Title

Detecting clear sky images

Authors

Pawar, Prathamesh
Cortés, Cristian
Murray, Keenan
et al.

Publication Date

2019-05-01

DOI

10.1016/j.solener.2019.02.069

Peer reviewed

Detecting clear sky images

Prathamesh Pawar[1], Cristian Cortés[1], Keenan Murray[1], Jan Kleissl[1]

[1] Center for Renewable Resources and Integration, Department of Mechanical and Aerospace Engineering, University of California, San Diego

Abstract:

Many solar forecast algorithms based on ground based sky imagery apply the red-blue ratio (RBR) method to classify image pixels as clear or cloudy, by comparing the current image with the corresponding image from a clear sky library (CSL). The CSL needs to be updated regularly due to changes in clear sky readings over time caused by aerosols and imager dome properties. This clear sky library is typically created by visually scrutinizing daily sky videos and selecting appropriate clear sky periods. This practice takes a significant amount of time and manual intervention can result in human errors. To avoid this, an automated CSL algorithm (ACSL) was developed which filters each image for clear sky features including maximum green pixel brightness, average RBR, and red channel difference by pixel with respect to the previous image. The root mean square difference (RMSD) between the image RBR of the manually created CSL and the ACSL for November and April 2013 at UC San Diego were observed to be less than 6% over the full range of solar zenith angles. The ACSL was found to be more representative of clear conditions than its manual counterpart.

Keywords: Algorithm, sky imager, clear sky, automated

1. Introduction

Various cloud detection and solar forecast algorithms applied to ground based sky imagery use a clear sky library (CSL) for cloud detection, which is generally created by visual inspection of sky images. After visually scrutinizing the images for clear sky periods, the periods chosen as clear sky are passed on to the CSL, which stores the red blue ratio (RBR) for each image pixel in a look-up table as a function of sun pixel angle (SPA), image zenith angle (IZA), and solar zenith angle (SZA) [1, 2]. IZA is the angle between any pixel projected on the dome and the vertical line drawn from the center of the imager, whereas SPA represents the angular distance from the sun [3]. These stored RBR values serve as a reference for classifying current image pixels as clear, thin, or thick clouds. But these RBR values change with time due to natural events such as changes in aerosol concentration or size distribution in the atmosphere, or instrument properties such as scratches or soiling on the imager dome [4]. For accurate cloud detection the CSL should therefore be as recent as possible, which is why the CSL needs to be updated regularly. Updating the CSL can consume significant time and introduce human errors. For example, reviewing daily sky

image videos takes several minutes per day and requires great attention of the viewer, but still small clouds around the horizon or near the sun are missed. Also, since the manual CSL cannot be created in real-time, there is a lag between when a sky imager is installed and when it can be used for forecasting. To avoid these issues, an automated clear sky library (ACSL) was developed (Section 2). ACSL images for November and April 2013 are compared against manual selection to test the proposed methodology (Section 3). Conclusions are provided in Section 4.

2. ACSL Methodology

In the remote sensing community, researchers have been using several variants of color channels to distinguish clear sky and cloud pixels [5]. In this work, the ACSL methodology establishes four multispectral tests to check for clouds in sky images; if all tests are satisfied a measurement period is classified as clear sky and stored in the CSL. The first two tests consider single images and analyze the maximum green value and mean RBR. The other two tests take advantage of the fact that cloud advection causes pixel in successive images with clouds to differ in their red channel values.

2.1. Single Image Tests

Test 1: The circumsolar region (SPA<15°) may confuse the algorithm because pixels in this area can be saturated due to forward scattering by aerosols and clouds [6]. Before removing the circumsolar region, the first ACSL test checks for clouds in the circumsolar region using the maximum green pixel value in the image. If the maximum green value is less than a threshold (15,000 counts), the sun is assumed to be obscured by a cloud and the image is classified as cloudy. This threshold was selected empirically by analyzing images from a UCSD Sky Imager (USI) camera with a full range of 49065 counts. The green pixel value is used over the RBR because the RBR luminous distribution is noisier and the brightness difference between clear sky and clouds is smaller as compared to individual RGB channels. Also, the color filter array (CFA) in digital cameras causes green pixel intensity to be about twice that of red or blue pixels [7].

Test 2: If the circumsolar region is clear, it is removed from the image, and cloudy conditions are tested by checking the average RBR of the remaining image pixels. If the mean RBR of the remaining pixels is greater than 0.6 the image is defined as cloudy following [8]. If the circumsolar region is cloudy or the mean RBR of pixels outside the circumsolar region is larger than 0.6, the image is not included in the CSL. Example full day results for both single image tests are illustrated in Appendix A.

2.2. Multiple Image Test on Red Channel Difference

2.2.1 Red Channel Difference

Images are grouped into sets of 11 images capturing 5 minutes of sky conditions. Since clouds affect the red channel the most [9] a Red Channel Difference (RCD) check is applied. ACSL calculates RCD between consecutive images as,

$$RCD_i(p,q) = RC_i(p,q) - RC_{i-1}(p,q);\ i = [2{:}M], \quad (1)$$

where, $p = 1{:}X$, $q = 1{:}Y$ are the pixel indices, i is the current time step, $M = \frac{N}{I} + 1$, $N = 5$ min, $I = 0.5$ min, and RCD is measured in counts.

The RCD matrix contains X x Y = 1691 pixels x 1691 pixels = 2.86 x 10^6 pixels and $M = 11$ such matrices exist for one set.[1]

2.2.2. Test 3: Image excluding the horizon

Let P be the fraction of pixels with an RCD value greater than a threshold (here RCD_t = 50 counts). Excluding the circumsolar and horizon region (see section 2.2.3), if P (RCD_i > 50 counts) > 0.06% the image is declared as cloudy (test 3). The value 50 (corresponding to 0.10% of the full range of 49065 counts) and the fraction 0.06% were chosen empirically. In a clear image without optical defects the fraction of pixels that differs by 50 counts should be zero. In real images, small areas can have difference above 50 counts due to optical effects of the imager dome such as sun glint. To understand how RCD is affected by varying sky conditions refer to Figure 1 (raw images are shown in Appendix A). Figure 1a was selected as clear sky whereas the other two images were discarded because they failed the clear sky checks. In Fig. 1a, the time sets were formed such that 15:05 to 15:10 is one group of images and i = 2 for the pair of images in Fig. 1a.

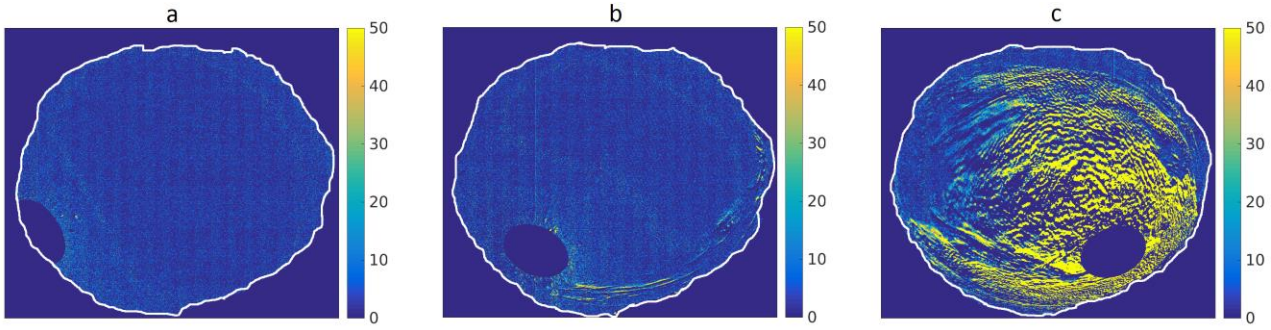

Figure 1 RCD between images from Nov 2, 2013 at a) 15:05:00 and 15:05:30 UTC. This image was classified as clear sky as P(RCD_i >50 counts) = 0.005% (< 0.06%); b) 17:10:00 and 17:10:30 UTC. This image was not selected as P(RCD_i >50 counts) = 0.152% (>0.06%) due to thin clouds; c) at 21:30:00 and 21:30:30 UTC. This image was not selected as clear sky, as the fraction is 16.9% due to widespread thick cloud cover. The white border line represents the geographical horizon.

[1] Note that the sky area represents only 1.94 x 10^6 pixels out of 2.86 x 10^6 total pixels, but the remaining non-sky pixels were included in the count for simplicity. Since these black pixels will never show a RCD above the threshold, the fractions in this paper have be to be adjusted for use with other camera systems.

Sample distributions of P as a function of RCD_t in the set of sky images presented in Figure 1 are shown in Figure 2a. Since the distribution for image 1a crosses the difference between consecutive images of 50 counts below the threshold of 0.06% image 1a passes the test. On the other hand, image 1b and 1c are rejected, because they are located above the horizontal dashed line for $RCD_t = 50$ counts, i.e. images 1b and 1c contain too many pixels that changed by more than 50 counts between two consecutive images.

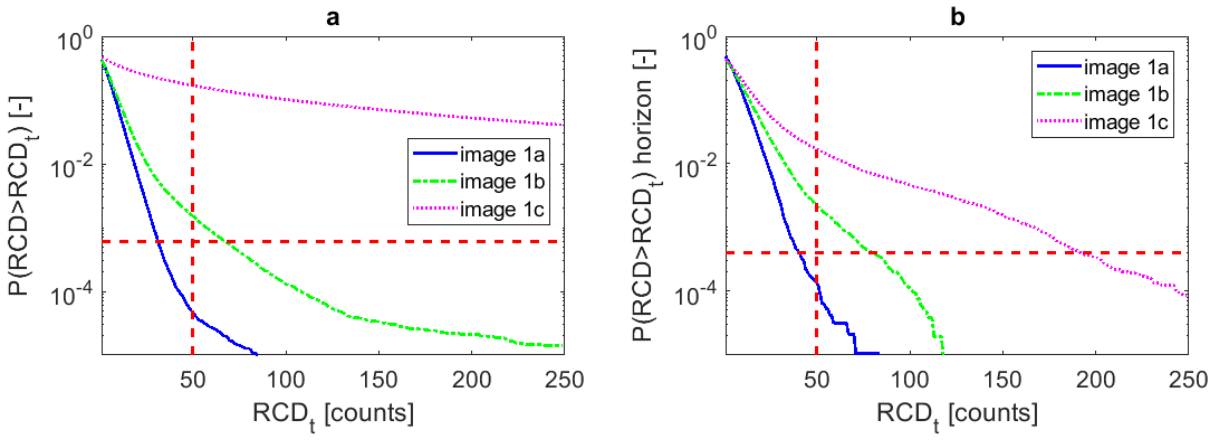

Figure 2 Distribution of P for IZA <80° (a) and IZA > 80° (b) for the 3 sky images in Figure 1. The dashed vertical line marks the chosen threshold for the red channel value difference between two consecutive images ($RCD_t = 50$). The dashed horizontal line marks the associated P threshold (0.06%).

2.2.3 Test 4: Horizon

The horizon is represented by the image area with IZA > 80° [1, 3]. The horizon requires special treatment, because clouds there move only a few pixels per image due to the optical perspective. In other words, given a cloud with certain velocity, it will affect more pixels in a particular time span if it is near the center of the image as compared to when it is near the horizon. Appendix B shows that cloud pixel speed is proportional to the cosine of IZA. In the horizon ring (consisting of 104,286 pixels or 5.4% of all pixels for the images in Figure 1) the change in the fraction of pixels between two consecutive images having a larger difference than the RCD threshold (here 50) has to be < 0.04% to classify the image as clear sky. The threshold for this test was also obtained empirically. Figure 2b shows the P distribution for three single images and results for a full day are presented in Appendix A. Figure 2b shows that clouds within the horizon region of images 1b and 1c caused P to be larger than the threshold 0.04% at $RCD_t = 50$ and both images are not classified as clear sky by test 4. Since image 1a does not contain clouds in the horizon region, its P distribution in Figure 2b decreases faster than the other two images, crossing $RCD_t = 50$ counts under 0.04%, and passing test 4.

2.3. Threshold selection

To implement the ACSL algorithm, five thresholds are needed. Four of them are chosen empirically (one for test 1, and three for tests 3, and 4, which share the RCD_t threshold) and one is obtained from [8] (test 2). The thresholds need to be recalculated when the algorithm is applied at other locations or for different camera configurations. In this section we discuss how to choose the thresholds to allow other users to customize the clear sky image detection to their needs.

The thresholds for single image tests only require simple refinements. The threshold for the maximum green pixel value test is a function of the camera count range. For most cameras, the maximum green pixels values are either the maximum number of counts (image saturated due to an unobscured sun) or less than the 30% of that magnitude (sun covered by the sun). For other cameras, we recommend specifying the threshold for test 1 as 30% of the saturation value for the green channel. For the average RBR test, the threshold was obtained following [8]. The threshold can be modified using samples of manually selected clear sky images.

The multiple image test thresholds need to be tuned for different cameras and even different locations. We recommend the following approach: (1) Calculate $P(RCD_i > RCD_t$, $RCD_t = 50$ counts) for both the entire image without the horizon and the horizon ring, using data from at least one entire clear sky day that was selected manually. (2) Average $P(RCD_i > RCD_t)$ over the day. (3) Define the thresholds by multiplying that average by a scale factor (for example 2).

RCD_t may also have to be adjusted, but it depends on the thresholds for P. Therefore, if an improvement were to be attempted, we recommend selecting a RCD_t, $P(RCD_i > RCD_t)$ pair based on a graph similar to Figure 2. The objective would be to select a RCD_t for which $P(RCD_i > RCD_t)$ is sensitive to the clearness of the sky image.

P was observed to be a (weak) function of solar zenith angle and the thresholds could be further refined to vary with solar zenith angle. Future work will target the development of a self-calibrating algorithm and consider incorporating additional tests into ACSL, such as (R-B)/(B+R) and saturation value S as suggested in [5].

2.4 Validation

A manually created CSL (MCSL) was generated by the authors by scrutinizing videos with a play back speed of around 20 minutes of clear sky data per second, i.e. 40 images/s. To prevent misclassification, the selected periods were double checked visually and by inspecting image RBR timeseries (as shown in Figure 7b later). The minimum time period chosen as clear sky manually was 35 minutes.

The Root Mean Squared Difference (RMSD) quantifies the agreement between automatic and manual clear sky libraries and is calculated as

$$RBR_{diff}(p,q) = RBR_{ACSL}(p,q) - RBR_{MCSL}(p,q); \quad (2)$$

$$RMSD = \sqrt{\frac{\sum_{p=1}^{P}\sum_{q=1}^{Q}\left(RBR_{diff}(p,q)\right)^2}{T}}, \quad (3)$$

where T = 1.94 x 10^6 is the number of pixels representing the actual sky area (including circumsolar region and horizon).

The relative root mean square deviation (rRMSD) was calculated by dividing the RMSD by the mean RBR.

3. Results and Discussion

If a 5 min sequence of image passes all four clear sky checks described in Section 2, the RBR of these images are added to the CSL. The ACSL was run for November and April 2013 which represent the winter and summer seasons that dominate the San Diego climate. For the same time period clear sky images were selected manually and added to a separate manual CSL (MCSL). The total number of clear sky images selected was 3,839 (3,224 for November + 615 for April) for the MCSL and 4,838 (4,200 for November +638 for April) for the ACSL. ACSL selected more images than MCSL because it can capture shorter coherent time periods (as short as 5 min) which are time-consuming to capture in the manual process.

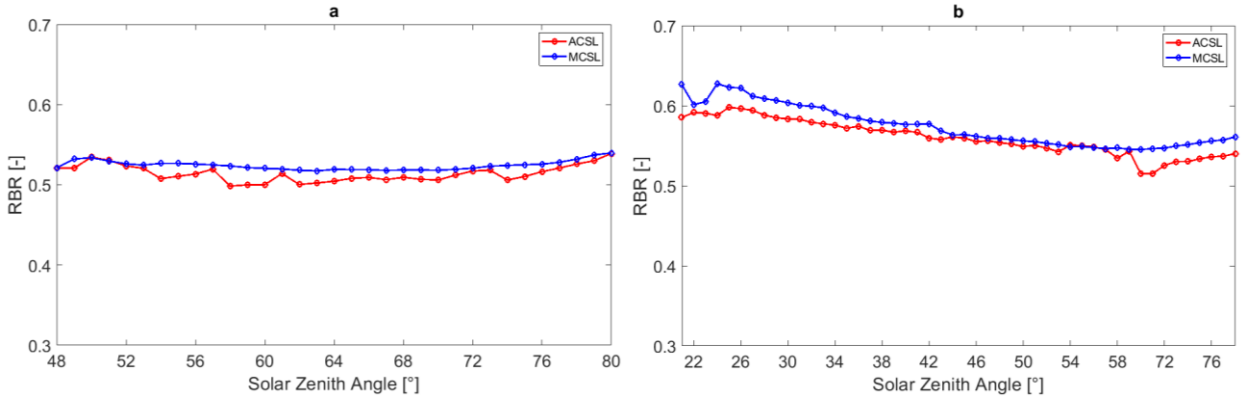

Figure 3: Comparison of mean RBR versus SZA for ACSL and MCSL for a) November & b) April 2013. RBR values are averaged over the entire image area for each SZA.

Mean RBR at different SZAs are compared in Figure 3 for both methods. The maximum difference (equivalent to the bias) between the averaged RBR is 5% at a SZA of 58° for November 2013. Figure 4 shows the RBR difference between ACSL and MCSL by IZA and SPA averaged over a range of SZAs during midday and near sunrise and sunset for November 2013 (results for April are similar). In Figure 4a and 4b, the RBR of the ACSL is always smaller than or equal to the RBR of MCSL. The RBR of the MCSL is larger since hazy conditions or small clouds may have been overlooked during manual selection of clear sky images. On the other hand, hazy conditions or small clouds are captured by ACSL due to the

RCD process. Since clear sky has lower RBR than cloudy or hazy sky [10], the ACSL does not "pollute" the CSL with cloudy images which would elevate the RBR in the CSL and cause cloud detection errors.

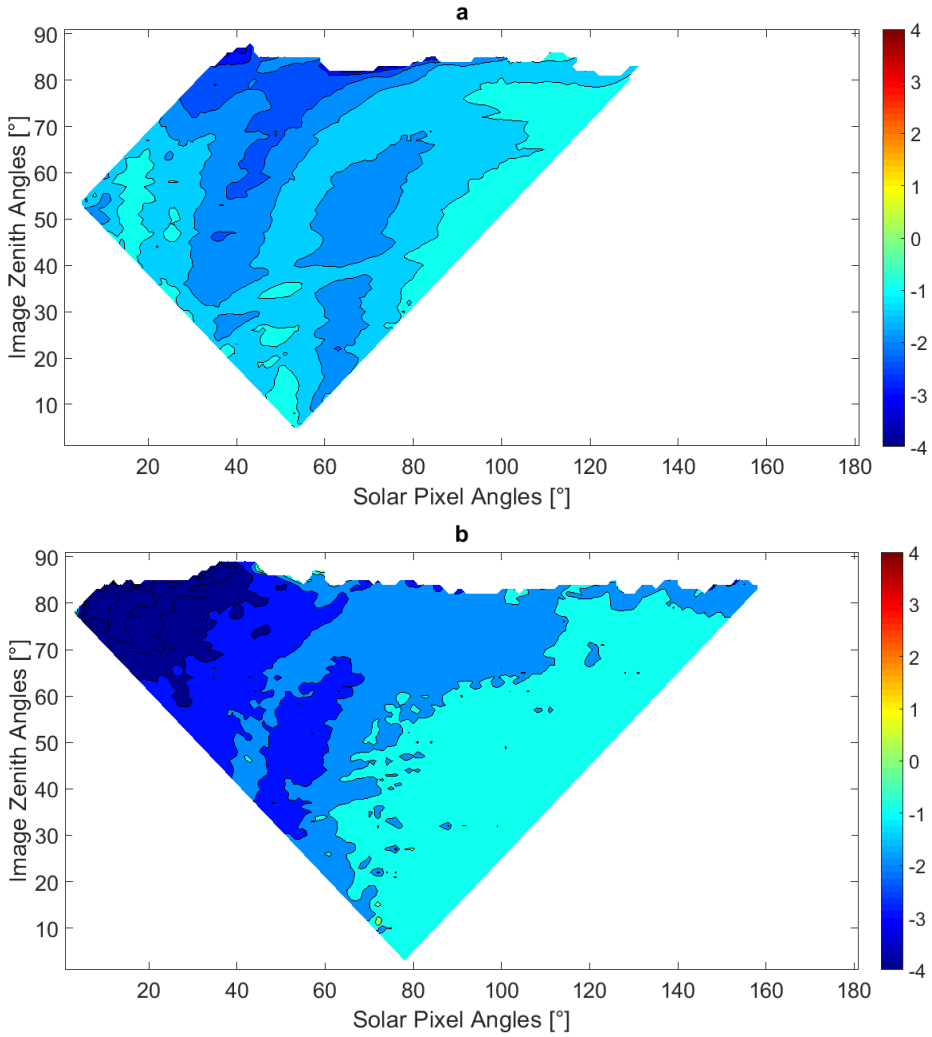

Figure 4: RBR difference in % of the mean RBR between ACSL and MCSL for a) SZAs 48° to 57° and b) SZAs 74° to 80° for November 2013.

Figure 5 shows pixel-by-pixel RMSD as a function of SZA. The RMSD between ACSL and MCSL ranges from 0.004 to 0.03 or 0.8 to 6% of the mean RBR.

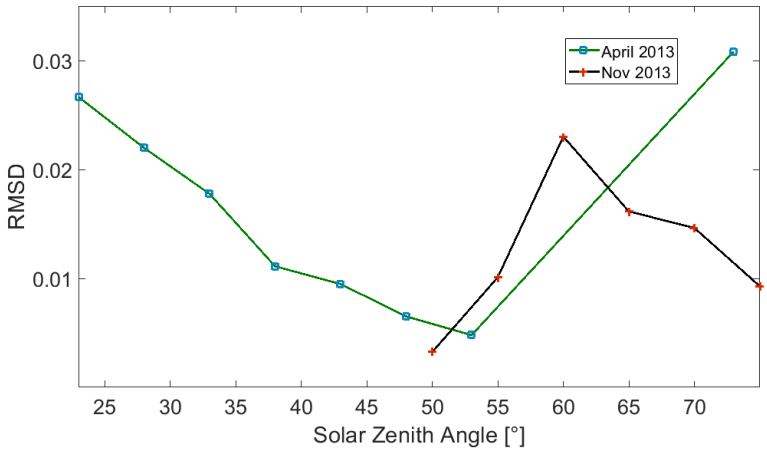

Figure 5: RMSD between RBR of ACSL and MCSL (Section 2.3) for Nov & April 2013. Each marker represents the center of a bin spanning a SZA range of 5°, and the RBR values are averaged over these SZAs before calculating the RMSD.

4. Conclusions

The ACSL performs well for the selected months. The RMSD of the RBR between MCSL and ACSL was found to be in the range of 0.004 to 0.03. The mean RBR of the ACSL was found to be lower than that of MCSL, which suggests that the ACSL algorithm is superior than the manual process, considering that RBR is lower for clearer sky. The ACSL contained more clear sky images than the MCSL, which means that it generates a more statistically converged CSL over a shorter time span. In that way, the ACSL will reflect the latest sky and imager conditions, like aerosol effects or imager dome scratches. The ACSL is robust to slow variations in sky and imager conditions because it works on relative differences, i.e. it compares two subsequent images which would experience similar aerosol or imager properties. Thus ACSL can be applied to any sky imager in general. For other imagers, thresholds will need to be adjusted for optimal performance due to differences in camera exposure, sensor properties, and local cloud properties.

For future work, tests 3 and 4 may be refined, as for clear days the *P* distribution depends on the solar zenith angle, which is not reflected in our fixed threshold. Defining thresholds as a function of the solar zenith angle may yield a more specific algorithm. Other cloud detection approaches in single sky images could also be applied. A self-calibrating algorithm is desirable for application to other cameras. The library of clear sky images is available to other researchers on request.

231 **References**

[1] M. S. Ghonima, B. Urquhart, C. W. Chow, J. E. Shields, A. Cazorla and J. Kleissl, "A method for cloud detection and opacity classification based on ground based sky imagery," *Atmospheric Measurement Techniques,* pp. 2881-2892, 2012.

[2] J. Shields, M. Karr, A. Burden, R. Johnson, V. Mikuls and J. Streeter, "Research toward Multi-Site Characterization of Sky Obscuration by Clouds," Marine Physical Laboratory, Scripps Institution of Oceanography, UC San Diego, San Diego, 2009.

[3] H. Yang, B. Kurtz, D. Nguyen, B. Urquhart, C. W. Chow, M. Ghonima and J. Kleissl, "Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego," *Solar Energy,* pp. 502-524, 2014.

[4] M. S. Ghonima, "Aerosol effects on Red Blue Ratio of Clear Sky Images, and Impact on Solar Forecasting," UC San Diego, San Diego, 2011.

[5] S. Dev, F. M. Savoy, Y. H. Lee and S. Winkler, "Rough-Set-Based Color Channel Selection," *IEEE Geoscience and Remote Sensing Letters,* vol. 14, no. 1, pp. 52-56, 2017.

[6] B. Urquhart, B. Kurtz, E. Dahlin, M. Ghonima, J. E. Shields and J. Kleissl, "Development of a sky imaging system for short-term solar power forecasting," *Atmospheric Measurement Techniques,* pp. 875-890, 2015.

[7] J. Yang, Q. Min, W. Lu, W. Yao, Y. Ma, J. Du, T. Lu and G. Liu, "An automated cloud detection method based on the green channel of total-sky visible images," *Atmospheric Measurement Techniques,* pp. 4671-4679, 2015.

[8] C. N. Long, J. M. Sabburg, J. Calbo' and D. Pages, "Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images," *Journal of Atmospheric and Oceanic Technology,* pp. 633-652, 2006.

[9] J. E. Shields, M. E. Karr, R. W. Johnson and A. R. Burden, "Day/night whole sky imagers for 24-h cloud and sky assessment: history and overview," *Applied optics,* pp. 1605-1616, 2013.

[10] C. W. Chow, B. Urquhart, M. Lave, A. Dominguez, J. Kleissl, J. Shields and B. Washomc, "Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed," *Solar Energy,* pp. 2881-2893, 2011.

[11] F. Mejia, B. Kurtz, K. Murray, L. Hinkelman, M. Sengupta, Y. Xie and J. Kleissl, "Coupling sky images with radiative transfer models: a new method to estimate cloud optical depth," *Atmospheric Measumerent Techniques,* vol. 9, pp. 4151-4165, 2016.

[12] C. Hughes, P. Denny, E. Jones and M. Glavin, "Accuracy of fish-eye lens models," *Applied optics,* pp. 3338-47, June 2010.

232

Appendix A

To illustrate the performance of the thresholds for each proposed test, additional details related to the examples contained in this paper are presented here. First, raw images of each example in Figure 1 are shown in Figure 6. Recall that only image 1a was classified as clear sky by ACSL.

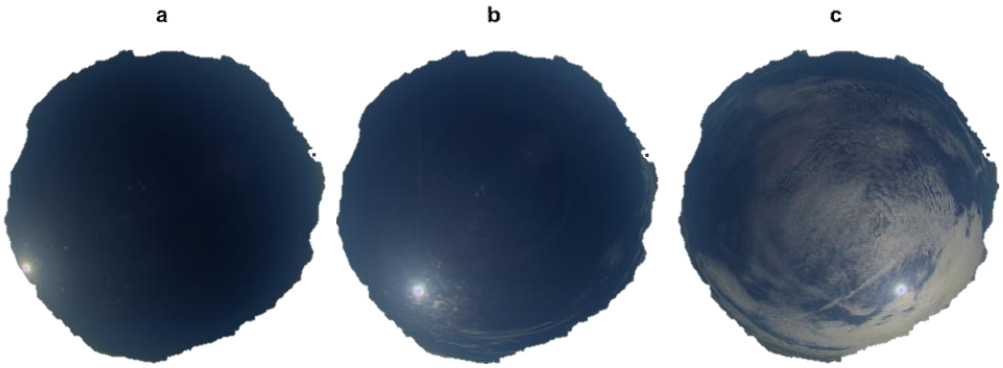

Figure 6 Image examples for November 2, 2013 with different sky conditions: (Left) Clear sky at 15:05:30 UTC. (Center) Few clouds near the horizon at 17:10:30 UTC. (Right) Cloudy at 21:30:30 UTC.

The first test is the maximum green value within each image which equals 49,065 (corresponding to a saturated pixel) for all three images, because the sun is not covered by clouds. Timeseries for November 2, 2013 are shown in Figure 7a, where 1,099 images pass the first test (82% of all images collected). Only images obtained before the sun entered the image (around 14:37 UTC), and images towards the end of the day, when the sun is covered by clouds (Figure 6c), are rejected by test 1.

The second test calculates the average RBR value of pixels located outside the circumsolar region (SPA>15°). For the images presented in Figure 6, the RBRs are 0.51 at 15:05:30 UTC, 0.52 at 17:10:30 UTC and 0.82 at 21:30:30 UTC. The complete time series for November 2, 2013 is presented in Figure 7b. 603 images pass test 2 (45% of all images collected). Images close to 00:00 UTC pass test 2 even though the sun is covered by clouds (as indicated in Figure 7a). This counterintuitive result can be explained by thick clouds covering the sun which reduces RBR below the threshold (0.6). See [11] for an analysis of RBR with cloud optical depth.

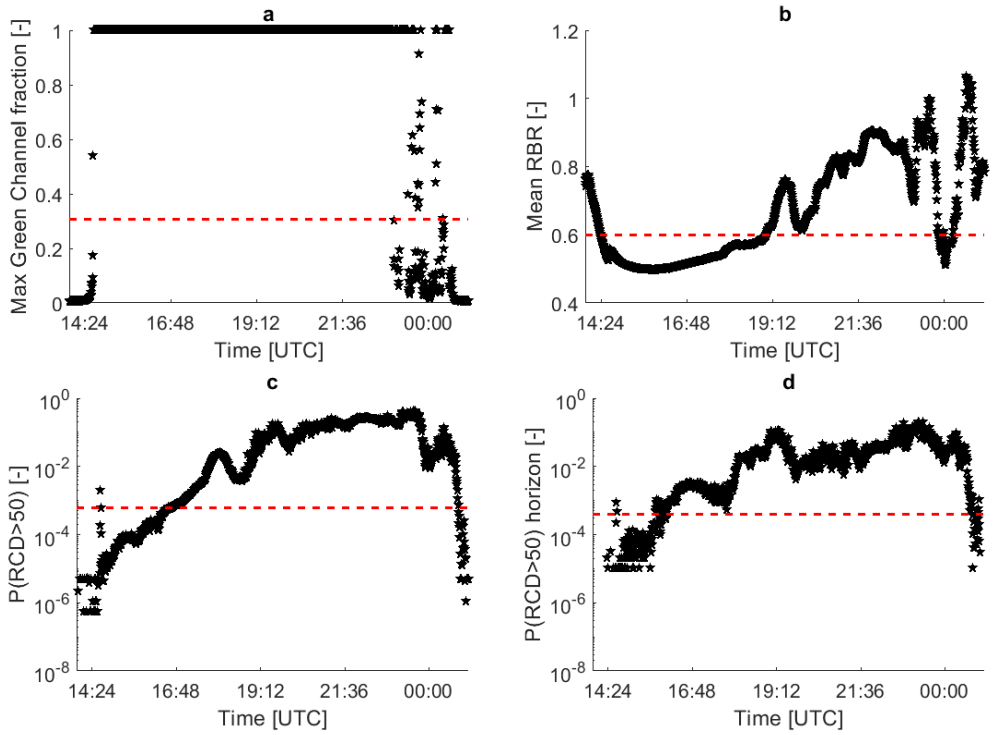

Figure 7 Full day timeseries of the maximum green channel fraction (a, test 1), Mean RBR (b, test 2), P(RCD>50) (c, test 3) and P(RCD>50) horizon (d, test 4) tests for November 2, 2013. The red dashed lines indicate the thresholds in each case.

RCD results for November 2, 2013 are presented in Figures 7c and 7d. 349 images pass test 3 and 263 pass test 4 (26% and 20% of all images obtained, correspondingly). In particular, images collected between 16:04:00 and 16:33:30 UTC have P(RCD>50) < 0.06% (pass test 3) in the center of the image, but P(RCD>50) > 0.04% in the horizon (fail test 4). From 16:34:00 UTC onwards images are rejected by both tests 3 and 4. The period from 16:04:00 to 16:33:30 marks a transition from clear conditions to clouds moving in from afar.

Considering all tests, the ACSL algorithm applied to images obtained on November 2, 2013 classified 146 images as clear sky (10.9% of all images), and all these images were taken between 14:37:30 UTC and 16:03:30 UTC.

Appendix B

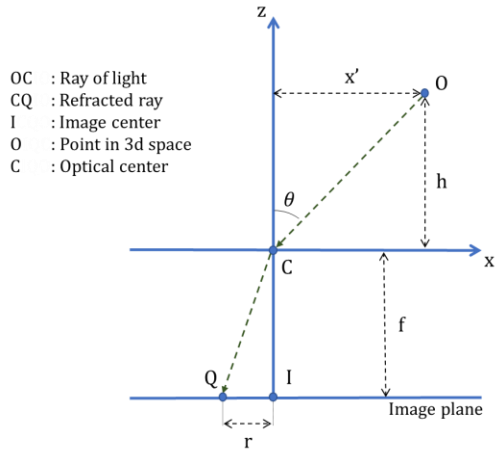

Figure 8: Schematic representation of an image for a fisheye lens.

Figure 8 shows the lens projection diagram for an object point O in space. For an equisolid angle projection, which is used for UCSD Sky Imager, the radial distance of the image point from the image center *I* is given by [12]

$$r = 2f\sin\left(\frac{\theta}{2}\right), \tag{4}$$

where θ is the IZA in radians and *f* is the focal length. Equation (4) can be derived from lens projection diagram and using simple trigonometry relations.

θ is computed from cloud base height *h* and the horizontal distance measured from the optical center *x'* as

$$\theta = \arctan\left(\frac{x'}{h}\right) \tag{5}$$

Assuming constant cloud speed at $v = dx/dt$ the velocity of an image point can then be written as,

$$\frac{\partial r}{\partial t} = \frac{\partial r}{\partial \theta}\frac{\partial \theta}{\partial x}\frac{dx}{dt} \tag{6}$$

Simplification of the above equations gives,

$$\frac{\partial r}{\partial t} = \frac{vf}{h^2}\cos^2\theta\cos\left(\frac{\theta}{2}\right) \tag{7}$$

Thus, it can be seen that the image point velocity, which will be proportional to the pixel velocity (pixels/s) in the image area, is proportional to the cosine of the IZA. Thus, for

a given cloud the number of pixels affected will be less if the cloud is near the horizon (larger IZA) as compared to if it was near the center of the image (lower IZAs).