Retrieving Cloud Characteristics from Ground-Based Daytime Color All-Sky Images

C. N. LONG

Pacific Northwest National Laboratory, Richland, Washington

J. M. SABBURG

Department of Biological and Physical Sciences, Faculty of Sciences, University of Southern Queensland, Toowoomba, Australia

J. CALBÓ

Department of Physics, and Institute of the Environment, University of Girona, Girona, Spain

D. PAGÈS

Institute of the Environment, University of Girona, Girona, Spain

ABSTRACT

A discussion is presented of daytime sky imaging and techniques that may be applied to the analysis of full-color sky images to infer cloud macrophysical properties. Descriptions of two different types of sky-imaging systems developed by the authors are presented, one of which has been developed into a commercially available instrument. Retrievals of fractional sky cover from automated processing methods are compared to human retrievals, both from direct observations and visual analyses of sky images. Although some uncertainty exists in fractional sky cover retrievals from sky images, this uncertainty is no greater than that attached to human observations for the commercially available sky-imager retrievals. Thus, the application of automatic digital image processing techniques on sky images is a useful method to complement, or even replace, traditional human observations of sky cover and, potentially, cloud type. Additionally, the possibilities for inferring other cloud parameters such as cloud brokenness and solar obstruction further enhance the usefulness of sky imagers.

1. Introduction

In recent years, atmospheric researchers have become increasingly interested in quantifying clouds. Clouds are a major meteorological phenomena related to the hydrological cycle and affect the energy balance on both local and global scales through interaction with solar and terrestrial radiation. It is broadly recognized that clouds (and cloud–aerosol interaction) are responsible for the largest uncertainties in climate models and climate predictions (Houghton et al. 2001). In addition, clouds affect our everyday lives, for example, by modi-

fying the amount of ultraviolet (UV) radiation that reaches the earth's surface (Calbó et al. 2005). Most cloud-related studies require some sort of cloud observations, such as the amount and type of clouds that are present. These macrophysical observations have been performed historically by human observers who recorded cloud cover and cloud type at several meteorological stations and at given time intervals (typically hourly at many U.S. sites, 3 hourly at many other sites worldwide). However, high costs associated with human observers have led observations toward automatic devices to detect and quantify cloud amount and type. There is, of course, satellite information, but satellite retrievals have known weaknesses in quantifying small and/or low cloud features due to their limited spatial resolution and unknown surface influences on the measured radiances.

One option for obtaining continuous information on

Corresponding author address: Dr. Charles N. Long, Atmospheric Radiation Measurement Program, Pacific Northwest National Laboratory, P.O. Box 999, MSIN: K9-38, Richland, WA 99352.
E-mail: Chuck.Long@pnl.gov

sky conditions is the use of sky-imaging devices. In two recent publications (Calbó et al. 2005; Parisi et al. 2004) there are overviews on atmospheric cloud detection, the importance of clouds with respect to the earth's climate and human health, and how ground-based sky imagers can complement the coverage of equivalent satellite instruments. These papers highlight the increased number of ground-based sky imagers being developed in several countries. This development is partly due to the dramatic improvements in technology in recent years, both with respect to the hardware, for example, charge-coupled devices (CCDs) and digital image processing techniques. Most of these imagers have been purposely constructed with a specific application in mind, for example, the first integrated sun-centered sky camera (SCSC) for solar UV research (Sabburg and Wong 1999), while other imagers have the general purpose of measuring cloud macroscopic characteristics (Pagès et al. 2002; Lu et al. 2004; Long and DeLuisi 1998). A family of Whole Sky Imagers (WSIs), developed by the Scripps Institution of Oceanography at the University of California, San Diego, are designed to measure radiances at distinct wavelength bands across the hemisphere (Johnson et al. 1989; Shields et al. 2003). The various models of the WSI include a high-quality temperature-controlled CDD, high-quality optics including spectral filtering, detailed mapping of the sky dome to CCD element, and careful calibration of dark current and stray light influences needed to make the scientific-quality spectral radiance measurements. The WSI data, besides having many other interesting scientific capabilities, can be used to estimate fractional sky cover (Tooman 2003; Johnson et al. 1989). Unfortunately, because of the high-quality components and sophisticated engineering involved, the significantly higher cost puts the WSI beyond the means of many individual researchers and research groups whose only interest lies in inferring daytime fractional sky cover. Commercially, there are very few nonradiance sky camera systems available. One of the better-known sky imagers to the atmospheric science community is the total sky imager (TSI) manufactured by Yankee Environmental Systems, Inc. (YES), Massachusetts.

There are advantages to developing specific-purpose imagers; for example, the user has a detailed knowledge of the prototype system and can choose components that best suit a specific task. However, this development process requires extensive knowledge of detector and optical systems, as well as development of the interface algorithms needed to construct actual images in a format using the readings of the detector array. In addition, given the implications inherent in the word "prototype," the reliability of a commercial instrument may likely be an advantage if the instrument operates for extended periods (years) with minimal operational problems. With this in mind, Sabburg and Long (2004) developed three new image-processing algorithms for the TSI, similar functionally to the algorithms used with the SCSC mentioned previously. The SCSC had few operational problems during the first year of operation, but a more robust design was needed to investigate the effects of clouds on UV enhancement over a time period greater than 2 yr. One solution was the purchase of a commercially available sky imager.

Our paper discusses image-processing techniques that can be used to obtain standard (i.e., fractional sky cover) and other sky characteristics from daytime ground-based, all-sky images. This paper provides information for researchers to use with images obtained from similar all-sky camera systems. Our examples use two sky imagers: the TSI and the whole sky camera (WSC) developed by Spain's University of Girona (Pagès et al. 2002). Technical information for both imagers are presented including elaboration on the TSI-type methodology previously reported by Long et al. (2001) and Pfister et al. (2003). Topics and concepts include shadowband and horizon masking, pixel classification (e.g., thresholding), geometric corrections, and other sky characteristics such as solar obstruction, cloud brokenness (CB), and cloud patterns. Results are presented to show how some of these concepts are applied to sample images from the TSI and WSC compared to direct sky observations as well as visual inspections of the images.

2. Sky imagers used

a. Hemispheric Sky Imager/total sky imager

The prototype Hemispheric Sky Imager (HSI), which was the precursor to the commercial TSI, was developed at the National Oceanic and Atmospheric Administration Surface Radiation Research Branch (SRRB) located in Boulder, Colorado (Long and DeLuisi 1998). The development was a joint effort between SRRB and the U.S. Department of Energy's Atmospheric Radiation Measurement (ARM) Program. This instrument was developed into a commercial product named the TSI in a cooperative effort with YES under a Small Business Innovative Research grant from the U.S. Department of Agriculture.

The basic design of both the HSI and TSI includes a digital camera mounted to look down on a curved mirror (Fig. 1) to provide a horizon-to-horizon view of the sky. The mirror rotates to keep a dull black strip on the mirror aligned with the solar azimuth angle to block the direct sun from the camera. The mirror rotation is ac-

FIG. 1. YES TSI deployed at the ARM Climate Research Facility Southern Great Plains site in Oklahoma.

tively controlled by a small onboard computer. The TSI captures images of the sky during daylight hours and can be set to do so as often as every 10 s (depending on ancillary image-processing computer processor speed). The sky images are 24-bit color JPEG format at 352 × 288 pixel resolution captured from a camera based on a commercially available digital camera. More information on TSI specifications is given in Table 1.

The raw sky images are processed to suggest fractional sky cover. Note that the TSI sky cover retrievals are generally valid only for solar elevation angles

TABLE 1. Technical specifications of the TSI.

Characteristic	Specification
Imager resolution	352 × 288, 24-bit color
Sampling rate	Variable, with a maximum of one image every 10 s
Operating temperature	−30° to 34°C
Weight	Approximately 50 lbs (23 kg)
Power requirements	110/220 VAC
Data storage	Disk on local computer or remote computer over a full-time TCP/IP connection

greater than 10° (zenith angles less than 80°), and images are processed for a maximum 160° field of view (FOV), ignoring the 10° of sky near the horizon. The user can configure the time between image captures, overall processing FOV, and adjustments to the basic image-processing limits to differentiate between cloudless, thin, and opaque cloud retrievals. The TSI software also allows setting a separate FOV for processing, centered at zenith and of lesser angular width than the overall setting. In addition, the software processes the relative brightness along the sun-blocking strip to act as a "sunshine meter," thus suggesting whether or not the sun is blocked by a cloud. An example of a TSI sky image and the corresponding cloud decision image is shown in Fig. 2 (top). The cloud decision image displays the 160° FOV of the retrieval, and a second retrieval circle centered on zenith covering a 100° FOV is outlined in green. (Other areas outlined in green are discussed in the next section.) The yellow dot on the sun-blocking strip mask represents the location of the sun in the image, with the color yellow indicating that the sun is not completely obscured by a cloud [similar to the World Meteorological Organization (WMO) specifications of sun obscuration (Pfister et al. 2003)]. If the sun is blocked by a cloud, the dot is colored white.

b. Whole sky camera

The WSC was developed at University of Girona (UdG) to collect a continuous record of the sky conditions at low cost and was used for research associated with radiative transfer phenomena in the atmosphere. The WSC consists of three different components: the CCD camera and optics, the graphics card and corresponding software, and the enclosure and various protections. A picture of the device and a schematic drawing showing its parts are presented in Fig. 3. The main component is a commercial CCD color video camera equipped with a "fish eye" zoom lens. The focal length is fixed to 1.6 mm to project a 180° FOV onto the 1/3-in. (8.3 mm) CCD matrix. Both the focus and the iris are manually operated and were fixed at the optimum values for sky viewing after initial tests. Various technical specifications of the system are listed in Table 2. The camera sends an analog signal to the graphics card, which captures and stores the image in digital format. Based upon the software package that comes with the graphics card, we developed software to control the capture and recording of the images, including switching on and off the capture function in relation to the time of sunrise and sunset. In addition, the same code controls the time interval between image capture and the storage format of the captured images. Currently,
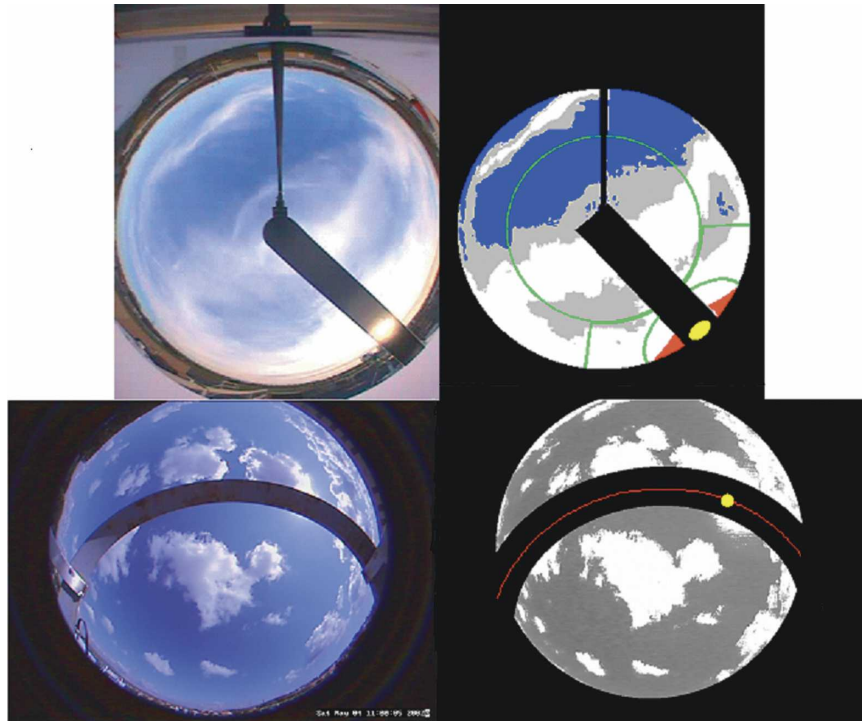
FIG. 2. (top left) TSI sky image and (top right) corresponding 160° FOV cloud decision image, (bottom left) WSC sky image, and (bottom right) corresponding 160° FOV cloud decision image. For the TSI cloud decision image, blue represents retrieved cloud-free pixels, gray represents thin cloud, white represents opaque cloud, and black represents masked pixels that are not counted in determining fractional sky cover. The green outlines denote special retrieval areas discussed in the text. The yellow dot on the sun-blocking strip mask denotes the sun location in the image. For the WSC processed image, retrieved cloud-free pixels are shown in gray, while cloudy pixels are in white. Black represents masked pixels that are not counted in determining fractional sky cover. The red line represents the path of the sun through the image FOV, and the yellow dot denotes the sun location in this image.

one bitmap (BMP) image is recorded every 15 min and one JPEG (JPG) image every minute.

The CCD camera is placed inside an enclosure, which is built in two layers. The first one covers the camera itself and contains two thermostats and connections for the power cable and signal cables. The second one protects the whole device from rain and other environmental factors. One thermostat controls a Peltier cell to refrigerate the air around the camera if needed. The Peltier cell has a maximum working temperature of 50°C. This cooling is needed when the camera is exposed to high summer temperatures (air temperature around 35°C) with accompanying direct sunlight, and the camera also generating some heat of its own. This Peltier system is turned on when the first thermostat records temperatures higher than 35°C. The second thermostat is for backup, that is, to protect the camera if temperatures reach 40°C. In this case, the camera is also switched off. Low temperatures are not a problem in the climate of Girona, where temperatures below

0°C are rare, and the heat produced by the camera is enough to avoid freezing temperatures. A hemispherical glass dome is used to protect the lens from the environment, while allowing a view of the sky. A shadowband, similar to those used for diffuse radiation measurements, is used to protect the CCD sensor from the direct sun. This sun-blocking arrangement makes it difficult for the detection of some sky characteristics, such as sun obstruction, because in this design the sun is always occulted behind the shadowband. However, the shadowband is required, because the CCD could be damaged by a continuous exposure to high radiation intensities. Due to the seasonal change of sun height, the position of the shadowband is manually adjusted periodically (approximately once a week). Since the summer of 2001, the WSC has been operated on the roof of a building of the University of Girona (41.97°N, 2.82°E; 100-m altitude), along with other meteorological and radiometric instrumentation. This site enjoys an open horizon, meaning that no obstacle in the horizon
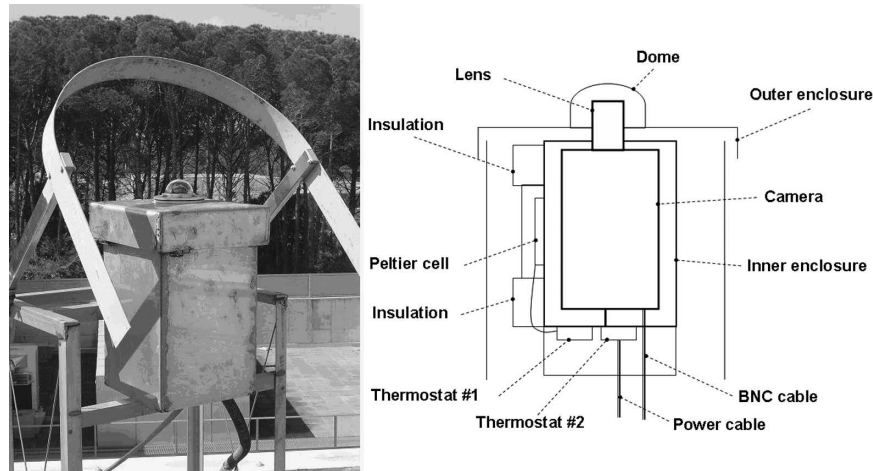
FIG. 3. (left) Picture of the WSC with shadowband, and (right) a schematic drawing of the main parts.

subtends an angle greater than 10°, so there are no obstructions over the 160° FOV currently analyzed. An example WSC image and the corresponding processed image are shown in Fig. 2 (bottom).

3. Image processing

a. Previous treatments

In processing raw sky images, several areas need to be identified before the actual cloud/clear pixel accounting is performed. The overall images themselves are generally rectangular in shape, whereas the mapping of the sky dome onto the image is circular. In addition, due to the natural radiative scattering characteristics of the atmosphere and the long pathlengths, unambiguous determination between clear sky and cloudy sky is difficult near the horizon. For this reason, both the WSC and TSI process only a 160° FOV of the sky image centered on zenith, resulting in a loss of about 17% of the hemispherical solid angle of the sky dome. The rest of the pixels in the image area outside

TABLE 2. Technical characteristics of the WSC.

Camera and optics
Color camera CCD: 1/3″
Minimum illumination: 0.5 lx at f 1.2
Image resolution: 752(H) × 582(V)
Working temperature: −10° to +50°C
Lens: fish-eye zoom, 1.6–3.4 mm at f 1.4
FOV: up to 180°
Enclosure and other
Protecting container with glass dome
Shadowband (radius, 635 mm; width, 73 mm)
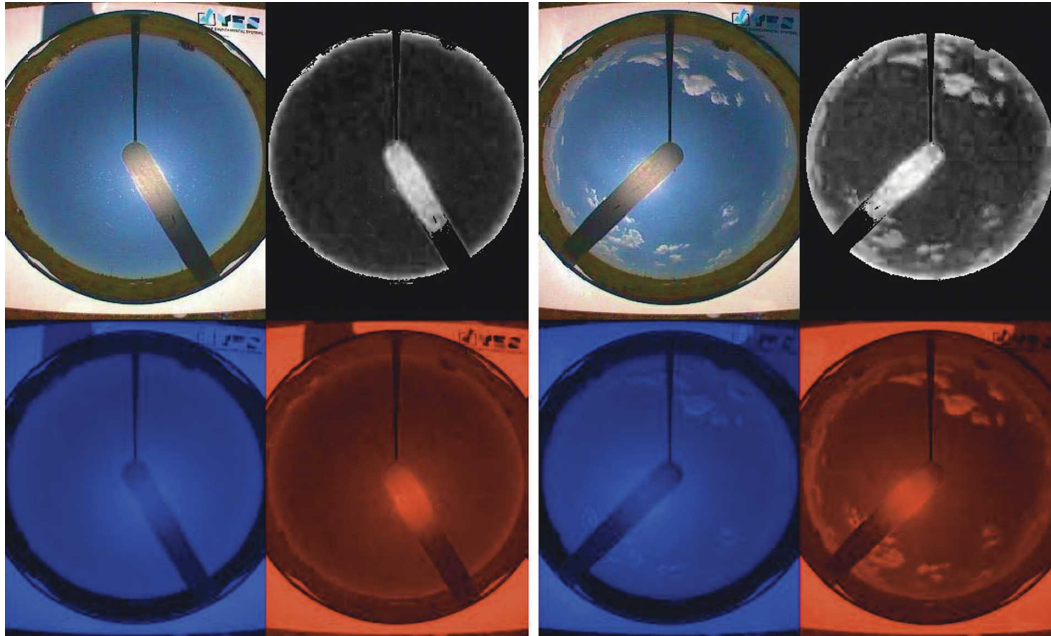Temperature controlled by two thermostats and a Peltier cell

this 160° FOV circle is set to a "mask" value corresponding to "black" and are ignored in the clear/cloudy pixel accounting for all TSI/HSI/WSC processing presented here.

Of the remaining 160° FOV, part of the area is covered by the shadowband (WSC) or the black sun-blocking strip (TSI) that is used to protect the camera from the direct sun. The TSI additionally must ignore the camera arm and the camera housing directly overhead. The pixels corresponding to these objects are also masked and not included in the clear/cloudy pixel accounting. For the TSI, because the black strip moves during the day with the azimuthal position of the sun, a different mask must be applied to each image depending on the time of day. However, the same masks are valid for all days in the year. The fraction of the 160° FOV image that is lost when masking the black strip, and camera arm and housing, is about 8%. For the WSC, the shadowband is manually adjusted every 2 to 5 days. We have defined a different mask to hide the shadowband for each day in the year, allowing the same mask to be applied to all images for a given day. The masks were obtained by drawing a strip of sufficient width (around 60 pixels in our case) that follows the daily trajectory of the solar disk derived from astronomical formulas. The fraction of area of the image that is hidden for these masks is between approximately 10% in winter and 15% in summer. See Fig. 2 for examples of how this masking affects the image processing.

In the cloudless sky, the area near the sun is most often whiter and brighter than the rest of the hemisphere due to the forward scattering by aerosols and haze. Even a slight haze or moderate aerosol loading

FIG. 4. (top left) Clear-sky image taken by the TSI, (top second from left) corresponding relative red/blue ratio "image," (lower left) separated blue, and (lower second from left) red pixel value amount images. (top third from left) Cloudy-sky image, (top right) corresponding relative red/blue ratio "image," (lower third from left) separated blue, and (lower right) red pixel value amount images.

will make a large angular area of the horizon whiter and brighter when the sun is low on the horizon. The human eye has an amazing ability to handle a range of light intensity spanning orders of magnitude. One of the problems in using commercial digital cameras such as those used in the TSI and WSC is the intensity range limitations of the camera detector. It is desirable to have images bright enough to detect thin clouds, yet this might lead to the part of the image near the sun and near the horizon for low sun appearing whiter in the images than they actually are, not because that is the color perceived by the human eye, but because the commercial CCD elements could produce an exaggerated relative signal. But even for high-quality detectors such as those used in the WSI, these areas of the image are naturally whiter than other parts of the cloudless sky in the image due to the forward scattering. With no a priori knowledge of the aerosol or haze loading that can be used in some way to predict an increased brightness, these pixels are often interpreted as "cloudy" in the sky-imager retrievals when a human observer would label them as "cloudless." The TSI software allows user-configurable settings to keep separate additional accounting of clear/thin/opaque determinations for these two "problem areas." The "sun circle" setting is in terms of an angular FOV centered on the sun position in the sky image. For the "horizon area" there

are two settings, one in terms of angular height above the horizon, and the other in angular width centered on the solar azimuth. These two special areas, along with a sample zenith circle retrieval discussed in section 2, are shown in the cloud decision image in Fig. 2 outlined in green. For this retrieval, the zenith circle is set to an FOV of 100°, the sun circle radius is set to 25°, and the horizon area is set to an elevation of 40° with a total angular width of 100°. While the TSI software itself only produces separate total, thin, and opaque pixel counts for these areas, an additional analysis, such as that described in Pfister et al. (2003), can be performed to help determine if these areas should be counted as cloudy or not [see section 4a(1)].

b. Pixel classification

For molecular scattering (clear skies, no aerosols), more blue light is scattered than red, which is why the clear sky appears blue to our eyes. A sample TSI image of clear sky is shown in Fig. 4. Below this sample sky image are two images that show the corresponding extracted red–green–blue (RGB) color channel blue and red pixel values that make up the sample image. The red pixel values are relatively small (dark) in the sky portion of the image because little red light is scattered by this clear atmosphere compared to the correspondingly greater blue scattering and greater blue pixel val-

ues, except near the horizon where the increased atmospheric pathlength makes the original sky image appear white to our eyes, and somewhat near the sun in the image. The corresponding relative red/blue ratio values are shown in the upper second from left image in Fig. 4. For clear sky the red/blue ratio is small, that is, dark in the image, but increasing near the sun and near the horizon. Additionally, although not shown here, the clear-sky relative red/blue ratio value for any given pixel changes with solar elevation. Thus a clear-sky limit for a given pixel should, for better results, be based as a function of solar elevation, the pixels' distance from zenith, and the pixels' distance from the sun location in the image.

Clouds, unlike the clear sky, generally scatter both the blue and red visible light more equally. A sample TSI image of a partly cloudy sky is also shown in Fig. 4. As in the Fig. 4 clear-sky case, below this sample cloudy-sky image are two images that show the corresponding extracted blue and red pixel values that make up the sample image. In this case, where there are clouds present, the red pixel values are much greater than where there are not clouds. The blue pixel image shows far less contrast in pixel values. The relative ratio of red/blue pixel values (Fig. 4, upper right) clearly shows that the ratio is greater for clouds than for clear sky.

The concept of the red/blue ratio for cloud algorithms was first developed at Scripps Institution of Oceanography with the WSI. A variable threshold algorithm similar to that used with the TSI is described in Koehler et al. (1991). A lower limit is set for a clear-sky ratio value for each pixel in the image, and the pixels for which the red/blue ratio exceeds the clear limit are counted as "cloudy." It must be noted that the particular limit function is climate and camera dependent. For example, if one takes three digital cameras (even of the same make and model), takes an image of the sky with the three cameras simultaneously, and then simultaneously displays all three images on the same screen, one will note that each image displays a slightly different color rendering. In addition, how white tinted the "blue" of the sky appears, which is considered to be "cloud free," relates to such factors as the typical aerosol loading and pressure depth of the atmosphere at a given location. One way to account for these effects is to tailor the clear/cloud limit specifically for a given camera and location, as is the case for WSC at the Girona site. Alternatively, as was the case with the SCSC, a solar sensor might be used to adjust the threshold limit based on the brightness of the sky. But because the TSI is a commercial instrument intended to be deployed at many locations, a generic baseline clear-sky function has been established using the pixel distance from center, distance from the sun, and solar zenith angle (SZA) as independent variables. The user then uses configurable settings to set the clear/thin and thin/opaque limits as desired, to a first approximation as a percentage offset from the baseline value for that pixel. The results of the above TSI type of processing are depicted in the cloud decision image shown in Fig. 2.

For the case of the WSC, pixel classification is based on the same grounds, but with a simpler approach. A single threshold is used in the whole nonmasked area of the images. A fixed threshold algorithm like that used with the WSC is presented in Johnson et al. (1988). Specifically, pixels with a red-to-blue signal ratio (R/B) greater that 0.6 are classified as cloudy, while lower values of R/B are labeled as cloud-free. The value of the threshold was set based on several tests performed on training images. The set of training images contained some 100 images and is a subset of the images used to assess the performance of the WSC automatic retrieval (see section 4b). Initially, some images were analyzed by using different thresholds and subsequently visually compared with the raw images. Second, a specific software package for image processing was used. In these latter tests, different areas of some images were classified subjectively as cloudy or cloudless. Then, R/B was computed for all pixels in these areas. We found that the value that distinguished the best between cloudy and cloudless areas was R/B = 0.6. With the use of a unique threshold, as expected, some problems occur with circumsolar pixels (especially in high aerosol conditions) and misdetection of thin clouds.

c. Fractional sky cover

Once the pixel-by-pixel determination of clear/cloud is made, the estimated fractional sky cover is then calculated, typically as the number of cloudy pixels divided by the total number of pixels in the 160° FOV (ignoring any masked pixels) for both the TSI and WSC. As a first approximation, fractional sky cover f is

$$f = \frac{N_{\text{cloudy}}}{N_{\text{total}}} = \frac{N_{\text{cloudy}}}{N_{\text{cloudy}} + N_{\text{clear}}}, \quad (1)$$

where N denotes the number of pixels of each type, according with the subindices. This is the approximation used in the commercial TSI retrievals. Note that this fractional sky cover is the fractional sky cover in the area of sky that has not been masked in the image. However, this should be a good approximation to the actual fractional sky cover over time, provided that (a) the masking process does not hide a large fraction of
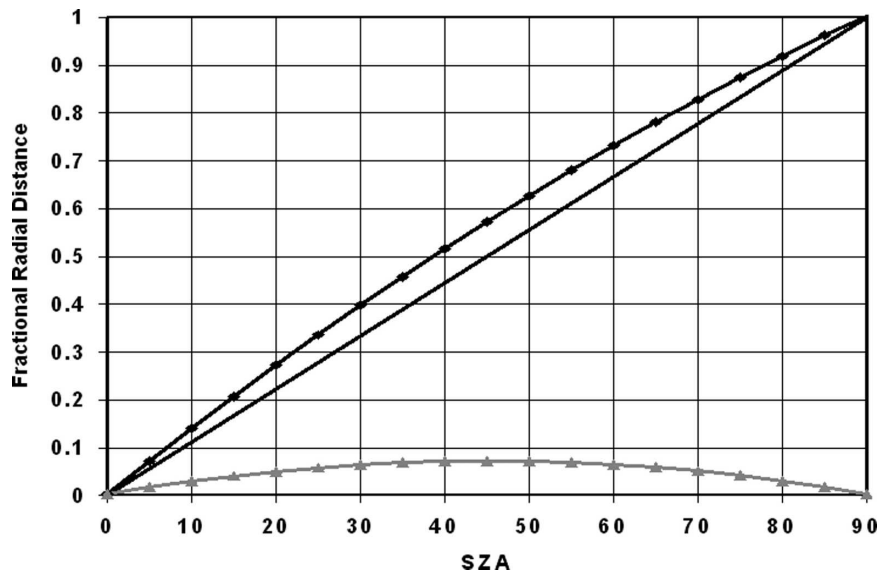
FIG. 5. Fractional radial distance vs SZA for HSI/TSI images. Black line is the HSI/TSI image radial distortion fitted function by SZA, and light gray line is the difference from the 1:1 line (dark gray).

the sky, and (b) the probability that clouds are systematically behind the masked sections is low. Condition (a) is usually met by typical all-sky imagers such as the TSI or the WSC. Condition (b) is difficult to demonstrate, but seems generally plausible given the typical behavior of clouds.

Fractional sky cover, which is typically defined from the point of view of a ground-based observer, is an angular measurement. That is, fractional sky cover is the ratio between the solid angle occulted by clouds and the total solid angle of the visible sky hemisphere (i.e., 2π sr if the horizon is absolutely free of obstacles). Therefore, Eq. (1) is correct only if all pixels correspond to the same solid angle in the sky. This condition is not usually met by typical all-sky images. Note that the TSI retrieval algorithm does not attempt to correct pixels for their view angle bias, that is, where each pixel in the imager has a slightly different solid angle view of the sky compared to adjacent pixels. The TSI retrieval does not attempt to bias or weight any pixels during processing for sky cover.

One relevant issue for image processing is geometrical distortion. Distortion is in general a complex issue, but for our purposes the primary issues have two forms. First, are equal angles (in the zenithal coordinate) in the real world represented as equal distances in the circularly mapped image? Second, is any object represented equally in the image, independently of its azimuthal position? If, from this definition, there is no significant distortion (as has been determined for the WSC), no correction is needed. For the HSI/TSI, with

the camera looking down from a finite distance onto a convex mirror, there is some small amount of radial distortion due to geometry. If there is significant distortion, images could be corrected by a transformation that adjusts the position of the pixels. The distortion and the suitable transformation depend on the exact optics and geometry of the specific device. Either a theoretical analysis of the optics or an empirical study of the device must be performed to estimate the distortion. An empirical analysis of the HSI/TSI radial distortion (and mapping as discussed below) is shown in Fig. 5. Here, the position of the sun in the image is compared to the calculated actual position of the sun in the sky, and the relative difference is plotted (along with the "true" line), as well as the relative difference from "true." In general, distortion only in the radial (zenithal) direction (from the image center) is to be expected, making it simple to perform the corresponding transformation: all pixels must be mapped into a new image slightly modifying their distance to the center.

Even when the image is not distorted or when it has already been corrected, the same areas close to the zenith, or close to the horizon, correspond to different solid angles. In other words, equal areas (or equal number of pixels) in the image correspond to different solid angles in reality. The correction to be applied for a single pixel covering an infinitesimal area, dA, corresponding to an infinitesimal solid angle, $d\Omega$, is $[\sin(\theta)]/\theta$, because the solid angle is proportional to the sine of the zenith angle θ, while the area is proportional to the
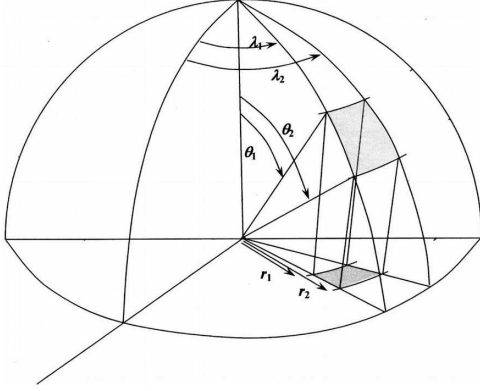
FIG. 6. Diagram showing the relationship between angular coordinates in the sky and polar coordinates in the image, used to obtain the relationship between solid angle and area of a portion of the sky in the sky image.

angle itself. The maximum difference between area and solid angle associated with this effect is about 35% for $\theta = 90°$. For the more usual maximum zenith angle analyzed in our images ($\theta = 80°$), the difference is about 30%. The corresponding differences for smaller zenith angles are 10% at 45°, 1% at 10°, and 0% at 0°.

To calculate the fractional sky cover, it is more convenient to apply similar corrections to portions or sectors of the sky rather than to single pixels. From Fig. 6 and from the definition of solid angle, it is easy to demonstrate that the solid angle of the portion of the sky that is highlighted is

$$\Omega = (\lambda_2 - \lambda_1)(\cos\theta_1 - \cos\theta_2), \qquad (2)$$

while the area of the corresponding projection in the image is (note angles in radians)

$$A = \frac{1}{2}(\lambda_2 - \lambda_1)q^2(\theta_2^2 - \theta_1^2) \qquad (3)$$

for a distortion-free image, that is, $r_1 = q\theta_1$ and $r_2 = q\theta_2$, with q being the proportionality constant between actual zenith angle and projected distances. Obviously, area and solid angle are not proportional to each other.

Combining Eqs. (2) and (3), and considering that $A = mN$, where m is the proportionality constant between the number of pixels and corresponding area, we can write

$$\Omega = \frac{2(\cos\theta_1 - \cos\theta_2)mN}{q^2(\theta_2^2 - \theta_1^2)}. \qquad (4)$$

If, in order to calculate the fractional sky cover, we divide the image (or the real sky) in a finite number of portions, we have

$$f = \frac{\Sigma\Omega_i^{\text{cloudy}}}{\Sigma\Omega_i^{\text{total}}}, \qquad (5)$$

where the subindex i refers to each of the portions, and the superindex means that Eq. (4) has been applied by counting only the cloudy pixels or all pixels in the portion, respectively. Note that the denominator should be equal to 2π sr if the image would cover the whole sky (180° FOV). Because we only use a 160° FOV and part of the sky image is hidden by the shadowband or the shadow strip, the effective total solid angle is always less than 2π sr. Combining Eqs. (4) and (5), we have

$$f = \frac{\Sigma\Omega_i^{\text{cloudy}}}{\Sigma\Omega_i^{\text{total}}} = \frac{\Sigma l_i \dfrac{m}{q^2} N_i^{\text{cloudy}}}{\Sigma l_i \dfrac{m}{q^2} N_i^{\text{total}}} = \frac{\Sigma l_i N_i^{\text{cloudy}}}{\Sigma l_i N_i^{\text{total}}}, \qquad (6)$$

where l_i is

$$l_i = \frac{2(\cos\theta_{i1} - \cos\theta_{i2})}{(\theta_{i2}^2 - \theta_{i1}^2)}. \qquad (7)$$

Equation (6) makes the difference between a correct estimation of fractional sky cover and the first approximation given by Eq. (1) apparent. Note that in Eq. (6), and everywhere in this mathematical development, both the number of pixels N or solid angles refer to the visible part of the image (i.e., masked pixels are not considered).

For the case of the WSC, we have divided the image of the sky into portions that cover $10° \times 10°$ (azimuth and zenith angles). The correction factor l_i for portions close to the zenith (i.e., between $\theta_1 = 0°$ and $\theta_2 = 10°$) is 0.997, while for the portions close to the horizon (i.e., between $\theta_1 = 70°$ and $\theta_2 = 80°$) it is 0.737. We use an example to evaluate the error in fractional sky cover associated to this effect. Assume that we have an image that has 10% cloudy pixels, that is $N_{\text{cloudy}} = 0.1 \times N_{\text{total}}$. By applying Eq. (1), we would get $f = 0.1$ independent of the position of the cloudy pixels in the image. Now assume that these cloudy pixels are in the portions close to the horizon. In this case, applying Eq. (6) with corrections given by Eq. (7), we obtain $f = 0.087$. If the same amount of cloudy pixels are placed close to the zenith, $f = 0.116$. Therefore, the relative error induced by neglecting this geometrical correction could be, for this case, about 15%, or an error of slightly greater than 1% in actual fractional sky cover. Although the absolute error may increase with increasing number of cloudy pixels, the relative error decreases when fractional sky cover increases.

d. Other sky characteristics: Solar obstruction, cloud brokenness, and cloud patterns

1) SOLAR OBSTRUCTION

One sky characteristic difficult to determine from a downward-pointing satellite sensor is solar obstruction.

We have seen in section 3a that a shadowband is needed to protect the WSC optics and CCD from the sun's direct radiation. One disadvantage of using a totally opaque shadowband is that it is more difficult to determine the state of solar obstruction. In the case of the upward-pointing SCSC, no shadowband was used, although an opaque disc blocked the sun when images were not being captured. In this case an image-processing algorithm was developed to overcome the "blooming" effect of the bright sun on the CCD, the net result being a further loss of FOV due to the area of the image, which was unable to be classified as sky or cloud. An additional image-processing algorithm was developed, which correlated the solar region of the image of interest with that of a previously obtained image on a clear day at a similar SZA (a reference image). The correlation was used to classify the solar region as indicative of solar disk obscured (DO), not obscured (DNO), or partially obscured (DPO). The uncertainty in determining solar obstruction was estimated to be less than ±16%, when considering either DNO or (DPO and DO) grouped together.

An alternative approach has been used in the design of the shadowband for the TSI. It is partially reflective and attached to the rotating mirror, as the camera is downward-pointing looking onto the mirror. It is possible to determine the relative brightness of pixels along the shadowband and to determine solar obstruction, or sunshine detection as it is referred to by YES. Although the exact method of how the TSI performs sunshine detection has not been revealed by YES, its precursor, the HSI (Long and DeLuisi 1998), uses the brightness of the white metallic area surrounding the mirror as a reference to calculate this parameter. The TSI algorithm looks along the shadowband for an increase in brightness. If the slope of the brightness increase falls above a user-defined threshold, then the sun disk is deemed unobscured by clouds, otherwise it is assumed to be obscured. An example is shown graphically in the right panel of Fig. 2 as the color of the dot in the sun-blocking mask. A yellow dot (as in this example) indicates that the sun is not significantly obscured by clouds, whereas a white dot indicates an obscured sun. According to Pfister et al. (2003), an all-sky imager (Allsky 1) that uses the HSI sunshine technique agreed with a collocated TSI in 89% of the cases investigated.

Pixels in the sun circle (an approximate 20° radius area centered on the sun position) and horizon area (see Fig. 2) often are interpreted by the processing to be cloudy even when they would not be labeled as such by a human observer. This can occur because of forward scattering from aerosol loading, thin cirrus clouds, or boundary layer haze, which is subvisible elsewhere in the image, but tends to "whiten" the affected pixels. As a result of these problems, Pfister et al. (2003) applied a procedure that uses the mean and standard deviation of the cloud fraction over an 11-min period centered on the image of interest to determine whether the sun circle and/or horizon area cloud pixels should be included in the total or not. If both the whole-sky cloud fraction and its variance, and the variance of the sun circle and/or horizon area are small for an 11-min period, then it is very likely that the sun circle and/or horizon area is free of clouds. Pfister et al. (2003) report that this procedure for the sun circle correlates well with the sunshine parameter (i.e., the solar obstruction) for the two imagers analyzed in their paper. Specifically, when the procedure indicates that there are no real clouds in a 20° radius around the sun, the algorithms identify unobscured sun in about 90% of cases. Because different brightness thresholds can be applied to either of the two imagers' cameras, a mismatch in the sunshine parameter will predominantly occur in situations that cannot be unambiguously classified as either obscured or unobscured sun, for example, situations when the sun disk is partly covered by clouds or when the sun disk is covered by optically thin clouds. This procedure is applied to all-sky image data of at least 1-min resolution, because it depends on the variability through time. As noted in Kassianov et al. (2005), the typical decorrelation time of the sky is about 10 to 15 min. Thus, sky images taken only every 5–10 min do not sample the natural evolution of the sky well enough to adequately track the variability in these problem areas for this "correction" process.

In the case of research reported by Sabburg and Long (2004), TSI images were compared to visual sky observations on 5 days (653 images) sampled randomly in the period December 2001 to September 2002. On clear days, the TSI sunshine indicator would incorrectly register solar obstruction near local solar noon when a shadow of the camera housing was cast onto the shadowband.

2) CLOUD BROKENNESS AND CLOUD DISTRIBUTION

Besides classifying solar obstruction by cloud across the solar region, algorithms have also been developed for the HSI, SCSC, and TSI sky images to measure cloud properties in other regions of the sky view. One of these properties is a measure of average cloud size and brokenness of the cloud coverage. For example, the cloud retrieval algorithm for the AllSky1 of Pfister et al. (2003) determines an "edge-to-area" ratio, defined as the number of pixels in the image on cloud/clear boundaries divided by the sum of cloudy pixels in the

image. A large edge-to-area ratio relates to broken clouds of small diameter, while a small edge-to-area ratio relates to extended clouds. Clear sky and completely covered sky have an edge-to-area ratio defined as zero (Pfister et al. 2003). The SCSC included the parameter CB, the ratio of the perimeter to area of a cloud (similar to the edge-to-area ratio). The perimeter and area measurements were based on the number of pixels classified as cloud, after thresholding the sky image to a binary image. The CB was measured in the range of zero to one (Sabburg and Wong 2000).

Sabburg and Long (2004), Sabburg and Wong (2000, 1999), and Pfister et al. (2003) describe algorithms defining the distribution of cloud around the solar region. In the case of the SCSC, this was done in terms of the angle subtended by an arc from the center of the sun to the outer edge of circular sectors of width 2.5°. The algorithm classified which sector contained the greatest area of cloud, of 10 possible concentric sectors, whose angles ranged from 12.5° to 37.5°. In the case of the commercial TSI software, cloud amount is also estimated separately in zenith circle, sun circle, and horizon regions. In an analysis technique written for the TSI, Sabburg and Long (2004) report cloud amounts that were calculated in four separate circular regions concentric about an estimated position of the sun as determined by the TSI processing algorithm. Regions were classified as inner (about 15° radius), middle (about 30° radius), outer (about 45° radius), and extreme (extending beyond the outer region that is within the image area). Errors in cloud amount near to the sun (i.e., in the inner and middle regions) were found to increase from approximately 60° SZA and increasing as the sun approaches the horizon. Causes for this error were previously discussed and can contribute to an overestimation of cloud amount, up to 60% in the inner (sun centered) region to a few percent in the middle region, depending on the aerosol, haze, or ice crystal loading of the atmosphere.

A new parameter, defined as cloud uniformity with respect to azimuth (in contrast to sun centered), was also developed for the TSI (Sabburg and Long 2004). The sky image is divided into four quadrants with vertical and horizontal crossbars centered on zenith with the quadrants defined as N to E, N to W, S to W, and S to E. Uniformity was recorded as 1 if cloud amount in the four quadrants of the image was within 20% of the total image cloud amount, otherwise it was listed as 0.

3) CLOUD PATTERNS

One of the most difficult research areas of sky-image processing has been that of cloud recognition. Parisi et al. (2004) make reference to some papers relating to this area of research in their overview of sky imagers. In this current paper, we make no attempt to classify cloud patterns (i.e., the parameter that best describes the "type" of cloud or cloud field present) for TSI or WSC images. Cloud pattern classification includes standard cloud types such as cumulus, stratus, and cirrus. Cases of haze, aerosol, and fog classification could also be included under this heading. We do address cloud field properties, such as brokenness, as described previously. In addition, the TSI image processing includes a separation of cloudiness into "thin" and "opaque" classifications, which is based on the amount of blue tint of the clouds in the image. This blue tint occurs when the clouds are optically thin, and thus one can see through them to the background blueness of the clear sky behind the cloud.

Although there have been some papers published on analysis of satellite images of the earth view (e.g., Harris and Barrett 1978; Ebert 1987) and overviewed by Parisi et al. (2004), there have only been a few published papers describing limited cloud-type recognition from ground-based sky imagers. For example, by the use of polarizing filters Horvath et al. (2002) have improved algorithms of radiometric cloud detection, particularly promising for very high altitude, thin (i.e., "bluish") cirrus clouds. Goodall and Hatton (2002) have performed some initial research with both visible and infrared images and have concentrated on the identification of towering cumulus and cumulonimbus clouds. Their results indicate that neural network processing has potential in cloud recognition.

For the SCSC, a parameter called "cloud texture" was defined as the standard deviation of the brightness of the cloudy pixels. Brightness was defined as the sum of the values of the three color components (RGB). This gave a value for the variation along the surfaces of the clouds as seen from the ground. From unpublished work by one of the authors of this current paper (Sabburg at the University of Southern Queensland), it is speculated that Lambertian reflection increases for light-textured cloud—for example, cirrus—compared to the specular reflection from a heavier-textured cloud surface—for example, cumulus. Thus it may be possible to use this parameter to assist in the classification of cumulus and cirrus clouds, which is a subject of ongoing research. Further unpublished work by Sabburg, originally developed for use by the Commonwealth Bureau of Meteorology to classify TSI data as stratiform cloud for UV index research, used color, brightness, and pixel transitions in an attempt to classify cloud data as stratus, cumulus, cirrus, or fog. Although the findings for opaque and stratiform (overcast) cloud classification
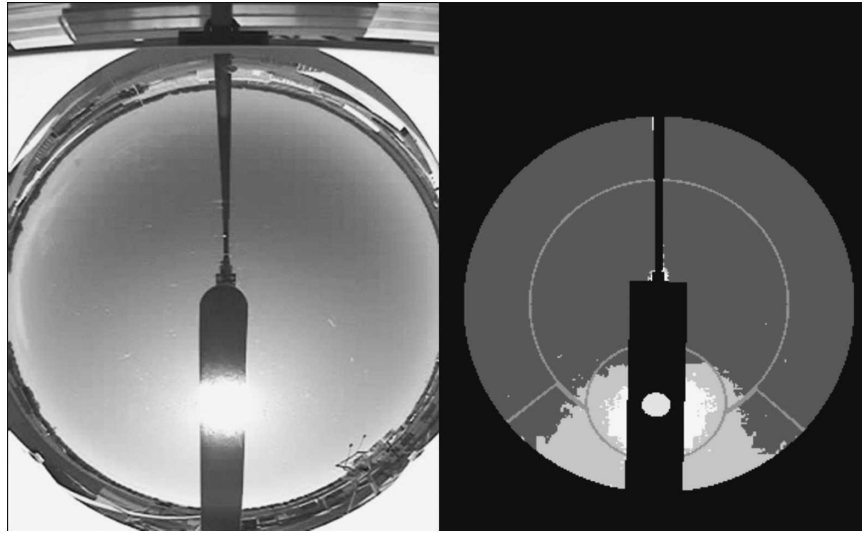
FIG. 7. (left) A sample gray-scaled HSI image and (right) corresponding gray-scaled cloud decision image taken at 1300 LT 4 Sep 2004 at the Pacific Northwest National Laboratory, Richland, WA.

were encouraging, it was not successful for cirrus, stratus, and cumulus clouds.

One technique that has been successfully used with CCD cameras for astronomical observations, but not clouds (Buil 1991), is that of Fourier transform or fast Fourier transform (FFT) analysis. The idea is that "signature" frequencies, corresponding to different types of cloud, may be produced from a Fourier transform of a cloudy-sky image [e.g., Garand (1988) applied this idea to satellite cloud images]. Additionally, two-dimensional FFT techniques have been applied by one of the authors of this current paper (Calbó at the University of Girona) on terrain topography to investigate the best grid size to be used in mesoscale meteorological modeling (Salvador et al. 1999). This last work indicates that the technique may be quite robust in dealing with any spatial characteristics, including cloud patterns, which will be the subject of further research.

4. Results

a. Total sky imager

1) TOTAL SKY IMAGER CORRECTION FOR SUN CIRCLE AND HORIZON AREA CLOUD DETECTION ERRORS

The original methodology for correcting the sun circle and horizon area (Fig. 2) cloudiness amount described in Pfister et al. (2003) has been refined and adapted for application to TSI data (Long 2005). In essence, the magnitude and variability of the cloud frac-

tion in the sun circle, horizon area, and the remainder of the image (total area minus the sun circle and horizon areas) are used to determine whether or not the cloud pixels in the sun circle and/or horizon area should be included in the total-image sky cover estimate. In the case of the sun circle, it must also be determined whether to count only half of the original cloud pixels (as will be discussed later in this section). The results are smoothed using a running 11-point running mean, that is, if 1-min data are being processed then the amount of adjustment applied is the average over 11 min centered on the point of interest.

Figure 7 shows a grayscale sample HSI image (left) and corresponding cloud decision image (right) taken at 1300 local time (LT), 4 September 2004, at the Pacific Northwest National Laboratory located in Richland, Washington. As shown in this relatively extreme example, both the sun circle and horizon areas contain pixels erroneously determined as "cloud" while the sky image shows what an observer would typically label as clear sky. On this day the morning was clear, with cloudiness moving in at about 1320 LT and lasting through about 1520 LT when skies cleared again. More cloudiness then moved slowly in again at around 1700 LT, slowly moving off through about 1840 LT. The sky and cloud decision images show that this day exhibited significant haze, producing the erroneous identification problem.

Figure 8 shows the retrieved total-sky cover for the 4 September 2004 daylight period, including the original retrieval, the "first guess" sun circle adjustment (Long
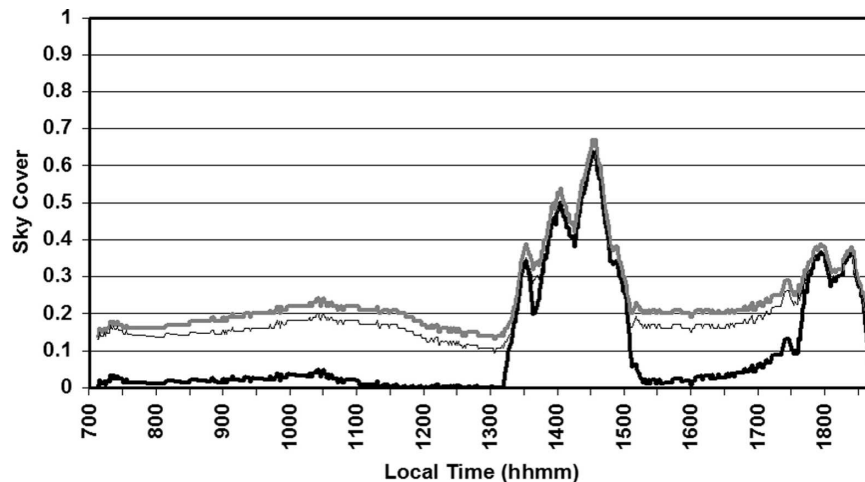
FIG. 8. Total sky cover retrieval for 4 Sep 2004 at the Pacific Northwest National Laboratory, Richland, WA. The gray line is the original retrieval, the thin black line is the retrieval including the "first guess" adjustment of the sun circle area, and the black line is the final result including all adjustments and smoothing.

2005), and the final adjusted retrieval. The first guess is intended to account for the probability of some error near the sun due to persistent forward scattering for times when the other tests do not subtract the sun circle cloud pixels. There is often some overestimation of cloud amount in the sun circle area, thus the reasoning behind the first-guess adjustment, which in general decreases the sun circle area cloud amount by up to half. As Fig. 8 shows, the adjustment methodology correctly decreased the initial erroneous sky cover values of nearly 20% during clear-sky periods downward in magnitude to near 0%, yet did not decrease the sky cover values for the times when clouds were present.

Figure 9 shows relative frequency histograms of various instruments and time periods as noted in the figure caption. In each case, the original retrievals (gray) show a bias away from the "clear" bin (on the left) toward higher values. This result is inconsistent with expectations for these sites, where it is common that the long-term frequency distribution includes about one-third clear sky, one-third overcast, and the remaining one-third distributed in between. This expected distribution is indeed the case when the adjustments detailed here are applied to the retrievals (black) for each case. In the top two plots, the third distribution (striped) is from the available 100° FOV "zenith circle" retrievals (see Fig. 2). This zenith area is far less susceptible to the misidentification problems we are addressing, since the entire horizon area is not included, and (as noted previously) the sun circle problem is generally less for higher sun elevations. As is seen, there is much better agreement with the adjusted distributions than with the origi-

nal. Similarly, the third distribution in the bottom plot, produced by the ARM Program using the shortwave flux analysis algorithm (Long and Ackerman 2000; Long and Gaustad 2004), agrees better with the adjusted values than the original. All these results suggest that the adjustment methodology significantly improves the sky-imager retrievals as intended.

2) TOTAL SKY IMAGER SOLAR OBSTRUCTION AND CLOUD DISTRIBUTION STUDIES

The TSI located at the campus of the University of Southern Queensland was used to undertake further analysis of the sky characteristics of solar obstruction and cloud uniformity (introduced in section 3d). The analysis of the performance of these characteristics is more extensive than that previously undertaken by Sabburg and Long (2004). The set of images (71 335 in total), captured every 5 min from early morning to late afternoon, is available for the period from June 2003 through December 2004. We chose for study a temporally evenly distributed set of images during this period of up to 10 images per day, resulting in a total of 2427 images covering the SZA range of 4° to 80°. This subset of images was also manually inspected by an independent researcher (with previous experience inspecting SCSC images). Each image was viewed on a computer screen and the researcher recorded the following results in a spreadsheet:

(a) for solar obstruction, "0" if the sun was either blocked or not visible due to cloud, otherwise "1."
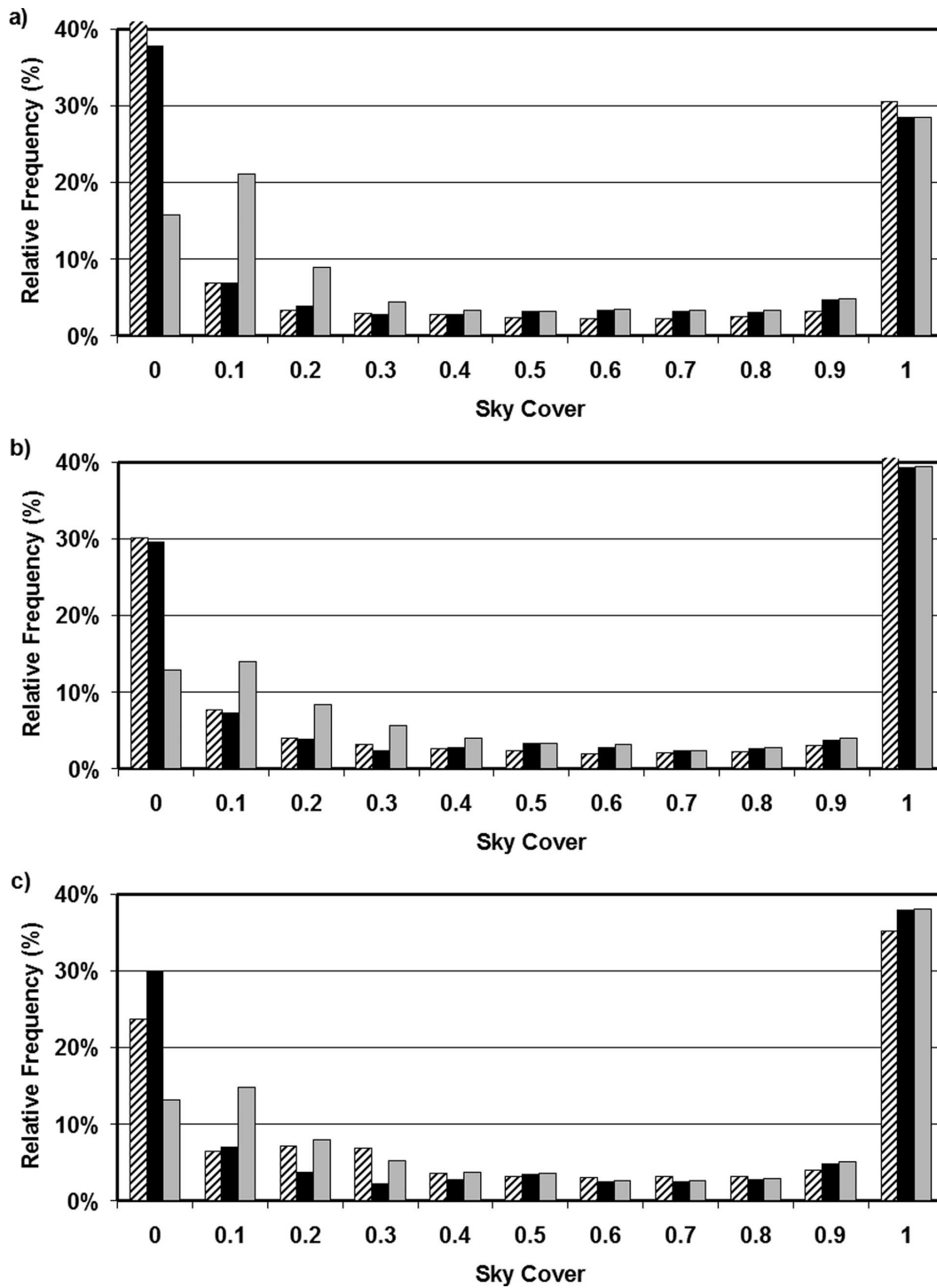(b) for cloud uniformity, "0" if cloud was not "uni-

FIG. 9. Sky cover frequency histograms for (a) more than 7 months of data at Pacific Northwest National Laboratory, and (b) for the HSI and (c) TSI deployed during the 3 months of the ARM Cloudiness Inter-Comparison (CIC) field experiment at the Southern Great Plains site. In all plots, gray represents the original retrievals, and black represents the adjusted retrievals. In (a) and (b), striped represents retrievals restricted to a 100° FOV. In (c), striped represents retrievals from the shortwave flux analysis methodology.

TABLE 3. All available data (SZA, 4° to 80°; TSI cloud fraction, 0%–100%) for (a) DO and (b) uniformity.

| | | (a) DO | | | | (b) Uniformity | | | |
| | | Inspection | | | | Inspection | | | |
		Yes	No	Undecided	Total	Yes	No	Undecided	Total
Algorithm	Yes	8	0	0	8	39	55	0	94
	No	67	2337	15	2419	124	2206	3	2333
	Undecided	0	0	0	0	0	0	0	0
	Total	75	2337	15	2427	163	2261	3	2427

formly" distributed in the image (i.e., less than 20% of the total cloud in each of four quadrants), otherwise "1."

If the visual inspector was not sure whether to record a "1" or "0," then a "9" indicated the decision was undecided in either of the two categories.

For analysis of results, the data were divided into three groups: all available data, restricted SZA range, and restricted cloud fraction range. The sum of the diagonal values of any one of the matrices shown in Table 3, divided by the total number of images, gives a corresponding indication of the performance of each of the algorithms as 97% and 93% for solar obstruction and cloud uniformity, respectively. On closer inspection the overall performance of 97% was found to be biased due to the exceptional classification (100%) when the disk was not obscured. When the disk was obscured, the algorithm did not perform nearly as well, and neither was the observer able to readily classify the state of the obstruction. This prompted the decision to test if the performance of the algorithm or visual inspection might be affected by the "whitening" phenomenon described in section 3a. As the whitening tends to decrease with decreasing SZA, it was thought that the performance might improve for smaller SZA (higher sun). However, analysis of the smaller SZA data exhibited no improvement in performance.

We also investigated whether the success of the comparison between algorithm and inspection improved for separate ranges of cloud fraction. The matrices in Table 4 show the performance of each of the algorithms for cloud fraction less than 50% (99% and 96%), and for a cloud fraction between 51% and 100% (85% and 76%, respectively). These results indicate an increased performance of 3% for both algorithms for cases with less then 50% of the sky-containing clouds. For the mostly cloudy cases, there is a decrease in performance of 11% and 17% for disk obstruction and uniformity, respectively. It could be concluded that the skill of the visual classification of these characteristics decreases with increased cloud fraction.

b. Whole sky camera

For the WSC, we analyze a set of images taken and processed during one year (November 2001 to October 2002). The only month with a significant number of missing images is August 2002. Images were captured and stored as BMP every 15 min, producing about 13 000 images, of which only about 10 700 taken at SZAs less than 80° were used for the present analysis. For each image, we calculated the fractional sky cover using both the geometrical correction derived in Eq. (6) and the unadjusted ratio of cloudy to total number of pixels [Eq. (1)]. These two values will be named hereafter as f_G and f_{NG}, respectively. Using the method described in section 3d, CB was calculated for the WSC images as the number of pixels in the perimeter of cloudy areas divided by the number of cloudy pixels. To investigate the effect of image format, we also computed the corresponding values for a subset of the same images, but stored in JPEG format.

Approximately one-third of the images, evenly distributed across all months, were also visually inspected to estimate the corresponding fractional sky cover. This visual inspection was performed by three researchers at the University of Girona (two of the coauthors of this paper and a third colleague). Each inspector looked at close to 1400 images, with a subset of 430 being inspected by all three for cross-comparison and to investigate possible systematic bias of this kind of subjective human estimation of sky cover. The visually determined fractional sky cover will be referred to as f_V. To estimate f_V we used some visual aids that allowed us to divide the sky dome into 16 sectors, giving a resolution of these estimations of fractional sky cover of 0.0625 (=1/16). We also have available about 150 human observations of the sky conditions that were performed from November 2001 to May 2002. Observations were made by the two University of Girona coauthors from the same site where the WSC is installed. Although the observers are not professionally trained, cloud observations were carefully made following WMO recommendations. We recorded fractional sky cover f_{obs} (in

TABLE 4. Data with TSI cloud fraction less than 50% and SZA 4° to 80° (2018 images) for (a) DO and (b) uniformity, and data with TSI cloud fraction 51%–100% and SZA 6° to 80° (409 images) for (c) DO and (d) uniformity.

		(a) DO				(b) Uniformity			
		Inspection				Inspection			
		Yes	No	Undecided	Total	Yes	No	Undecided	Total
Algorithm	Yes	1	0	0	1	3	38	0	41
	No	13	1997	7	2017	44	1930	3	1977
	Undecided	0	0	0	0	0	0	0	0
	Total	14	1997	7	2018	47	1968	3	2018
		(c) DO				(d) Uniformity			
		Inspection				Inspection			
		Yes	No	Undecided	Total	Yes	No	Undecided	Total
Algorithm	Yes	7	0	0	7	36	17	0	53
	No	54	340	8	402	80	276	0	356
	Undecided	0	0	0	0	0	0	0	0
	Total	61	340	8	409	116	293	0	409

oktas) and also cloud type and other sky characteristics (e.g., sun obstruction). These observations will be used here to be compared, both with the visual estimations from the images and the computed values.

Based on the common set of images, we found that no relevant systematic bias was exhibited when the three trained researchers looked at the same images. Despite the absence of bias, there is some dispersion of values. In 5%–10% of the images, we found differences among two of the estimates of f greater than 0.5. Most of these cases correspond to early morning images, that is, dark images with possible dew on the WSC glass dome. The agreement between estimates is very high for totally cloudless or absolutely overcast skies. For the rest of images, that is, for f in the range 0.05–0.95, the root-mean-square error (rmse) is close to 0.11. This figure can be taken as a measure of the uncertainty when f is determined from WSC images by visual inspection, although for some range of f values, the uncertainty may be larger (see Table 5). The rmse was computed here from the differences between each estimate and the average of the three values. As a consequence of these analyses we decided that, for comparison with automatic estimations, f_V would be equal to the average of the three values when available, and equal to the single value when only one researcher had inspected an image.

The effect of the geometric correction through Eq. (6) is, as expected, relatively small. We found a determination coefficient $r^2 = 0.9998$ between f_G and f_{NG}. The largest absolute differences are less than 0.04 in fractional sky cover. Corresponding relative errors are always less than 10% (and usually less than 5%), except for almost cloudless skies, when absolute differences of 0.01 may result in relative errors greater than 10%. Despite this minor effect, we have used f_G in further analyses. Similarly, f from the JPEG images is almost identical to results from the BMP images in fractional sky cover, with a mean bias deviation (MBD) of 0.003 and rmse of 0.02.

Results of the comparison between f_G and f_V are presented in the box charts of Fig. 10. All differences $f_G - f_V$ have been grouped in bins according to f_V (top plot). Each bin (except the first and the last ones, which correspond to cloudless and overcast skies, respectively) has a width of 0.10. We can see that the automatic estimation of fractional sky cover is in general

TABLE 5. Mean and standard deviation of the visual estimates of f, for several intervals of f, from the set of images analyzed by all three researchers.

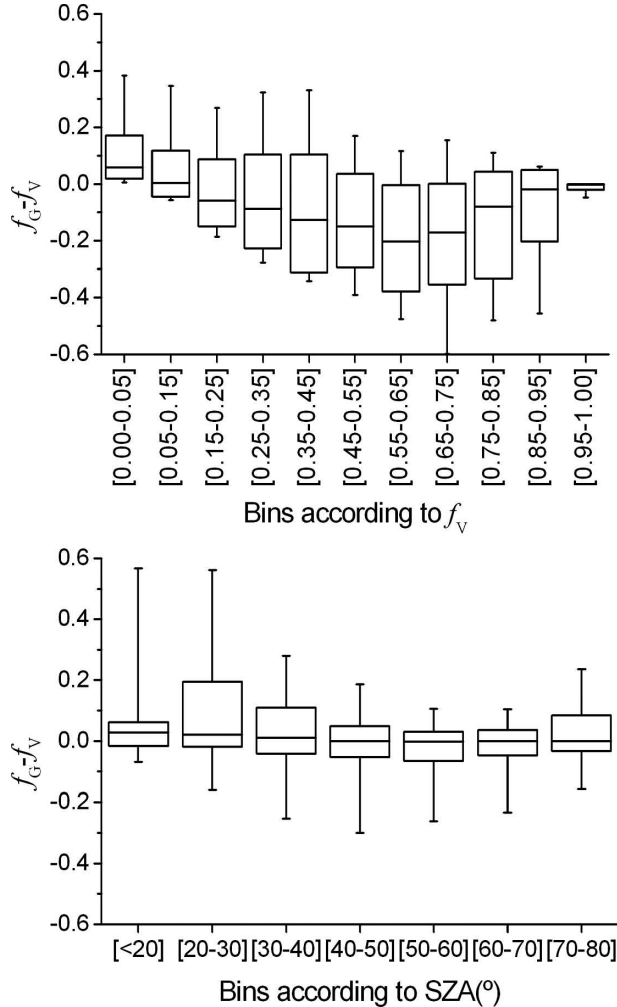| Interval of f	Mean	Std dev
0.00	0.00	0.00
0.00–0.06	0.04	0.03
0.06–0.12	0.10	0.05
0.12–0.19	0.16	0.07
0.19–0.25	0.22	0.08
0.25–0.31	0.30	0.09
0.31–0.37	0.36	0.14
0.37–0.44	0.41	0.07
0.44–0.50	0.48	0.13
0.50–0.56	0.56	0.21
0.56–0.62	0.61	0.16
0.62–0.69	0.68	0.19
0.69–0.75	0.73	0.16
0.75–0.81	0.79	0.15
0.81–0.87	0.85	0.13
0.87–0.94	0.92	0.07
0.94–1.00	0.97	0.03
1.00	1.00	0.00

FIG. 10. Differences between automatic estimation of fractional sky cover f_G and visual estimation f_V: (upper) depending on fractional sky cover, and (lower) depending on SZA. Boxes show the median and the percentiles 25 and 75. Additional error bars represent the percentiles 10 and 90.

lower than the human visual estimation. Most medians correspond to negative differences, with absolute values always less than 0.20 and in general less than 0.10. The dispersion of differences is larger for broken-cloud conditions (f_V in the range 0.35–0.85). This behavior is likely due to an overestimation of the human estimate, which consists of counting large sectors of the sky that can be "patched" with small clear areas as cloudy, while the automatic estimation is based upon pixel counts only. Automatic and manual estimations are virtually identical as far as overcast conditions are concerned. For cloudless skies, the automatic estimate hardly ever results in $f_G = 0.00$. This is due to circumsolar areas that are considered cloudy by the automatic method, given that these areas appear white when there is some

amount of haze or aerosols, as mentioned previously. For the whole dataset, MBD between f_G and f_V is −0.001, and rmse is 0.21.

Figure 10 (bottom) shows the same differences $f_G - f_V$ as in the top plot, but versus SZA. The medians of the differences are practically 0 for all SZAs, which is a good characteristic of the image processing. It should be noted that about two-thirds of the data in the top plot resides in the first two and last bins (i.e., the nearly clear and overcast bins), which are about evenly distributed by SZA in the bottom plot. Dispersion is somewhat larger for smaller SZAs. There are two possible reasons for this behavior. First, there are fewer images taken at these lower SZA values, which correspond to noontime of summer months. Second, these same noontime summer data are when aerosol optical depths at Girona are usually higher than in other hours and seasons (González et al. 1998). Thus, the effect of the "white" circumsolar area under high aerosol conditions is enhanced at smaller SZA, resulting in an overestimate of f. This also may be related to the fact that we are using a single threshold (R/B = 0.6) for all images and suggests that a slightly higher value of this ratio should be used for summer conditions in Girona.

In comparing f_G and f_{obs}, the best agreement is found for overcast skies, followed by cloudless or almost cloudless (less than 1 okta) skies. In the latter case, as expected, f_G tends to be greater than f_{obs}, because of the already explained difficulty of obtaining $f_G = 0$ by using our automatic retrieval. In all other cases (2 oktas $\leq f_{obs} \leq 7$ oktas) f_G is systematically biased toward lower values or, conversely, the human observations systematically indicate higher cloud amounts. The MBD between f_G and f_{obs} is −0.12, and the rmse is 0.28. The maximum differences correspond to cases when the human observation has reported cirriform clouds: from the 150 observations there are seven cases with $f_G - f_{obs} < -0.5$. In these seven cases either Ci or Cs were reported. The number of direct observations of the sky is not large enough to derive robust conclusions from the commented differences. However, this comparison seems to confirm the tendency that has already been detected when comparing f_G with f_V. In summary, the automatic estimate derived from WSC images generally results in lower fractional sky cover values than the human estimates from either direct observations or visual inspection of corresponding sky images.

The most frequent value of CB ranges from 0.10 to 0.15, with most CB values less than 0.35. This is true for sky conditions corresponding to f_G in the range 0.05–0.95. Obviously, when the sky is virtually cloudless ($f_G < 0.05$) or almost overcast ($f_G > 0.95$), CB tends to be 0. Logically, most frequent CB values are higher for
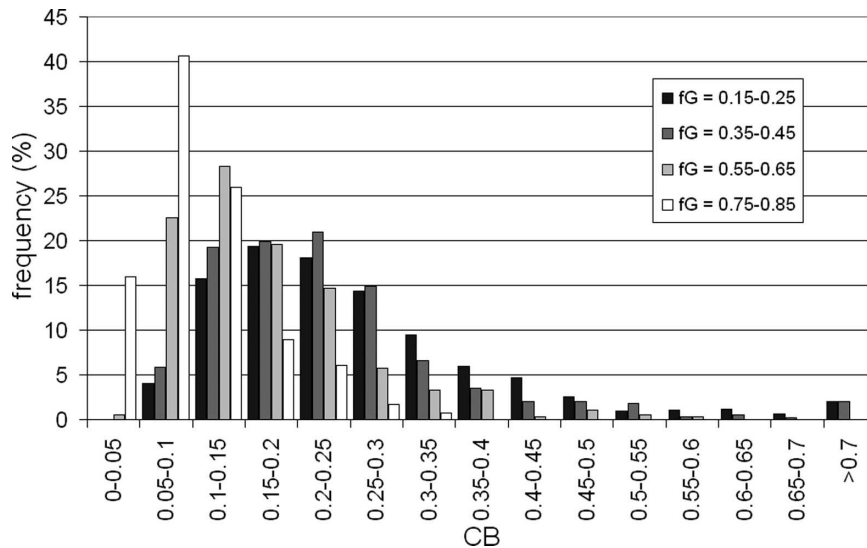
FIG. 11. Frequency distributions of CB values for four different ranges of f_G as noted in the figure legend.

scattered cloudy conditions and lower when fractional sky cover is small or large. For example, for f_G in the range 0.25–0.45, typical CB is 0.3, while for $f_G > 0.65$, CB is less than 0.1 in general. The major importance of CB, however, is the dispersion of values for a given value of fractional sky cover. Different CB values indicate different sky conditions: smaller CB means compact clouds, while larger CB means patchy clouds. Figure 11 shows the frequency distributions of CB for four different ranges of f_G. We can see that, for f_G between 0.35 and 0.45, there is maximum dispersion of CB values, indicating that these clouds may correspond either to a few clouds covering a part of the sky or a larger number of broken cloudy areas likely occupying the whole sky dome. Figure 11 also confirms that dispersion of CB values is smaller when f_G is greater. The CB values computed from JPEG images tend to be smaller (by a factor of 2) than CB values from BMP images. This is due to the more physically representative "smoothing" of the JPEG format (discussed below). However, the relative frequency distribution of CB values from JPEG images is quite similar to the distribution obtained from BMP images that is shown in Fig. 11.

While it is true that a BMP image does better capture each individual CCD element value, commercial CCD arrays such as we are using have element-to-element sensitivity differences that affect the clear/cloud classification. It is not "normal" for one isolated pixel to be "cloudy" when all of its surrounding pixels are not. Using an element-by-element map, such as a BMP image, often results in isolated pixels erroneously being

classified as cloud. While this results in only a small error in total sky cover, it can have a significant effect on a parameter such as CB as noted above. JPEG compression, which by its nature is a slightly "smoothed" rendering at typical default JPEG compression settings (75–80), tends to compensate for the CCD element-to-element sensitivity differences giving sky cover retrievals that much better reflect the way nature behaves in the sky.

5. Summary

In this paper we have shown that the application of automatic digital image-processing techniques on sky images is a useful method to complement, or even replace, traditional human observations of sky cover, and there is likely potential for inferring cloud type. Although some uncertainty exists in fractional sky cover retrievals from sky images, previous work has shown this uncertainty is no greater than that attached to human observations for the commercially available sky imager and processing technique (i.e., for the TSI) discussed here. Even for the WSC imager, the uncertainty is still acceptable and comparable to human observational uncertainty. We note that cloud cover has been traditionally recorded as eighths (oktas) or tenths of the sky dome by human observers. This means that an uncertainty of at least 0.125 or 0.10, respectively, is to be expected. Unlike human observations, current sky condition descriptions from digital images do not include cloud typing in the traditional way (i.e., using cloud genera). However, other equally interesting (for radia-

tion studies, e.g.) sky characteristics can be implied, such as cloud brokenness and distinction between optically thin and thick clouds.

The main advantages of sky imagers compared to human observations are threefold. First, sky imagers can provide an almost continuous observation of the sky. Classically, cloud observations are made every 3 h, or on an hourly basis at some meteorological stations. Second, sky imagers can provide long-term sky condition information at relatively low cost, compared to the cost of human observers. Finally, where human observations of clouds are subjective, decreasing their precision, observation of clouds by automatic devices such as sky imagers is objective and highly reproducible.

One of the imagers presented in this paper was designed and built by researchers at the University of Girona. This imager (WSC) has been continuously taking sky images since the Northern Hemisphere summer of 2001. One year of such images has been analyzed by using a simple process that consists of initial masking of parts of the image and using a single threshold to distinguish between cloudless and cloudy pixels. When computing the fractional sky cover, an expression that accounts for the differences between the actual solid angles and the corresponding image areas has been considered. The effect of this correction is minor, but nevertheless the correction has been applied because it does not present particular difficulties in processing the images. With this imager and simpler processing methodology, fractional sky cover can be estimated with an uncertainty of about 0.2. More specifically, imager-derived sky cover tends to be greater than the corresponding human observations for amounts less than 0.2, but less than the human observations for amounts ranging between 0.2 and 0.8. The two values virtually always agree for overcast conditions.

We see two directions for future research stemming from the current work: improving the hardware and improving the image processing. One possibility for device improvement is mounting a fish-eye camera on a solar tracker and shading the lens with a shading sphere, instead of a shadow strip or shadowband. With this approach, the area of the sky obscured could be reduced. However, it is also realized that the smaller the "dome" or mirror surface used, the greater the portion of the sky image adversely affected by obstructions such as rain- or dewdrops. Regarding the image processing, we will investigate further parameters such as CB, and how these parameters relate with classical cloud types. In addition, suggestions made in this paper about the use of Fourier analyses techniques applied to ground-based sky images will be further explored.

REFERENCES

Buil, C., 1991: *CCD Astronomy—Construction and Use of an Astronomical CCD Camera.* Willmann-Bell, 321 pp.

Calbó, J., D. Pagès, and J. A. González, 2005: Empirical studies of cloud effects on UV radiation: A review. *Rev. Geophys.,* **43,** RG2002, doi:10.1029/2004RG000155.

Ebert, E., 1987: A pattern recognition technique for distinguishing surface and cloud types in the polar regions. *J. Climate Appl. Meteor.,* **26,** 1412–1427.

Garand, L., 1988: Automated recognition of oceanic cloud patterns. Part I: Methodology and application to cloud climatology. *J. Climate,* **1,** 20–39.

González, J. A., J. Calbó, and J. Mejías, 1998: Ground-based evaluation of aerosol transmittance for cloudless and scattered cloudy skies. *Proc. SPIE,* **3493,** 148–155.

Goodall, P., and D. Hatton, 2002: Meteorological observations by computer analysis of video images. *Extended Abstracts, Technical Conf. on Meteorological and Environmental Instruments and Methods of Observation,* Bratislava, Slovakia, WMO, CD-ROM, 1.1(10).

Harris, R., and R. C. Barrett, 1978: Toward an objective nephanalysis. *J. Appl. Meteor.,* **17,** 1258–1266.

Horvath, G., A. Barta, J. Gal, B. Suhai, and O. Haiman, 2002: Ground-based full-sky imaging polarimetry of rapidly changing skies and its use for polarimetric cloud detection. *Appl. Opt.,* **41,** 543–559.

Houghton, J. T., Y. Ding, D. J. Griggs, M. Noguer, P. J. van der Linden, X. Dai, K. Maskell, and C. A. Johnson, Eds., 2001: *Climate Change 2001: The Scientific Basis.* Cambridge University Press, 881 pp.

Johnson, R. W., T. L. Koehler, and J. E. Shields, 1988: A multistation set of whole sky imagers and a preliminary assessment of the emerging data base. *Proc. Cloud Impacts on DOD Operations and Systems—1988 Workshop,* Silver Spring, MD, Department of Defense, 159–162.

——, W. S. Hering, and J. E. Shields, 1989: Automated visibility and cloud cover measurements with a solid-state imaging system. SIO Reference 89-7, GL-TR-89-0061, Marine Physical Laboratory, Scripps Institution of Oceanography, University of California, San Diego, 118 pp.

Kassianov, E., C. N. Long, and M. Ovtchinnikov, 2005: Cloud sky cover versus cloud fraction: Whole-sky simulations and observations. *J. Appl. Meteor.,* **44,** 86–98.

Koehler, T. L., R. W. Johnson, and J. E. Shields, 1991: Status of the whole sky imager database. *Proc. Cloud Impacts on DOD Operations and Systems—1991 Conference,* El Segundo, CA, Department of Defense, 77–80.

Long, C. N., 2005: Accounting for circumsolar and horizon cloud

determination errors in sky image inferral of sky cover. *Proc. 15th Atmospheric Radiation Measurement Science Team Meeting,* Daytona Beach, FL, Department of Energy ARM Program.

——, and J. J. DeLuisi, 1998: Development of an automated hemispheric sky imager for cloud fraction retrievals. *Proc. 10th Symp. on Meteorological Observations and Instrumentation*, Phoenix, AZ, Amer. Meteor. Soc., 171–174.

——, and T. P. Ackerman, 2000: Identification of clear skies from broadband pyranometer measurements and calculation of downwelling shortwave cloud effects. *J. Geophys. Res.,* **105** (D12), 15 609–15 626.

——, and K. L. Gaustad, 2004: The shortwave (SW) clear-sky detection and fitting algorithm: Algorithm operational details and explanations. Atmospheric Radiation Measurement Program Tech. Rep. ARM TR-004.1, 24 pp. [Available online at http://www.arm.gov/publications/tech_reports/arm-tr-004-1.pdf.]

——, D. W. Slater, and T. Tooman, 2001: Total Sky Imager (TSI) model 880 status and testing results. Atmospheric Radiation Measurement Program Tech. Rep. ARM TR-006, 36 pp. [Available online at http://www.arm.gov/publications/tech_reports/arm-tr-006.pdf.]

Lu, D., J. Huo, and W. Zhang, 2004: All-sky visible and infrared images for cloud macro characteristics observation. *Proc. 14th Int. Conf. on Clouds and Precipitation,* Vol. 2, Bologna, Italy, ICCP, IAMAS, 1127–1129.

Pagès, D., J. Calbó, C. N. Long, J.-A. González, and J. Badosa, 2002: Comparison of several ground-based cloud detection techniques. *Extended Abstracts, European Geophysical Society XXVII General Assembly,* Nice, France, European Geophysical Society, CD-ROM.

Parisi, A. V., J. Sabburg, and M. J. Kimlin, 2004: *Scattered and Filtered Solar UV Measurements*. Advances in Global Change Research Series, Kluwer Academic, 195 pp.

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

Sabburg, J., and J. Wong, 1999: Evaluation of a ground-based sky camera system for use in surface irradiance measurement. *J. Atmos. Oceanic Technol.,* **16,** 752–759.

——, and ——, 2000: Evaluation of a sky/cloud formula for estimating UV-B irradiance under cloudy skies. *J. Geophys. Res.*, **105** (D24), 29 685–29 692.

——, and C. N. Long, 2004: Improved sky imaging for studies of enhanced UV irradiance. *Atmos. Chem. Phys.,* **4,** 2543–2552.

Salvador, R., J. Calbó, and M. M. Millán, 1999: Horizontal grid size selection and its influence on mesoscale model simulations. *J. Appl. Meteor.,* **38,** 1311–1329.

Shields, J. E., R. W. Johnson, M. E. Karr, A. R. Burden, and J. G. Baker, 2003: Daylight visible/NIR whole-sky imagers for cloud and radiance monitoring in support of UV research programs. *Proc. SPIE,* **5156,** 155–166.

Tooman, T. P., 2003: Whole Sky Imager retrieval guide. Atmospheric Radiation Measurement Program Tech. Rep. ARM TR-011.1, 109 pp. [Available online at http://www.arm.gov/publications/tech_reports/arm-tr-011-1.pdf.]