Gorodetskaya, I. V., Kneifel, S., Lhermitte, S., Tricht, K. V., and Lipzig, N. P. M. van: How does the spaceborne radar blind zone affect derived surface snowfall statistics in polar regions?, J. Geophys. Res. Atmospheres, 119, 13,604-613,620, https://doi.org/10.1002/2014JD022079, 2014.

Mace, G., Marchand, R., Keywood, M., Protat, A., Humphries, R., Fiddes, S., McCluskey, C., Siems, S., Huang, Y., May, P., and Ma, P.-L.: Cloud and Precipitation Experiment at Kennaook (Cape-K) Science Plan, United States, https://doi.org/10.2172/1986088, 2023.

Mace, G. G. and Benson, S.: The Vertical Structure of Cloud Occurrence and Radiative Forcing at the SGP ARM Site as Revealed by 8 Years of Continuous Data, J. Clim., 21, 2591–2610, https://doi.org/10.1175/2007JCLI1987.1, 2008.

Manandhar, P., Temimi, M., and Aung, Z.: Short-term solar radiation forecast using total sky imager via transfer learning, Energy Rep., 9, 819–828, https://doi.org/10.1016/j.egyr.2022.11.087, 2023.

Mather, J.: Atmospheric Radiation Measurement (ARM) Management Plan, DOE ARM User Facility, Pacific Northwest National Laboratory, Richland, WA, United States, https://doi.org/10.2172/1253897, 2024.

Minnis, P., Trepte, Q. Z., Sun-Mack, S., Chen, Y., Doelling, D. R., Young, D. F., Spangenberg, D. A., Miller, W. F., Wielicki, B. A., Brown, R. R., Gibson, S. C., and Geier, E. B.: Cloud Detection in Nonpolar Regions for CERES Using TRMM VIRS and Terra and Aqua MODIS Data, IEEE Trans. Geosci. Remote Sens., 46, 3857–3884, https://doi.org/10.1109/TGRS.2008.2001351, 2008.

Morris, V.: Total Sky Imager (TSI) Handbook, https://doi.org/10.2172/1020716, 2005.

Morris, V. R.: Ceilometer Instrument Handbook, DOE/SC-ARM-TR-020, DOE Office of Science, Office of Biological and Environmental Research, https://doi.org/10.2172/1036530, 2016.

Park, S., Kim, Y., Ferrier, N. J., Collis, S. M., Sankaran, R., and Beckman, P. H.: Prediction of Solar Irradiance and Photovoltaic Solar Energy Product Based on Cloud Coverage Estimation Using Machine Learning Methods, Atmosphere, 12, 395, https://doi.org/10.3390/atmos12030395, 2021.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, J. Mach. Learn. Res., 12, 2825–2830, 2011.

Pfister, G., McKenzie, R. L., Liley, J. B., Thomas, A., Forgan, B. W., and Long, C. N.: Cloud Coverage Based on All-Sky Imaging and Its Impact on Surface Solar Irradiance, J. Appl. Meteorol., 42, 1421–1434, https://doi.org/10.1175/1520-0450(2003)042%253C1421:CCBOAI%253E2.0.CO;2, 2003.

Raut, B. A., Muradyan, P., Sankaran, R., Jackson, R. C., Park, S., Shahkarami, S. A., Dematties, D., Kim, Y., Swantek, J., Conrad, N., Gerlach, W., Shemyakin, S., Beckman, P., Ferrier, N. J., and Collis, S. M.: Optimizing cloud motion estimation on the edge with phase correlation and optical flow, Atmospheric Meas. Tech., 16, 1195–1209, https://doi.org/10.5194/amt-16-1195-2023, 2023.

Riley, E. A., Kleiss, J. M., Riihimaki, L. D., Long, C. N., Berg, L. K., and Kassianov, E.: Shallow cumuli cover and its uncertainties from ground-based lidar–radar data and sky images, Atmospheric Meas. Tech., 13, 2099–2117, https://doi.org/10.5194/amt-13-2099-2020, 2020.

Sarangi, S. K., Sarangi, C., Patel, N., Madhavan, B. L., Ningombam, S. S., Ravindra, B., and Ratnam, M. V.: Cloud fraction estimation using random forest classifier on sky images, Atmospheric Meas. Tech., 18, 5637–5648, https://doi.org/10.5194/amt-18-5637-2025, 2025.

Savoy, F. M., Dev, S., Lee, Y. H., and Winkler, S.: Geo-referencing and stereo calibration of ground-based Whole Sky Imagers using the sun trajectory, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), IGARSS 2016 - 2016 IEEE International Geoscience and Remote Sensing Symposium, 7473–7476, https://doi.org/10.1109/IGARSS.2016.7730949, 2016.

Schreder-CMS: New All Sky Imaging Solutions, http://www.schreder-cms.com/en_pdf/ASI-16_InfoBroch.pdf, 2017.

Silber, I., Verlinde, J., Eloranta, E. W., and Cadeddu, M.: Antarctic cloud macrophysical, thermodynamic phase, and atmospheric inversion coupling properties at McMurdo Station. Part I: Principal data processing and climatology, J. Geophys. Res. Atmospheres, 123, 6099–6121, https://doi.org/10.1029/2018JD028279, 2018.

Silber, I., Flynn, D., Cromwell, E., and Ermold, B.: All-sky imager cloud mask, https://doi.org/10.5439/3005849, 2026a.

Silber, I., Flynn, D., Cromwell, E., and Ermold, B.: asiskycover (b0), https://doi.org/10.5439/3005850, 2026b.

Silber, I., Flynn, D., Cromwell, E., and Ermold, B.: asiskycover (b1), https://doi.org/10.5439/1890629, 2026c.

Sisterson, D. L., Peppler, R. A., Cress, T. S., Lamb, P. J., and Turner, D. D.: The ARM Southern Great Plains (SGP) Site, Meteorol. Monogr., 57, 6.1-6.14, https://doi.org/10.1175/AMSMONOGRAPHS-D-16-0004.1, 2016.

Song, S., Yang, Z., Goh, H., Huang, Q., and Li, G.: A novel sky image-based solar irradiance nowcasting model with convolutional block attention mechanism, Energy Rep., 8, 125–132, https://doi.org/10.1016/j.egyr.2022.02.166, 2022.

Verlinde, J., Zak, B. D., Shupe, M. D., Ivey, M. D., and Stamnes, K.: The ARM North Slope of Alaska (NSA) Sites, Meteorol. Monogr., 57, 8.1-8.13, https://doi.org/10.1175/AMSMONOGRAPHS-D-15-0023.1, 2016.

Wang, M., Zhou, S., Yang, Z., and Liu, Z.: CloudA: A Ground-Based Cloud Classification Method with a Convolutional Neural Network, J. Atmospheric Ocean. Technol., 37, 1661–1668, https://doi.org/10.1175/JTECH-D-19-0189.1, 2020.

Wang, Y. and Zhao, C.: Can MODIS cloud fraction fully represent the diurnal and seasonal variations at DOE ARM SGP and Manus sites?, J. Geophys. Res. Atmospheres, 122, 329–343, https://doi.org/10.1002/2016JD025954, 2017.

Werkmeister, A., Lockhoff, M., Schrempf, M., Tohsing, K., Liley, B., and Seckmeyer, G.: Comparing satellite- to ground-based automated and manual cloud coverage observations – a case study, Atmospheric Meas. Tech., 8, 2001–2015, https://doi.org/10.5194/amt-8-2001-2015, 2015.

Widener, K. B., Bharadwaj, N., and Johnson, K.: Ka-Band ARM Zenith Radar (KAZR) Instrument Handbook, ARM-TR-106, DOE Office of Science, Office of Biological and Environmental Research, https://doi.org/10.2172/1035855, 2012.

Wood, R., Wyant, M., Bretherton, C. S., Rémillard, J., Kollias, P., Fletcher, J., Stemmler, J., Szoeke, S. de, Yuter, S., Miller, M., Mechem, D., Tselioudis, G., Chiu, J. C., Mann, J. A. L., O'Connor, E. J., Hogan, R. J., Dong, X., Miller, M., Ghate, V., Jefferson, A., Min, Q., Minnis, P., Palikonda, R., Albrecht, B., Luke, E., Hannay, C., and Lin, Y.: Clouds, Aerosols, and Precipitation in the Marine Boundary Layer: An Arm Mobile Facility Deployment, Bull. Am. Meteorol. Soc., 96, 419–440, https://doi.org/10.1175/BAMS-D-13-00180.1, 2015.

Xie, W., Liu, D., Yang, M., Chen, S., Wang, B., Wang, Z., Xia, Y., Liu, Y., Wang, Y., and Zhang, C.: SegCloud: a novel cloud image segmentation model using a deep convolutional neural network for ground-based all-sky-view camera observation, Atmospheric Meas. Tech., 13, 1953–1961, https://doi.org/10.5194/amt-13-1953-2020, 2020.

Zhang, D., Ermold, B., and Morris, V.: ceil, https://doi.org/10.5439/1181954, 1996.

Zhao, W., Marchand, R., and Fu, Q.: The diurnal cycle of clouds and precipitation at the ARM SGP site: Cloud radar observations and simulations from the multiscale modeling framework, J. Geophys. Res. Atmospheres, 122, 7519–7536, https://doi.org/10.1002/2016JD026353, 2017.