Automated nighttime cloud detection using keograms when aurora is present

Alex English[1], David J Stuart[1], Donald L. Hampton[2], and Seebany Datta-Barua[1]

[1]Illinois Institute of Technology
[2]University of Alaska Fairbanks

December 22, 2022

Abstract

We present a metric for detecting clouds in auroral all-sky images based on single-wavelength keograms made with a collocated meridian spectrograph. The coefficient of variation, the ratio of the sample standard deviation to the sample mean taken over viewing angle, is the metric for cloud detection. After calibrating and flat-field correcting keogram data, then excluding dark sky intervals, the effectiveness of the coefficient of variation as a detector is tested compared to true conditions as determined by Advanced Very High Resolution Radiometer (AVHRR) satellite imagery of cloud cover. The cloud mask, an index of cloud cover, is selected at the corresponding nearest time and location to the site of a meridian spectrograph at Poker Flat Research Range (PFRR). We use events that are completely cloud-free or completely cloudy according to AVHRR to compute the false alarm and missed detection statistics for the coefficient of variation of the greenline 557.7 nm emission and of the redline 630.0 nm emission. For training data of the years 2014 and 2016, we find a greenline threshold of 0.51 maximizes the percent of events correctly identified at 75%. When applied to testing data of the years 2015 and 2017, the 0.51 threshold yields an accuracy of 77%. There is a relatively shallow and wide minimum of mislabeled events for thresholds spanning about 0.2 to 0.8. For the same events, the minimum is narrower for the redline, spanning roughly 0.3-0.5, with a threshold of 0.46 maximizing detector accuracy at 78-79%.

Automated Nighttime Cloud Detection using Keograms when Aurora is Present

Alex English[1], David J. Stuart[1], Donald L. Hampton[2], and Seebany Datta-Barua[1]

[1]Illinois Institute of Technology
[2]University of Alaska Fairbanks

Key Points:

- Keogram coefficient of variation is used to determine if the sky is cloudy or clear, and verified with NOAA satellite imagery from 2014-2017
- At 557.7 nm, a 0.51 threshold gives 75% accuracy but is comparable to results between 0.2-0.8
- At 630.0 nm, 0.46 is 78% accurate and comparable within 0.3-0.5

Corresponding author: Seebany Datta-Barua, `sdattaba@iit.edu`

Abstract

We present a metric for detecting clouds in auroral all-sky images based on single-wavelength keograms made with a collocated meridian spectrograph. The coefficient of variation, the ratio of the sample standard deviation to the sample mean taken over viewing angle, is the metric for cloud detection. After calibrating and flat-field correcting keogram data, then excluding dark sky intervals, the effectiveness of the coefficient of variation as a detector is tested compared to true conditions as determined by Advanced Very High Resolution Radiometer (AVHRR) satellite imagery of cloud cover. The cloud mask, an index of cloud cover, is selected at the corresponding nearest time and location to the site of a meridian spectrograph at Poker Flat Research Range (PFRR). We use events that are completely cloud-free or completely cloudy according to AVHRR to compute the false alarm and missed detection statistics for the coefficient of variation of the greenline 557.7 nm emission and of the redline 630.0 nm emission. For training data of the years 2014 and 2016, we find a greenline threshold of 0.51 maximizes the percent of events correctly identified at 75%. When applied to testing data of the years 2015 and 2017, the 0.51 threshold yields an accuracy of 77%. There is a relatively shallow and wide minimum of mislabeled events for thresholds spanning about 0.2 to 0.8. For the same events, the minimum is narrower for the redline, spanning roughly 0.3-0.5, with a threshold of 0.46 maximizing detector accuracy at 78-79%.

Plain Language Summary

Clouds in the sky are a problem for scientists trying to view space beyond. For upper atmospheric scientists, clouds can obscure or scatter auroral light in all-sky images (ASI), making it hard to identify, locate, and track auroral shapes. This paper shows a way to simply and automatically detect clouds using a north-to-south line scan of a single color of light from the sky over time, known as a keogram. We compute the ratio of the variation in pixel intensity to the average pixel intensity, for each north-to-south scan. Excluding dark sky periods, a large ratio means that the sky is cloudless, and a small ratio that the sky is cloudy. We find the method works with about a 75-80% correct rate using red or green auroral light. With this method we can eliminate data during cloudy conditions for any auroral studies that require clear sky conditions.

1 Introduction

Aurorae occur at the polar regions of the Earth, and are colloquially known as the northern and southern lights. These visual light emissions result from the interactions between charged particles in the Earth's magnetosphere and upper atmospheric species. Because of their relationship to interactions with the magnetosphere, researchers have been interested in classifying types of aurorae (M. T. Syrjäsuo & Donovan, 2004) and correlating them with other events. Researchers have noted that the passage of aurorae are associated with radio frequency scintillations at high latitudes (Semeter et al., 2017; Mrak et al., 2018; Loucks et al., 2017; D. L. Hampton et al., 2013). The quality of ground-based auroral images is limited by the presence of clouds in the sky. For individual case studies, researchers can visually inspect and often determine by eye the presence of clouds. However, this is not practical for large surveys of events.

Auroral scientists are not unique in being interested in detecting the presence or absence of clouds. For many practical and scientific applications, satellite imagery at various wavelengths is a standard tool for coverage spanning continent-scale areas. Multi-decade clear sky (i.e., not cloudy) identification can be done by non-optical means of comparing the measured irradiance to top of the atmosphere irradiance, compared to a clear-sky transmittance threshold (Correa et al., 2022). Such studies are longer term or generally lower resolution than might be needed for nightly auroral studies at a single site. For local conditions, ground-based methods can provide measures of cloud cover for day or night.

Many of the daytime methods leverage or are interested in solar illumination. Clear sky detection based on broadband irradiance is one avenue of cloud detection in use for decade-scale studies (Long & Ackerman, 2000). At optical wavelengths, low-cost cameras may be used by solar power station operators who want an automated method for estimating or forecasting power generation (Alonso-Montesinos, 2020). Daylight polarization can be used to determine clear sky versus cloudy sky, and the optical thickness of the cloud layer, if present (W. Li et al., 2022). A number of researchers have successfully developed methods for sorting cloud data automatically using the red and blue intensity relationships of all sky images, total sky imagers, or whole sky cameras (Q. Li et al., 2011; Long et al., 2006). Other groups have developed hybrid or adaptive thresholding algorithms (F.-F. Li et al., 2022). Another method was developed using three cloud features to categorize the ASC images into four cloud cover categories, rather than image threshold techniques (X. Li et al., 2022). These studies use daytime images illuminated by sunlight, and may be interested in classifying cloud types or regions of the sky with cloud cover. For auroral investigations, we are interested in tools usable at night and less interested in cloud types.

Recent interest in machine learning has shown that aurorae can be classified with trained algorithms (Clausen & Nickisch, 2018). One of the classification categories in this process is "cloudy" (Sado et al., 2022). Astronomers have also used machine learning methods to determine cloud cover at night for protecting telescope equipment (Mommert, 2020). While these methods hold promise, they can be computationally expensive and time consuming for training and validating at a single site for multiyear studies, necessitating a method that provides sorting of a multitude of night-time images in an efficient and consistent manner. One such method was used as part of an auroral detection and tracking method, in which aurorae were detected using the ratio of maximum to mean brightness of an all-sky image, after using synoptic cloud index measurements to eliminate cloudy periods (M. Syrjäsuo & Donovan, 2002). In this work we are interested in leveraging the nighttime single-wavelength one-dimensional images themselves to detect and discard thex cloudy intervals in the night sky, without need for separate cloud measurement.

In image processing, blurring and other distortions in a received image are modeled as convolution of a kernel with an original signal. The distortions of a camera itself may be characterized as a convolution of a point-spread function defining the camera's characteristics. In astronomy, the point spread function of the camera can often be determined using known stars. If the point-spread function is known, the image can be deconvolved to recover the original signal. For example, a theoretical determination of the point spread function due to clouds and fog for imaging objects 20 km from the imager was conducted by (Jaruwatanadilok et al., 2003) based on radiative transfer theory. In some disciplines, the point spread function may be recovered via blind deconvolution techniques. In this work, the presence of a filtering function due to atmospheric scattering is the focus, rather than defining the precise form of it. The concept of atmospheric filtering is mentioned by Guo et al. (2022) who investigated neural network-based restoration of images distorted by atmospheric turbulence. We do not need to go so far as to restore images blurred by clouds in a large multi-year database of auroral imagery, but we can leverage the effect of clouds on one-dimensional single-wavelength images over time to determine their presence.

In this work, we present a simple metric for efficiently and automatically detecting clouds if auroral light is present. This method is intended for subsequent automation of auroral all-sky image analysis. Section 2 motivates and introduces our proposed detection metric. Section 3 describes the method and data sets used to test and validate our proposed detection technique, with details on pre-processing in Appendix A. Section 4 shows the key results, and conclusions are summarized in Section 5.

2 Conceptual approach

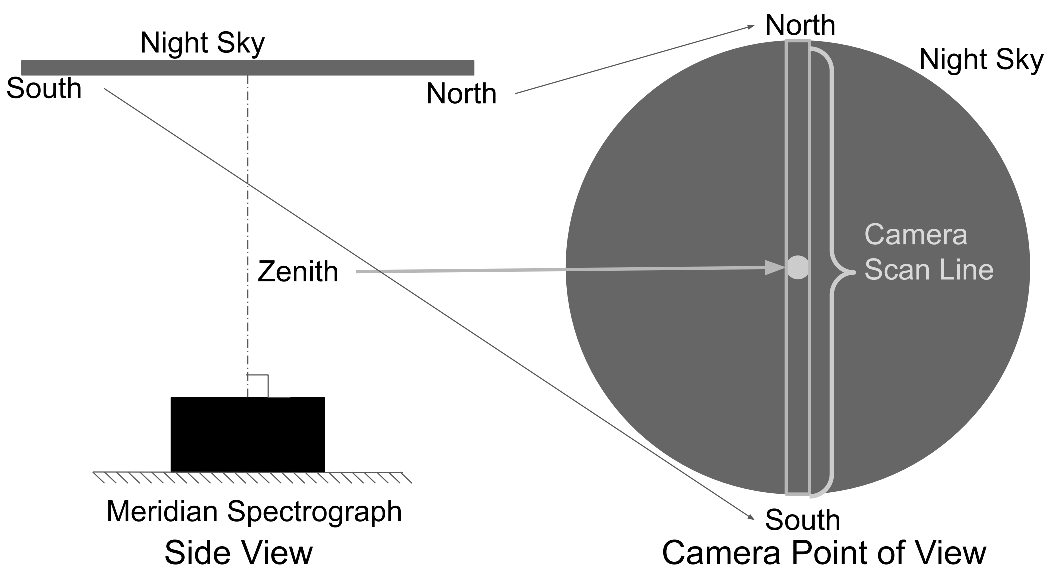

Figure 1. Schematic of keogram imaging system. The left shows a side view of a meridian spectrograph looking up local zenith and the right shows a view of the night sky from the perspective of a camera as the meridian spectrograph takes a one-pixel-wide scan from horizon to horizon through local zenith.

A keogram is a time sequence of one-dimensional images taken over the course of a night. A keogram may be taken with a meridian spectrograph or constructed from the field-of-view of an all-sky imager by extracting one subset of pixels. The diagram in Figure 1 illustrates a side view of an imaging system (left) and a sky view of an all-sky imager's field of view (right). The meridian spectrograph takes one-pixel-wide images of the sky at intervals throughout the course of the night. The pixel intensities are recorded as a function of the elevation angle from the northern to southern horizon passing through local zenith. At auroral latitudes a north-to-south scan is most likely to sample any auroral light because of the orientation of the auroral oval generally gives aurorae that are oriented east-west.

A sample keogram (calibrated and corrected, as described in later sections) taken at one wavelength is shown in Figure 2a. The x axis is time, and each column is a line-scan image from north (0 deg) to south (180 deg) of light intensity (Rayleighs, shown by color) taken at one instant. Our objective is to use the keogram to detect whether clouds are present or not at each moment. By inspection we observe that Interval 1 identified in Figure 2a corresponds to a dark sky with no aurora. A plot of the intensity as a function of elevation at the example instant identified with a red vertical line is shown in Figure 2b. The intensities are uniformly low at 04:00 UT. A histogram of these intensities over all angles at this instant is then shown in Figure 2c. The histogram of this snapshot taken over all viewing angles has a small both sample mean μ and standard deviation σ.

Interval 2 identified in Figure 2a contains a segment of an auroral band in the northern part of the sky. For this example time, the intensity as a function of viewing angle is shown in Figure 2d, consisting of one narrow region of high intensity at the viewing angle to the aurora. The sky is clear because we can see the narrow angular extent of the band of the aurora, and is verified by manually viewing an all-sky image. The histogram is shown in Figure 2e, and there is a spread of intensities due to distinctly brighter or dimmer auroral features.

Interval 3 of Figure 2a corresponds to a period during which there are aurorae, but the presence of clouds has dimmed and scattered the auroral light (again, apparent by manually viewing the all-sky image). Clouds smear the light intensities spatially to give a more uniform brightness at all viewing angles, as shown in Figure 2f. As a result, the distribution of keogram intensities is narrowly clustered around a non-zero mean.

Cloud cover has the effect of blurring the auroral light in the keogram. A commonly used image processing concept is useful here. Images taken are often post-processed to reduce noise or smooth out other unwanted effects by filtering. Comparing Figures 2d and 2f, we note that clouds between the auroral source and the imager have the effect

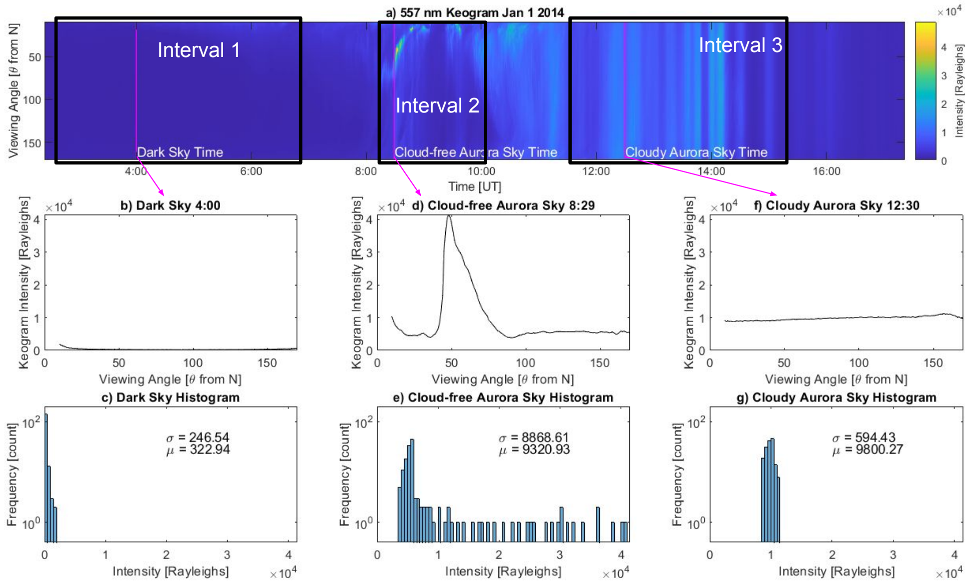

Figure 2. (a) Keogram for 01 Jan 2014 for 557.7 nm wavelength with three sky conditions highlighted: (1) Dark sky (2) Cloud-free time with aurora, and (3) Cloudy aurora time. The red lines in each interval mark example timestamps for the remaining plots. (b) Intensity versus viewing angle and (c) histogram of keogram intensity for the dark sky example time. (d) Intensity versus viewing angle and (e) histogram of the intensities at the cloud-free aurora time. (f) Intensity versus viewing angle and (g) histogram of the intensities at the cloudy aurora time.

of smoothing out the intensities spatially, and effectively act as an imaging filter that blurs the image. The mathematical process of filtering is given by convolution of a filter that modifies an original signal. Clouds in the sky act as a filter that, convolved with light sources that would otherwise be present in a keogram at a cloud-free instant, produces a smoothed set of intensities received at the ground. In the case of the example shown in Figures 2f-2g, the filtered signal results in a histogram whose distribution is narrowed, as all viewing angles have similar intensity.

At each instant t the keogram Y is a one-dimensional image of received intensities at a single wavelength over N discrete spatial coordinate elements θ_n. Assuming the keogram instrument is calibrated for uniform gain in all directions and undesired broadband and noise sources (e.g., from light pollution) have largely been removed, the residuals ϵ in the corrected keogram Y may be assumed to be zero-mean with a standard deviation of σ_ϵ. The intensity Y at a given wavelength in this case may be written as:

$$\begin{aligned} Y(t,\theta_n) &= (a * g)(t,\theta_n) + \epsilon \qquad (1)\\ &= \sum_{m=-N}^{N} a[n-m]g[m] + \epsilon_n \qquad (2)\end{aligned}$$

where a represents any light sources behind the clouds, i.e., aurorae; g the filtering function (sometimes called the kernel or point-spread function) of the clouds that scatters the light source, the symbol * represents the convolution operation, and ϵ is a random variable representing the residuals and noise after calibration. Equation 2 defines convolution for discrete signals over viewing angle at time t. The signal a at N discrete angles can be zero-padded for the convolution operation.

For a cloud-free sky (subscript "cf") we can represent the cloud kernel as a Kronecker delta function $g_{cf}[m] = \delta_{0m}$, which does no spreading of the intensity, so the summation simplifies as:

$$\begin{aligned} Y_{cf}(t,\theta_n) &= \sum_{m=-N}^{N} a[n-m]g_{cf}[m] + \epsilon_n \qquad (3)\\ &= a[n] + \epsilon_n \qquad (4)\end{aligned}$$

For zero-mean noise, the mean intensity $\bar{Y}$ over all viewing angles θ_n at time t is the mean intensity $\bar{a}$ of a over all elevations:

$$\begin{aligned} \bar{Y}_{cf}(t) &= \frac{1}{N}\sum_{n=1}^{N} a[n] + \epsilon_n \qquad (5)\\ &= \bar{a}(t) \qquad (6)\end{aligned}$$

The sample variance would be the sum of the variance σ_a^2 of a over all elevations and of the noise, assuming the light sources and noise to be independent, which can be seen by substituting Eqs. 4 and 6 into Eq. 7:

$$\begin{aligned} \sigma_{cf}^2 &= \frac{1}{N-1}\sum_{n=1}^{N}\left(Y(t,\theta_n) - \bar{Y}(t)\right)^2 \qquad (7)\\ &= \frac{1}{N-1}\sum_{n=1}^{N}\left(a[n] + \epsilon_n - \bar{a}\right)^2 \qquad (8)\\ &= \sigma_a^2 + \sigma_\epsilon^2 \qquad (9)\end{aligned}$$

While a specific cloud kernel is not known and might depend on the type of cloud, we can imagine the extreme case of a cloud that spreads the intensity evenly across all N elevations, whose filter would be $g_c[n] = 1/N$. In this case, the intensity would be:

$$\begin{aligned} Y_c(t,\theta_n) &= \sum_{m=1}^{N} a[n-m]g_c[n] + \epsilon_n \qquad (10)\\ &= \bar{a}(t) + \epsilon_n \qquad (11)\end{aligned}$$

The angle-averaged intensity would be $\bar{Y} = \bar{a}$ as in the cloud-free case. However, the variance with angle would be given by:

$$\sigma_c^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(Y(t,\theta_n) - \bar{Y}(t)\right)^2 \tag{12}$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}(\cancel{\bar{a}(t)} + \epsilon_n - \cancel{\bar{a}(t)})^2 \tag{13}$$

$$= \sigma_\epsilon^2 \tag{14}$$

leaving only the variance of the noise.

However, if the sky is dark, there is no light source to be blurred, meaning $a = 0$, the cloud kernel whether g_c or g_{cf} has little effect on the intensity Y_d of a dark sky.

$$Y_d(t,\theta_n) = (0 * g)(t,\theta_n) + \epsilon \tag{15}$$

$$= \epsilon_n \tag{16}$$

$$\bar{Y}_d(t) = \bar{\epsilon}(t) = 0 \tag{17}$$

$$\sigma_d^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(Y(t,\theta_n) - \bar{Y}(t)\right)^2 \tag{18}$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}(\epsilon_n - \bar{\epsilon}(t))^2 \tag{19}$$

$$= \sigma_\epsilon^2 \tag{20}$$

The mean and variance of a dark clear sky would be indistinguishable from that of a dark cloudy sky. On the other hand, they are not of interest for auroral studies. For this reason we exclude dark sky intervals such as Interval 1 from consideration, by setting a minimum mean value $\bar{Y}$ of the samples that must be exceeded.

Given that there is auroral light in the keogram at time t, our objective is to determine whether the image at that time is cloudy or not. The coefficient of variation $c(t)$ is the sample standard deviation σ of $Y(t)$ normalized by the mean $\bar{Y}$, shown in Eq. 21. It is a measure of how much variation there is at each time over all elevation angles θ of the keogram.

$$c(t) \equiv \frac{\sigma(t)}{\bar{Y}(t)} \tag{21}$$

The example relationship between (a) a keogram, (b) its standard deviation, (c) mean, and (d) coefficient of variation can be seen in Figure 3. In the cloud-free aurora-present case (Interval 2), $c = \sigma_{cf}/\bar{a} \sim 1$, but for the cloudy sky case (Interval 3) $c \approx \sigma_c/\bar{a} << 1$. The dark sky case (Interval 1) also has $c \approx \sigma_\epsilon/\bar{\epsilon} \sim 1$, but is artificially large because $\bar{Y}$ is so low. After filtering out dark-sky intervals, for which a small $\bar{Y}$ would artificially inflate c, we propose the coefficient of variation as a metric for detecting cloudy auroral-lit intervals in keograms (i.e., distinguishing Interval 2 from 3 in Figure 2.

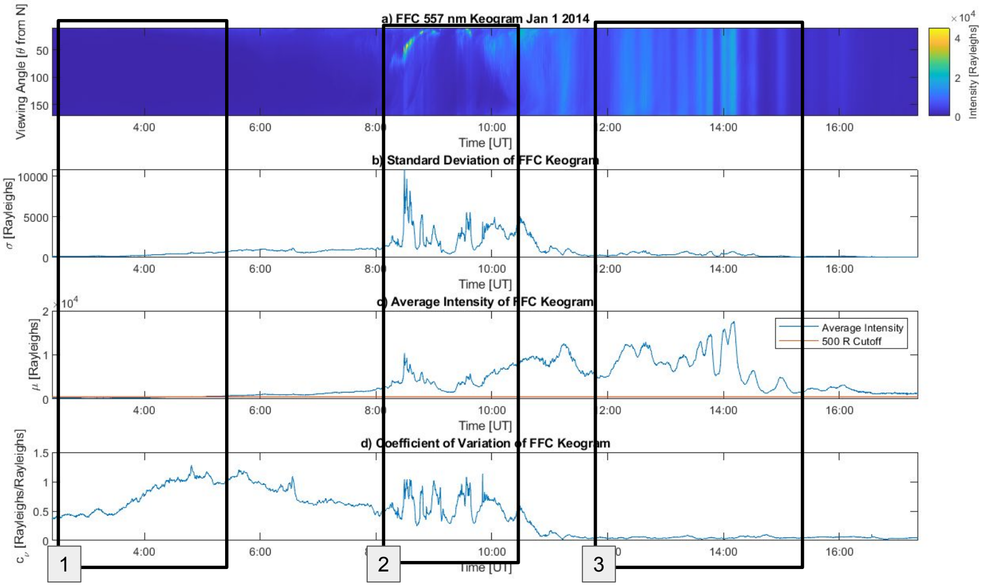

Figure 3. (a) Keogram Y of 1 January 2014 pre-processed as described in Appendix A with the corresponding sample (b) standard deviation, (c) mean, and (d) coefficient of variation c with specific times highlighted to explain what the keogram looks like in various sky conditions: 1) Dark sky 2) cloud-free with aurora 3) cloudy with aurora.

3 Method

In order to test the effectiveness of the coefficient of variation as a detection metric for clouds, we use a database of keograms collected at Poker Flat Research Range (PFRR), Alaska, from 2014-2017 (source listed in Open Research Section). After calibrating and correcting the keograms, we compute the coefficient of variation for each over time and compare them to NOAA satellite image-derived cloud mask data over PFRR. The satellite imagery provides a truth reference for whether clouds were present or not. We use standard detection theory to identify the distributions of coefficient of variation for two populations (cloudy and cloud-free). We test different thresholds of the detection metric to compute the number of events that are correctly identified or mislabeled. We use the events in years 2014 and 2016 as the training data, to find a threshold that produces the fewest mislabeled events (missed detections and false alarms), which is maximizes the accuracy (correct detections and true negatives). Then we apply the same threshold for keogram data for the years 2015 and 2017, to test whether the threshold found yields reproducible results on different data.

At PFRR, a meridian spectrograph operates with filters at 6 wavelengths to record keograms from sunset to sunrise nightly, except during the summer months which have near-continuous daylight. The keogram image intensities are given in camera counts at 6 different wavelengths: 427.8 nm, 486.1 nm, 520 nm, 557.7 nm, 630.0 nm, and 670 nm. Intensities at each wavelength are accumulated over approximately 12.5-second intervals. The wavelengths used in this study for computing the coefficient of variation are 557.7 nm (green) and, separately, 630.0 nm (red). The processing of the raw data, conversion to intensity in Rayleighs, removal of background light, and flat-field correction to produce $Y(t, \theta_n)$ are described in Appendix A.

Figure 3a represents the flat-field corrected keogram Y (identical to Figure 2a). By inspection Interval 1 has dark sky with no aurora present. Dark sky times are defined using the mean intensity of the keogram $\bar{Y}(t)$ at that time point, shown in Figure 3c. The average intensity is very low when there is no aurora in the sky in Interval 1 in Figure 3c, and increases as aurora becomes present. We choose 500 R in the 557.7 nm keogram (marked with a red line in Figure 3c) as the threshold to automatically determine darkness. If $\bar{Y}(t) < 500$ R, then the sky is determined to be dark and thus cannot be used to determine cloud presence. The dark sky test based on the green emission is used whether the red or green cloud detection metric is used.

The National Oceanic and Atmospheric Association (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and High-resolution Infra-Red Sounder (HIRS) Pathfinder Atmospheres Extended (PATMOS-x) Climate Data Record (CDR) database is used as the reference true cloud condition. The AVHRR+HIRS Cloud Properties in the PATMOS-x CDR provides data for cloud properties, brightness, and temperatures collected by the AVHRR and HIRS instruments on board the NASA Polar Operational Environmental Satellites (POES) NOAA-15, NOAA-18, and NOAA-19, and European MetOp-2 platforms (Oceanic & Administration, n.d.).

Within the PATMOS-x CDR, the cloud mask is an index describing how cloudy the sky is at a given geographic latitude, longitude, and time. The cloud mask is on a scale of 0-3 as follows: 0 for clear, 1 for probably clear, 2 for probably cloudy, 3 for cloudy. An example of the cloud mask data over Alaska is shown in Figure 4. These data are used as the truth reference, to train and test the keogram cloud detection method.

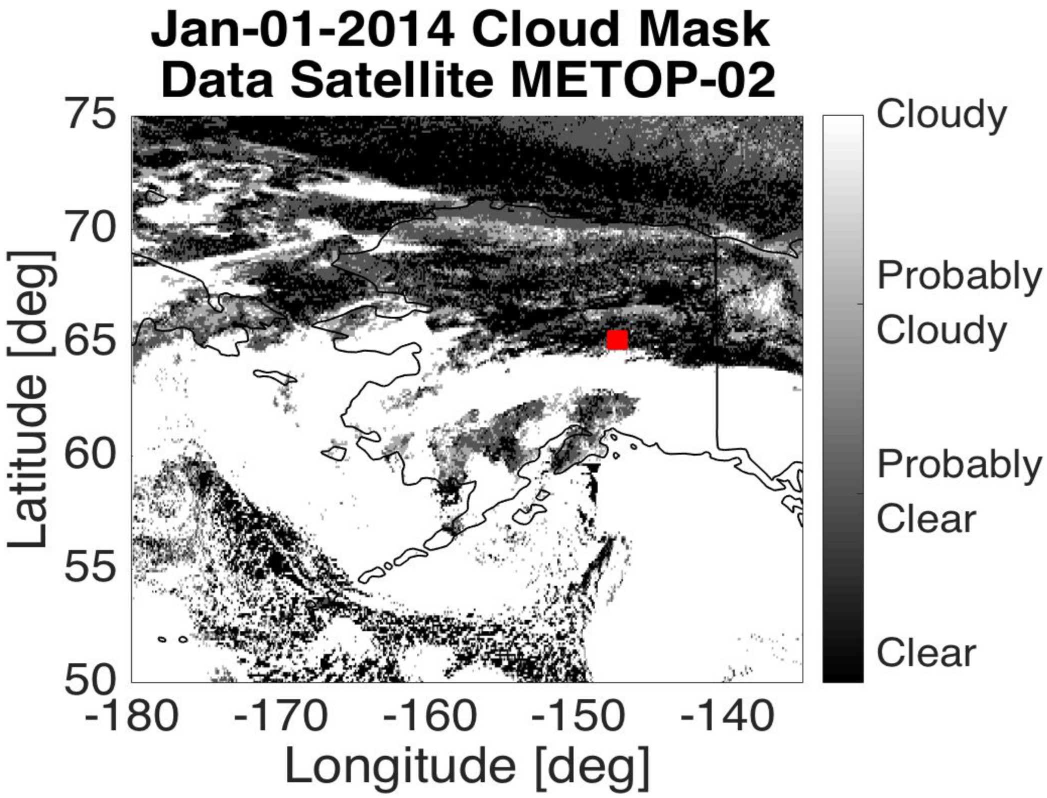

Figure 4. NOAA cloud mask data over Alaska with Poker Flat Research Range marked with a red square.

Provisional cloud mask files, available daily for 2014 through the first half of 2017, are used. From each cloud mask file, the times, cloud mask, and latitude and longitude of points within 8 km of PFRR are saved.

For each NOAA data point, we determine the keogram 557.7 nm snapshot that is closest in time and at least within 20 s of the time the keogram data was taken. Because satellite data are recorded imaging over a swath, if there is more than one NOAA data point within 20 s of the same keogram timestamp, the NOAA pixel that is geographically closest to PFRR is used, so that there is only one NOAA cloud mask associated with one keogram timestamp.

The true condition is determined from the NOAA cloud mask, corresponding to 0 when cloud-free, and 3 when cloudy. The cloud masks of 1 and 2 are not considered in this work. The keogram cloud categorization is determined from the coefficient of variation c being either less than the threshold (cloudy) or greater than or equal to the threshold (cloud-free). Each coefficient of variation and cloud mask pair are categorized into one of four groups: 1) the keogram-derived coefficient of variation c and NOAA cloud mask both indicate cloud-free conditions; 2) the keogram and NOAA cloud mask both indicate cloudy; 3) the keogram categorization predicts cloud-free but the NOAA categorization shows that the sky is cloudy (missed detection); and 4) the keogram categorization predicts cloudy and the NOAA categorization cloud-free (false alarm).

The training data for keogram-based cloud detection are all cloud masks over PFRR that have a 557.7 nm keogram measurement present at the corresponding time, in 2014 and 2016. We find a threshold with the lowest percent of mislabeled events (both missed detections and false alarms), starting from a threshold of $c = 0.01$ incrementing by 0.01 to $c = 1$. We then apply the best threshold found to the testing data of 2015 and 2017, and compute the mislabeling rates for that set of events. The accuracy of the detector is defined as 100 percent minus the mislabeled percent.

4 Results

In the training data of 2014 and 2016, there are a total of 794 events for which there are cloud mask and keogram data at the corresponding times and location. Of these, 434 of the events have cloud mask of 0 or 3 (cloudy or clear). Among these 434 events, 295 of the events are bright enough to exceed the dark sky threshold. The percentage of events mislabeled (the sum of false alarms and missed detections) as a function of the 557.7 nm keogram coefficient of variation threshold is shown in Figure 5a. The plot shows that the threshold with the lowest percent of events that are mislabeled is 0.51, with about 21% of events mislabeled. For about 13% of the events, NOAA cloud mask indicates clear sky but the keogram coefficient of variation indicates cloudy. For 8% of the events the keogram is cloud-free but the cloud mask indicates cloudy. The percent for which both the cloud mask and keogram agree the sky is cloud-free is 26%. For about 53% of the events they both indicate cloudy conditions. Histograms plotted in Figure 5b show the distribution of the coefficient of variation for cloudy events (blue) and for clear sky events (red). A vertical red line marks the threshold of 0.51. The blue bars exceeding that threshold are the ones that are missed detections of clouds. The red bars below the threshold line are the false alarms, in which using the detector c value indicates cloudy sky but the true condition is clear. On Figure 5a, we can see that above a threshold of about 0.2, there is a wide shallow minimum area up to about 0.8. This indicates that the greenline detection statistics may not be very sensitive to the specific choice of threshold within this range.

For the testing data set of 2015 and the first half of 2017, there are a total of 529 events, 266 of which have a cloud mask of 0 or 3 (cloudy or clear, respectively). Of these events, 196 of them are above the dark sky threshold. We compute the percent of events mislabeled as either false alarms or missed detections for a range of thresholds, as shown in Figure 5c. The threshold of 0.51, which was found to yield the lowest mislabeling rate with the training data, is marked with a red circle. For this data set, while 0.51 is near a local minimum, it is not the global minimum. For the testing data, 25% of the events are mislabeled (with 10% identified as cloudy with the cloud mask but detected cloud-free with our method, and 15% cloud-free but determined to be cloudy by our method). The histograms of the coefficient of variation for cloud-free events (red) and cloudy events (blue) are shown in Figure 5d, with the 0.51 threshold marked with a vertical line. There are fewer events in this data set than the training data, and this appears in the histograms with fewer counts in the modal intervals than in the training data, as well as some bins, e.g., in the clear distribution at $c = 1.2$ that are completely unpopulated. This sampling likely accounts for the appearance of multiple local minima in Figure 5c. For this data set the global minimum occurs at $c = 0.37$ with a 23% mislabeled event rate. This is comparable to the mislabeled rate for the 0.51 threshold. The testing data set has one-third fewer events for assessment than the training set. We expect that with more complete sampling, e.g., including the second half of 2017 for which at this time provisional cloud mask data are not yet available, we would likely again find a wide region of minimum mislabeling error spanning from around 0.2 to 0.8.

For the same set of training and testing events, in which the dark sky has been eliminated using the requirement that the average green emission exceed 500 R, we test the effectiveness of using the 630.0 nm emission coefficient of variation. The training mislabeling percentage results and histograms are shown in Figures 6a and b. The testing results are shown in Figures 6c and d. The threshold yielding the minimum combined rate of false alarms and missed detections of about 21% using 630.0 nm is 0.46. Apply-ing the same threshold to the testing data yields a comparable 22% mislabeled rate. In the case of the redline mislabeling statistics (Figures 6a and c), the minimum percentage mislabeled region does not appear to be quite as wide and shallow as for the green emission, only dropping below 25% above a threshold of about 0.3, and increasing close to monotonically for thresholds higher than about 0.5. It is possible that for a given set of events, the redline emission has the potential to improve accuracy by a few percent relative to the greenline emission, but may be more sensitive to choice of threshold.

An effective detector metric is one that separates the distributions between two different populations most widely. We demonstrated the coefficient of variation metric using the greenline emission, which are associated with discrete aurora at a range of higher energy precipitation populations. It will likely perform less well for diffuse aurora which are spatially more widespread. We also tested the coefficient of variation on the redline emission, and we found it performed a few percent better for the same sets of training and testing events. On the other hand, to ensure the same set of events, we relied on the greenline emission to define "dark," so the results may differ for a darkness threshold based on only the redline emissions, which would need to be chosen.

This method's reliance on a one-dimensional line scan across the sky also does not indicate cloud conditions in different regions of the sky. The keogram line scan should ideally be oriented orthogonally to the typical orientation of aurorae at a given location, if possible. It could in principle be extended to all-sky images with a sequence of 1D bands or as an all-sky distribution of intensity. This method has been tested for fully clear and fully cloudy events, which as events, likely provide the best separation between the populations. For partly cloudy or mostly cloudy events (cloud masks 1-2), we expect the mislabeled rate to be higher than the 25% found in this work. Our processing did not test for or eliminate moonlight because we assume that is eliminated in the background removal described in the Appendix.

Whether this method might be useful for airglow observations is an open question. In particular uniform airglow might be mistaken for cloud cover, but for studies investigating atmospheric waves or traveling disturbances as they manifest in airglow e.g., (Ramkumar et al., 2021), the variation in the airglow intensities might be sufficient to be able to distinguish a "wavy" from a uniform sky intensity, which could filter out a stratus-type cloud layer. The coefficient of variation would tend to mislabel waves whose wavefronts are aligned with the 1D linescan direction chosen. In principle the point spread function might be derived for different cloud types based on radiative transfer modeling, such that cloudy data might someday be recoverable.

While detection theory with traditional metrics and thresholds does not have the recent popularity of some machine learning methods applied to all-sky images (Zhong et al., 2020; Clausen & Nickisch, 2018; Sado et al., 2022), its advantages are simplicity and computational ease. For a few percent accuracy penalty, the coefficient of variation metric could potentially be implemented in real-time at remote observing sites with limited computational power. In addition, while beyond the scope of this work, theoretical or empirical fits to the sample histogram distributions could be used to demonstrate a probability of false alarm or missed detection, should an application have a "not-to-exceed" requirement on the probability of either.

5 Conclusion

The method of using a keogram-based coefficient of variation to determine whether a timestamp is cloudy or not during nighttime while aurora is present has been developed and verified. A coefficient of variation threshold for the 557.7 nm wavelength of 0.51 was shown based on cloud mask truth data from 2014 and 2016 to give the lowest percent of mislabeled events by the keogram method when referenced to NOAA cloud mask data, at 21% in the training data and 25% in the validation data. After using the 557.7 nm greenline emission to omit dark sky periods, the 630.0 nm coefficient of variation threshold of 0.46 was found to give a 21% mislabeled (79% accuracy) in the 2014 and 2016 training data set and 78% accuracy in the validation data set.

This method is computationally efficient and useful working with multi-year surveys of imaging data. Future work includes testing this method on air glow keograms, and how well the coefficient of variation test statistic could also be used on all-sky images to determine which portions of the images are cloudy and cloud free.

Appendix A Keogram Processing

This section describes the method of obtaining, calibrating, and flat-field correcting the keograms before cloud detection analysis. Raw keogram netcdf files at 557.7 nm and 630.0 nm wavelengths are first downloaded for every night in 2014-2017 from the

Geophysical Institute and PFRR optics data archive website (Geophysical Institute and Poker Flat Research Range, n.d.) (D. Hampton, n.d.) and then processed using the method outlined in Figure A1.

The downloaded keograms are the raw sensor data S_λ in camera counts for $\lambda = 557.7, 630.0$ nm wavelengths. For a given wavelength λ, a measurement model of the photon flux measurement S in camera counts as a function of time t and elevation angle θ is shown in Eq. A1.

$$S_\lambda(t,\theta_n) = G(\theta_n)\left[(a * g)(t,\theta_n) + b(t,\theta_n)\right] + \beta(t,\theta_n) + \nu \tag{A1}$$

The sources of photons in a keogram measurement S are auroral light a, which may be scattered by clouds, represented as kernel g, undesired broadband emissions from light pollution b (which may also be reflected and scattered by the bottomside of the clouds but is absorbed into b), keogram sensor bias β, and noise ν. The spectrograph sensor response to received light at each viewing angle is represented as a gain function G and multiplied element-wise to the quantity in brackets.

We remove error sources b and β by subtracting a background keogram of base intensity from the measured keogram. The keogram spectrograph makes a second measurement $\tilde{S}_\lambda$, the background keogram, by filtering at a nearby wavelength, whose components are shown in Eq. A2. Broadband emissions b are still present at the same strength, but the narrow auroral emissions a drop. The same sensor gain G and bias b are present, and random noise $\hat{\nu}$ remains.

$$\tilde{S}_\lambda(t,\theta_n) = G(\theta_n)\left[b(t,\theta_n)\right] + \beta(t,\theta_n) + \tilde{\nu} \tag{A2}$$

The background keogram $\tilde{S}_\lambda$ is then subtracted from the measured keogram S, giving a baseline keogram ΔS_λ in Eq. A3. Broadband light b and common bias β are removed, leaving direct auroral light a, cloud scattering g, and differenced noise (ν - $\tilde{\nu}$).

$$\Delta S_\lambda(t,\theta_n) = S_\lambda(t,\theta_n) - \tilde{S}_\lambda(t,\theta_n) = G(\theta_n)\left[(a * g)(t,\theta_n)\right] + \nu - \tilde{\nu} \tag{A3}$$

Then each keogram is cropped to remove excess sunlight from the times near dusk or dawn, and near the horizons. Sunlight intensity during twilight is a function of the sun's angle below the horizon. To crop the keogram in time to remove light saturation, a sun elevation angle cutoff of 12° below the horizon (solar zenith angle of 102°) is used. Sunlight also appears at the horizon first. The regions within 10° of the northern and southern horizons are discarded, leaving a keogram spanning $\theta = [10°, 170°]$.

The unbiased cropped keogram ΔS_λ in camera units is converted to photon flux M_λ in Rayleighs (R) using the camera calibration factor k_λ, by Eq. A4.

$$M_\lambda(t, \theta_n) = k_\lambda \Delta S_\lambda(t, \theta_n) \tag{A4}$$

where k_λ is the wavelength-specific calibration factor. The calibration factor is $k_{557.7} =$ 6.2 R/count, and $k_{630.0} = 7.8$ R/count for 13 s exposures.

The calibrated keogram M_λ for a specific date each year is used to estimate the flat field gain G, one for each year. The gain can vary over time due to aging of the instrument and changes to the enclosure through which the instrument views the sky. When processing images, variations $G(\theta)$ in a sensor response as a function of viewing angle must be taken into account. Sometimes both a dark field (unlit) image and a flat-field (i.e., uniformly lit) image are captured before data collection, to be used later to calibrate the image for the sensor response. For this meridian spectrograph, the dark field is effectively the background keogram at the nearby wavelength $\tilde{S}_\lambda$. A flat field image is typically taken by uniformly lighting a camera and taking an image. However, uniformly-lit images were not separately collected with the meridian spectrograph and, in any case, the gain response changes over the years.

Therefore, to estimate $G(\theta)$, we select time intervals during which the camera is naturally as uniformly lit as possible. These occur when there is heavy cloud cover over auroral light. Figure A2a shows the calibrated keogram at 557.7 nm before flat-field correction for 1 Jan 2014. Between 12:00 and 14:00 UT, we note by inspection that there is heavy cloud cover over auroral light. During this time, variations in intensity with elevation angle are continuous over time, and the variations appear as faint horizontal streaks of dimming/brightening. To remove the sensor's direction-dependent response, we can use this type of time interval (cloudy and uniformly lit) as a period of flat-field imaging. We identify this time interval by using the coefficient of variation of the calibrated keogram (see Figure A2b), because the lower the coefficient of variation is, the more uniformly lit the keogram is. We identify times with a coefficient of variation $c <= 0.15$ (black dashed line in Figure A2b) as uniformly lit enough to be used in reconstruction of the flat field. The flat-field timestamps t_i meeting this criterion on 1 Jan 2014 are identified in Figure A2b with orange dots.

At each time t_i for which the coefficient of variation is below 0.15, the individual keogram snapshot measurement in units of R after calibration is

$$M_\lambda(t_i, \theta_n) = G_\lambda(t_i, \theta_n)\,[(a * g)(t_i, \theta_n)] + \epsilon \tag{A5}$$

where $\epsilon = \nu - \tilde{\nu}$ is random and zero-mean with some standard deviation σ_ϵ. The mean intensity over all N elevation angles will be:

$$\begin{aligned} \bar{M}_\lambda(t_i) &= \frac{1}{N}\sum_{n=1}^{N} M_\lambda(t_i, \theta_n) && \text{(A6)} \\ &= \frac{1}{N}\sum_{n=1}^{N} G_\lambda(t_i, \theta_n)[(a * g)(t_i, \theta_n)] && \text{(A7)} \end{aligned}$$

The sensor gain G_λ at time t_i is found by dividing each keogram intensity at viewing angle θ_n by the average intensity $\bar{M}$ of the keogram over angle.

$$G_\lambda(t_i, \theta_n) = \frac{M_\lambda(t_i, \theta_n)}{\bar{M}_\lambda(t_i)} \tag{A8}$$

where the average appearing in the denominator is taken over all angles θ_n. The time series of $G_\lambda(t_i, \theta_n)$ is then averaged for each viewing angle θ_n, by summing over time and dividing by the number of uniformly lit time points N_t, to make an estimate $\hat{G}_\lambda$ of the flat-field gain as the time-averaged mean $\bar{G}_\lambda$.

$$\hat{G}_\lambda(\theta_n) = \bar{G}_\lambda(\theta_n) = \frac{1}{N_t}\sum_{i=1}^{N_t} G(t_i, \theta_n) \tag{A9}$$

In this work, the flat field gain is determined by averaging over all cloudy intervals in one date chosen for flat-field correction per year: 1 Jan 2014, 11 Jan 2015, 1 Jan 2016, 1 Jan 2017. The flat field gains $\bar{G}_{557.7}$ for 557.7 nm for each year 2014-2017 are plotted as a function of elevation in Figure A2c. Flat-field gains are similarly constructed for the 630.0 nm keograms as well. From this figure, we note that the camera sensor gain is changing over the years. For this reason taking a flat field image in the present day is not likely to work as well for correcting images dating back to 2014, and that constructing a flat field gain for each year analyzed is useful.

The flat field gain $\bar{G}_\lambda$ is used to modify the calibrated keogram images M_λ from Eq. A5 to be the corrected images Y_λ using Eq. A10, where "/" represents element-wise division along the viewing angle θ_n dimension.

$$Y_\lambda(t, \theta_n) = \frac{\Delta M(t, \theta_n)}{\bar{G}(\theta_n)} \tag{A10}$$

The flat-field-corrected keogram $Y_{557.7}$ for 1 Jan 2014 is shown in Figure A2d, as well as Figures 2 and 3. Notice that the horizontal stripes of brightness variation are greatly reduced compared to Figure A2a. This flat-field-corrected form of keogram Y is then used for detecting cloudy intervals, as given in Eqs. 1-21.

Once used in those equations for detecting cloudy intervals (also via the coefficient of variation), the coefficient of variation computed from Y differs slightly from that of M, as shown in Figure A2e with blue (c before flat-field correction) and red (c after flat-field correction). The blue curve is identical to that shown in Figure A2b, and the red curve is identical to the curve shown in Figure 3d. The effect of flat-field correcting the keogram is to enhance the contrast in the coefficient of variation between clear sky intervals (e.g., 08:00-10:00 UT) and cloudy intervals (e.g., 12:00-14:00 UT).

Open Research Section

The keogram data used in this effort are publicly available at `http://optics.gi.alaska.edu/amisrarchive/PKR/DMSP/NCDF/`. The National Oceanic and Atmospheric Administration cloud mask data are publicly available at `https://www.ncei.noaa.gov/products/climate-data552-records/avhrr-hirs-cloud-properties-patmos`. The source code used to process the data and produce the plots shown in this paper will be made publicly available upon acceptance for publication.

Acknowledgments

This work was supported by National Science Foundation (NSF) award 1651465, National Aeronautics and Space Administration (NASA) grant 80NSSC21K1354. The support of the Illinois Space Grant Consortium under the auspices of the NASA Space Grant is greatly appreciated. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NSF or NASA.

References

Alonso-Montesinos, J. (2020). Real-time automatic cloud detection using a low-cost sky camera. *Remote Sensing*, *12*(9). Retrieved from https://www.mdpi.com/2072-4292/12/9/1382 doi: 10.3390/rs12091382

Clausen, L. B. N., & Nickisch, H. (2018). Automatic classification of auroral images from the oslo auroral themis (oath) data set using machine learning. *Journal of Geophysical Research: Space Physics*, *123*(7), 5640-5647. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JA025274 doi: https://doi.org/10.1029/2018JA025274

Correa, L. F., Folini, D., Chtirkova, B., & Wild, M. (2022). A method for clear-sky identification and long-term trends assessment using daily surface solar radiation records. *Earth and Space Science*, *9*(8), e2021EA002197. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EA002197 (e2021EA002197 2021EA002197) doi: https://doi.org/10.1029/2021EA002197

Geophysical Institute and Poker Flat Research Range. (n.d.). *Alaska optics.* Retrieved from http://optics.gi.alaska.edu/optics/ (accessed: 11 Sept 2022)

Guo, Y., Wu, X., Qing, C., Su, C., Yang, Q., & Wang, Z. (2022). Blind restoration of images distorted by atmospheric turbulence based on deep transfer learning. *Photonics*, *9*(8). Retrieved from https://www.mdpi.com/2304-6732/9/8/582 doi: 10.3390/photonics9080582

Hampton, D. (n.d.). *Keogram data archive.* Retrieved from http://optics.gi.alaska.edu/amisr_archive/PKR/DMSP/NCDF/ (accessed: 11 Sept 2022)

Hampton, D. L., Azeem, S. I., Crowley, G., Santana, J., & Reynolds, A. (2013, December). GPS phase scintillation correlated with auroral forms. *American Geophysical Union, Fall Meeting 2013*, Abstract SA12A-02.

Jaruwatanadilok, S., Ishimaru, A., & Kuga, Y. (2003). Optical imaging through clouds and fog. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(8), 1834-1843. doi: 10.1109/TGRS.2003.813845

Li, F.-F., Zuo, H.-M., Jia, Y.-H., Wang, Q., & Qiu, J. (2022). Hybrid cloud detection algorithm based on intelligent scene recognition. *Journal of Atmospheric and Oceanic Technology*, *39*(6), 837–847. doi: 10.1175/jtech-d-21-0159.1

Li, Q., Lu, W., & Yang, J. (2011). A hybrid thresholding algorithm for cloud detection on ground-based color images. *Journal of Atmospheric and Oceanic Tech-*

nology, *28*(10), 1286–1296. doi: 10.1175/jtech-d-11-00009.1

Li, W., Cao, Y., Zhang, W., Ning, Y., & Xu, X. (2022). Cloud detection method based on all-sky polarization imaging. *Sensors*, *22*(16). Retrieved from `https://www.mdpi.com/1424-8220/22/16/6162` doi: 10.3390/s22166162

Li, X., Wang, B., Qiu, B., & Wu, C. (2022). An all-sky camera image classification method using cloud cover features. *Atmospheric Measurement Techniques*, *15*(11), 3629–3639. Retrieved from `https://amt.copernicus.org/articles/15/3629/2022/` doi: 10.5194/amt-15-3629-2022

Long, C. N., & Ackerman, T. P. (2000). Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *Journal of Geophysical Research: Atmospheres*, *105*(D12), 15609-15626. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900077` doi: https://doi.org/10.1029/2000JD900077

Long, C. N., Sabburg, J. M., Calbó, J., & Pagès, D. (2006). Retrieving cloud characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric and Oceanic Technology*, *23*(5), 633 - 652. Retrieved from `https://journals.ametsoc.org/view/journals/atot/23/5/jtech1875_1.xml` doi: 10.1175/JTECH1875.1

Loucks, D., Palo, S., Pilinski, M., Crowley, G., Azeem, I., & Hampton, D. (2017). High-latitude gps phase scintillation from e region electron density gradients during the 20–21 december 2015 geomagnetic storm. *Journal of Geophysical Research: Space Physics*, *122*(7), 7473-7490. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JA023839` doi: https://doi.org/10.1002/2016JA023839

Mommert, M. (2020, mar). Cloud identification from all-sky camera data with machine learning. *The Astronomical Journal*, *159*(4), 178. Retrieved from `https://dx.doi.org/10.3847/1538-3881/ab744f` doi: 10.3847/1538-3881/ab744f

Mrak, S., Semeter, J., Hirsch, M., Starr, G., Hampton, D., Varney, R. H., ... Pankratius, V. (2018). Field-aligned gps scintillation: Multisensor data fusion. *Journal of Geophysical Research: Space Physics*, *123*(1), 974-992. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JA024557` doi: 10.1002/2017JA024557

Oceanic, N., & Administration, A. (n.d.). *Avhrr + hirs cloud properties - patmos-x cdr.* Retrieved from https://www.ncei.noaa.gov/products/climate-data-records/avhrr-hirs-cloud-properties-patmos

Ramkumar, T. K., Malik, M. A., Ganaie, B. A., & Bhat, A. H. (2021). Airglow-imager based observation of possible influences of subtropical mesospheric gravity waves on f-region ionosphere over jammu & kashmir, india. *Scientific Reports*, *11*(1), 10168. Retrieved from https://doi.org/10.1038/s41598-021-89694-3 doi: 10.1038/s41598-021-89694-3

Sado, P., Clausen, L. B. N., Miloch, W. J., & Nickisch, H. (2022). Transfer learning aurora image classification and magnetic disturbance evaluation. *Journal of Geophysical Research: Space Physics*, *127*(1), e2021JA029683. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JA029683 (e2021JA029683 2021JA029683) doi: https://doi.org/10.1029/2021JA029683

Semeter, J., Mrak, S., Hirsch, M., Swoboda, J., Akbari, H., Starr, G., … Pankratius, V. (2017). Gps signal corruption by the discrete aurora: Precise measurements from the mahali experiment. *Geophysical Research Letters*, *44*(19), 9539-9546. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073570 doi: https://doi.org/10.1002/2017GL073570

Syrjäsuo, M., & Donovan, E. (2002). Analysis of auroral images: detection and tracking. *Geophysica*, *38*, 3–14.

Syrjäsuo, M. T., & Donovan, E. F. (2004). Diurnal auroral occurrence statistics obtained via machine vision. *Annales Geophysicae*, *22*(4), 1103–1113. Retrieved from https://angeo.copernicus.org/articles/22/1103/2004/ doi: 10.5194/angeo-22-1103-2004

Zhong, Y., Ye, R., Liu, T., Hu, Z., & Zhang, L. (2020). Automatic aurora image classification framework based on deep learning for occurrence distribution analysis: A case study of all-sky image data sets from the yellow river station. *Journal of Geophysical Research: Space Physics*, *125*(9), e2019JA027590. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JA027590 (e2019JA027590 2019JA027590) doi: https://doi.org/10.1029/2019JA027590

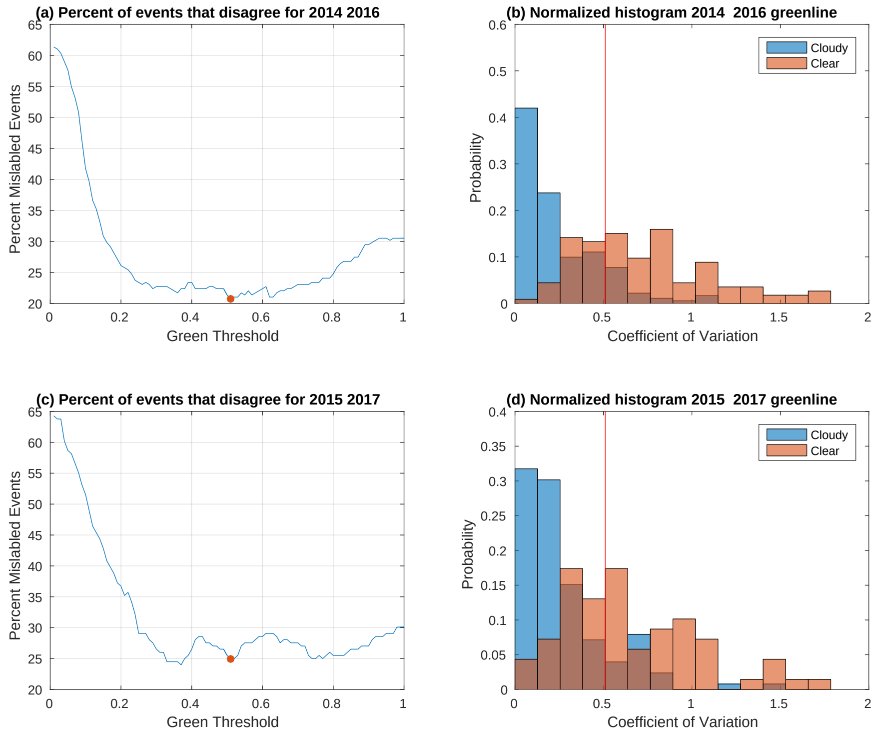

Figure 5. (a) Results from comparing 2014 and 2016 events using greenline coefficient of variation thresholds from 0.01 to 1 with steps of 0.01. The threshold that produces the lowest percent of mislabeled events is marked with a red circle. (b) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation for 2014 and 2016. The vertical line marks the threshold coefficient of variation of 0.51. (c) Results from comparing 2015 and 2017 events using thresholds starting from 0.01 to 1 with steps of 0.01. The best threshold found with the training data of 0.51 is marked with a red circle. (d) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation. The vertical line marks the threshold coefficient of variation of 0.51.

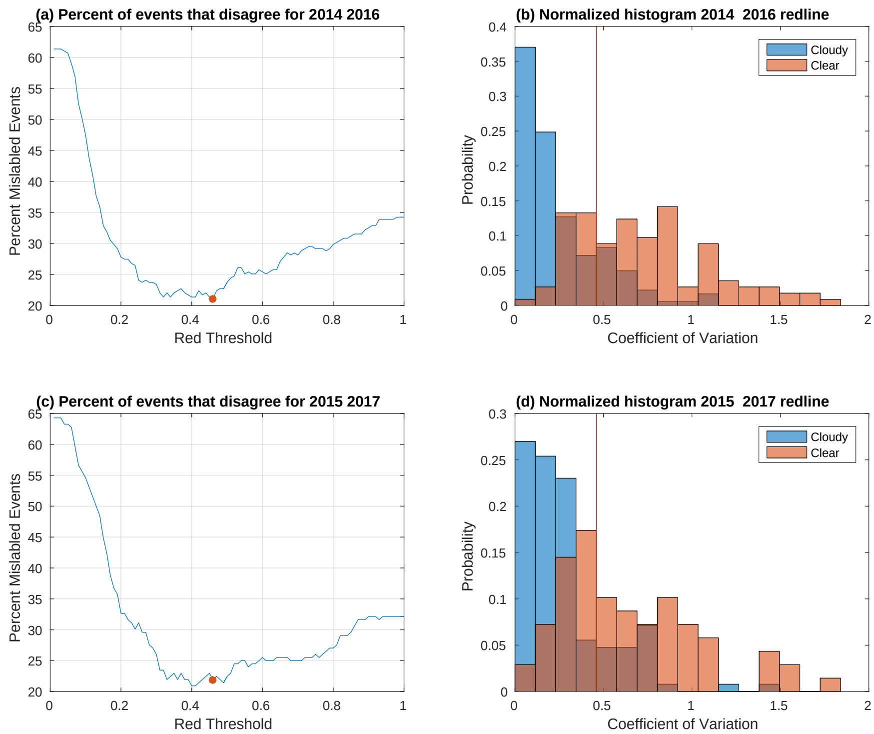

Figure 6. (a) Results from comparing 2014 and 2016 events using redline coefficient of variation thresholds from 0.01 to 1 with steps of 0.01. The threshold that produces the lowest percent of mislabeled events is marked with a red circle. (b) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation for 2014 and 2016. The vertical line marks the threshold coefficient of variation of 0.46. (c) Results from comparing 2015 and 2017 events using thresholds starting from 0.01 to 1 with steps of 0.01. The best threshold found with the training data of 0.46 is marked with a red circle. (d) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation. The vertical line marks the threshold coefficient of variation of 0.46.

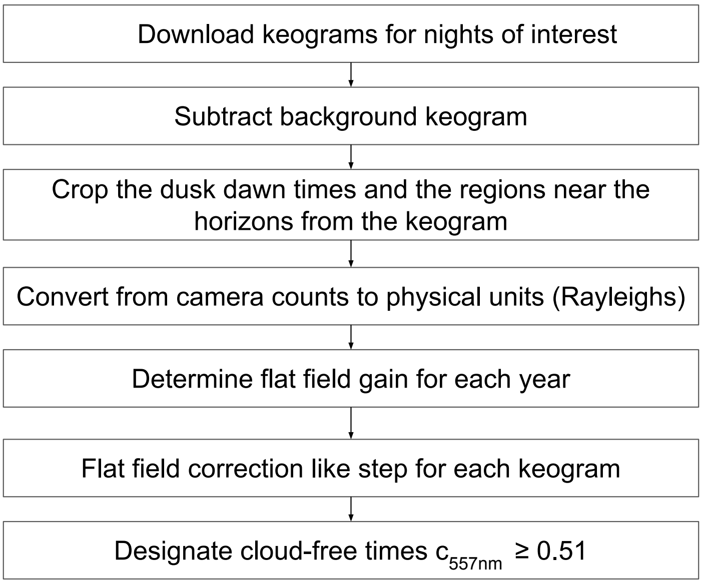

Figure A1. Method of processing raw keograms.

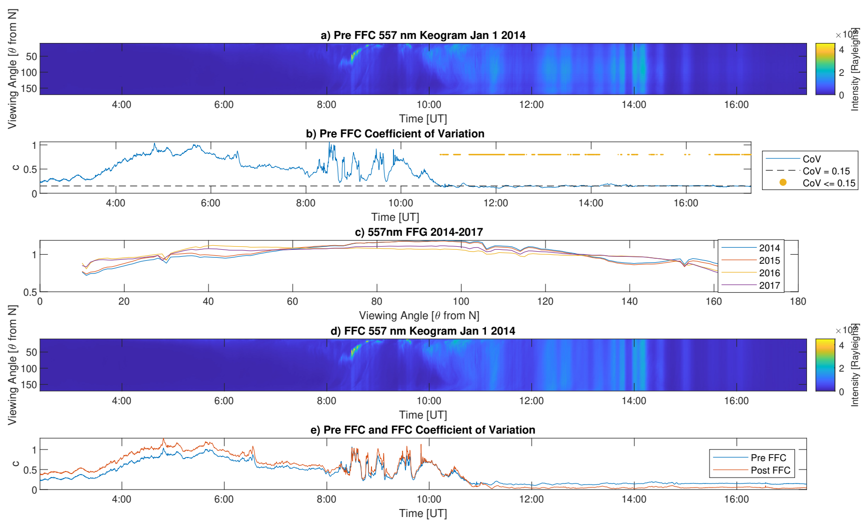

Figure A2. (a) Calibrated but not flat-field-corrected keogram M of Jan 1 2014 with the corresponding sample (b) coefficient of variation with the time points where the c is less than or equal to 0.15, (c) annual flat field gains for 557.7 nm for years 2014-2017, (d) flat-field-corrected keogram for 2014 using the 2014 flat-field gain, and (e) the coefficient of variation before and after flat field correction.

Automated Nighttime Cloud Detection using Keograms when Aurora is Present

Alex English[1], David J. Stuart[1], Donald L. Hampton[2], and Seebany Datta-Barua[1]

[1]Illinois Institute of Technology

[2]University of Alaska Fairbanks

Key Points:

- Keogram coefficient of variation is used to determine if the sky is cloudy or clear, and verified with NOAA satellite imagery from 2014-2017
- At 557.7 nm, a 0.51 threshold gives 75% accuracy but is comparable to results between 0.2-0.8
- At 630.0 nm, 0.46 is 78% accurate and comparable within 0.3-0.5

Corresponding author: Seebany Datta-Barua, sdattaba@iit.edu

Abstract

We present a metric for detecting clouds in auroral all-sky images based on single-wavelength keograms made with a collocated meridian spectrograph. The coefficient of variation, the ratio of the sample standard deviation to the sample mean taken over viewing angle, is the metric for cloud detection. After calibrating and flat-field correcting keogram data, then excluding dark sky intervals, the effectiveness of the coefficient of variation as a detector is tested compared to true conditions as determined by Advanced Very High Resolution Radiometer (AVHRR) satellite imagery of cloud cover. The cloud mask, an index of cloud cover, is selected at the corresponding nearest time and location to the site of a meridian spectrograph at Poker Flat Research Range (PFRR). We use events that are completely cloud-free or completely cloudy according to AVHRR to compute the false alarm and missed detection statistics for the coefficient of variation of the greenline 557.7 nm emission and of the redline 630.0 nm emission. For training data of the years 2014 and 2016, we find a greenline threshold of 0.51 maximizes the percent of events correctly identified at 75%. When applied to testing data of the years 2015 and 2017, the 0.51 threshold yields an accuracy of 77%. There is a relatively shallow and wide minimum of mislabeled events for thresholds spanning about 0.2 to 0.8. For the same events, the minimum is narrower for the redline, spanning roughly 0.3-0.5, with a threshold of 0.46 maximizing detector accuracy at 78-79%.

Plain Language Summary

Clouds in the sky are a problem for scientists trying to view space beyond. For upper atmospheric scientists, clouds can obscure or scatter auroral light in all-sky images (ASI), making it hard to identify, locate, and track auroral shapes. This paper shows a way to simply and automatically detect clouds using a north-to-south line scan of a single color of light from the sky over time, known as a keogram. We compute the ratio of the variation in pixel intensity to the average pixel intensity, for each north-to-south scan. Excluding dark sky periods, a large ratio means that the sky is cloudless, and a small ratio that the sky is cloudy. We find the method works with about a 75-80% correct rate using red or green auroral light. With this method we can eliminate data during cloudy conditions for any auroral studies that require clear sky conditions.

1 Introduction

Aurorae occur at the polar regions of the Earth, and are colloquially known as the northern and southern lights. These visual light emissions result from the interactions between charged particles in the Earth's magnetosphere and upper atmospheric species. Because of their relationship to interactions with the magnetosphere, researchers have been interested in classifying types of aurorae (M. T. Syrjäsuo & Donovan, 2004) and correlating them with other events. Researchers have noted that the passage of aurorae are associated with radio frequency scintillations at high latitudes (Semeter et al., 2017; Mrak et al., 2018; Loucks et al., 2017; D. L. Hampton et al., 2013). The quality of ground-based auroral images is limited by the presence of clouds in the sky. For individual case studies, researchers can visually inspect and often determine by eye the presence of clouds. However, this is not practical for large surveys of events.

Auroral scientists are not unique in being interested in detecting the presence or absence of clouds. For many practical and scientific applications, satellite imagery at various wavelengths is a standard tool for coverage spanning continent-scale areas. Multi-decade clear sky (i.e., not cloudy) identification can be done by non-optical means of comparing the measured irradiance to top of the atmosphere irradiance, compared to a clear-sky transmittance threshold (Correa et al., 2022). Such studies are longer term or generally lower resolution than might be needed for nightly auroral studies at a single site. For local conditions, ground-based methods can provide measures of cloud cover for day or night.

Many of the daytime methods leverage or are interested in solar illumination. Clear sky detection based on broadband irradiance is one avenue of cloud detection in use for decade-scale studies (Long & Ackerman, 2000). At optical wavelengths, low-cost cameras may be used by solar power station operators who want an automated method for estimating or forecasting power generation (Alonso-Montesinos, 2020). Daylight polarization can be used to determine clear sky versus cloudy sky, and the optical thickness of the cloud layer, if present (W. Li et al., 2022). A number of researchers have successfully developed methods for sorting cloud data automatically using the red and blue intensity relationships of all sky images, total sky imagers, or whole sky cameras (Q. Li et al., 2011; Long et al., 2006). Other groups have developed hybrid or adaptive thresholding algorithms (F.-F. Li et al., 2022). Another method was developed using three cloud features to categorize the ASC images into four cloud cover categories, rather than image threshold techniques (X. Li et al., 2022). These studies use daytime images illuminated by sunlight, and may be interested in classifying cloud types or regions of the sky with cloud cover. For auroral investigations, we are interested in tools usable at night and less interested in cloud types.

Recent interest in machine learning has shown that aurorae can be classified with trained algorithms (Clausen & Nickisch, 2018). One of the classification categories in this process is "cloudy" (Sado et al., 2022). Astronomers have also used machine learning methods to determine cloud cover at night for protecting telescope equipment (Mommert, 2020). While these methods hold promise, they can be computationally expensive and time consuming for training and validating at a single site for multiyear studies, necessitating a method that provides sorting of a multitude of night-time images in an efficient and consistent manner. One such method was used as part of an auroral detection and tracking method, in which aurorae were detected using the ratio of maximum to mean brightness of an all-sky image, after using synoptic cloud index measurements to eliminate cloudy periods (M. Syrjäsuo & Donovan, 2002). In this work we are interested in leveraging the nighttime single-wavelength one-dimensional images themselves to detect and discard thex cloudy intervals in the night sky, without need for separate cloud measurement.

In image processing, blurring and other distortions in a received image are modeled as convolution of a kernel with an original signal. The distortions of a camera itself may be characterized as a convolution of a point-spread function defining the camera's characteristics. In astronomy, the point spread function of the camera can often be determined using known stars. If the point-spread function is known, the image can be deconvolved to recover the original signal. For example, a theoretical determination of the point spread function due to clouds and fog for imaging objects 20 km from the imager was conducted by (Jaruwatanadilok et al., 2003) based on radiative transfer theory. In some disciplines, the point spread function may be recovered via blind deconvolution techniques. In this work, the presence of a filtering function due to atmospheric scattering is the focus, rather than defining the precise form of it. The concept of atmospheric filtering is mentioned by Guo et al. (2022) who investigated neural network-based restoration of images distorted by atmospheric turbulence. We do not need to go so far as to restore images blurred by clouds in a large multi-year database of auroral imagery, but we can leverage the effect of clouds on one-dimensional single-wavelength images over time to determine their presence.

In this work, we present a simple metric for efficiently and automatically detecting clouds if auroral light is present. This method is intended for subsequent automation of auroral all-sky image analysis. Section 2 motivates and introduces our proposed detection metric. Section 3 describes the method and data sets used to test and validate our proposed detection technique, with details on pre-processing in Appendix A. Section 4 shows the key results, and conclusions are summarized in Section 5.

2 Conceptual approach

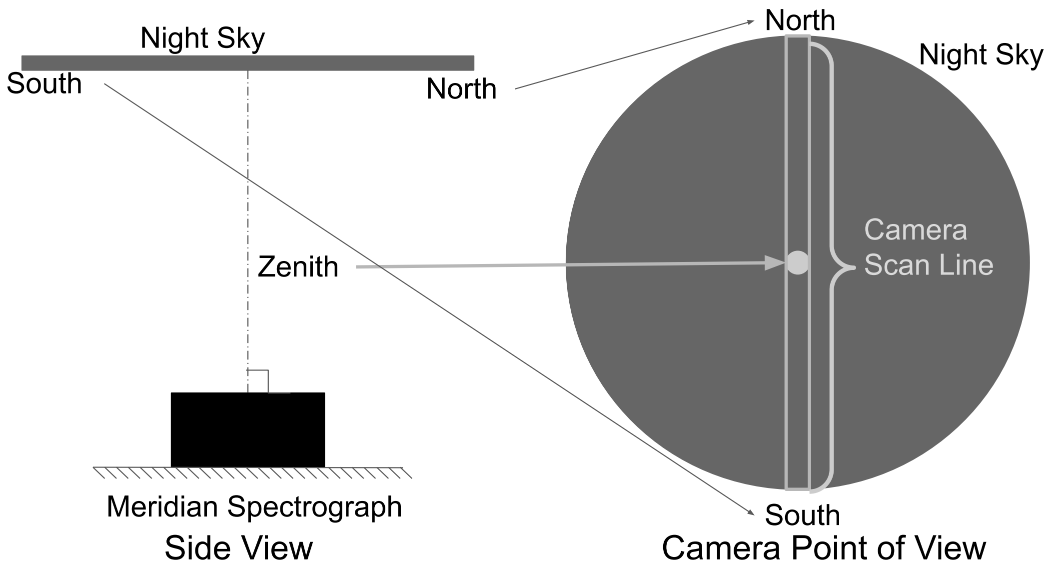

Figure 1. Schematic of keogram imaging system. The left shows a side view of a meridian spectrograph looking up local zenith and the right shows a view of the night sky from the perspective of a camera as the meridian spectrograph takes a one-pixel-wide scan from horizon to horizon through local zenith.

A keogram is a time sequence of one-dimensional images taken over the course of a night. A keogram may be taken with a meridian spectrograph or constructed from the field-of-view of an all-sky imager by extracting one subset of pixels. The diagram in Figure 1 illustrates a side view of an imaging system (left) and a sky view of an all-sky imager's field of view (right). The meridian spectrograph takes one-pixel-wide images of the sky at intervals throughout the course of the night. The pixel intensities are recorded as a function of the elevation angle from the northern to southern horizon passing through local zenith. At auroral latitudes a north-to-south scan is most likely to sample any auroral light because of the orientation of the auroral oval generally gives aurorae that are oriented east-west.

A sample keogram (calibrated and corrected, as described in later sections) taken at one wavelength is shown in Figure 2a. The x axis is time, and each column is a line-scan image from north (0 deg) to south (180 deg) of light intensity (Rayleighs, shown by color) taken at one instant. Our objective is to use the keogram to detect whether clouds are present or not at each moment. By inspection we observe that Interval 1 identified in Figure 2a corresponds to a dark sky with no aurora. A plot of the intensity as a function of elevation at the example instant identified with a red vertical line is shown in Figure 2b. The intensities are uniformly low at 04:00 UT. A histogram of these intensities over all angles at this instant is then shown in Figure 2c. The histogram of this snapshot taken over all viewing angles has a small both sample mean μ and standard deviation σ.

Interval 2 identified in Figure 2a contains a segment of an auroral band in the northern part of the sky. For this example time, the intensity as a function of viewing angle is shown in Figure 2d, consisting of one narrow region of high intensity at the viewing angle to the aurora. The sky is clear because we can see the narrow angular extent of the band of the aurora, and is verified by manually viewing an all-sky image. The histogram is shown in Figure 2e, and there is a spread of intensities due to distinctly brighter or dimmer auroral features.

Interval 3 of Figure 2a corresponds to a period during which there are aurorae, but the presence of clouds has dimmed and scattered the auroral light (again, apparent by manually viewing the all-sky image). Clouds smear the light intensities spatially to give a more uniform brightness at all viewing angles, as shown in Figure 2f. As a result, the distribution of keogram intensities is narrowly clustered around a non-zero mean.

Cloud cover has the effect of blurring the auroral light in the keogram. A commonly used image processing concept is useful here. Images taken are often post-processed to reduce noise or smooth out other unwanted effects by filtering. Comparing Figures 2d and 2f, we note that clouds between the auroral source and the imager have the effect

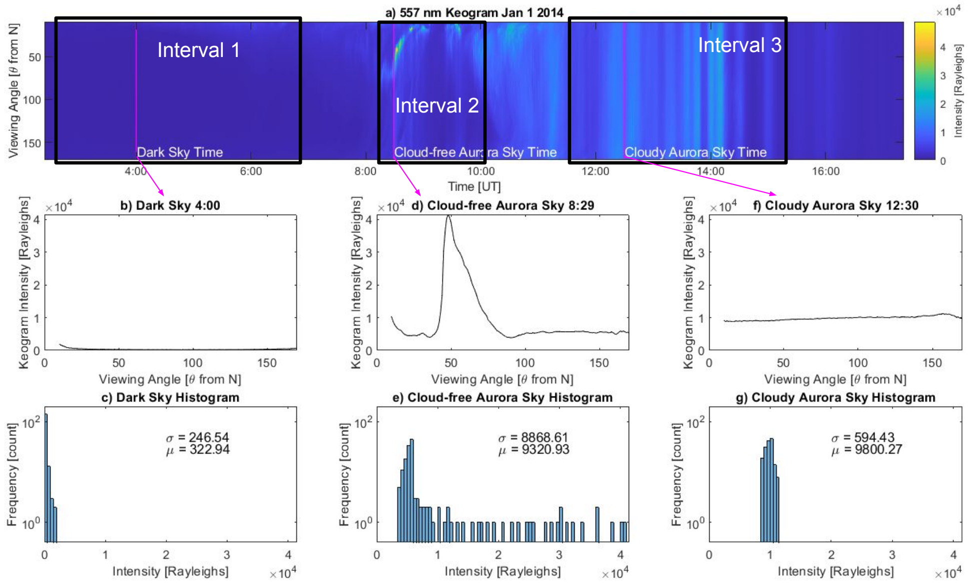

Figure 2. (a) Keogram for 01 Jan 2014 for 557.7 nm wavelength with three sky conditions highlighted: (1) Dark sky (2) Cloud-free time with aurora, and (3) Cloudy aurora time. The red lines in each interval mark example timestamps for the remaining plots. (b) Intensity versus viewing angle and (c) histogram of keogram intensity for the dark sky example time. (d) Intensity versus viewing angle and (e) histogram of the intensities at the cloud-free aurora time. (f) Intensity versus viewing angle and (g) histogram of the intensities at the cloudy aurora time.

of smoothing out the intensities spatially, and effectively act as an imaging filter that blurs the image. The mathematical process of filtering is given by convolution of a filter that modifies an original signal. Clouds in the sky act as a filter that, convolved with light sources that would otherwise be present in a keogram at a cloud-free instant, produces a smoothed set of intensities received at the ground. In the case of the example shown in Figures 2f-2g, the filtered signal results in a histogram whose distribution is narrowed, as all viewing angles have similar intensity.

At each instant t the keogram Y is a one-dimensional image of received intensities at a single wavelength over N discrete spatial coordinate elements θ_n. Assuming the keogram instrument is calibrated for uniform gain in all directions and undesired broadband and noise sources (e.g., from light pollution) have largely been removed, the residuals ϵ in the corrected keogram Y may be assumed to be zero-mean with a standard deviation of σ_ϵ. The intensity Y at a given wavelength in this case may be written as:

$$\begin{aligned} Y(t,\theta_n) &= (a*g)(t,\theta_n)+\epsilon & (1)\\ &= \sum_{m=-N}^{N} a[n-m]g[m]+\epsilon_n & (2)\end{aligned}$$

where a represents any light sources behind the clouds, i.e., aurorae; g the filtering function (sometimes called the kernel or point-spread function) of the clouds that scatters the light source, the symbol * represents the convolution operation, and ϵ is a random variable representing the residuals and noise after calibration. Equation 2 defines convolution for discrete signals over viewing angle at time t. The signal a at N discrete angles can be zero-padded for the convolution operation.

For a cloud-free sky (subscript "cf") we can represent the cloud kernel as a Kronecker delta function $g_{cf}[m]=\delta_{0m}$, which does no spreading of the intensity, so the summation simplifies as:

$$\begin{aligned} Y_{cf}(t,\theta_n) &= \sum_{m=-N}^{N} a[n-m]g_{cf}[m]+\epsilon_n & (3)\\ &= a[n]+\epsilon_n & (4)\end{aligned}$$

For zero-mean noise, the mean intensity $\bar{Y}$ over all viewing angles θ_n at time t is the mean intensity $\bar{a}$ of a over all elevations:

$$\begin{aligned} \bar{Y}_{cf}(t) &= \frac{1}{N}\sum_{n=1}^{N} a[n]+\epsilon_n & (5)\\ &= \bar{a}(t) & (6)\end{aligned}$$

The sample variance would be the sum of the variance σ_a^2 of a over all elevations and of the noise, assuming the light sources and noise to be independent, which can be seen by substituting Eqs. 4 and 6 into Eq. 7:

$$\begin{aligned} \sigma_{cf}^2 &= \frac{1}{N-1}\sum_{n=1}^{N}\left(Y(t,\theta_n)-\bar{Y}(t)\right)^2 & (7)\\ &= \frac{1}{N-1}\sum_{n=1}^{N}\left(a[n]+\epsilon_n-\bar{a}\right)^2 & (8)\\ &= \sigma_a^2+\sigma_\epsilon^2 & (9)\end{aligned}$$

While a specific cloud kernel is not known and might depend on the type of cloud, we can imagine the extreme case of a cloud that spreads the intensity evenly across all N elevations, whose filter would be $g_c[n]=1/N$. In this case, the intensity would be:

$$\begin{aligned} Y_c(t,\theta_n) &= \sum_{m=1}^{N} a[n-m]g_c[n]+\epsilon_n & (10)\\ &= \bar{a}(t)+\epsilon_n & (11)\end{aligned}$$

The angle-averaged intensity would be $\bar{Y} = \bar{a}$ as in the cloud-free case. However, the variance with angle would be given by:

$$\sigma_c^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(Y(t,\theta_n) - \bar{Y}(t)\right)^2 \tag{12}$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}(\cancel{\bar{a}(t)} + \epsilon_n - \cancel{\bar{a}(t)})^2 \tag{13}$$

$$= \sigma_\epsilon^2 \tag{14}$$

leaving only the variance of the noise.

However, if the sky is dark, there is no light source to be blurred, meaning $a = 0$, the cloud kernel whether g_c or g_{cf} has little effect on the intensity Y_d of a dark sky.

$$Y_d(t,\theta_n) = (0 * g)(t,\theta_n) + \epsilon \tag{15}$$

$$= \epsilon_n \tag{16}$$

$$\bar{Y}_d(t) = \bar{\epsilon}(t) = 0 \tag{17}$$

$$\sigma_d^2 = \frac{1}{N-1}\sum_{n=1}^{N}\left(Y(t,\theta_n) - \bar{Y}(t)\right)^2 \tag{18}$$

$$= \frac{1}{N-1}\sum_{n=1}^{N}(\epsilon_n - \bar{\epsilon}(t))^2 \tag{19}$$

$$= \sigma_\epsilon^2 \tag{20}$$

The mean and variance of a dark clear sky would be indistinguishable from that of a dark cloudy sky. On the other hand, they are not of interest for auroral studies. For this reason we exclude dark sky intervals such as Interval 1 from consideration, by setting a minimum mean value $\bar{Y}$ of the samples that must be exceeded.

Given that there is auroral light in the keogram at time t, our objective is to determine whether the image at that time is cloudy or not. The coefficient of variation $c(t)$ is the sample standard deviation σ of $Y(t)$ normalized by the mean $\bar{Y}$, shown in Eq. 21. It is a measure of how much variation there is at each time over all elevation angles θ of the keogram.

$$c(t) \equiv \frac{\sigma(t)}{\bar{Y}(t)} \tag{21}$$

The example relationship between (a) a keogram, (b) its standard deviation, (c) mean, and (d) coefficient of variation can be seen in Figure 3. In the cloud-free aurora-present case (Interval 2), $c = \sigma_{cf}/\bar{a} \sim 1$, but for the cloudy sky case (Interval 3) $c \approx \sigma_c/\bar{a} << 1$. The dark sky case (Interval 1) also has $c \approx \sigma_\epsilon/\bar{\epsilon} \sim 1$, but is artificially large because $\bar{Y}$ is so low. After filtering out dark-sky intervals, for which a small $\bar{Y}$ would artificially inflate c, we propose the coefficient of variation as a metric for detecting cloudy auroral-lit intervals in keograms (i.e., distinguishing Interval 2 from 3 in Figure 2.

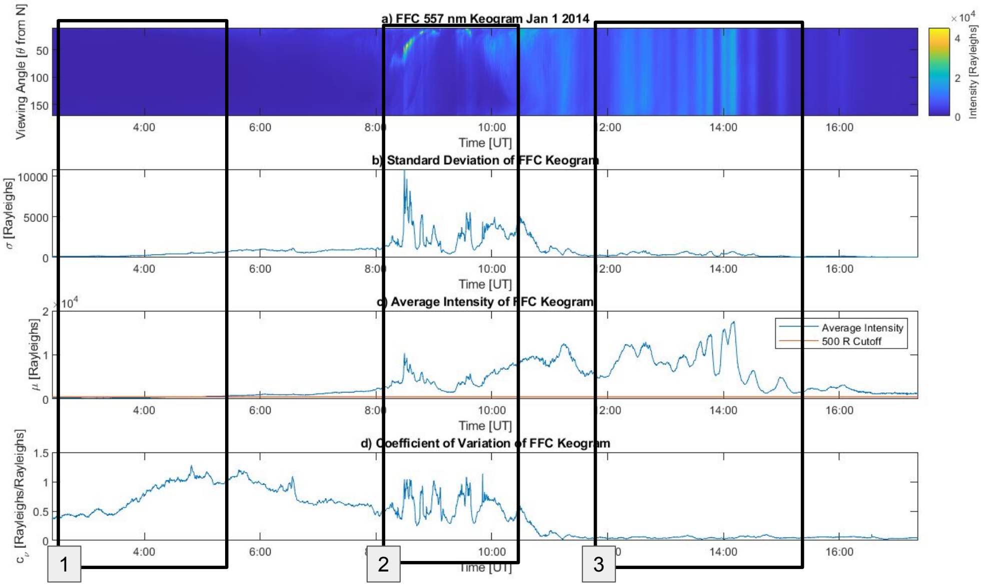

Figure 3. (a) Keogram Y of 1 January 2014 pre-processed as described in Appendix A with the corresponding sample (b) standard deviation, (c) mean, and (d) coefficient of variation c with specific times highlighted to explain what the keogram looks like in various sky conditions: 1) Dark sky 2) cloud-free with aurora 3) cloudy with aurora.

3 Method

In order to test the effectiveness of the coefficient of variation as a detection metric for clouds, we use a database of keograms collected at Poker Flat Research Range (PFRR), Alaska, from 2014-2017 (source listed in Open Research Section). After calibrating and correcting the keograms, we compute the coefficient of variation for each over time and compare them to NOAA satellite image-derived cloud mask data over PFRR. The satellite imagery provides a truth reference for whether clouds were present or not. We use standard detection theory to identify the distributions of coefficient of variation for two populations (cloudy and cloud-free). We test different thresholds of the detection metric to compute the number of events that are correctly identified or mislabeled. We use the events in years 2014 and 2016 as the training data, to find a threshold that produces the fewest mislabeled events (missed detections and false alarms), which is maximizes the accuracy (correct detections and true negatives). Then we apply the same threshold for keogram data for the years 2015 and 2017, to test whether the threshold found yields reproducible results on different data.

At PFRR, a meridian spectrograph operates with filters at 6 wavelengths to record keograms from sunset to sunrise nightly, except during the summer months which have near-continuous daylight. The keogram image intensities are given in camera counts at 6 different wavelengths: 427.8 nm, 486.1 nm, 520 nm, 557.7 nm, 630.0 nm, and 670 nm. Intensities at each wavelength are accumulated over approximately 12.5-second intervals. The wavelengths used in this study for computing the coefficient of variation are 557.7 nm (green) and, separately, 630.0 nm (red). The processing of the raw data, conversion to intensity in Rayleighs, removal of background light, and flat-field correction to produce $Y(t, \theta_n)$ are described in Appendix A.

Figure 3a represents the flat-field corrected keogram Y (identical to Figure 2a). By inspection Interval 1 has dark sky with no aurora present. Dark sky times are defined using the mean intensity of the keogram $\bar{Y}(t)$ at that time point, shown in Figure 3c. The average intensity is very low when there is no aurora in the sky in Interval 1 in Figure 3c, and increases as aurora becomes present. We choose 500 R in the 557.7 nm keogram (marked with a red line in Figure 3c) as the threshold to automatically determine darkness. If $\bar{Y}(t) < 500$ R, then the sky is determined to be dark and thus cannot be used to determine cloud presence. The dark sky test based on the green emission is used whether the red or green cloud detection metric is used.

The National Oceanic and Atmospheric Association (NOAA) Advanced Very High Resolution Radiometer (AVHRR) and High-resolution Infra-Red Sounder (HIRS) Pathfinder Atmospheres Extended (PATMOS-x) Climate Data Record (CDR) database is used as the reference true cloud condition. The AVHRR+HIRS Cloud Properties in the PATMOS-x CDR provides data for cloud properties, brightness, and temperatures collected by the AVHRR and HIRS instruments on board the NASA Polar Operational Environmental Satellites (POES) NOAA-15, NOAA-18, and NOAA-19, and European MetOp-2 platforms (Oceanic & Administration, n.d.).

Within the PATMOS-x CDR, the cloud mask is an index describing how cloudy the sky is at a given geographic latitude, longitude, and time. The cloud mask is on a scale of 0-3 as follows: 0 for clear, 1 for probably clear, 2 for probably cloudy, 3 for cloudy. An example of the cloud mask data over Alaska is shown in Figure 4. These data are used as the truth reference, to train and test the keogram cloud detection method.

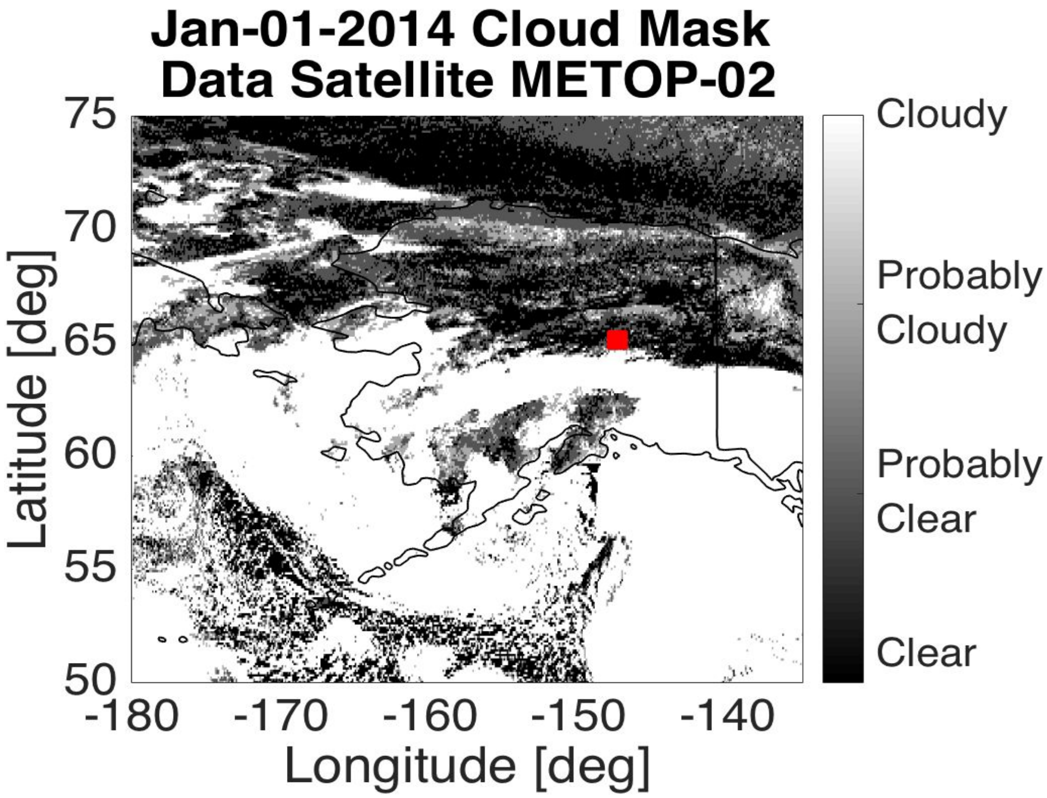

Figure 4. NOAA cloud mask data over Alaska with Poker Flat Research Range marked with a red square.

Provisional cloud mask files, available daily for 2014 through the first half of 2017, are used. From each cloud mask file, the times, cloud mask, and latitude and longitude of points within 8 km of PFRR are saved.

For each NOAA data point, we determine the keogram 557.7 nm snapshot that is closest in time and at least within 20 s of the time the keogram data was taken. Because satellite data are recorded imaging over a swath, if there is more than one NOAA data point within 20 s of the same keogram timestamp, the NOAA pixel that is geographically closest to PFRR is used, so that there is only one NOAA cloud mask associated with one keogram timestamp.

The true condition is determined from the NOAA cloud mask, corresponding to 0 when cloud-free, and 3 when cloudy. The cloud masks of 1 and 2 are not considered in this work. The keogram cloud categorization is determined from the coefficient of variation c being either less than the threshold (cloudy) or greater than or equal to the threshold (cloud-free). Each coefficient of variation and cloud mask pair are categorized into one of four groups: 1) the keogram-derived coefficient of variation c and NOAA cloud mask both indicate cloud-free conditions; 2) the keogram and NOAA cloud mask both indicate cloudy; 3) the keogram categorization predicts cloud-free but the NOAA categorization shows that the sky is cloudy (missed detection); and 4) the keogram categorization predicts cloudy and the NOAA categorization cloud-free (false alarm).

The training data for keogram-based cloud detection are all cloud masks over PFRR that have a 557.7 nm keogram measurement present at the corresponding time, in 2014 and 2016. We find a threshold with the lowest percent of mislabeled events (both missed detections and false alarms), starting from a threshold of $c = 0.01$ incrementing by 0.01 to $c = 1$. We then apply the best threshold found to the testing data of 2015 and 2017, and compute the mislabeling rates for that set of events. The accuracy of the detector is defined as 100 percent minus the mislabeled percent.

4 Results

In the training data of 2014 and 2016, there are a total of 794 events for which there are cloud mask and keogram data at the corresponding times and location. Of these, 434 of the events have cloud mask of 0 or 3 (cloudy or clear). Among these 434 events, 295 of the events are bright enough to exceed the dark sky threshold. The percentage of events mislabeled (the sum of false alarms and missed detections) as a function of the 557.7 nm keogram coefficient of variation threshold is shown in Figure 5a. The plot shows that the threshold with the lowest percent of events that are mislabeled is 0.51, with about 21% of events mislabeled. For about 13% of the events, NOAA cloud mask indicates clear sky but the keogram coefficient of variation indicates cloudy. For 8% of the events the keogram is cloud-free but the cloud mask indicates cloudy. The percent for which both the cloud mask and keogram agree the sky is cloud-free is 26%. For about 53% of the events they both indicate cloudy conditions. Histograms plotted in Figure 5b show the distribution of the coefficient of variation for cloudy events (blue) and for clear sky events (red). A vertical red line marks the threshold of 0.51. The blue bars exceeding that threshold are the ones that are missed detections of clouds. The red bars below the threshold line are the false alarms, in which using the detector c value indicates cloudy sky but the true condition is clear. On Figure 5a, we can see that above a threshold of about 0.2, there is a wide shallow minimum area up to about 0.8. This indicates that the greenline detection statistics may not be very sensitive to the specific choice of threshold within this range.

For the testing data set of 2015 and the first half of 2017, there are a total of 529 events, 266 of which have a cloud mask of 0 or 3 (cloudy or clear, respectively). Of these events, 196 of them are above the dark sky threshold. We compute the percent of events mislabeled as either false alarms or missed detections for a range of thresholds, as shown in Figure 5c. The threshold of 0.51, which was found to yield the lowest mislabeling rate with the training data, is marked with a red circle. For this data set, while 0.51 is near a local minimum, it is not the global minimum. For the testing data, 25% of the events are mislabeled (with 10% identified as cloudy with the cloud mask but detected cloud-free with our method, and 15% cloud-free but determined to be cloudy by our method). The histograms of the coefficient of variation for cloud-free events (red) and cloudy events (blue) are shown in Figure 5d, with the 0.51 threshold marked with a vertical line. There are fewer events in this data set than the training data, and this appears in the histograms with fewer counts in the modal intervals than in the training data, as well as some bins, e.g., in the clear distribution at $c = 1.2$ that are completely unpopulated. This sampling likely accounts for the appearance of multiple local minima in Figure 5c. For this data set the global minimum occurs at $c = 0.37$ with a 23% mislabeled event rate. This is comparable to the mislabeled rate for the 0.51 threshold. The testing data set has one-third fewer events for assessment than the training set. We expect that with more complete sampling, e.g., including the second half of 2017 for which at this time provisional cloud mask data are not yet available, we would likely again find a wide region of minimum mislabeling error spanning from around 0.2 to 0.8.

For the same set of training and testing events, in which the dark sky has been eliminated using the requirement that the average green emission exceed 500 R, we test the effectiveness of using the 630.0 nm emission coefficient of variation. The training mislabeling percentage results and histograms are shown in Figures 6a and b. The testing results are shown in Figures 6c and d. The threshold yielding the minimum combined rate of false alarms and missed detections of about 21% using 630.0 nm is 0.46. Apply-ing the same threshold to the testing data yields a comparable 22% mislabeled rate. In the case of the redline mislabeling statistics (Figures 6a and c), the minimum percentage mislabeled region does not appear to be quite as wide and shallow as for the green emission, only dropping below 25% above a threshold of about 0.3, and increasing close to monotonically for thresholds higher than about 0.5. It is possible that for a given set of events, the redline emission has the potential to improve accuracy by a few percent relative to the greenline emission, but may be more sensitive to choice of threshold.

An effective detector metric is one that separates the distributions between two different populations most widely. We demonstrated the coefficient of variation metric using the greenline emission, which are associated with discrete aurora at a range of higher energy precipitation populations. It will likely perform less well for diffuse aurora which are spatially more widespread. We also tested the coefficient of variation on the redline emission, and we found it performed a few percent better for the same sets of training and testing events. On the other hand, to ensure the same set of events, we relied on the greenline emission to define "dark," so the results may differ for a darkness threshold based on only the redline emissions, which would need to be chosen.

This method's reliance on a one-dimensional line scan across the sky also does not indicate cloud conditions in different regions of the sky. The keogram line scan should ideally be oriented orthogonally to the typical orientation of aurorae at a given location, if possible. It could in principle be extended to all-sky images with a sequence of 1D bands or as an all-sky distribution of intensity. This method has been tested for fully clear and fully cloudy events, which as events, likely provide the best separation between the populations. For partly cloudy or mostly cloudy events (cloud masks 1-2), we expect the mislabeled rate to be higher than the 25% found in this work. Our processing did not test for or eliminate moonlight because we assume that is eliminated in the background removal described in the Appendix.

Whether this method might be useful for airglow observations is an open question. In particular uniform airglow might be mistaken for cloud cover, but for studies investigating atmospheric waves or traveling disturbances as they manifest in airglow e.g., (Ramkumar et al., 2021), the variation in the airglow intensities might be sufficient to be able to distinguish a "wavy" from a uniform sky intensity, which could filter out a stratus-type cloud layer. The coefficient of variation would tend to mislabel waves whose wavefronts are aligned with the 1D linescan direction chosen. In principle the point spread function might be derived for different cloud types based on radiative transfer modeling, such that cloudy data might someday be recoverable.

While detection theory with traditional metrics and thresholds does not have the recent popularity of some machine learning methods applied to all-sky images (Zhong et al., 2020; Clausen & Nickisch, 2018; Sado et al., 2022), its advantages are simplicity and computational ease. For a few percent accuracy penalty, the coefficient of variation metric could potentially be implemented in real-time at remote observing sites with limited computational power. In addition, while beyond the scope of this work, theoretical or empirical fits to the sample histogram distributions could be used to demonstrate a probability of false alarm or missed detection, should an application have a "not-to-exceed" requirement on the probability of either.

5 Conclusion

The method of using a keogram-based coefficient of variation to determine whether a timestamp is cloudy or not during nighttime while aurora is present has been developed and verified. A coefficient of variation threshold for the 557.7 nm wavelength of 0.51 was shown based on cloud mask truth data from 2014 and 2016 to give the lowest percent of mislabeled events by the keogram method when referenced to NOAA cloud mask data, at 21% in the training data and 25% in the validation data. After using the 557.7 nm greenline emission to omit dark sky periods, the 630.0 nm coefficient of variation threshold of 0.46 was found to give a 21% mislabeled (79% accuracy) in the 2014 and 2016 training data set and 78% accuracy in the validation data set.

This method is computationally efficient and useful working with multi-year surveys of imaging data. Future work includes testing this method on air glow keograms, and how well the coefficient of variation test statistic could also be used on all-sky images to determine which portions of the images are cloudy and cloud free.

Appendix A Keogram Processing

This section describes the method of obtaining, calibrating, and flat-field correcting the keograms before cloud detection analysis. Raw keogram netcdf files at 557.7 nm and 630.0 nm wavelengths are first downloaded for every night in 2014-2017 from the

Geophysical Institute and PFRR optics data archive website (Geophysical Institute and Poker Flat Research Range, n.d.) (D. Hampton, n.d.) and then processed using the method outlined in Figure A1.

The downloaded keograms are the raw sensor data S_λ in camera counts for $\lambda = 557.7, 630.0$ nm wavelengths. For a given wavelength λ, a measurement model of the photon flux measurement S in camera counts as a function of time t and elevation angle θ is shown in Eq. A1.

$$S_\lambda(t, \theta_n) = G(\theta_n) \left[(a * g)(t, \theta_n) + b(t, \theta_n)\right] + \beta(t, \theta_n) + \nu \tag{A1}$$

The sources of photons in a keogram measurement S are auroral light a, which may be scattered by clouds, represented as kernel g, undesired broadband emissions from light pollution b (which may also be reflected and scattered by the bottomside of the clouds but is absorbed into b), keogram sensor bias β, and noise ν. The spectrograph sensor response to received light at each viewing angle is represented as a gain function G and multiplied element-wise to the quantity in brackets.

We remove error sources b and β by subtracting a background keogram of base intensity from the measured keogram. The keogram spectrograph makes a second measurement $\tilde{S}_\lambda$, the background keogram, by filtering at a nearby wavelength, whose components are shown in Eq. A2. Broadband emissions b are still present at the same strength, but the narrow auroral emissions a drop. The same sensor gain G and bias b are present, and random noise $\hat{\nu}$ remains.

$$\tilde{S}_\lambda(t, \theta_n) = G(\theta_n) \left[b(t, \theta_n)\right] + \beta(t, \theta_n) + \tilde{\nu} \tag{A2}$$

The background keogram $\tilde{S}_\lambda$ is then subtracted from the measured keogram S, giving a baseline keogram ΔS_λ in Eq. A3. Broadband light b and common bias β are removed, leaving direct auroral light a, cloud scattering g, and differenced noise (ν - $\tilde{\nu}$).

$$\Delta S_\lambda(t, \theta_n) = S_\lambda(t, \theta_n) - \tilde{S}_\lambda(t, \theta_n) = G(\theta_n) \left[(a * g)(t, \theta_n)\right] + \nu - \tilde{\nu} \tag{A3}$$

Then each keogram is cropped to remove excess sunlight from the times near dusk or dawn, and near the horizons. Sunlight intensity during twilight is a function of the sun's angle below the horizon. To crop the keogram in time to remove light saturation, a sun elevation angle cutoff of 12° below the horizon (solar zenith angle of 102°) is used. Sunlight also appears at the horizon first. The regions within 10° of the northern and southern horizons are discarded, leaving a keogram spanning $\theta = [10°, 170°]$.

The unbiased cropped keogram ΔS_λ in camera units is converted to photon flux M_λ in Rayleighs (R) using the camera calibration factor k_λ, by Eq. A4.

$$M_\lambda(t, \theta_n) = k_\lambda \Delta S_\lambda(t, \theta_n) \tag{A4}$$

where k_λ is the wavelength-specific calibration factor. The calibration factor is $k_{557.7} =$ 6.2 R/count, and $k_{630.0} = 7.8$ R/count for 13 s exposures.

The calibrated keogram M_λ for a specific date each year is used to estimate the flat field gain G, one for each year. The gain can vary over time due to aging of the instrument and changes to the enclosure through which the instrument views the sky. When processing images, variations $G(\theta)$ in a sensor response as a function of viewing angle must be taken into account. Sometimes both a dark field (unlit) image and a flat-field (i.e., uniformly lit) image are captured before data collection, to be used later to calibrate the image for the sensor response. For this meridian spectrograph, the dark field is effectively the background keogram at the nearby wavelength $\tilde{S}_\lambda$. A flat field image is typically taken by uniformly lighting a camera and taking an image. However, uniformly-lit images were not separately collected with the meridian spectrograph and, in any case, the gain response changes over the years.

Therefore, to estimate $G(\theta)$, we select time intervals during which the camera is naturally as uniformly lit as possible. These occur when there is heavy cloud cover over auroral light. Figure A2a shows the calibrated keogram at 557.7 nm before flat-field correction for 1 Jan 2014. Between 12:00 and 14:00 UT, we note by inspection that there is heavy cloud cover over auroral light. During this time, variations in intensity with elevation angle are continuous over time, and the variations appear as faint horizontal streaks of dimming/brightening. To remove the sensor's direction-dependent response, we can use this type of time interval (cloudy and uniformly lit) as a period of flat-field imaging. We identify this time interval by using the coefficient of variation of the calibrated keogram (see Figure A2b), because the lower the coefficient of variation is, the more uniformly lit the keogram is. We identify times with a coefficient of variation $c <= 0.15$ (black dashed line in Figure A2b) as uniformly lit enough to be used in reconstruction of the flat field. The flat-field timestamps t_i meeting this criterion on 1 Jan 2014 are identified in Figure A2b with orange dots.

At each time t_i for which the coefficient of variation is below 0.15, the individual keogram snapshot measurement in units of R after calibration is

$$M_\lambda(t_i, \theta_n) = G_\lambda(t_i, \theta_n)\,[(a * g)(t_i, \theta_n)] + \epsilon \tag{A5}$$

where $\epsilon = \nu - \tilde{\nu}$ is random and zero-mean with some standard deviation σ_ϵ. The mean intensity over all N elevation angles will be:

$$\begin{aligned} \bar{M}_\lambda(t_i) &= \frac{1}{N}\sum_{n=1}^{N} M_\lambda(t_i, \theta_n) && \text{(A6)} \\ &= \frac{1}{N}\sum_{n=1}^{N} G_\lambda(t_i, \theta_n)[(a * g)(t_i, \theta_n)] && \text{(A7)} \end{aligned}$$

The sensor gain G_λ at time t_i is found by dividing each keogram intensity at viewing angle θ_n by the average intensity $\bar{M}$ of the keogram over angle.

$$G_\lambda(t_i, \theta_n) = \frac{M_\lambda(t_i, \theta_n)}{\bar{M}_\lambda(t_i)} \tag{A8}$$

where the average appearing in the denominator is taken over all angles θ_n. The time series of $G_\lambda(t_i, \theta_n)$ is then averaged for each viewing angle θ_n, by summing over time and dividing by the number of uniformly lit time points N_t, to make an estimate $\hat{G}_\lambda$ of the flat-field gain as the time-averaged mean $\bar{G}_\lambda$.

$$\hat{G}_\lambda(\theta_n) = \bar{G}_\lambda(\theta_n) = \frac{1}{N_t}\sum_{i=1}^{N_t} G(t_i, \theta_n) \tag{A9}$$

In this work, the flat field gain is determined by averaging over all cloudy intervals in one date chosen for flat-field correction per year: 1 Jan 2014, 11 Jan 2015, 1 Jan 2016, 1 Jan 2017. The flat field gains $\bar{G}_{557.7}$ for 557.7 nm for each year 2014-2017 are plotted as a function of elevation in Figure A2c. Flat-field gains are similarly constructed for the 630.0 nm keograms as well. From this figure, we note that the camera sensor gain is changing over the years. For this reason taking a flat field image in the present day is not likely to work as well for correcting images dating back to 2014, and that constructing a flat field gain for each year analyzed is useful.

The flat field gain $\bar{G}_\lambda$ is used to modify the calibrated keogram images M_λ from Eq. A5 to be the corrected images Y_λ using Eq. A10, where "/" represents element-wise division along the viewing angle θ_n dimension.

$$Y_\lambda(t, \theta_n) = \frac{\Delta M(t, \theta_n)}{\bar{G}(\theta_n)} \tag{A10}$$

The flat-field-corrected keogram $Y_{557.7}$ for 1 Jan 2014 is shown in Figure A2d, as well as Figures 2 and 3. Notice that the horizontal stripes of brightness variation are greatly reduced compared to Figure A2a. This flat-field-corrected form of keogram Y is then used for detecting cloudy intervals, as given in Eqs. 1-21.

Once used in those equations for detecting cloudy intervals (also via the coefficient of variation), the coefficient of variation computed from Y differs slightly from that of M, as shown in Figure A2e with blue (c before flat-field correction) and red (c after flat-field correction). The blue curve is identical to that shown in Figure A2b, and the red curve is identical to the curve shown in Figure 3d. The effect of flat-field correcting the keogram is to enhance the contrast in the coefficient of variation between clear sky intervals (e.g., 08:00-10:00 UT) and cloudy intervals (e.g., 12:00-14:00 UT).

Open Research Section

The keogram data used in this effort are publicly available at `http://optics.gi.alaska.edu/amisrarchive/PKR/DMSP/NCDF/`. The National Oceanic and Atmospheric Administration cloud mask data are publicly available at `https://www.ncei.noaa.gov/products/climate-data552-records/avhrr-hirs-cloud-properties-patmos`. The source code used to process the data and produce the plots shown in this paper will be made publicly available upon acceptance for publication.

Acknowledgments

This work was supported by National Science Foundation (NSF) award 1651465, National Aeronautics and Space Administration (NASA) grant 80NSSC21K1354. The support of the Illinois Space Grant Consortium under the auspices of the NASA Space Grant is greatly appreciated. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the author and do not necessarily reflect the views of NSF or NASA.

References

Alonso-Montesinos, J. (2020). Real-time automatic cloud detection using a low-cost sky camera. *Remote Sensing*, *12*(9). Retrieved from https://www.mdpi.com/2072-4292/12/9/1382 doi: 10.3390/rs12091382

Clausen, L. B. N., & Nickisch, H. (2018). Automatic classification of auroral images from the oslo auroral themis (oath) data set using machine learning. *Journal of Geophysical Research: Space Physics*, *123*(7), 5640-5647. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018JA025274 doi: https://doi.org/10.1029/2018JA025274

Correa, L. F., Folini, D., Chtirkova, B., & Wild, M. (2022). A method for clear-sky identification and long-term trends assessment using daily surface solar radiation records. *Earth and Space Science*, *9*(8), e2021EA002197. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021EA002197 (e2021EA002197 2021EA002197) doi: https://doi.org/10.1029/2021EA002197

Geophysical Institute and Poker Flat Research Range. (n.d.). *Alaska optics.* Retrieved from http://optics.gi.alaska.edu/optics/ (accessed: 11 Sept 2022)

Guo, Y., Wu, X., Qing, C., Su, C., Yang, Q., & Wang, Z. (2022). Blind restoration of images distorted by atmospheric turbulence based on deep transfer learning. *Photonics*, *9*(8). Retrieved from https://www.mdpi.com/2304-6732/9/8/582 doi: 10.3390/photonics9080582

Hampton, D. (n.d.). *Keogram data archive.* Retrieved from http://optics.gi.alaska.edu/amisr_archive/PKR/DMSP/NCDF/ (accessed: 11 Sept 2022)

Hampton, D. L., Azeem, S. I., Crowley, G., Santana, J., & Reynolds, A. (2013, December). GPS phase scintillation correlated with auroral forms. *American Geophysical Union, Fall Meeting 2013*, Abstract SA12A-02.

Jaruwatanadilok, S., Ishimaru, A., & Kuga, Y. (2003). Optical imaging through clouds and fog. *IEEE Transactions on Geoscience and Remote Sensing*, *41*(8), 1834-1843. doi: 10.1109/TGRS.2003.813845

Li, F.-F., Zuo, H.-M., Jia, Y.-H., Wang, Q., & Qiu, J. (2022). Hybrid cloud detection algorithm based on intelligent scene recognition. *Journal of Atmospheric and Oceanic Technology*, *39*(6), 837–847. doi: 10.1175/jtech-d-21-0159.1

Li, Q., Lu, W., & Yang, J. (2011). A hybrid thresholding algorithm for cloud detection on ground-based color images. *Journal of Atmospheric and Oceanic Tech-*

nology, *28*(10), 1286–1296. doi: 10.1175/jtech-d-11-00009.1

Li, W., Cao, Y., Zhang, W., Ning, Y., & Xu, X. (2022). Cloud detection method based on all-sky polarization imaging. *Sensors*, *22*(16). Retrieved from `https://www.mdpi.com/1424-8220/22/16/6162` doi: 10.3390/s22166162

Li, X., Wang, B., Qiu, B., & Wu, C. (2022). An all-sky camera image classification method using cloud cover features. *Atmospheric Measurement Techniques*, *15*(11), 3629–3639. Retrieved from `https://amt.copernicus.org/articles/15/3629/2022/` doi: 10.5194/amt-15-3629-2022

Long, C. N., & Ackerman, T. P. (2000). Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *Journal of Geophysical Research: Atmospheres*, *105*(D12), 15609-15626. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2000JD900077` doi: https://doi.org/10.1029/2000JD900077

Long, C. N., Sabburg, J. M., Calbó, J., & Pagès, D. (2006). Retrieving cloud characteristics from ground-based daytime color all-sky images. *Journal of Atmospheric and Oceanic Technology*, *23*(5), 633 - 652. Retrieved from `https://journals.ametsoc.org/view/journals/atot/23/5/jtech1875_1.xml` doi: 10.1175/JTECH1875.1

Loucks, D., Palo, S., Pilinski, M., Crowley, G., Azeem, I., & Hampton, D. (2017). High-latitude gps phase scintillation from e region electron density gradients during the 20–21 december 2015 geomagnetic storm. *Journal of Geophysical Research: Space Physics*, *122*(7), 7473-7490. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2016JA023839` doi: https://doi.org/10.1002/2016JA023839

Mommert, M. (2020, mar). Cloud identification from all-sky camera data with machine learning. *The Astronomical Journal*, *159*(4), 178. Retrieved from `https://dx.doi.org/10.3847/1538-3881/ab744f` doi: 10.3847/1538-3881/ab744f

Mrak, S., Semeter, J., Hirsch, M., Starr, G., Hampton, D., Varney, R. H., ... Pankratius, V. (2018). Field-aligned gps scintillation: Multisensor data fusion. *Journal of Geophysical Research: Space Physics*, *123*(1), 974-992. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017JA024557` doi: 10.1002/2017JA024557

Oceanic, N., & Administration, A. (n.d.). *Avhrr + hirs cloud properties - patmos-x cdr.* Retrieved from https://www.ncei.noaa.gov/products/climate-data-records/avhrr-hirs-cloud-properties-patmos

Ramkumar, T. K., Malik, M. A., Ganaie, B. A., & Bhat, A. H. (2021). Airglow-imager based observation of possible influences of subtropical mesospheric gravity waves on f-region ionosphere over jammu & kashmir, india. *Scientific Reports*, *11*(1), 10168. Retrieved from https://doi.org/10.1038/s41598-021-89694-3 doi: 10.1038/s41598-021-89694-3

Sado, P., Clausen, L. B. N., Miloch, W. J., & Nickisch, H. (2022). Transfer learning aurora image classification and magnetic disturbance evaluation. *Journal of Geophysical Research: Space Physics*, *127*(1), e2021JA029683. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021JA029683 (e2021JA029683 2021JA029683) doi: https://doi.org/10.1029/2021JA029683

Semeter, J., Mrak, S., Hirsch, M., Swoboda, J., Akbari, H., Starr, G., ... Pankratius, V. (2017). Gps signal corruption by the discrete aurora: Precise measurements from the mahali experiment. *Geophysical Research Letters*, *44*(19), 9539-9546. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017GL073570 doi: https://doi.org/10.1002/2017GL073570

Syrjäsuo, M., & Donovan, E. (2002). Analysis of auroral images: detection and tracking. *Geophysica*, *38*, 3–14.

Syrjäsuo, M. T., & Donovan, E. F. (2004). Diurnal auroral occurrence statistics obtained via machine vision. *Annales Geophysicae*, *22*(4), 1103–1113. Retrieved from https://angeo.copernicus.org/articles/22/1103/2004/ doi: 10.5194/angeo-22-1103-2004

Zhong, Y., Ye, R., Liu, T., Hu, Z., & Zhang, L. (2020). Automatic aurora image classification framework based on deep learning for occurrence distribution analysis: A case study of all-sky image data sets from the yellow river station. *Journal of Geophysical Research: Space Physics*, *125*(9), e2019JA027590. Retrieved from https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019JA027590 (e2019JA027590 2019JA027590) doi: https://doi.org/10.1029/2019JA027590

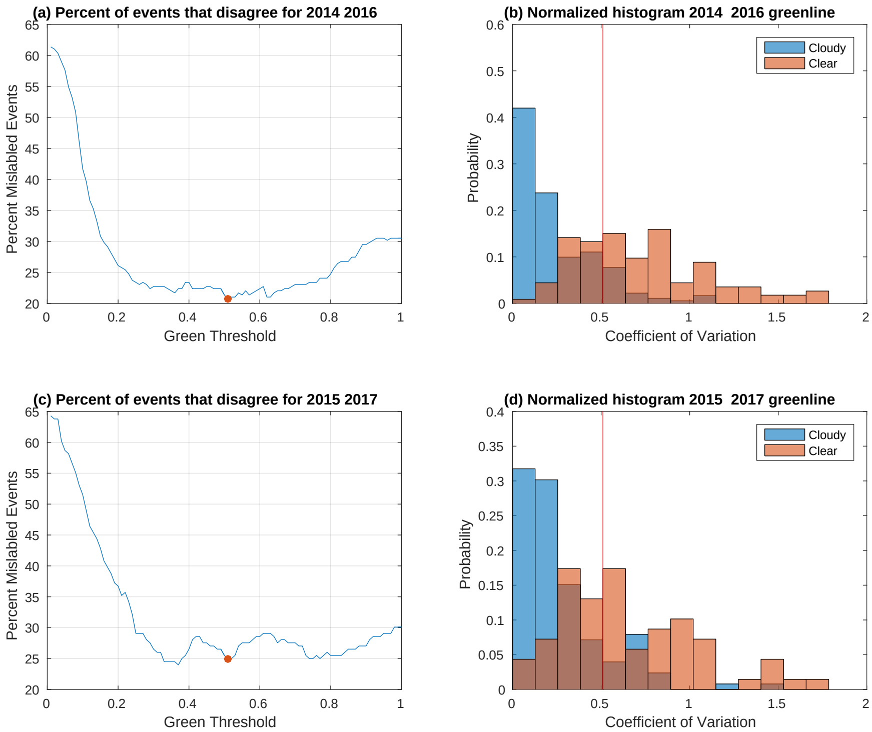

Figure 5. (a) Results from comparing 2014 and 2016 events using greenline coefficient of variation thresholds from 0.01 to 1 with steps of 0.01. The threshold that produces the lowest percent of mislabeled events is marked with a red circle. (b) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation for 2014 and 2016. The vertical line marks the threshold coefficient of variation of 0.51. (c) Results from comparing 2015 and 2017 events using thresholds starting from 0.01 to 1 with steps of 0.01. The best threshold found with the training data of 0.51 is marked with a red circle. (d) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation. The vertical line marks the threshold coefficient of variation of 0.51.

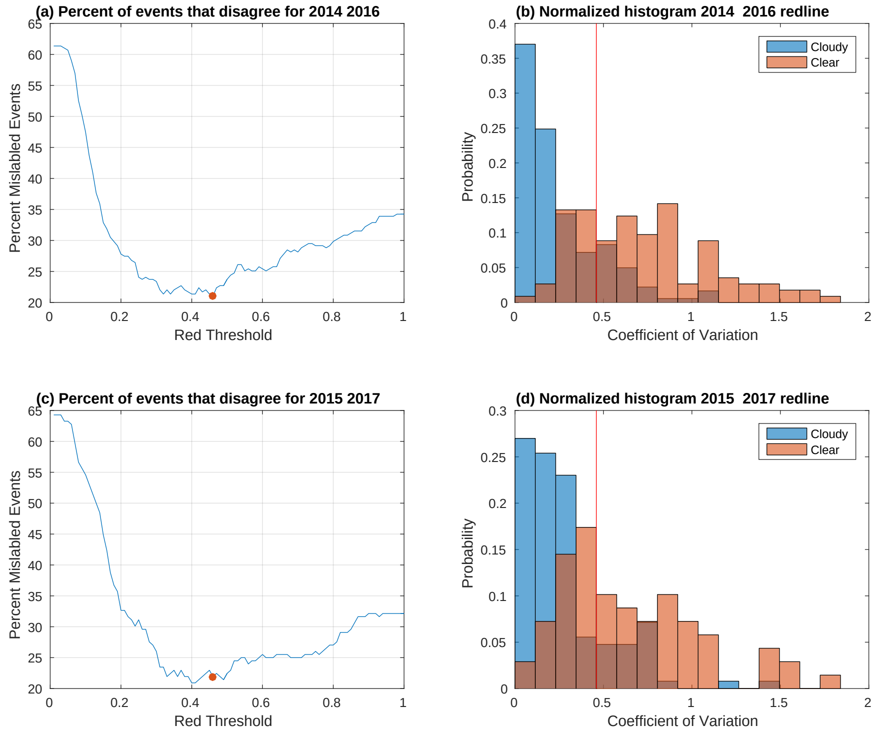

Figure 6. (a) Results from comparing 2014 and 2016 events using redline coefficient of variation thresholds from 0.01 to 1 with steps of 0.01. The threshold that produces the lowest percent of mislabeled events is marked with a red circle. (b) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation for 2014 and 2016. The vertical line marks the threshold coefficient of variation of 0.46. (c) Results from comparing 2015 and 2017 events using thresholds starting from 0.01 to 1 with steps of 0.01. The best threshold found with the training data of 0.46 is marked with a red circle. (d) Histogram of the cloudy (blue) and cloud free (red) NOAA categorized events and their respective keogram coefficients of variation. The vertical line marks the threshold coefficient of variation of 0.46.

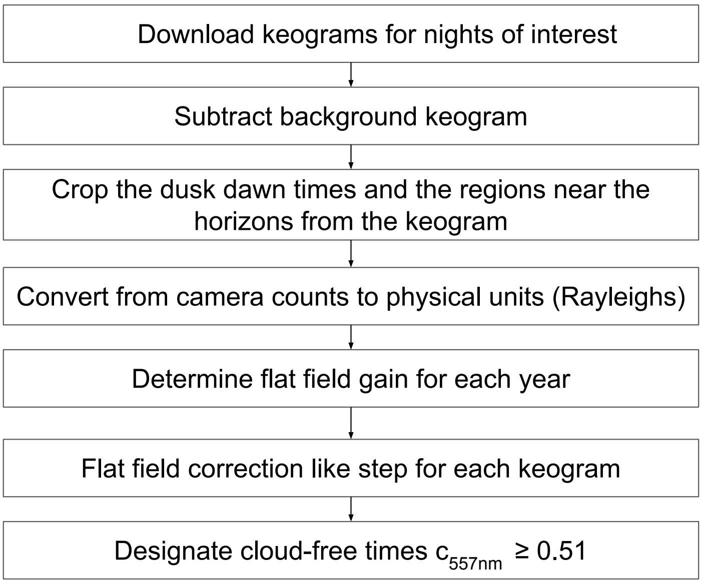

Figure A1. Method of processing raw keograms.

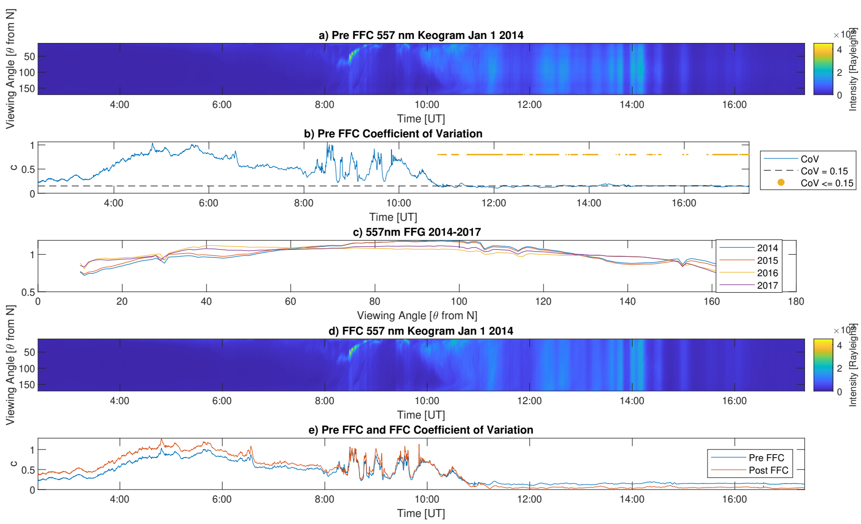

Figure A2. (a) Calibrated but not flat-field-corrected keogram M of Jan 1 2014 with the corresponding sample (b) coefficient of variation with the time points where the c is less than or equal to 0.15, (c) annual flat field gains for 557.7 nm for years 2014-2017, (d) flat-field-corrected keogram for 2014 using the 2014 flat-field gain, and (e) the coefficient of variation before and after flat field correction.