**Atmospheric
Measurement
Techniques**

A method for cloud detection and opacity classification based on ground based sky imagery

M. S. Ghonima[1], B. Urquhart[1], C. W. Chow[1], J. E. Shields[2], A. Cazorla[3], and J. Kleissl[1]

[1]Department of Mechanical and Aerospace Engineering, University of California, San Diego, USA
[2]Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, USA
[3]Department of Chemistry and Biochemistry, University of California, San Diego, USA

Correspondence to: J. Kleissl (jkleissl@ucsd.edu)

Abstract. Digital images of the sky obtained using a total sky imager (TSI) are classified pixel by pixel into clear sky, optically thin and optically thick clouds. A new classification algorithm was developed that compares the pixel red-blue ratio (RBR) to the RBR of a clear sky library (CSL) generated from images captured on clear days. The difference, rather than the ratio, between pixel RBR and CSL RBR resulted in more accurate cloud classification. High correlation between TSI image RBR and aerosol optical depth (AOD) measured by an AERONET photometer was observed and motivated the addition of a haze correction factor (HCF) to the classification model to account for variations in AOD. Thresholds for clear and thick clouds were chosen based on a training image set and validated with set of manually annotated images. Misclassifications of clear and thick clouds into the opposite category were less than 1 %. Thin clouds were classified with an accuracy of 60 %. Accurate cloud detection and opacity classification techniques will improve the accuracy of short-term solar power forecasting.

1 Introduction

Clouds play an important role in Earth's climate; however there are still large uncertainties in the cloud-climate feedback (Solomon et al., 2007). Cloud reflection and in some cases enhancement of incoming solar radiation is an active research area (Cess et al., 1995; Kindel et al., 2011; Luoma et al., 2012). For solar power applications cloud transmissivity is the critical parameter and it is a function of cloud characteristics such as vertical and horizontal extent, droplet concentration and size distribution. Aerosols also affect the radiation budget firstly by scattering and absorbing solar radiation and secondly by acting as cloud condensation nuclei thereby modifying the radiative properties of clouds (Twohy et al., 2005; Kim and Ramanathan, 2008).

Cloud properties such as cloud optical depth and cloud fraction can be estimated from satellite images (Rossow and Schiffer, 1999; Zhao and Di Girolamo, 2006). Satellites sample on a global scale; however, their resolution is coarse at 1 km for the geostationary (GOES-12-15 series) satellites and 250 m resolution with only 1–2 images taken per day for the polar orbiting MODIS satellite. Another major shortcoming of satellite retrieved data, in the field of solar resource assessment, is the inability of satellites to determine solar obstruction accurately for a specific site due to uncertainties in cloud height and depth retrievals. Thus, ground based sky imagers (henceforth, we will refer to all types of ground based sky imagers as SIs) were developed in order to address the need for atmospheric imaging at higher spatial and temporal resolution. Since the first digital SIs were developed at University of California, San Diego (Johnson et al., 1989; Shields et al., 1993, Shields et al., 1998a; Shields et al., 1998b; Shields et al., 2009), various groups have designed different SIs. The most popular design consists of a digital camera coupled with an upward looking fisheye-lens to provide field of view (FOV) of about 180°, (Seiz et al., 2007; Souza-Echer et al., 2006; Calbó et al., 2008; Cazorla et al., 2008b; Román et al., 2012). Another SI system design uses a downward looking camera on top and a spherical mirror (Pfister et al., 2003; Long et al., 2006; Neto et al., 2010; Chow et al., 2011).

Cloud detection using SIs is generally based on a thresholding technique that utilizes the camera's red-green-blue (RGB) channel magnitudes to determine the red-blue ratio (RBR) (Shields et al., 1993). The Shields et al. (1993) algorithm uses fixed ratio thresholds to identify opaque clouds; thin clouds are detected through a comparison with a clear sky background RBR library as a function of solar angle, look angle and site location. Souza-Echer et al. (2006) used saturation in the hue, saturation and luminance (HSL) color-space with fixed thresholds for cloud detection. Cazorla et al. (2008b) classified clouds based on neural networks. Neto et al. (2010) utilized the multidimensional Euclidean geometric distance (EGD) and Bayesian methods to classify image pixels based on cloud and sky patterns. Shields et al. (2010) added an adaptive thresholding technique to account for variations in haze amount in real time. Finally, Li et al. (2011) developed a hybrid thresholding technique (HYTA) that is based on both fixed and adaptive thresholding techniques for cloud detection.

SIs have also been used to detect aerosols (Cazorla et al., 2008a, 2009; Huo and Lü, 2010). Cazorla et al. (2008a) obtained Aerosol Optical Depth (AOD) at different wavelengths from pixel counts in the red and blue channels of the SI input to a neural network. The presence of aerosols modifies the ratio of red to blue scattered light and can adversely impact the performance of cloud classification algorithms. The main purpose of this paper is to create dynamic thresholding techniques for cloud detection that account for aerosol variations. Clear sky, optically thin and thick cloud pixels are classified on a pixel by pixel basis for each image. Compared to other algorithms in the literature, our method provides an accurate mean to classify the pixels of a sky image captured by a commercially produced sky imager into three different classes as it takes aerosol conditions into account. Section 2 presents the experimental set up. Section 3 outlines the method by which the images are classified and Sect. 4 presents the results and discussion of the classification. Finally, Sect. 5 provides concluding remarks.

2 Experimental setup

2.1 Sky camera setup and environment

The University of California, San Diego (UCSD) is located 0.5 km from the Pacific Ocean in a temperate climate averaging 5 kWh m^{-2} day^{-1} of global horizontal irradiation. Maritime shallow cumulus clouds are the most common form of clouds, however during the summer mornings, marine layer stratus overcast clouds are prevalent. Maritime aerosols such as sea salt are dominant, but also urban-industrial aerosols originating locally and sometimes from the Los Angeles metropolitan area impact the San Diego atmosphere (Ault et al., 2009). In the absence of clouds, the AOD at 500 nm averages about 0.1 and typically ranges from 0.02 to 0.3.

A Total Sky Imager 440A (TSI Yankee Environmental Systems) was installed on the UCSD campus (32.885° N, 117.240° W, and 124 m m.s.l.) in August 2009. The TSI consists of a camera that looks down on a spherical mirror reflecting the sky. The mirror contains a dull black rubber shadowband which tracks the sun in order to reduce the dynamic range of the sampled sky signal, thus increasing radiometric resolution in the portion of the sky which is of interest. Images are taken by the TSI every 30 s. Sky images on selected days between January and July 2011 were used, representing a range of cloud and atmospheric conditions.

The TSI outputs 24-bit (8 bit for each RGB channel) JPEG images with a resolution of 640 by 480 pixels, of which the mirror occupies 420 by 420 pixels. A small loss of information occurs due to JPEG compression. Pixels of the image corresponding to the shadowband and the camera arm are identified automatically and excluded. Pixels at a FOV > 140° are also excluded due to distortion.

2.2 Image metrics for cloud and aerosol characterization

Compared to the clean cloudless atmosphere, both clouds and aerosols enhance red versus blue intensity increasing the RBR (Shields et al., 1993) and the red-blue difference (RBD, Heinle et al., 2010).

$$\text{RBR} = \frac{R}{B} = 1 + \frac{R - B}{B} \tag{1}$$

$$\text{RBD} = R - B \tag{2}$$

Both RBR and RBD take into account the chrominance ($C_r C_b$), reflected by the difference ($R - B$); RBR is also a function of the intensity or luminosity (Y) of the image due to normalization by B while RBD is not a function of Y. Images captured by the TSI are automatically compressed to JPEG with a downsampling ratio of $4:2:0$, in which $C_r C_b$ are sampled on each alternate line and Y is not subsampled. As a result of this downsampling, chrominance has lower resolution than luminosity and RBR will have a higher resolution (i.e. more unique values) than RBD.

Another parameter proposed by Yamashita et al. (2005) and Li et al. (2011) is the normalized red-blue ratio:

$$\text{NBR} = \frac{B - R}{B + R} = 1 - \frac{2}{\frac{B}{R} + 1}. \tag{3}$$

Equation (3) shows that NRBR can be written as a nonlinear monotonically decreasing function of RBR. For our cloud decision algorithm, we will be using an offset to a clear sky pixel RBR magnitude for cloud detection and opacity classification thus there will be no difference in accuracy between RBR and NRBR. Finally, for this paper we will use the RBR parameter as it has a higher resolution than the RBD and will provide similar results to NRBR in our cloud detection and opacity classification (CDOC) algorithm.

## 2.3 Effect of atmospheric properties on spectral features

In a clean, cloudless atmosphere, Rayleigh scattering of incoming solar radiation dominates. Since the magnitude of Rayleigh scattering is inversely proportional to the fourth power of the wavelength, visible light in the blue spectrum is predominately scattered. Consequently, in a clear atmosphere and outside the circumsolar region, Rayleigh scattering causes the input to the blue channel of the TSI camera to be higher than that of the red and green channels. In a cloudless atmosphere with high AOD, incoming solar radiation is scattered due to both Mie and Rayleigh scattering. Since Mie scattering is less dependent on wavelength, more light at larger wavelengths is scattered. This in turn causes the magnitude of the red and green channels to increase relative to the blue channel, especially near the circumsolar region as the forward lobe Mie scattering is dominant. Near and inside the circumsolar region, the RGB channels of the image saturate due to the high intensity of the direct solar beam and forward scattering of aerosols.

Thin clouds are challenging to detect as their RBR is similar to clear sky especially in haze (atmosphere with high AOD). Optically thick clouds, on the other hand result in similar signals across the RGB wavelengths. Since thick clouds have a RBR of around one, they can be easily identified in a clear atmosphere (RBR ~ 0.5) even under high AOD. It should be noted that the measured RBRs are also affected by camera specifications such as spectral responsitivity of the sensing device. Thus, the RBR will vary between different SI instruments.

3 Methods

3.1 Effect of aerosol optical depth on clear sky red blue ratio

In order to determine the effect of AOD variation on the channel magnitudes of the TSI, we compared the RBR of the TSI images with the Aerosol Optical Depth (AOD) measurement taken at 500 nm by an Aerosol Robotic NETwork (AERONET; Holben et al., 1998; Smirnov et al., 2000) sun photometer located less than 3 km away, at the Scripps Institution of Oceanography, UCSD. Cloudless sky condition images on 35 days between January and June 2011 with solar zenith angles (SZAs) less than 70° were considered. Absence of clouds on these days was confirmed using visual inspection of the images. While TSI images are taken every 30 s, AERONET timesteps are irregular at 0.25 air mass intervals for SZA < 70°. To generate a representative RBR value for an image, an average was taken from all the pixels that lie in a circular band between the 35° and 45° scattering angles. The mean RBR of the pixels is then compared with the nearest AOD measurement (no more than 5 min time difference,
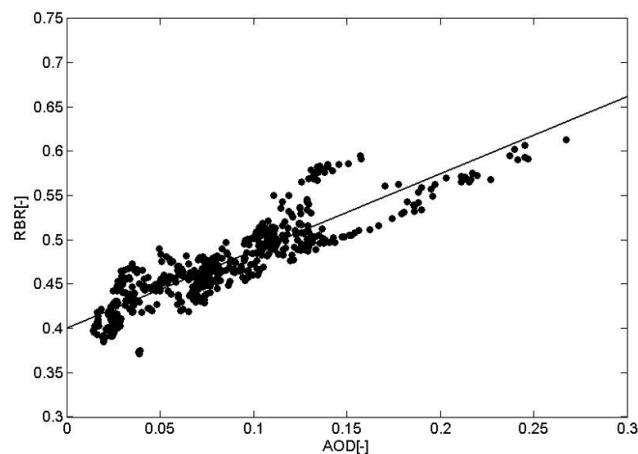
Fig. 1. Scatter graph of RBR from a total sky imager versus AOD from AERONET for data on 35 clear days in January–June 2011 (dots), RBR is extracted from sun-pixel (scattering) angles of 35° to 45°. The line is a linear regression fit (Eq. 4).

depending on SZA; Ghonima, 2011). There is a strong correlation between RBR and 500 nm AOD (τ_{500}) with a coefficient of determination (COD) of 0.797 for a linear regression of

$$RBR = 0.87\tau_{500} + 0.40. \tag{4}$$

The direct relationship can be explained by the fact that increasing AOD in the atmosphere increases Mie scattering. As a result of the increased Mie scattering, light will be scattered more evenly across the spectrum hereby increasing the RBR. The correlation was higher for the RBR versus τ_{500} correlation was higher for the RBRthan it was for the red channel versus τ_{500}, because normalizing the red channel by the blue channel helps to remove variations caused by SZA and image zenith angles (IZA) dependence, resulting in a more stable metric for comparison. Figure 1 demonstrates that AOD affects the RBR and furthermore that the AOD can be determined from the RBR of a TSI, enabling haze corrections to the CDOC algorithm thresholds.

3.2 Cloud detection and opacity classification algorithm

In our algorithm, pixels in the images collected by the TSI are classified into three classes (clear, thin or thick) based on the difference between a pixel's actual RBR and the corresponding expected RBR if the pixel were clear. A haze correction factor (HCF) is added to account for the effects of variations in AOD on RBR.

3.2.1 Clear sky library

In a clear sky, the RBR is largest near the sun and decreases with increasing sun-pixel angle (SPA, Figs. 2, 4c). The RBR also increases near the horizon (large IZAs) due to increased
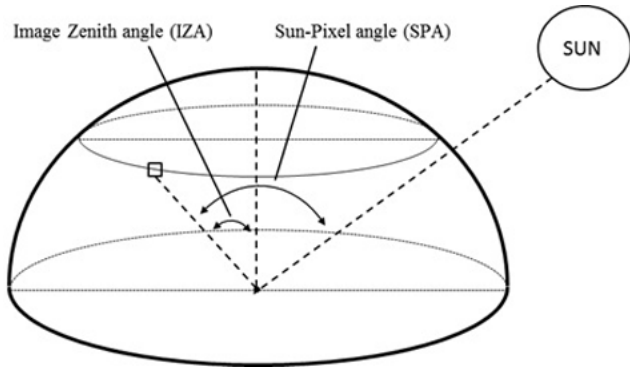
Fig. 2. Spherical mirror of the TSI and solar geometry. The thick black line shows the TSI's spherical mirror. The square on the left shows a particular pixel for illustration purposes. The sun-pixel angle is the angle between the solar direct beam and the pixel. The thin black circle is drawn through the pixel and denotes a line of constant Image Zenith Angle (IZA). IZA is the angle between the pixel and a vertical line through the center of the imager.

optical path and larger aerosol concentrations near the surface (Gueymard and Thevenard, 2009). Consequently, for cloud detection these dependencies should be removed. A clear sky library (CSL, Shields et al., 1993) provides reference RBR for each pixel and time from historical clear day images. In our CSL the RGB intensities for each pixel is stored in a matrix as a function of IZA (Fig. 2), SPA, and solar zenith angle (SZA) from historical images on a clear day (Fig 3a). Given the large sun–earth distance, the SPA is nearly identical to the scattering angle that the photon experiences at the scattering molecule or particle.

The CSL is updated on every clear day throughout the year because changing solar position and its projection on the TSI mirror and aerosol climatology affects RGB magnitudes and RBR. For example, the variation in RBR of the CSL computed on different days is highest near the solar region (small SPA) with a standard deviation of 0.1 versus magnitudes of ∼ 0.5 in Fig. 3b). Therefore, when the cloud decision algorithm is applied for a certain day it utilizes the CSL generated on the closest date.

### 3.2.2 Cloud detection and opacity classification metrics

An example sky image from 20 February 2011 and RBR image are shown in Figs. 4a and b, respectively. For each integer SZA, the RBR from the CSL is obtained as a function of SPA and IZA for each pixel (Fig. 4c). A pixel is classified as a thick cloud if the difference (Diff Fig. 4d) between the pixel RBR and CSL

$$\text{Diff} = \text{RBR} - \text{CSL} \tag{5}$$

is greater than the thick cloud threshold (see Sect. 3.3).
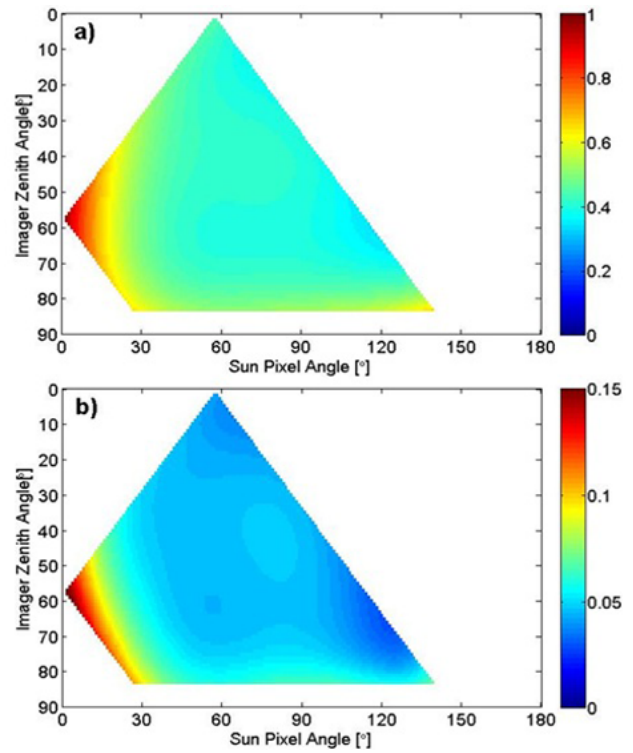
Fig. 3. (a) CSL's RBR generated on 27 January at SZA = 56° **(b)** Standard Deviation of RBR (color bar, unitless) between CSLs generated on 27 January, 12 February, 27 April and 4 May 2011 at SZA = 56°.
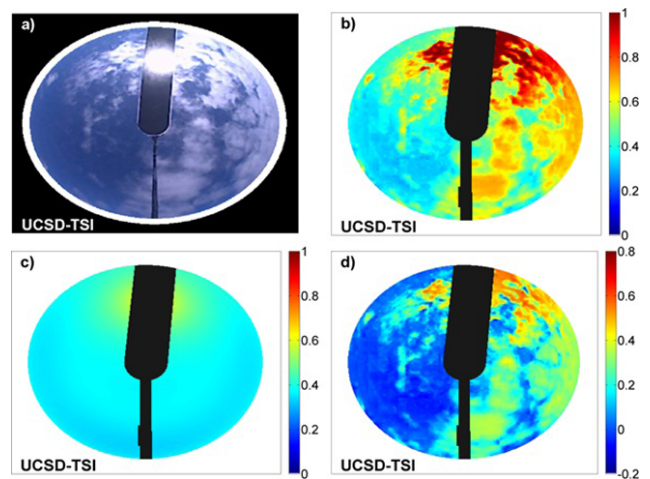
Fig. 4. (a) True-color image captured by the TSI on 12:10 PST, 20 February 2011 **(b)** RBR of the image for FOV < 140° **(c)** Clear Sky RBR generated from the CSL on 12 February 2011. **(d)** RBR difference (Diff) between image RBR and CSL RBR.

Figure 4c shows the CSL extracted from 12 February for SZA = 43°. Consistent with Fig. 3a, the CSL is fairly homogeneous across the image with the exception of the solar region (small SPAs, large RBR) and large SPAs and IZAs
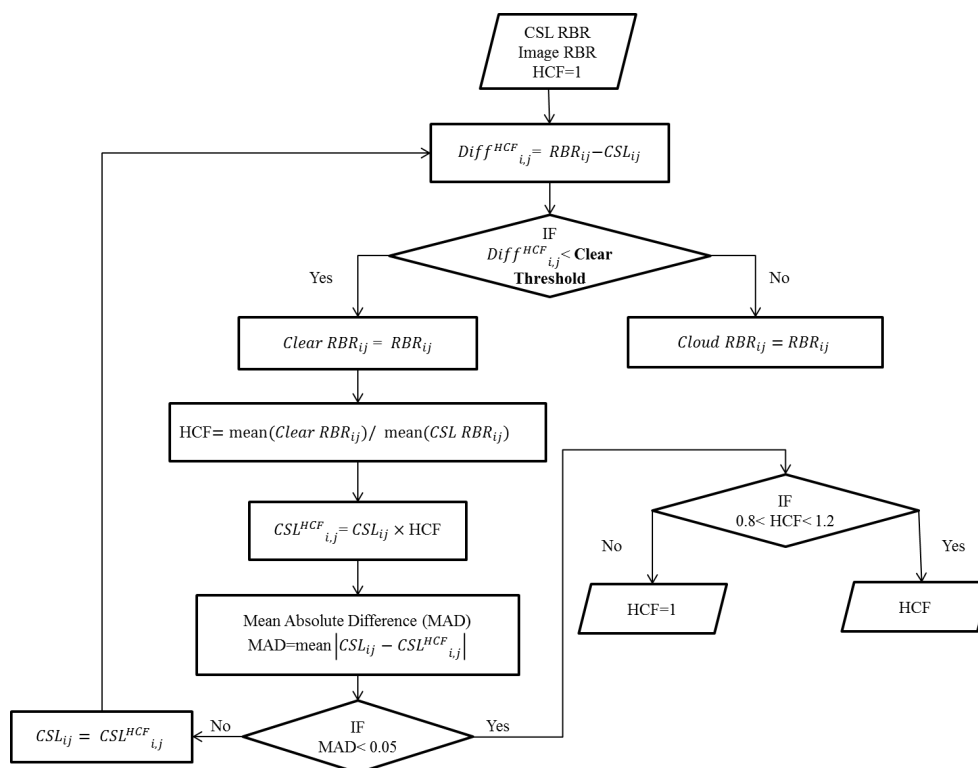
Fig. 5. Flow chart for determining HCF which is executed pixel-by-pixel. The first box represents the initialization of HCF, RBR, and CSL. Since the selection of clear pixels also depends on HCF (see 3rd box from the top), the HCF must be obtained iteratively. (i, j) denote the pixel number in the image.

(small RBR). Once the CSL is subtracted, Fig. 4d shows that all clear areas assume a similar Diff value and opaque clouds in all areas of the image can now be clearly distinguished from clear sky. Consequently, Diff allows the use of a uniform threshold for comparison of all clouds with respect to the clear sky ratio across all pixels.

Figure 1 showed that AOD significantly affects the RBR, but this is not accounted for in the CSL. By dynamically correcting the CSL for aerosol content, more consistent thresholds can be chosen to distinguish between clear pixels and thin clouds. For example, if the CSL was generated on a day with small AOD and was applied to a clear day with large AOD, Diff would be positive throughout the image, which may lead to false overcast cloud detection. Thus, a haze correction factor (HCF, Shields et al., 2010) to the CSL is introduced to account for variation of AOD. A single HCF is used across each CSL as Shields et al. (2010) found that the change in RBR with AOD is approximately independent of SPA and IZA, except for a small dependence in the solar aureole and near the horizon.

HCF is determined iteratively at each time step (Fig. 5). First, the CSL is initialized and HCF is set to 1 (first box). The 3rd box (decision diamond) describes how clear pixels are selected based on Diff and the "yes" branch shows how clear sky RBR is obtained from these clear pixels. Pixels are

determined to be clear with 96 % confidence if Diff is below a threshold which is calculated based on the probability density function (PDF) of clear pixels (see Sect. 3.3). Next, HCF is calculated by dividing the mean of the clear pixels' RBR by the corresponding CSL's mean RBR. The CSL is then multiplied by the HCF to obtain the aerosol-corrected CSL (CSL$^{\text{HCF}}$). Depending on the difference in AOD between the day under consideration and the day when the CSL was generated, HCF can either be greater or less than one. The iteration continues until convergence below an error threshold. If no clear pixels can be identified (e.g. for overcast skies) or if the correction is too large (more than 20 %) then HCF = 1. Now the difference between the image and the CSL corrected by the HCF can be calculated as

$$\text{Diff}^{\text{HCF}} = \text{RBR} - (\text{CSL} \times \text{HCF}). \tag{6}$$

Another method to control for the AOD effect on RBR proposed by Shields et al. (2010) is the perturbation ratio, which is the ratio of the current pixel RBR to the CSL pixel RBR.

$$\text{Prt} = \text{RBR}/\text{CSL} \tag{7}$$

$$\text{Prt}^{\text{HCF}} = \text{RBR}/(\text{CSL} \times \text{HCF}) \tag{8}$$
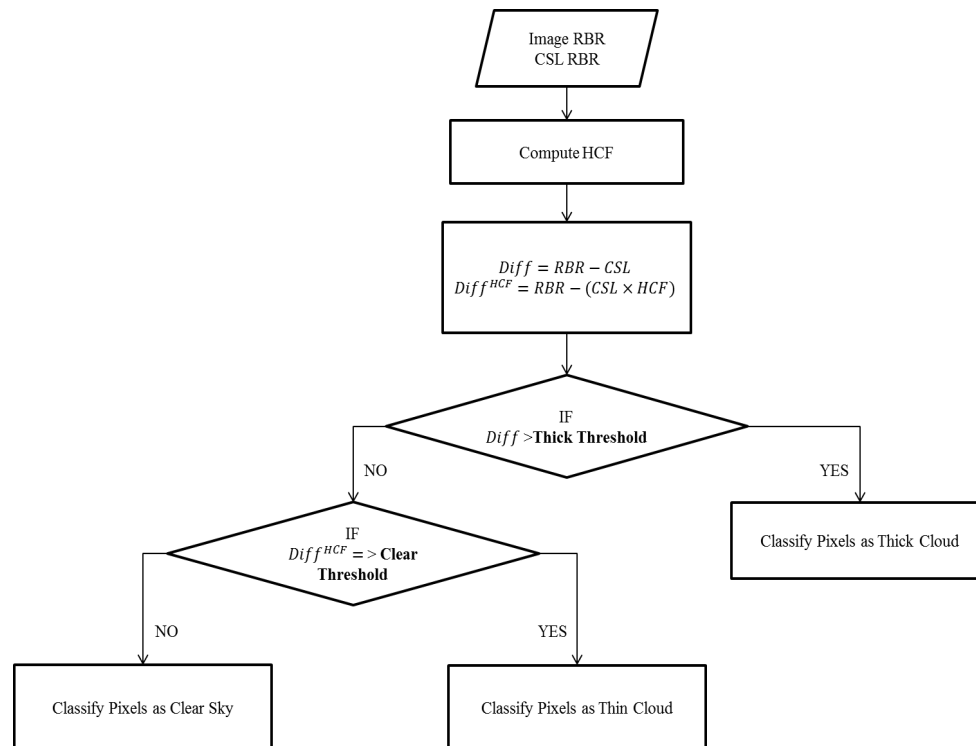
Fig. 6. Flowchart of the CDOC algorithm which is executed pixel-by-pixel. Note that thick clouds are determined based on the Diff in Eq. (5) while the distinction between clear sky and thin clouds is based on the Diff$^{\text{HCF}}$ in Eq. (6). A similar process is applied for the perturbation ratio (Eqs. 7 and 8).

There are other differences however. In the Shields et al. (2010) algorithm, the perturbation ratio was used only for thin clouds, and not for thick clouds; also the spatial variance in this perturbation ratio was used to help distinguish between heavy haze and thin cloud (Shields et al., 2010, personal communication, 2011). In Sect. 4.1, we will compare the performance of the CDOC algorithm based on Diff and Prt.

3.3 Training data and threshold determination

We generated a training set that consisted of 60 images collected on 5 different days between January and June 2011. Twelve images were sampled from each day spaced at 5 min intervals to avoid excessive overlaps in the clouds sampled. The days were chosen to represent the different cloud and atmospheric conditions encountered in coastal southern California. Images with completely clear skies were excluded from the training set. Each pixel in the image was manually classified into clear, thin, and thick (opaque) clouds by drawing polygons on the image.

The training set of manually annotated images was utilized to determine the thick cloud and clear sky Diff threshold values through trial and error. The objective was to maximize the overall accuracies for all 3 classes subject to the constraint (refer to Sect. 4.1) that the clear sky and thick

cloud accuracy is greater than 80 %. For photovoltaic solar power generation, attenuation of solar radiation by thick cloud causes the most significant impact while thin clouds have a relatively small effect. As a result, the clear sky and thick cloud thresholds are chosen to maximize thick cloud and clear sky accuracy rather than thin cloud accuracy.

The CDOC algorithm is as follows: first, the image RBR and CSL RBR are input (Fig. 6). Second, the HCF is determined as outlined in Sect. 3.2. Next, the Diff, Diff$^{\text{HCF}}$ or Prt, Prt$^{\text{HCF}}$ are computed. Finally, based on the thick cloud and clear sky thresholds pixels are classified into clear sky, thin cloud and thick cloud classes.

4 Results and discussion

4.1 Training set

In order to understand the potential accuracies for the different methods, based on the training set of manually annotated images, a PDF was generated for clear, thin, and thick cloudy pixels using the metrics of Diff and Prt with the HCF applied for the clear and thin cloud pixels (Fig. 7). For the metrics to be selective one would expect distinct and sharp peaks with little overlap in the distributions. All PDFs of Diff$^{\text{HCF}}$ and Prt$^{\text{HCF}}$ for the different classes followed a near Gaussian
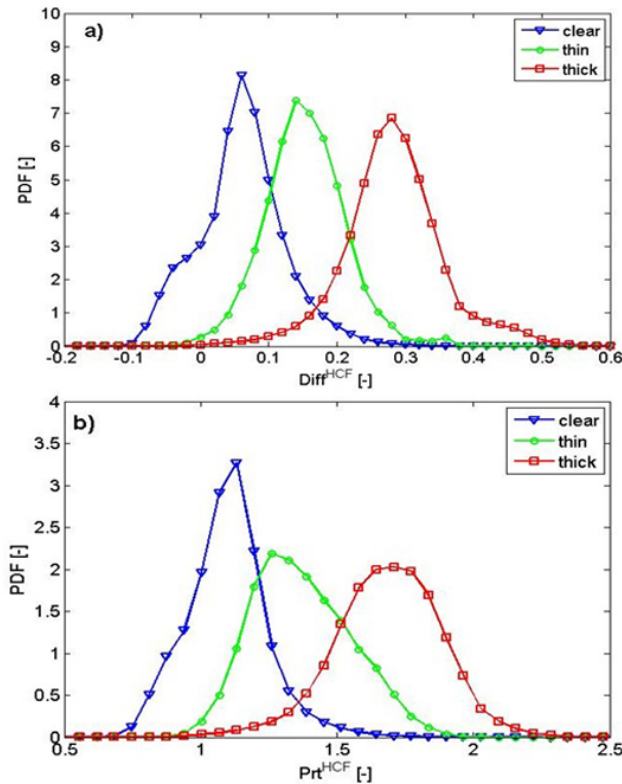
Fig. 7. PDF of the training set for each class with HCF applied for **(a)** $\mathrm{Diff}^{\mathrm{HCF}}$ (Eq. 6); **(b)** $\mathrm{Prt}^{\mathrm{HCF}}$ (Eq. 8).
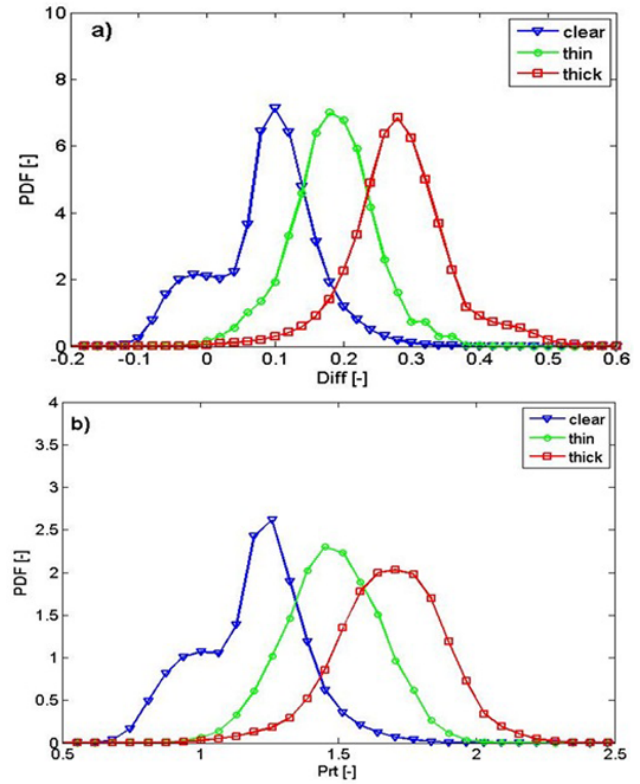
Fig. 8. PDF of the training set for each class without HCF applied for **(a)** Diff (Eq. 5); **(b)** Prt (Eq. 7).

distribution. If the HCF is not applied (Fig. 8), there is more variance in the clear sky PDF and a second peak appears due to the variations in aerosol content within the training set. The thick clouds have a distinct PDF with a much larger RBR than clear sky or thin clouds. The Prt metric results in greater overlap between the clear sky and thin cloud pixel distributions causing more misclassifications.

Visual inspection revealed that the algorithm sometimes misclassified thick clouds in the circumsolar region on overcast days. The reason for the misclassification is that pixels in the circumsolar region saturate on clear days. Consequently, the RBR in the CSL is close to 1, which is similar to the RBR of thick clouds. Hence the small difference or ratio between CSL the thick cloud RBR and the CSL results in clouds being misclassified as clear or thin. In order to correct the misclassification, we decreased the thick cloud threshold value in the circumsolar region ($0° < \mathrm{SPA} < 35°$).

In order to evaluate the performance of the CDOC algorithm, we use a confusion matrix (Kohavi and Provost, 1998) for the three classes (1): clear pixels, (2): thin cloud pixels, and (3): thick cloud pixels, thus there are nine possible outcomes for the CDOC algorithm (Table 1). Kohavi and Provost (1998) define accuracy as the sum of correct classification made by the algorithm, i.e. (TC11 + TC22 + TC33) divided by the sum of all categories. This metric is indepen-

dent of the number of clear, thin, or thick clouds observed and will be used for evaluating the performance of our algorithm and to determine the threshold values. For the CDOC algorithms based on Diff and Prt, we generated confusion matrices with and without the application of the HCF (Tables 2, 3).

The cloud decision algorithm based on Diff outperforms the algorithm based on Prt. The Diff algorithm has a high accuracy in classifying thick cloud and clear sky pixels. However, the accuracy is smaller for pixels with thin clouds. The HCF improved the Diff thin pixel accuracy by 5 points (Table 2). Especially noteworthy is the very low likelihood of Diff clear/thick cloud confusion; less than 2 % of clear pixels were classified as thick clouds and less than 3 % of thick clouds were classified as clear. The low accuracy for thin clouds is at least partly related to the biases in the manual classification due to human error by the observer; visually it is hard to delineate the "cloud edges" of thin clouds. Moreover, thin clouds usually have gaps of clear skies and do not have uniform textures. This is reflected in the overlap in Diff and Prt values between the classes that is evident in the PDFs (Figs. 7, 8). Thus, with a fixed threshold we are bound to misclassify pixels that lie in the overlap region. We will base our cloud decision algorithm on $\mathrm{Diff}^{\mathrm{HCF}}$ as it has yielded more accurate results.

Table 1. Confusion matrix for CDOC. For example, true class 11 (TC11) denotes the percentage of clear pixels that were correctly classified as clear pixels, false class 12 (FC12) denotes the percentage of clear pixels that were classified as thin cloud pixels, false class 13 (FC13) denotes the percentage of clear pixels that were classified as thick cloud pixels, false class 21 (FC21) denotes the percentage of thin cloud pixels that were classified as clear pixels, and so on.

Manual	Algorithm Classification		
Classification	Clear (1)	Thin (2)	Thick (3)
Clear (1)	TC11	FC12	FC13
Thin (2)	FC21	TC22	FC23
Thick (3)	FC31	FC32	TC33

Table 2. Confusion matrix for CDOC of training set based on the Diff and Diff^{HCF} metric (Eq. 5, 6 respectively) All values are in [%].

Manual	Algorithm Classification					
Classification	Clear (1)		Thin (2)		Thick (3)	
	Diff	Diff^{HCF}	Diff	Diff^{HCF}	Diff	Diff^{HCF}
Clear (1)	**81.4**	**80.1**	16.8	18.1	1.8	1.8
Thin (2)	22.5	17.3	**59.8**	**65.0**	17.7	17.7
Thick (3)	2.0	2.3	17.6	17.3	**80.4**	**80.4**

While the results in Tables 2, 3 were obtained for challenging conditions with a mixture of cloud types, the CDOC algorithm has high accuracies for classifying clear sky images as will be shown in the next section (however, note that much of the more challenging circumsolar region is not considered due to the shadowband, Fig. 4). Thus, manual inspection is only required for one or two clear days to initialize the CSL. Afterwards, days during which no clouds are detected by the algorithm can readily be added to the CSL.

4.2 Validation set

In order to evaluate the performance of the CDOC algorithm, an "out-of-sample" set of 30 manually annotated images was chosen. To avoid biasing the selection of images towards particular sky conditions, we used images collected within 30 min of solar noon for 12 to 16 February and 16 to 20 April. These periods were chosen because – at this site – they represented a large range in aerosol content from 0.1–0.13 in the April set compared to the 0.017–0.059 in the February set. Reviewing the algorithm's classification accuracies at different SZAs (39–65°) in the training set at a different site (not shown), there was no considerable change in accuracy with SZA. Thus, images captured around solar noon were chosen for the validation set.

Table 4 shows the CDOC performance metrics by image and certain images from the set are illustrated in Fig. 9. For

Table 3. Confusion matrix for CDOC of training set based on the Prt metric (Eqs. 7, 8 respectively). All values are in [%].

Manual	Algorithm Classification					
Classification	Clear (1)		Thin (2)		Thick (3)	
	Prt	Prt^{HCF}	Prt	Prt^{HCF}	Prt	Prt^{HCF}
Clear (1)	**65.0**	**66.1**	31.4	30.2	3.6	3.7
Thin (2)	11.3	7.2	**49.6**	**53.7**	39.1	39.1
Thick (3)	2.8	2.8	16.6	16.6	**80.6**	**80.6**

overcast skies (100 % thick, Table 4, Images 7–10, 13–15, 25, 29; Fig. 9a, b), the CDOC algorithm accurately classified more than 95 % of the pixels. However, in some cases even after the lower thick cloud threshold was applied in the solar region (Sect. 3.3), thick cloud pixels were incorrectly classified as thin clouds due to the high RBR of the CSL (Fig. 4c). In the case of clear skies, the algorithm was on average over 99 % accurate (Table 4, Images 1–6, 16–18, 22; Fig. 9c, d), but a few pixels were misclassified as thin clouds due to objects or corrosion present on the mirror, especially in the solar region.

For the case of skies with few thin clouds (Table 4, Image 12; Fig. 9e, f), 91 % of pixels were correctly classified. Discrepancies between visual and automated classification can be explained by inaccuracies of the visual classifier. For broken skies with a mixture of thick and thin clouds (Fig. 9g, h, i, j) the algorithm performed close to that of manual classification (Table 4, Images 19, 24).

A confusion matrix was generated to determine the performance of the algorithm for the validation set (Table 5). We note that the accuracy for clear sky pixel identification and thick cloud pixel identification is higher than that of the test sample (Table 2) because for randomly chosen images there are more cases of completely or predominantly clear or overcast images which simplify CDOC (Table 4). Thin clouds pixels have a lower classification accuracy compared to the other classes as their Diff^{HCF} value falls into a transition region between clear sky and thick cloud pixels which create difficulties as discussed in Sect. 4.1.

4.3 Comparison to fixed thresholding technique

We compared the CDOC algorithm against classifying the RBR image based on fixed uniform thresholds used in the original TSI algorithm (Long et al., 2006). The TSI algorithms now shipped with the instrument fit a predetermined function to vary the clear/think/thick RBR thresholds across the image depending on the sun-pixel distance. To apply the technique, we first used the training set to determine the optimal RBR thresholds to yield the highest accuracies for the three classes (Table 6). Then, we applied these thresholds to the validation set (Table 7).

Comparing both methods, we see that the CDOC method is superior to the fixed threshold method as higher accuracies

Table 4. Results of manual classification and CDOC algorithm for the validation images as well as AOD measurements at 500 nm averaged during the time period of the sky images. Note that for overcast skies (18–20 April) there are no AOD measurements.

Date in 2011	AOD 500 nm	Image #	Manual Classification			CDOC Algorithm		
			Clear (%)	Thin (%)	Thick (%)	Clear (%)	Thin (%)	Thick (%)
16 Apr	0.10	1	100.0	0.0	0.0	98.3	1.4	0.3
		2	100.0	0.0	0.0	99.5	0.2	0.3
		3	100.0	0.0	0.0	99.7	0.2	0.1
17 Apr	0.13	4	100.0	0.0	0.0	99.2	0.6	0.3
		5	100.0	0.0	0.0	95.3	4.4	0.3
		6	100.0	0.0	0.0	99.2	0.5	0.3
18 Apr	–	7	0.0	0.0	100.0	0.0	3.6	96.4
		8	0.0	0.0	100.0	0.0	6.8	93.2
		9	0.0	0.0	100.0	0.0	2.9	97.1
19 Apr	–	10	0.0	0.6	99.5	0.0	1.7	98.3
		11	5.1	10.5	88.7	0.0	11.4	88.6
		12	94.3	5.6	0.7	84.7	14.3	1.0
20 Apr	–	13	0.0	0.0	100.0	0.0	2.6	97.4
		14	0.0	0.0	100.0	0.0	0.1	99.9
		15	0.0	0.0	100.0	0.0	0.6	99.4
12 Feb	0.017	16	100.0	0.0	0.0	100.0	0.0	0.0
		17	100.0	0.0	0.0	100.0	0.0	0.0
		18	100.0	0.0	0.0	100.0	0.0	0.0
13 Feb	0.031	19	27.1	28.6	44.7	29.4	26.9	43.7
		20	64.3	8.4	27.4	61.4	16.7	21.9
		21	12.4	16.8	71.7	5.3	19.6	75.1
14 Feb	0.065	22	100.0	0.0	0.0	99.9	0.0	0.0
		23	41.0	39.9	19.8	44.7	31.9	23.3
		24	20.6	25.7	54.0	21.2	33.8	45.0
15 Feb	0.059	25	0.0	0.0	100.0	0.0	0.0	100.0
		26	13.9	1.0	85.2	15.0	9.6	75.4
		27	96.2	1.3	2.5	97.8	1.6	0.6
16 Feb	0.057	28	4.0	6.4	90.6	0.0	6.8	93.2
		29	0.0	2.9	98.9	0.0	0.6	99.4
		30	2.4	30.9	72.4	0.0	18.3	81.7

were obtained for all three classes for both the training and validation set. That is, in the validation for clear, thin, and thick cloud, respectively, the CDOC algorithm provided 96.0 %, 60.0 %, and 96.3 % accuracy, as compared with 89.3 %, 56.1 %, and 91.5 %, even though we had optimized the fixed RBR thresholds. One of the short comings of the fixed thresholds method is that the threshold values need to be modified throughout the year to account for changes in AOD and degradation and/or soiling of the mirror. Also, the classification accuracy is sensitive to the thresholds chosen. Our CDOC algorithm, on the other hand, classifies pixels by comparing them to a CSL that is modified through the year to account for changes in aerosol content, solar position and instrument degradation.

Table 5. Confusion matrix for CDOC of validation set based on the Diff^{HCF} metric (Eq. 6). All values are in [%] and add up to 100 % across rows.

Manual	Algorithm Classification		
Classification	Clear (1)	Thin (2)	Thick (3)
Clear (1)	**96.0**	3.1	0.9
Thin (2)	11.6	**60.0**	28.4
Thick (3)	0.3	3.4	**96.3**
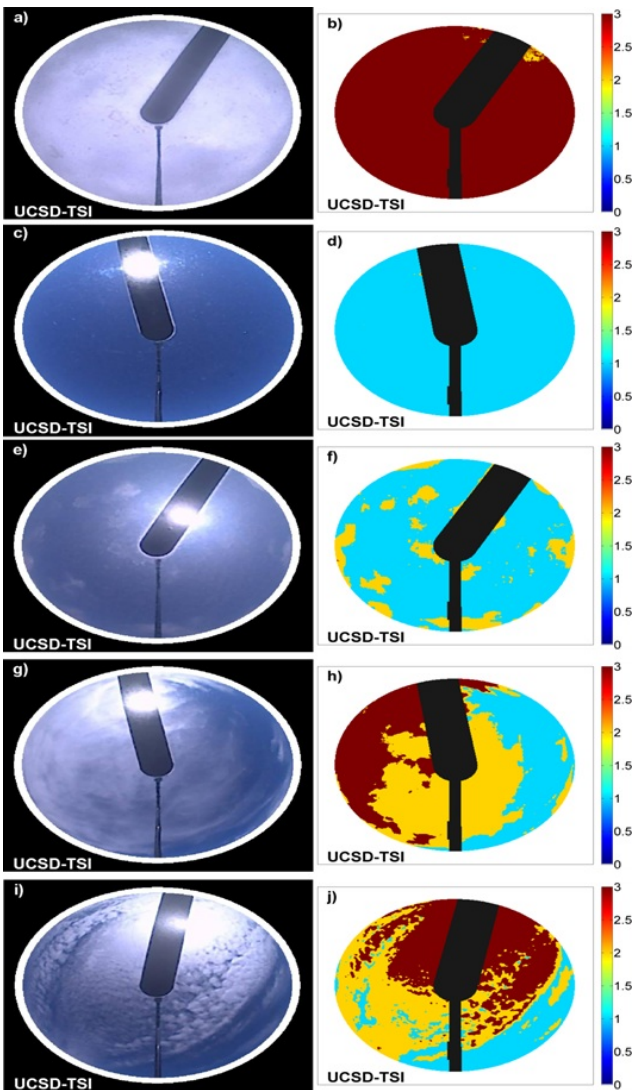
Fig. 9. Total sky image **(a, c, e, g, i)** and CDOC **(b, d, f, h, j)** for: (1) overcast skies **(a, b)** taken on 20 April 2011 and corresponding to image 15 in Table 4; (2) clear skies **(c, d)** taken on 12 February 2011 and corresponding to image 16 in Table 4; (3) few thin clouds **(e, f)** taken on 19 April 2011, and corresponding to image 12 in Table 4; (4) partly cloudy skies **(g, h)**, **(i, j)** taken on 13 February 2011, and 14 February 2011, and corresponding to images 19 and 24 in Table 4, respectively. For the classification images, a value of 3 on the color scale represents thick clouds, 2 represents thin clouds, and 1 represents clear skies.

5 Conclusions

The purpose of this study was to develop a methodology to automatically classify clear skies, thin cloud, and thick cloud image pixels obtained from a ground-based sky imager. This method was applied to Total Sky Imager (TSI) imagery. The red-blue ratio (RBR) on cloudless days was shown to be well-correlated to aerosol optical depth (AOD). As a result, a haze correction factor (HCF) was introduced to account for

Table 6. Confusion matrix for CDOC of training set based on fixed RBR thresholds. All values are in [%] and add up to 100 % across rows.

Manual	Algorithm Classification		
Classification	Clear (1)	Thin (2)	Thick (3)
Clear (1)	**80.7**	16.7	2.6
Thin (2)	41.8	**45.0**	13.2
Thick (3)	1.6	17.5	**80.9**

Table 7. Confusion matrix for CDOC of validation set based on fixed RBR thresholds. All values are in [%] and add up to 100 % across rows.

Manual	Algorithm Classification		
Classification	Clear (1)	Thin (2)	Thick (3)
Clear (1)	**89.3**	9.5	1.2
Thin (2)	34.9	**56.1**	9.0
Thick (3)	0.6	7.9	**91.5**

AOD effects on the RBR. By applying the correction factor we were able to better distinguish between haze and thin clouds in the atmosphere. In order to classify the images we compared each pixel's RBR to the corresponding clear sky's RBR that was auto-calibrated using the HCF. CDOC was found to be more accurate when based on the difference in RBR from a clear sky RBR rather than the ratio of RBR to clear sky RBR. Comparing automated and visually classified images, the algorithm was found to be very accurate in classifying thick cloud and clear sky pixel in a variety of sky conditions. Thin cloud pixel classification accuracy was lower, due to a small range of Diff values over which it was classified as well as difficulties in marking and defining thin cloud boundaries.

The method developed provides a significant improvement over the TSI's original software in pixel classification accuracy. The haze correction factor method avoids the need to constantly adjust the threshold values for cloud classification. This paper introduces a method of applying some aspects of the HCF method developed by Shields et al. (1993, 2010) in a manner that can be applied with a TSI instruments. We also found that with our algorithm we got better results using the difference between the image RBR and the CSL's RBR to identify the pixels as thick, thin or clear, rather than the ratios.

The CDOC algorithm will be implemented to improve short-term solar forecast accuracy by improving cloud detection as well providing the added information of cloud opacity.

References

Ault, A. P., Moore, M. J., Furutani, H., and Prather, K. A.: Impact of Emissions from the Los Angeles Port Region on San Diego Air Quality during Regional Transport Events, Environ. Sci. Technol., 43, 3500–3506, 2009.

Calbó, J. and Sabburg, J.: Feature extraction from whole-sky ground-based images for cloud-type recognition, J. Atmos. Ocean. Technol., 25, 3–14, 2008.

Cazorla, A., Olmo, F. J., and Alados-Arboledas, L.: Using a Sky Imager for aerosol characterization, Atmos. Environ., 42, 2739–2745, 2008a.

Cazorla, A., Olmo, F. J., and Alados-Arboledas, L.: Development of a sky imager for cloud cover assessment, J. Opt. Soc. Am, A-Opt. Image Sci. Vis., 25, 29–39, 2008b.

Cazorla, A., Shields, J. E., Karr, M. E., Olmo, F. J., Burden, A., and Alados-Arboledas, L.: Technical Note: Determination of aerosol optical properties by a calibrated sky imager, Atmos. Chem. Phys., 9, 6417–6427, doi:10.5194/acp-9-6417-2009, 2009.

Cess, R. D., Zhang, M. H., Minnis, P., Corsetti, L., Dutton, E. G., Forgan, B. W., Garber, D. P., Gates, W. L., Hack, J. J., Harrison, E. F., Jing, X., Kiehl, J. T., Long, C. N., Morcrette, J.-J., Potter, G. L., Ramanathan, V., Subasilar, B., Whitlock, C. H., Young, D. F., and Zhou, Y.: Absorption of Solar Radiation by Clouds: Observations Versus Models, Science, 267, 496–499, 1995.

Chow, C. W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., and Washom, B.: Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed, Sol. Energy, 85, 2881–2893, 2011.

Ghonima, M. S.: Aerosol effects on Red Blue Ratio of Clear Sky Images, and Impact on Solar Forecasting, M.S. thesis/masters, University of California, San Diego, United States – California, 2011.

Gueymard, C. A. and Thevenard, D.: Monthly average clear-sky broadband irradiance database for worldwide solar heat gain and building cooling load calculations, Sol. Energy 83, 1998–2018, 2009.

Heinle, A., Macke, A., and Srivastav, A.: Automatic cloud classification of whole sky images, Atmos. Meas. Tech., 3, 557–567, doi:10.5194/amt-3-557-2010, 2010.

Holben, B., Eck, T., Slutsker, I., Tanre, D., Buis, J., Setzer, A., Vermote, E., Reagan, J., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., and Smirnov, A.: AERONET – A federated instrument network and data archive for aerosol characterization, Remote Sens. Environ., 66, 1–16, 1998.

Huo, J. and Lü, D. R.: Preliminary retrieval of aerosol optical depth from all-sky images, Adv. Atmos. Sci., 27, 421–426, 2010.

Johnson, R. W., Hering, W. S., and Shields, J. E.: Automated Visibility and Cloud Cover Measurements with a Solid State Imaging System, Marine Physical Laboratory , Scripps Institution

of Oceanography, University of California, San Diego, DTIS (Stinet) File ADA216906, 1989.

Kim, D. and Ramanathan, V.: Solar radiation budget and radiative forcing due to aerosols and clouds, J. Geophys. Res.-Atmos., 113, D02203, doi:10.1029/2007JD008434, 2008.

Kindel, B. C., Pilewskie, P., Schmidt, K. S., Coddington, O., and King, M. D.: Solar spectral absorption by marine stratus clouds: Measurements and modeling, J. Geophys. Res.-Atmos., 116, D10203, doi:10.1029/2010JD015071, 2011.

Kohavi, R. and Provost, F.: Glossary of terms: Special issue on application of machine learning and the knowledge discovery process, Mach. Learn., 30, 271–274, 1998.

Li, Q., Lu, W., and Yang, J.: A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images, J. Atmos. Ocean. Technol., 28, 1286–1296, 2011.

Long, C., Sabburg, J., Calbó, J., and Pages, D.: Retrieving cloud characteristics from ground-based daytime color all-sky images, J. Atmos. Ocean. Technol., 23, 633–652, 2006.

Luoma, J., Kleissl, J., and Murray, K.: Optimal inverter sizing considering cloud enhancement, Sol. Energy, 86, 421–429, 2012.

Neto, S. L. M., von Wangenheim, A., Pereira, E. B., and Comunello, E.: The Use of Euclidean Geometric Distance on RGB Color Space for the Classification of Sky and Cloud Patterns, J. Atmos. Ocean. Technol., 27, 1504–1517, 2010.

Pfister, G., McKenzie, R., Liley, J., Thomas, A., Forgan, B., and Long, C.: Cloud coverage based on all-sky imaging and its impact on surface solar irradiance, J. Appl. Meteorol., 42, 1421–1434, 2003.

Rossow, W. and Schiffer, R.: Advances in understanding clouds from ISCCP, Bull. Am. Meteorol. Soc., 80, 2261–2287, 1999.

Seiz, G., Shields, J., Feister, U., Baltsavias, E. P., and Gruen, A.: Cloud mapping with ground-based photogrammetric cameras, Int. J. Remote Sens., 28, 2001–2032, 2007.

Shields, J. E., Johnson, R. W., and Koehler, T. L.: Automated whole sky imaging systems for cloud field assessment, Fourth Symposium on Global Change Studies, American Meteorological Society, 1993.

Shields, J. E., Karr, M., Tooman, T., Sowle, D., and Moore, S.: The whole sky imager – a year of progress. Presented at the Eighth Atmospheric Radiation Measurement (ARM) Science Team Meeting, Tucson, Arizona, 23–27 March 1998.

Shields, J. E., Johnson, R. W., Karr, M. E., and Wertz, J. L.: Automated day/night whole sky imagers for field assessment of cloud cover distributions and radiance distributions, in: Proc. 10th Symp. on Meteorological Observations and Instrumentation, 11–16 January, Boston, MA, 1998.

Shields, J. E., Karr, M. E., Burden, A. R., Johnson, R. W., Mikuls, V. W., Streeter, J. R., and Hodgkiss, W. S.: Research toward Multi-Site Characterization of Sky Obscuration by Clouds, Final Report for Grant N00244-07-1-009, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California San Diego, Technical Note 274, DTIS (Stinet) File ADA126296, 2009.

Shields, J. E., Karr, M. E., Burden, A. R., Mikuls, V. W., Streeter, J. R., Johnson, R. W., and Hodgkiss, W. S.: Scientific Report on Whole Sky Imager Characterization of Sky Obscuration by Clouds for the Starfire Optical Range, Scientific Report for AFRL Contract FA9451-008-C-0226, Marine Physical Laboratory, Scripps Institution of Oceanography, University of Califor-

nia San Diego, DTIS (Stinet) File ADB367708, 2010.

Román, R., Antón, M., Cazorla, A., de Miguel, A., Olmo, F. J., Bilbao, J., and Alados-Arboledas, L.: Calibration of an all-sky camera for obtaining sky radiance at three wavelengths, Atmos. Meas. Tech., 5, 2013–2024, doi:10.5194/amt-5-2013-2012, 2012.

Smirnov, A., Holben, B., Eck, T., Dubovik, O., and Slutsker, I.: Cloud-screening and quality control algorithms for the AERONET database, Remote Sens. Environ., 73, 337–349, 2000.

Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K. B., Tignor, M., and Miller, H. L.: Climate Change 2007: The Physical Science Basis – Contribution of Working group I to the Fourth Assessment Report of the IPCC, Cambridge University Press, 2007.

Souza-Echer, M., Pereira, E., B., Bins, L., and Andrade, M.: A simple method for the assessment of the cloud cover state in high-latitude regions by a ground-based digital camera, J. Atmos. Ocean. Technol., 23, 437–447, 2006.

Twohy, C., Petters, M., Snider, J., Stevens, B., Tahnk, W., Wetzel, M., Russell, L., and Burnet, F.: Evaluation of the aerosol indirect effect in marine stratocumulus clouds: Droplet number, size, liquid water path, and radiative impact, J. Geophys. Res.-Atmos., 110, D08203, doi:10.1029/2004JD005116, 2005.

Yamashita M., Yoshimura M., and Iwao, K.: Monitoring and Discrimination for Sky Conditions Using Multi-temporal Whole Sky Imageries, ACRS Proceedings, Hanoi, 2005.

Zhao, G. and Di Girolamo, L.: Cloud fraction errors for trade wind cumuli from EOS-Terra instruments, Geophys. Res. Lett., 33, L20802, doi:10.1029/2006GL027088, 2006.