

## A New Method to Estimate Cloud Cover Fraction over El Leoncito Observatory from an All-Sky Imager Designed for Upper Atmosphere Studies

C. MARTINIS,<sup>1</sup> J. WILSON,<sup>2</sup> P. ZABLOWSKI,<sup>1</sup> J. BAUMGARDNER,<sup>1</sup> J. L. ABALLAY,<sup>3</sup> B. GARCIA,<sup>4,5</sup>  
P. RASTORI,<sup>6</sup> AND L. OTERO<sup>6</sup>

*Received 2012 October 01; accepted 2012 November 29; published 2013 January 17*

**ABSTRACT.** A method for determining cloud cover fraction over El Leoncito Observatory (31.8°S, 69.3°W) is presented. Data from an all-sky imaging system, designed to measure nightglow originating from the mesosphere and thermosphere, is used to determine the fraction of the sky covered by clouds. More than 9,000 hr of observations from May 2006 to December 2010 are used to show that El Leoncito is clear approximately 75–80% of the time. No significant seasonal variations are observed. The optical ground-based data are compared with data from the Moderate Resolution Imaging Spectro-radiometer (MODIS) instrument on board the *TERRA* and *AQUA* satellites.

*Online material:* color figures

### 1. INTRODUCTION

The accurate determination of cloudiness conditions at an astronomical observatory is a key parameter for characterizing the quality of the site. Several different methods of determining cloudiness conditions, or the fraction of clouds, above an astronomical site have been used. One common method has been the use of images taken by instruments placed on different satellites. Thus, cloud sky coverage above astronomical sites has been commonly determined from satellite observations (Erasmus & van Rooyen 2006). Satellite images have the advantage of being easily accessible, provide a long-term database for many sites around the globe, and the data from each site can be reduced using comparable techniques. As we will discuss later, there are some challenging issues that need to be addressed in order to properly interpret the satellite results.

Another method of determining the fraction of clouds involves the use of ground-based all-sky cameras. Usually a set of reference clear sky images are created which are then compared to subsequent images taken at the same sidereal times (Shamir & Nemiroff 2005). To be effective, this technique needs

a sufficient database of relatively clear images, something that is not always available.

All-sky images are a compelling resource for determining the transparency of the atmosphere from moment to moment and night to night, due to the varying diminution of stellar images that ‘contaminate’ images of airglow. In essence, each spectroscopically stable star acts as a calibration lamp in outer space that is viewed through the atmosphere, thus revealing the effects of clouds, aerosols, or airmass-dependent atmospheric absorption. Not only do individual specific stars appear in multiple images at multiple airmasses, but individual images contain multiple usable stellar images that allow for consistent calibration across a single image and between images as different stars rise and set during the night.

One example of using all-sky images to estimate cloud coverage was presented by Aceituno et al. (2011), who were primarily trying to determine the surface brightness of the night sky at several bands over two observatories in Spain. They estimated cloud coverage by comparing the number of detected stars with the number of star detections expected in clear conditions based on a star catalog, and the ratio was taken as an estimate of the clear fraction of the sky. Although no other details of the method were described by Aceituno et al., we think that the method is probably not quantitative since a stellar image that is dimmed but not completely obscured by thin clouds would presumably count simply as a detected star, and the result would give clear sky conditions in the area.

All-sky cameras have also been deployed at the Thirty Meter Telescope candidate sites to determine cloud statistics. Skidmore et al. (2008) described two methods: an automated method based on photometric analysis of the images in order to determine atmospheric extinction, and a manual one to determine cloud statistics. The latter was based on active visual

---

<sup>1</sup> Center for Space Physics, Astronomy Department, Boston University, 725 Commonwealth Avenue, Boston, MA 02215.

<sup>2</sup> Space Science Center, University of New Hampshire, 8 College Road, Durham, NH 03824.

<sup>3</sup> Complejo Astronómico El Leoncito, Avda. España 1512 Sur, San Juan 5400, Argentina.

<sup>4</sup> Instituto de Tecnologías en Detección y Astropartículas, Parque Gral. San Martín, Mendoza 5500, Argentina.

<sup>5</sup> Universidad Tecnológica Nacional-Facultad Regional Mendoza, Rodriguez 273, Mendoza 5500, Argentina.

<sup>6</sup> Division lidar, CEILAP, UNIDEF, UMI-IFAECI-CNRS, J.B. de La Salle, 4397—B1603ALO Villa Martelli, Buenos Aires, Argentina.

inspection of movies made with red filter (RG695) and blue filter (BG38) images. Unfortunately the human element in this method results in many types of errors, including incorrect key strokes, subjective categorizations of features, and misidentification of airglow structures as thin clouds or geophysical variations (RG695 filters are typically used for mesospheric studies to detect gravity waves from OH emissions). Major improvement occurred when the method was compared to measurements from the Line of Sight Sky Absorption Monitor (LOSSAM) method, developed for the VLT Astronomical Site Monitor at Cerro Paranal (Sarazin 2000).

Here we describe a method using data from the all-sky imaging system at El Leoncito which is typically used to study atmospheric emissions coming from the upper atmosphere. The system takes images at several wavelengths, with a cadence of  $\sim 10$  minutes for a given filter. This data rate does not allow the formation of a database covering all possible sidereal times, and most of the images contain ever-changing airglow features, so the reference image method described earlier cannot be applied. Instead, one of the filters used with peak transmission at 644.4 nm serves as a background monitor because there are no emissions coming from the atmosphere at this wavelength. The images from this filter are used as a real-time sky monitor on a nightly basis at El Leoncito.<sup>7</sup>

The sheer volume of data that can be obtained with automated remote all-sky camera systems is both a boon and a problem for researchers. On the one hand, dozens or more stellar images appear in a single CCD image when skies are clear, and dozens to hundreds of CCD images can be obtained on a single night, so sky conditions can theoretically be determined with very good statistical certainty. On the other hand, measuring hundreds to thousands of stellar images per night in a data set spanning months to years presents a daunting task, and is nearly impossible to contemplate doing manually. We thus have a need for an automated method that can use stellar images in all-sky images to determine the clearness or cloudiness of all-sky images, both to facilitate the selection of high-quality all-sky images in our large data sets, and to characterize the average sky quality of future proposed astronomical or atmospheric observatory sites.

Using the relatively long term (5+ years) all-sky imaging database, a method to compute a clearness index has been developed. This method is based on the average diminution of stellar images as a function of ZA on clear nights. The method is completely automatic and there is no subjective categorization of cloud cover conditions.

In addition to characterizing the fraction of cloud coverage at El Leoncito Observatory, this work is being used to evaluate this observatory as the site for the Cherenkov Telescope Array (CTA), a project of nearly 1000 scientists from around

the world building two gamma-ray observatories, one each in the Northern and Southern Hemispheres. Each observatory or array will be composed of different telescope types. At the center of each array,  $\sim 25$  m diameter telescopes will sample the visible light produced by low-energy gamma rays ( $< 10$  GeV). Surrounding these telescopes, medium sized telescopes ( $\sim 12$  m diameter) will collect light from medium energy gamma rays (100 GeV to 1 TeV). Completing the array, many small telescopes ( $\sim 6$  m diameter) will be installed around the medium size ones, and they will collect light from high-energy gamma rays ( $> 1$  TeV).<sup>8</sup> Several location characteristics will be studied to choose the site: (1) atmospheric conditions; (2) soil characteristics; and (3) infrastructure. Item (1) involves temperature, wind speed, humidity, pressure, aerosols, night-sky brightness and nighttime cloud coverage. Determination of nighttime cloud coverage is the topic of this study.

Section 2 will give a brief description of the imaging system and its use for upper atmosphere studies. Section 3 is a detailed description of the method developed to determine nighttime clear sky conditions. Results are presented in § 4. In § 5 a comparison with space-based observations is performed, and, finally, a summary is presented in § 6.

## 2. ALL-SKY IMAGING SYSTEM AND AIRGLOW MEASUREMENTS

An all-sky imager (ASI) system built to measure mesosphere and thermosphere airglow emissions (Martinis et al. 2006) is used to study clear sky conditions at El Leoncito Observatory. The front objective of the system is a 16 mm f.l. F/2.8 fish-eye lens. This lens forms a 40 mm diameter image of the sky near the flat surface of a 70 mm f.l. field lens. This field lens directs the light into a 340 mm f.l. F/3.4 collimating lens. At this pupil, a filter wheel is used to select one of six interference filters. A typical filter would have a bandwidth of  $\sim 1.2$  nm, half-power full width (HPFW), to allow rays up to  $4.5^\circ$  to the optical axis to pass through the high-transmission (90%) region of the filter curve. After the filter, a 135 mm f.l. F/1.5 lens reimages the 40 mm diameter image to  $\sim 22$  mm diameter onto a 26 mm square CCD. The CCD camera used is an ANDOR DW 436 BV, back-illuminated, thinned CCD of  $2048 \times 2048$  pixels. The CCD is normally binned  $3 \times 3$  before read-out to produce a  $682 \times 682$  array (a sky image 540 pixels in diameter). The detector assembly is housed in a thermoelectrically cooled chamber capable of cooling to  $-60^\circ$  below the ambient temperature. The read-out noise is typically  $3 e^-$  (rms), the gain is  $\sim 1.3 e^-/\text{dn}$ , and the dark signal is  $\sim 0.05 e^-/\text{s}/\text{binned pixel}$ . More details can be found in Baumgardner et al. (1993) and Matta et al. (2009). The field of view of the instrument is  $180^\circ$ , yielding an average plate scale of  $0.34^\circ$  per pixel. A typical stellar image has a point spread function of  $\sim 3$  pixels (HPFW).

<sup>7</sup> See <http://www.casleo.gov.ar/imager/imager.php>.

<sup>8</sup> Further information can be found at <http://www.cta-observatory.org/>.

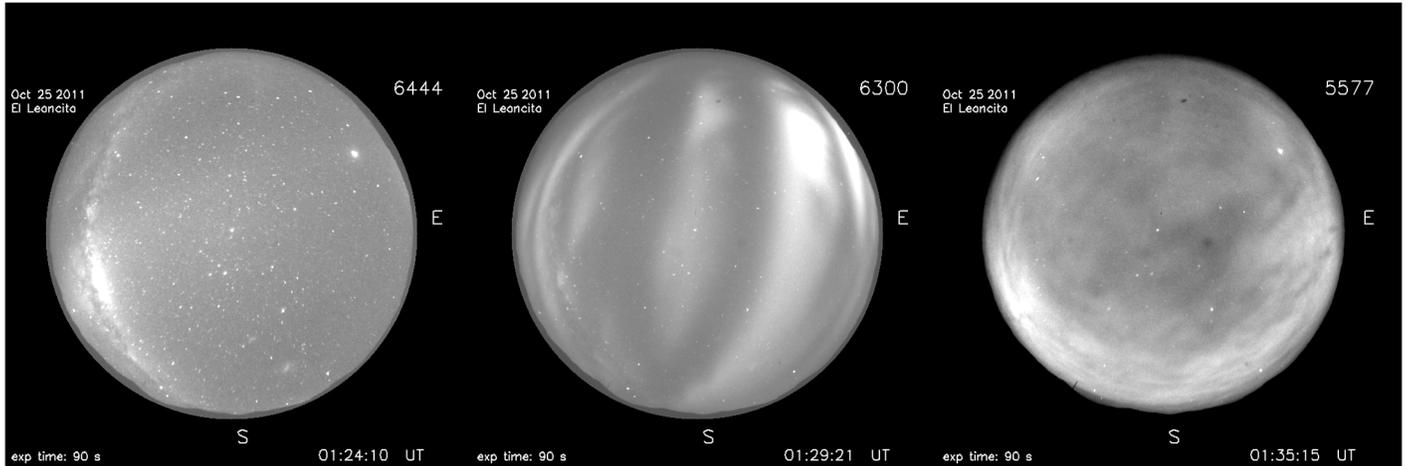


FIG. 1.—Images taken on 2011 October 25. Background filter (*left*) shows no structuring in the airglow. Image at 630.0 nm (*center*) shows strong variations in airglow due to thermospheric processes. Image to the *right*, at 557.7 nm, shows the presence of mesospheric gravity waves. Images from [www.buimaging.com](http://www.buimaging.com).

The system takes images at several wavelengths for the duration of the night on nearly every night in a given month. Thermospheric emissions of neutral oxygen are measured at wavelengths of 630.0 and 557.7 nm (emissions due to dissociative recombination of  $O_2^+$ ), and 777.4 nm (due to radiative recombination of  $O^+$ ). Mesospheric processes are studied by measuring emissions at 589.3 nm (neutral Na emissions), 557.7 nm (via the three-body Barth reaction), and lines with  $\lambda > 695.0$  nm (OH Meinel-band emissions). A background filter at 644.4 nm, chosen to avoid thermospheric and mesospheric emission lines, is used as a control filter. Exposure times are 90 s for all the filters except hydroxyl, which has an exposure time of 30 s. An entire cycle of observations takes  $\sim 10$  minute, so there are usually  $\sim 50$  images per filter per night during the summer and  $\sim 80$  images per filter per night during winter. The only nights when data is not taken are those within a day of full Moon, so data are collected on 27–28 nights per month. All the data are automatically transferred to a Boston University server at the end of the observing night.<sup>9</sup>

The features seen in the images depend strongly on the wavelengths of the filters used. Figure 1 shows an example of images taken at 644.4, 630.0, and 557.7 nm on the night of 2011 October 25. Structures can be seen in 630.0 and 557.7 nm, while the background filter shows a relatively uniform brightness. The structures observed with the 557.7 nm filter are typical mesospheric gravity waves, while the ones observed at 630.0 nm are airglow depletions associated with equatorial spread F (Martinis et al. 2006), an ionospheric irregularity that significantly affects the propagation of radio signals and disrupts communication and navigational systems. Figure 1 shows that the 644.4 nm is suitable for carrying out cloudiness

measurements due to the lack of ‘contamination’ coming from the processes occurring in the mesosphere and thermosphere. The next section describes the method we have developed to characterize cloudiness conditions at El Leoncito Observatory.

### 3. METHOD

#### 3.1 Using Stars in All-Sky Images

All-sky atmospheric emission images have some unique properties and processing requirements that may be unfamiliar to researchers working with astronomical images. The  $180^\circ$  field of view (FOV) of all-sky cameras measures light coming from the entire sky, but distortion effects near the edges of the FOV limit the effective range of the measurements to  $80^\circ$  zenith angle (ZA), although for some applications data can still be analyzed at  $ZA > 80^\circ$ . The spatial coverage of the FOV depends on the particular emission being considered. For example, 630.0 nm airglow emission peaks at  $\sim 250$  km during low solar activity, and thus a  $160^\circ$  FOV will cover an area of  $\sim 6.3 \times 10^6$  km<sup>2</sup>. For 557.7 nm emissions, which come from a much lower altitude of 96 km, the area covered is only  $\sim 9 \times 10^5$  km<sup>2</sup>. For this study the brightness of stars at 644.4 nm is being analyzed. This light does not come from a particular atmospheric height; however, the absorbers of this light, e.g., clouds, are found at heights of  $\sim 10$  km, meaning a  $160^\circ$  FOV corresponds to a comparatively tiny area of  $\sim 1 \times 10^4$  km<sup>2</sup>.

To allow for the characterization of clearness/cloudiness in different parts of the sky using our all-sky images, we divide the sky into five zones: a circular central zone and four surrounding sectors. With this configuration, shown in Figure 2, clouds to the north, south, east, west and zenith sectors can be studied. The central zone is bounded by the circle at ZA of  $45^\circ$ , and the surrounding sectors are limited to  $80^\circ$  ZA. For this study, we focus only on measurements inside the central

<sup>9</sup>Quick-look images and movies can be seen at [www.buimaging.com](http://www.buimaging.com).

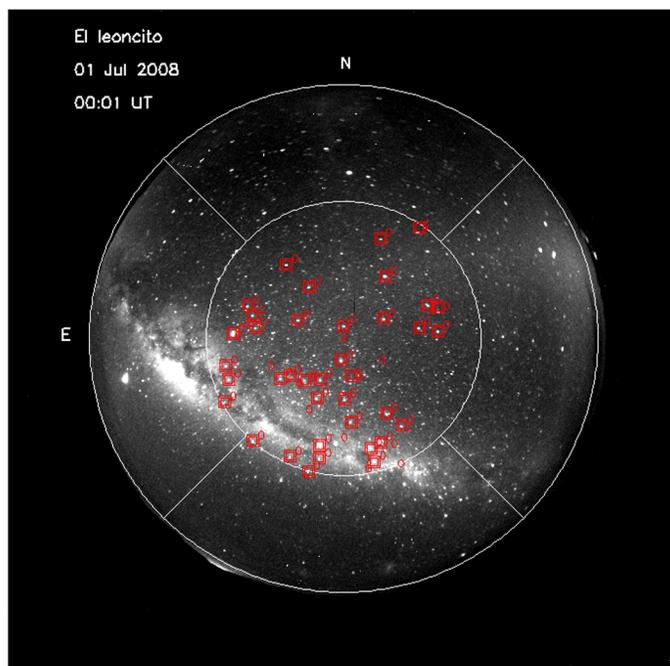


FIG. 2.—Graphical representation of star mapping and zone plotting. The *inner circle* represents a  $45^\circ$  zenith angle. The *outer circle* is used to define  $80^\circ$  zenith angle. The four trapezoidal boxes separate other zones. The number of zones can be specified by the user. *Red boxes* centered in each star are drawn to compute star flux in data number (DN). See the online edition of the *PASP* for a color version of this figure.

zone and leave the analysis of the surrounding areas to future studies. The automatic method described here does not allow for the distinction between patchy clouds (star images varying from unobstructed to completely obstructed) and continuous layers of cirrus (star images partly obscured). In the future, an additional level of analysis may allow the identification of the different type of clouds.

A data cadence of  $\sim 10$  minutes does not allow for the creation of a database of clear skies covering all possible sidereal times, so our method obtains an average clear sky relationship between star brightness and ZA that is compared with individual nights to determine a clearness index. The stars in the FOV are identified using the Yale Catalogue of Bright Stars (Hoffleit & Warren 1991) which includes all stars of stellar magnitude 6.5 or brighter (9110 stars). For this work we chose a limiting magnitude of 3.5 to single out the brighter stars that are easier to measure while still including a sufficient number of stars to obtain statistically meaningful results. Typically  $\sim 150$  stars are identified in a single image, and only those stars that are measured more than 100 times over the course of several nights are included in the analysis.

Several steps are required to obtain star fluxes in an all-sky image. First, the position of all of the stars being considered is computed for the camera location, the geometrical mapping of the optical system, and the UT time and date. A  $5 \times 5$  pixel box

centered on the predicted location of each star is sampled, and then the actual location of the star image within each box is determined. The boxes are re-centered on the star images, and the total data number in each box is calculated. A one pixel wide frame is then drawn around each box (see red boxes in Fig. 2), and the median value of this frame is taken to be the background level, which is then subtracted from each pixel in the original  $5 \times 5$  box to obtain a new total star brightness. The stellar images in the system are under sampled, meaning there are too few pixels with which to fit a Gaussian to the stellar image profile, so such a method of measuring the total data number in a star image would not improve the results. Finally, a plot of flux in data numbers versus ZA is then generated for each star on each night.

### 3.2 Clearness Index Determination

The next major step in our method is to single out and measure those star images which were taken through a clear, cloud-free atmosphere so that we can determine what star brightness to expect on clear nights and then compare those to all available star images to determine the degree of cloudiness affecting their fluxes. The obvious first problem is that each star in the sky has a different brightness from every other star—even on clear nights—so the first step is to normalize or rescale every star all to a single brightness value at a fixed ZA. We rescale the brightness of each star to a value of 1000 data number at a ZA of  $45^\circ$ . This is done by calculating a multiplicative linear factor specific for each star that raises or lowers its brightness at a given ZA. In this way, if a star's brightness is reduced by 50% from its scaled clear-sky brightness, we derive an equal cloudiness factor for that portion of the sky regardless of whether the star is intrinsically bright or dim.

We next use an iterative culling method to weed out star fluxes which are either too low (behind clouds) or unusually bright due to measurement errors, cosmic ray hits, or possible star misidentifications. Once all the star brightnesses have been rescaled and plotted as a function of ZA, a hyperbolic curve is fitted to the data. In general, we observe that star brightnesses are a hyperbolic function of the zenith angle, so we fit a hyperbolic curve using multiple nights to ensure that there is enough clear-sky data to create an obvious trend in a plot of scaled brightness versus ZA. In practice, more clear-sky data points are available than cloudy ones, so there tends to be a preponderance of data points along the brightness versus zenith angle curve that corresponds to clear sky conditions. One or more months of data can be used to calculate the clear-sky curve. The disadvantage of combining several months is that sky transparency could be affected by seasonal variations. The initial hyperbolic fit to brightness-versus-ZA points inevitably falls below the clear-sky level, since there are some dimmer points corresponding to occasional cloud cover. Thus, for the purpose of singling out the brightness values for clear conditions, we discard the data falling below 2 standard deviations from the

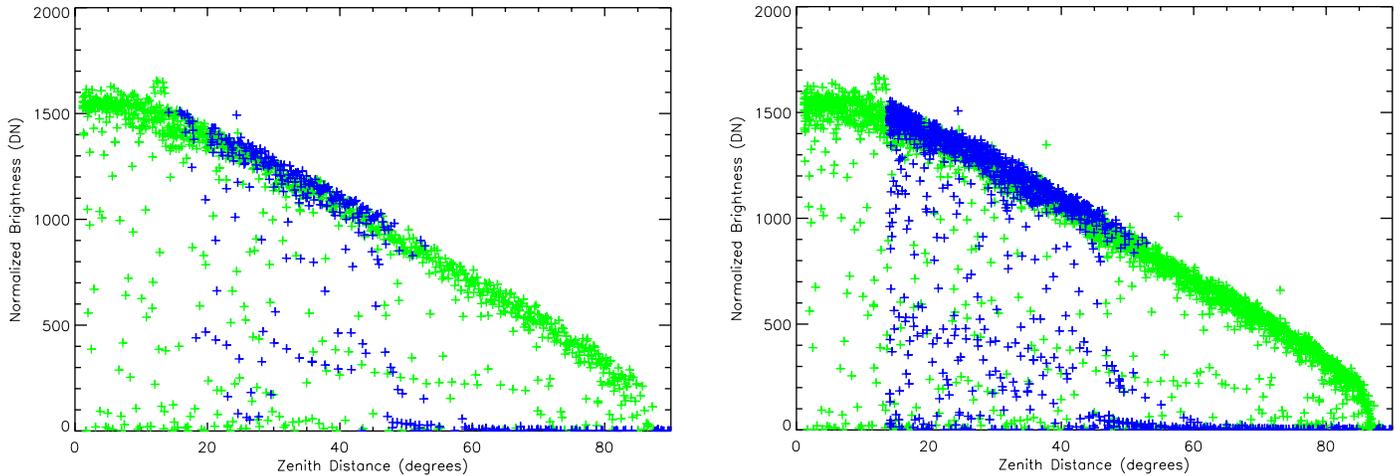


FIG. 3.—*Left*: Example of the scattering in the data leading to the formation of a ‘clear sky band’ using 1 month worth of data for two stars, Sirius (*blue crosses*) and Fomalhaut (*green crosses*). Values have been normalized to a data number of 1000 at  $45^\circ$  ZA; *Right*: Similar to the left panel, but now using data from 3 months. See the online edition of the *PASP* for a color version of this figure.

initial fitted curve (and data above 3 standard deviations from the curve) and then repeat the fitting process. This process is iterated ten times, which proves in practice to settle on the clear-sky brightness values. The hyperbolic curve fitted to the data is then used as a reference to calculate the clearness index values using all star images in the data.

It should be noted that the fitted curve represents a combination of atmospheric extinction and instrumental vignetting, which is the off-axis response of the imaging system that reduces brightness at the rim of the fish-eye lens. Vignetting can be calculated at the lab before shipping the instrument or on site, if needed, and if one were to remove vignetting effects, the resulting curve would characterize atmospheric extinction on cloudless nights.

Allowances must be made for small variations in star brightnesses from the clear sky curve on clear nights, because an individual star’s brightness versus ZA does not follow a perfect single curve. This could be due to the fact that images are not corrected for pixel-to-pixel sensitivity variation, because the glass dome covering the system occasionally could have blemishes (dirt), or because the calculation of a star’s brightness can be affected by the presence of dimmer stars nearby that wander in and out of the measuring box, which changes the measured brightness. Figure 3 shows the result of computing data number versus ZA for only two stars, Sirius and Fomalhaut. The left panel shows points calculated with 1 month of data, while points in the right panel are the result of including 3 months of data. For the 3 months of measurements, 2327 data points for Fomalhaut and 1406 points for Sirius were calculated. In both cases, stars observed during clear skies are not clustered around a thin line, but are scattered inside a “band”. Points below this area therefore represent cloudy sky conditions.

Figure 4 demonstrates much of what we discussed above. The left panel in Figure 4 shows the result of computing the

normalized fluxes of all stars brighter than magnitude 3.5 during the month of 2010 June. The red line represents the fit to the rescaled brightness values versus ZA using the hyperbolic function. Because we are only considering the central zone in each all-sky image, only the portion of the fitted curve between  $0^\circ$  and  $45^\circ$  is used to compute the clearness index for specific nights. The red vertical line represents  $1\sigma$  variability computed from a Gaussian fit to the residuals from the fitted curve, i.e., a point-by-point absolute difference.

The right panel of Figure 4 shows the histogram of residuals to the fit in the left panel, and the dashed line is the Gaussian fit to the histogram of the residuals. These are the differences between the brightness of each stellar image from the fitted clear sky curve. Note in the right panel of Figure 4 that the fit is made just to obtain an accurate value of the width of the curve, not its peak or tail values. This distribution is asymmetric, with more points to the left, representing the cloudy nights. The width of the Gaussian is calculated as its standard deviation and this value is used to represent the width of a clear sky band, as shown with the vertical red line in the left panel of Figure 4. Points below this band represent partially cloudy and cloudy skies. Applying this method to all the months of a year, a monthly ‘width’ can be obtained. This width varies from  $\sim 60$  data number to  $\sim 100$  data number, representing a scattering from  $\sim 12\%$  to  $\sim 20\%$  at  $45^\circ$ .

To quantify how clear average sky conditions are over El Leoncito, a clearness index is defined as the median value of the ratios of all star fluxes inside the central zone to the fitted line at the corresponding zenith distance. This is shown in Figure 5, where sky conditions during 2010 May 13 (left) and 2010 June 12 (right) are shown. The individual points represent the ratio of an individual rescaled star flux to the fitted curve. The asterisks are the median values for each time (i.e., each image). The June 12 case shows clear sky conditions during the entire night, with

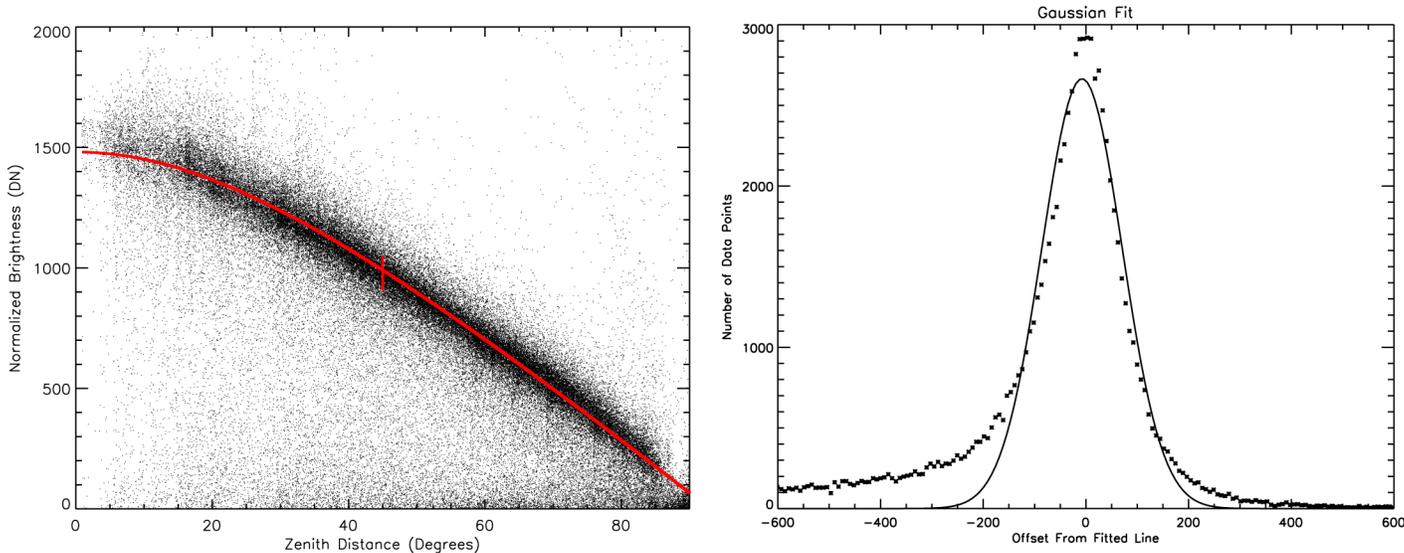


FIG. 4.—*Left*: Normalized data points for 2010 June. The *solid red line* is the fitted curve to the data points representing clear sky conditions. The *vertical red line* indicates the characteristic scattering ‘width’ for clear sky conditions; *Right*: Histogram of clustered points around the clear sky curve fitted to the data. The width is extracted from the fitted Gaussian. A value of 78 data number is obtained, and thus clear sky conditions could be defined by the fitted curve  $\pm 78$  data number. See the online edition of the *PASP* for a color version of this figure.

clearness index values close to one. Notice that individual points can have values larger than one; this simply means that a particular star flux was higher than the clear sky curve value at that zenith angle. The night of May 13 shows clear skies, i.e., clearness index  $\sim 1$ , only at the end of the night.

#### 4. RESULTS

Figure 6 shows a sample of the type of basic output that can be produced. Each asterisk, calculated from individual star

ratios, as shown in Figure 5, represents average conditions inside  $45^\circ$  ZA. In this sample, only July 6 and July 7 did not show clear skies the entire night. Data for July 9 show that until  $\sim 0400$  the clear sky index fluctuates between  $\sim 0.8$  and  $\sim 0.9$ ; this could be evidence of partially cloudy conditions, something that will be quantified in this section. Issues that potentially influence the results include the presence of bright objects or features, like Jupiter, the crescent Moon, or the bulge of the Milky Way, which can affect the individual star flux calculation. The resulting fluxes will be affected due to the enhanced background

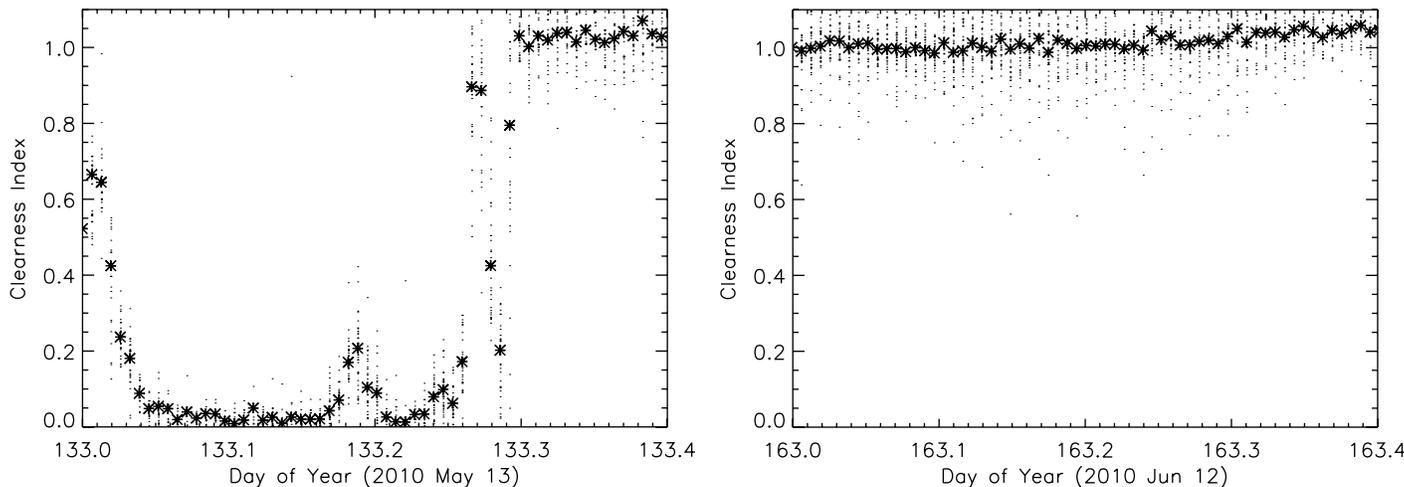


FIG. 5.—Examples of clearness index calculation for the nights of May 13 (*left*) and June 12 (*right*). The individual points at each time (i.e., each image) represent the ratio of a given star flux to the fitted clear sky curve at the respective zenith angle. At each time the median value is calculated and shown as an asterisk. June 12 shows a completely clear sky night, while May 13 shows that only at the end of the night the sky was cloud-free.

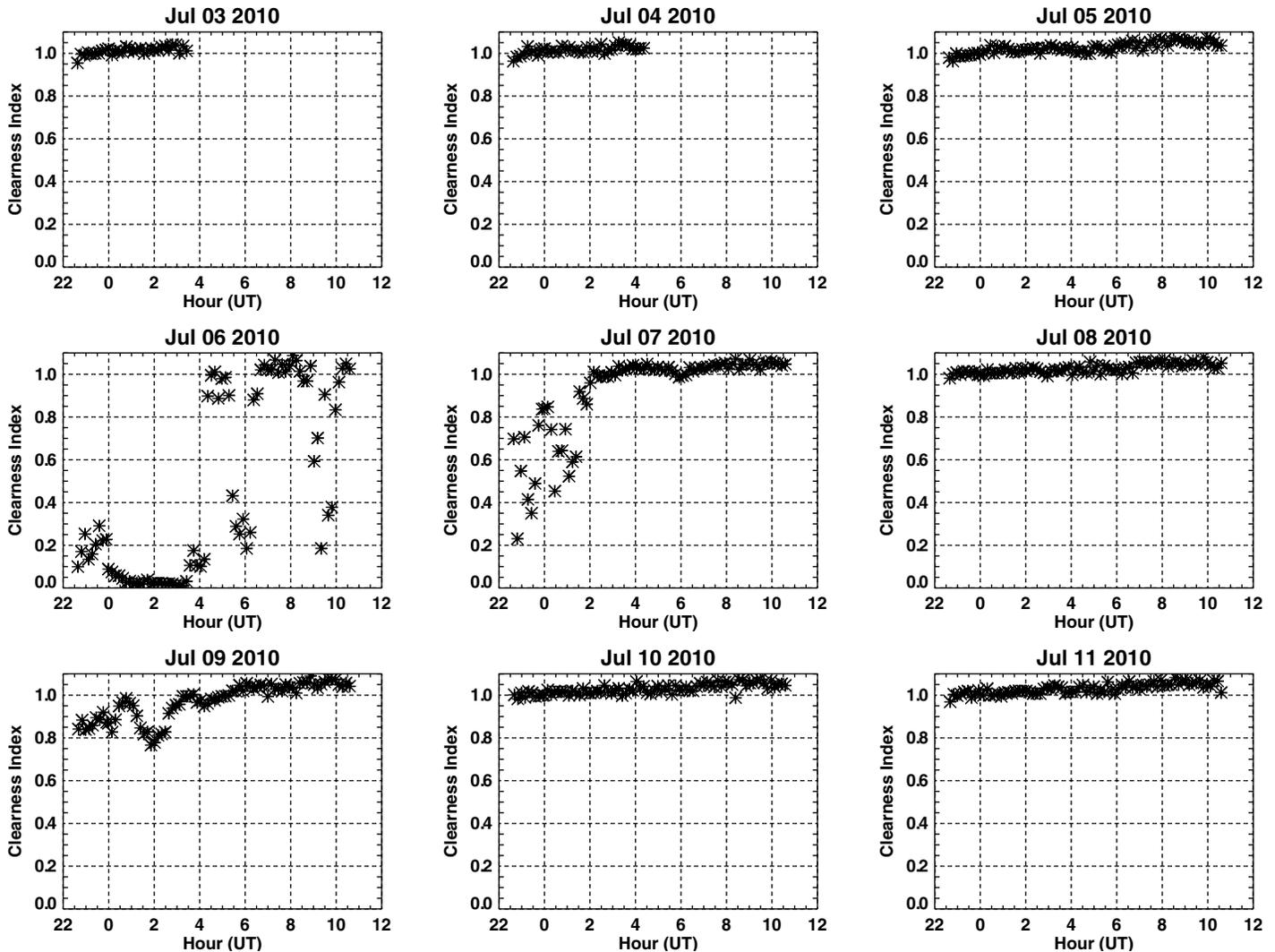


FIG. 6.—Summary of the results for several days in 2010 July. The *horizontal* axis represents universal time. The *vertical* axis represents clearness index. July 3, 4, 5, 8, 9, 10, 11 show values close to 1 during the entire night. July 9 indicates that clear skies occurred only after  $\sim 0400$  UT. July 6 shows clear skies at  $\sim 0500$  and between  $\sim 0700$  and  $0900$ . July 7 is cloudy until  $0200$  UT.

produced by the bright features that can interfere with star flux calculations. Another problem encountered is the presence of condensation on the dome, either outside or inside, due to large gradients in temperature. This results in very small values of the clearness index, even though the sky can be perfectly clear. Approximately 500 hr of observations were affected by frosted dome conditions during the 2006–2010 period, representing less than 5% of the data.

Having determined a clearness index, the next step is to quantify the definition of a clear sky. There is not necessarily one unique definition of a clear sky, since it can depend on the purpose of each individual study. In one a very stringent (photometric) definition, clear skies can be identified when the rms fluctuation of the flux of a star is less than 2% (Patat et al. 2011).

A less restricted definition of clear skies that is more appropriate for spectroscopic work includes any night when cloud cover is less than 20%. This criterion has been used in studies investigating cloudiness characteristics of astronomical observatories (Xia 2012; Moles 2010; Sanchez et al. 2007), and it is also adopted here to characterize hourly and seasonal behavior of clear sky conditions at El Leoncito.

Given the errors inherent in measuring stellar images, and therefore in determining the clearness index for an image, it follows that many images of 100% clear sky conditions will yield clearness indexes below or above 1. Figure 7 shows the histogram of all data points for the year 2010 with data binned every 0.01 clearness index intervals. A total of 13,006 points (or images) are included in the histogram. Data are clustered

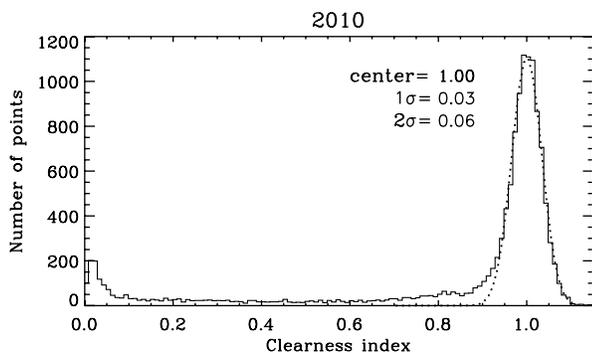


FIG. 7.—Histogram of data points binned every 0.01 clearness index intervals for the year 2010. Data are clustered near clearness index equal to 1. Criteria to define 100% clear skies use all the points inside the fitted Gaussian (dotted line) at  $1\sigma$  level (index  $>0.97$ ) or at  $2\sigma$  level (index  $>0.94$ ). This gives 100% clear skies 63% of the time for the first definition and 68% of the time for the second definition.

around a clearness index of 1, an indication of the good sky conditions found at El Leoncito. Assuming that the spread around the value of 1.0 is due to expected star measurement errors, a Gaussian fit to this histogram, shown by the dashed line in Figure 7, can be used to define the range of indices when 100% clear-sky conditions are found. Since the half-width of the Gaussian (0.03) approximately represents the  $1\text{-}\sigma$  spread of clear sky values, then clear skies include all the points with a clearness index greater than 0.97, representing 63% of the time for the year 2010. If one computes the average width for the 12 months of the year, the result is equivalent to the one obtained for Figure 7. Similar width is found for the other years used in this study.

It may be possible to have a more generous definition of 100% clear skies. An inspection of the fitted curve in Figure 7 shows that data points outside the  $1\text{-}\sigma$  level are still well described by the lower wing of the Gaussian, and therefore may still represent 100% clear skies. Thus one could expand the expected error in the clear sky index to  $2\sigma$  (0.06), meaning 100% clear skies would correspond to points with an index

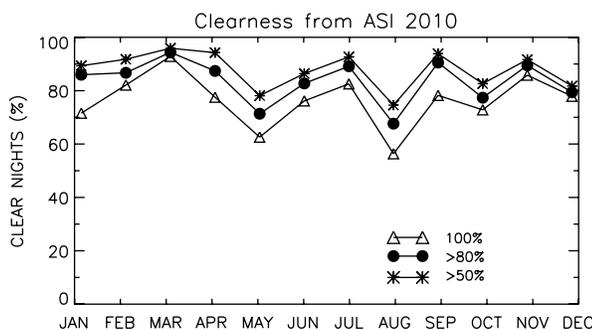


FIG. 8.—Percentage of time with 100%, 80%, and 50% clear skies for the year 2010. No clear seasonal pattern is observed. May and August are the cloudiest months.

TABLE 1  
SUMMARY OF HOURS OBSERVED AND PERCENTAGE OF TIME WITH 100% AND 80% CLEAR SKIES, DEFINED AT THE  $1\sigma$  AND  $2\sigma$  LEVELS, RESPECTIVELY

Year	Hours observed	% clear sky		% clear sky	
		100	>80	100	>80
		$2\sigma$		$1\sigma$	
		% of time		% of time	
2006	1,337	70	82	63	80
2007	1,990	69	79	63	78
2008	2,115	65	75	56	74
2009	1,957	69	80	62	78
2010	2,020	68	82	63	81

above 0.94, which would mean the skies were clear 68% of the time in 2010.

We can also define index ranges for different percentages of cloudiness. For example, if “practically clear skies” can be defined as times when clouds cover less than 20% of the sky, then for the year 2010, points with clear sky index greater than 0.74 (0.80 minus 0.06 uncertainty) represent these conditions, and they occurred 82% of the time.

Figure 8 shows the percentage of time with 100%, 80%, and 50% clear skies for the year 2010. As seen, 100% clear skies occurred 68% of the time, while 80% clear skies and 50% clear skies occurred 82% and 88% of the time, respectively. March was a particularly good month in terms of clear-sky nights: out of the 163 hr with observations, the number of hours with 100% clear skies was 151 (~93%).

Table 1 shows a summary of the results for the 2006–2010 period, assuming widths at  $1\sigma$  and  $2\sigma$  levels. A total of 9348 hr were used, with 7442 hr showing at least 80% clear skies. The total number of hours per year is the sum of the nightly number of hours. Except for the year 2006 when measurements started in May, our instrument recorded a relatively uniform number of hours throughout each, with no particular bias. This brings the average percentage of time with clear skies for the entire 2006–2010 period to ~79.6%. Even if the criterion is constrained to the conservative  $1\text{-}\sigma$  uncertainty definition of clear skies, the respective percentages of clear skies for the years 2006 through 2010 are 80%, 78%, 74%, 78%, and 81%, respectively, with 80% clear skies occurring ~78.2% of the time in the grand average.

Thus, all-sky images showing less than 20% of cloud coverage are catalogued as clear-sky images from which hourly statistics can be computed for each month. Figure 9 shows an example for June 2010. The top plot has the number of hours observed (gray bars) and the number of hours with clear skies (black bars). The ratio of the two, expressed in percentage, is shown at the bottom.

Figure 10 shows monthly averages for all the times in the year 2010, in a similar format then the one shown in Figure 9.

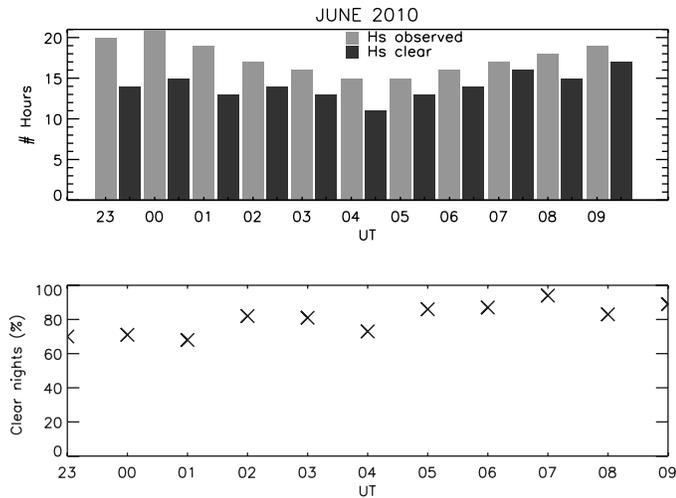


FIG. 9.—*Top*: Number of hours observed and number of clear skies determined by considering images with cloud coverage less than 80% for the month of June 2010; *Bottom*: Percentage of clear nights as a function of time. This kind of plot can be used to characterize nighttime variability of cloudiness conditions, e.g., to check how cloud coverage changes during a given night.

The error bar on the June 2010 data point represents  $1\sigma$  variability.

Figure 11 shows the monthly behavior for the 5 years included in the study. The plot indicates larger variability during the summer months. Solid black circles represent the 5 year average for each month. Overall an average of  $\sim 78\%$  clear skies is obtained. The results from Figures 10 and 11 show no strong seasonal dependence of cloud coverage over El Leoncito: skies are relatively clear throughout the year.

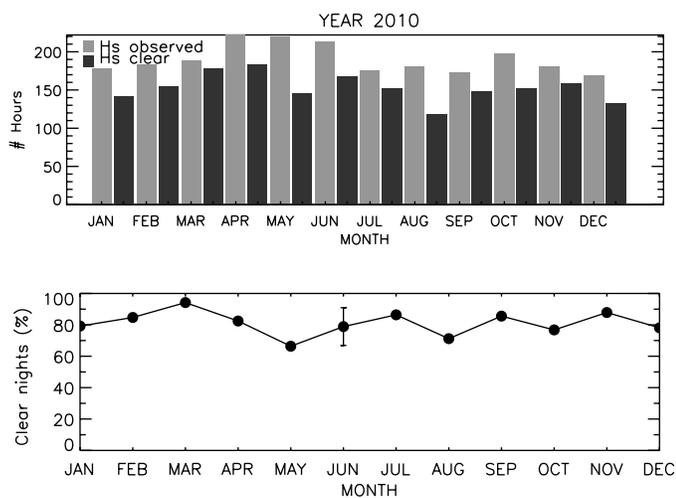


FIG. 10.—*Top*: Number of hours observed and number of clear hours as a function of month of the year; *Bottom*: Average clear sky conditions for the year 2010. The error bar on the June data point represents  $\pm 1\sigma$  standard deviation.

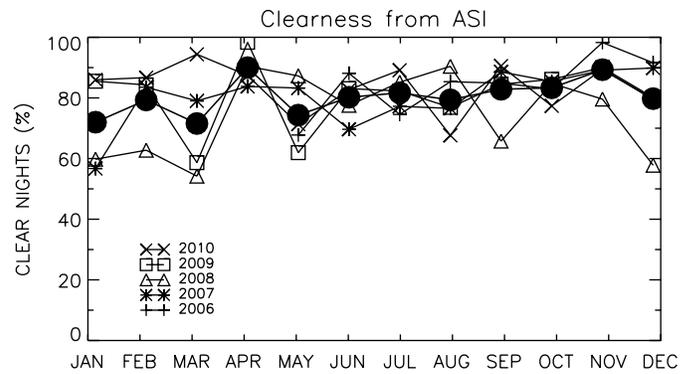


FIG. 11.—Average results for the 5 years analyzed. The plot shows more variability during December–February months. Solid black circles represent the 5 year average for each month. No significant seasonal variation is observed.

## 5. COMPARISON WITH SPACE-BASED SATELLITE DATA

Sky transparency is one of the key parameters used to judge candidate locations for an astronomical observatory. Satellite observations have commonly been used for this purpose (Erasmus & van Rooyen 2006), as such data is easily accessible and often cover a long period of time over many sites around the world. However, the analysis of satellite data can be hindered by factors such as winter snow coverage, complex topographies, and variable vegetation. The Andes are a good example of a location with both rough topography and variable snow cover, and data with low spatial resolution can produce inaccurate results (Kurlandczyk & Sarazin 2007).

Data from non-geostationary satellites suffer the additional shortcoming of not providing continuous nighttime measurements, since they do not linger over one area. In light of the limitations of satellite data, ground-based data provide obvious advantages. Since ground-based data come from upward-looking cameras, they are not affected by topography or changes in ground vegetation or albedo. In addition, ground-based imagers can also provide data continuously over an entire night.

To test the ground truth of our all-sky camera technique, we compare our sky analysis of El Leoncito with coincident satellite data. The TERRA and AQUA satellites each house a Moderate Resolution Imaging Spectro-radiometer (MODIS) that can be used to compute cloudiness conditions over El Leoncito. Both satellites have Sun-synchronous retrograde orbits, with a period of 98 minute at 654–685 km height.<sup>10</sup> The cloud coverage used for this analysis belongs to the TERRA and AQUA Level 2 Products (MOD06\_L2) which were retrieved from the database at <http://ladsweb.nascom.nasa.gov/>. Level 2 products have a spatial resolution between 1 and 5 km. We use 2 years of nightly AQUA data taken between 0510 and 0635 UT and

<sup>10</sup> More details on the satellite instrument can be found at <http://modis.gsfc.nasa.gov/>.

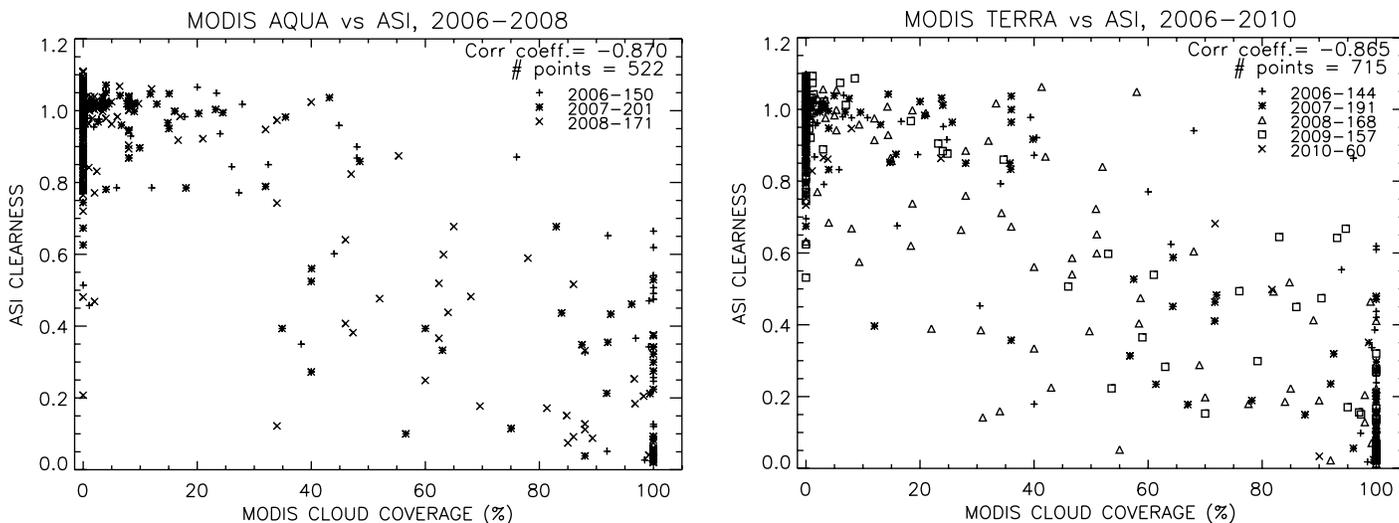


FIG. 12.—*Left:* AQUA satellite data compared with ASI between 2006 and 2008; *Right:* TERRA data compared with the ASI between 2006 and 2010. Both panels show that data are mostly clustered around the clear sky values, i.e., ASI clearness  $\sim 1$  and MODIS cloud coverage  $\sim 0\%$ , and around cloudy sky values, i.e., ASI clearness  $\sim 0$  and MODIS  $\sim 100\%$ . The correlation coefficients shown are calculated assuming a normal distribution (see text).

4 years *TERRA* data from 0240 to 0410 UT, and plot the results against our all-sky data in Figure 12. Notice that while our all-sky imager data is expressed as a clearness index, the satellites measures the percentage of cloudiness—so a clearness index of 1 corresponds to 0% cloudiness and vice-versa.

Interestingly, the satellite cloudiness data is not necessarily well correlated with our ground-based data. A simple analysis of the plots in Figure 12 gives a correlation coefficient of  $-0.87$ , but significant scatter is evident in the data, even when completely clear or cloudy conditions are measured. A more accurate analysis that considers the nature of the data distribution shows that neither series has a normal distribution (most of the data are clustered near very small or very large values),

meaning the “typical” or Pearson’s correlation coefficient is not appropriate (Kendall & Gibbon 1990). The more representative Spearman’s correlation coefficient, which does not require data series to have a normal distribution, is computed. The correlation obtained now is  $-0.69$  with a  $p$ -value = .999, using a  $t$ -distribution. The square of the Spearman’s correlation coefficient (0.48) gives the percent of variation in one variable that is due to the other; meaning that in this case less than 50% of the variability seen in the ground-based data is explained by the satellite data.

This poor correlation between the satellite and all-sky imager data is probably due to several factors, including: different wavelengths used to measure clouds (644.4 nm vs. IR channels), different areas sampled by the two methods, complex topography conditions that affects the satellite results, and limited spatial resolution from the spaced-based diagnostics (typically only 1–2 pixels covering the area sampled by our camera).

The two datasets are more consistent in describing average year-to-year variations in cloudiness. Figure 13 shows yearly average values from MODIS *TERRA* and all-sky imager data. These data include all the observed nights with the two methods. Since MODIS measures cloud coverage percentage, we subtracted this percentage value from 100% in order to show clear sky conditions; i.e., if MODIS is giving 20% cloudiness, we re-sampled that value to indicate 80% clear sky). Representative error bars are shown for all-sky imager data in 2007 and for MODIS data in 2008. MODIS data are consistently lower than all-sky imager data, but similar trends are observed with the cloudiest conditions occurring in 2008 in both datasets.

Although short-term disagreements between the two datasets are expected as discussed above, it is not obvious why the

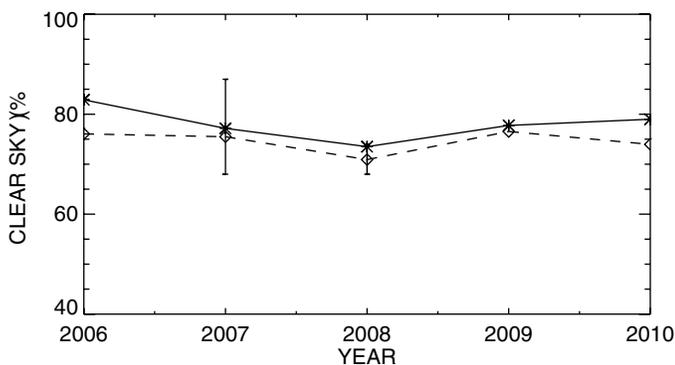


FIG. 13.—Comparison of yearly values using both techniques. Solid line represents clearness index multiplied by 100. Dashed line is MODIS results converted into clearness percentage, i.e., (100- MODIS data). Notice how 2008 stands out as a particular cloudier year. Representative error bars for the ASI and for MODIS are shown for the years 2007 and 2008, respectively.

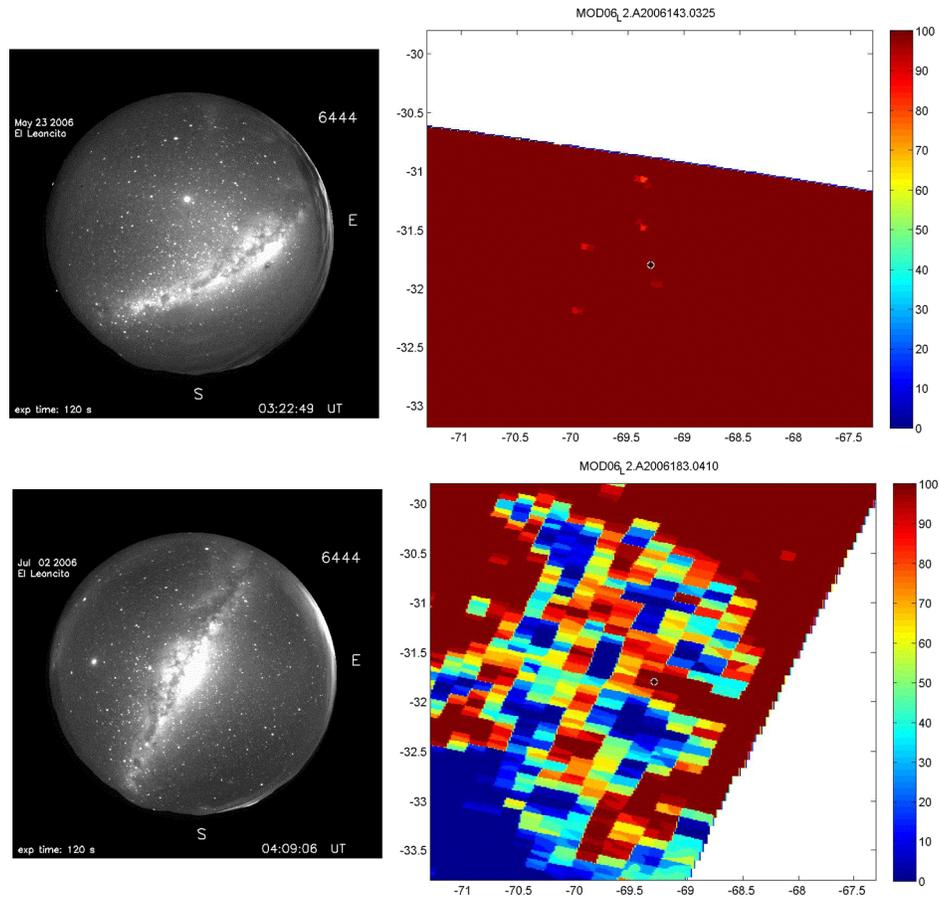


FIG. 14.—*Top*: A case on 2006 May 23, when the ASI data (*left*) showed clear skies, while the satellite data (*right*) showed completely cloudy skies. The time difference between both images is 3 minutes. *Bottom*: A similar case, 2006 July 2, with a time difference of only 1 minute. Only 1–2 pixels centered at the location of El Leoncito Observatory ( $-69.3, -31.8$ ) are used in the satellite-based cloudiness determination. The ASI data show for both cases indices  $\sim 1$  (i.e., clear sky), but the satellite data indicate  $\sim 100\%$  (i.e., cloudy). See the online edition of the *PASP* for a color version of this figure.

satellite data consistently underestimate clear sky conditions relative to ground-based data. Even if all-sky imager clearness values were computed using the  $1\text{-}\sigma$  threshold as in Figure 7, the satellite-based data would still give lower rates of clear sky conditions. One possibility is that high-elevation terrain in the Andes results in excess cloud reflectance in MODIS on very dry days when water vapor absorption at  $1.38\ \mu\text{m}$  is very low. We do see several cases when the all-sky imager data was showing completely clear sky conditions while the satellite data gave values corresponding to cloudy conditions, and some examples are shown in Figure 14 for 23 May 2006 and 2 July 2006. In both cases, the ASI shows completely clear skies in both cases, with clearness index  $>0.9$ , while MODIS data show 100% cloudy skies for one case and 96% for the other.

## 6. SUMMARY

A method to measure clear sky conditions above El Leoncito Observatory from an all-sky imager used to study the upper

atmosphere has been described. The method is automatic, and even though it does not use the traditional wavelength bands to determine cloud coverage conditions, it provides valuable information of a key parameter to determine the quality of sky conditions above a site. Overall, sky conditions at El Leoncito are clear between 75% and 80% of the time, values comparable to good astronomical observatories. In addition, El Leoncito shows relatively uniform clear-sky conditions the entire year, without strong seasons of rain or large cloud coverage conditions. This is a highly desirable situation if one wants to sample different regions of the sky all year around.

Additional studies can be made with the set of all-sky imager data if vignetting is removed; this would allow the study of extinction curves from zenith to the horizon on a nightly basis, valuable information for characterizing the sky quality.

Our results have been compared with space-based data from the *TERRA* and *AQUA* satellites that measure cloud coverage. The differences suggest that the satellite measurements may be consistently over-estimating cloudiness at El Leoncito.

The method described here can be applied to improve the climatology of clear sky conditions at any site with existing multiyear archived all-sky data of terrestrial airglow.

We thank the continuing support of Ricardo Gil-Hutton, Director of the Complejo Astronomico El Leoncito (CASLEO).

The entire Observatory staff has provided invaluable help during all the years of operation of the all-sky imager. We thank Ana Georgina Elias from the Universidad Nacional de Tucuman and Professor Kenneth Janes, Boston University, for valuable discussion and suggestions.

## REFERENCES

- Aceituno, J., Sanchez, S. F., Aceituno, F. J., Galadi-Enriquez, D., Negro, J. J., Soriguer, R. C., & Sanchez, G. 2011, *PASP*, 123, 907
- Baumgardner, J., Flynn, B., & Mendillo, M. 1993, *Opt. Eng.*, 32, 3028
- Erasmus, D., & van Rooyen, R. 2006, Final Report to ESO 2006, (73526/TSD/04/6179/GW/LET; Garching: ESO)
- Hoffleit, D., & Warren, W. H., Jr 1991, *The Bright Star Catalogue*, 5th Revised Ed. (Strasbourg: Astronomical Data Center, NSSDC/ADC) <http://cdsarc.u-strasbg.fr/viz-bin/Cat?V/50>
- Kendall, M., & Gibbons, J. D. 1990, *Rank Correlation Methods* (New York: Oxford University Press)
- Kurlandczyk, H., & Sarazin, M. 2007, *Proc. SPIE*, 6745, 674507
- Martinis, C., Baumgardner, J., Smith, S. M., Colerico, M., & Mendillo, M. 2006, *Ann. Geophys.*, 24, 1375
- Matta, M., Smith, S., Baumgardner, J., Wilson, J., Martinis, C., & Mendillo, M. 2009, *Icarus*, 204, 409
- Moles, M., Sánchez, S. F., Lamadrid, J. L., Cenarro, A. J., Cristóbal-Hornillos, D., Maicas, N., & Aceituno, J. 2010, *PASP*, 122, 363
- Patat, F., Moehler, S., O'Brien, K., Pompei, E., Bensby, T., Carraro, G., de Ugarte Postigo, A., Fox, A., et al. 2011, *A&A*, 527, A 91
- Sánchez, S. F., Aceituno, J., Thiele, U., Pérez-Ramírez, D., & Alves, J. 2007, *PASP*, 119, 1186
- Sarazin, M. 2000. The ASM Line of Sight Sky Absorption Monitor (Garching: ESO) <http://www.eso.org/gen-fac/pubs/astclim/lasilla/asm/lossam/>
- Shamir, L., & Nemiroff, R. J. 2005, *PASP*, 117, 972
- Skidmore, W., Schock, M., Magnier, E., Walker, D., Feldman, D., Riddle, R., Els, S., et al. 2008, *Proc. SPIE*, 7012, 701224-3
- Xia, X. 2012, *Ann. Geophys.*, 30, 573