

¹ Department of Geography, Federal University of Goiás, Campus of Jataí, Jataí, Goiás, Brazil

² Department of Natural Resources, School of Agronomic Science, University of State of São Paulo, Botucatu, São Paulo, Brazil

³ Department of Atmospheric Sciences, Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, São Paulo, Brazil

A new algorithm to estimate sky condition based on 5 minutes-averaged values of clearness index and relative optical air mass

H. F. Assunção¹, J. F. Escobedo², and A. P. Oliveira³

With 8 Figures

Received February 10, 2006; revised July 24, 2006; accepted October 12, 2006

Published online February 21, 2007 © Springer-Verlag 2007

Summary

This work describes a new algorithm to characterize sky condition in intervals of 5 min using four categories of sun exposition: apparent sun with cloud reflection effects; apparent sun without cloud effects; sun partially concealed by clouds; and sun totally concealed by clouds. The algorithm can also be applied to estimate hourly and daily sky condition in terms of the traditional three categories: clear, partially cloudy and cloudy day. It identifies sky conditions within a confidence interval of 95% by minimizing local climate and measurement effects. This is accomplished by using a logistic cumulative probability function to characterize clear sky and Weibull cumulative probability function to represent cloudy sky. Both probability functions are derived from frequency distributions of clearness index, based on 5 minutes-averaged values of global solar irradiance observed at the surface during a period of 6 years in Botucatu, Southeastern of Brazil. The relative sunshine estimated from the new algorithm is statistically comparable to the one derived from Campbell-Stocks sunshine recorder for both daily and monthly values. The new method indicates that the highest frequency of clear sky days occurs in Botucatu during winter (66%) and the lowest during the summer (38%). Partially cloudy condition is the dominant feature during all months of the year.

1. Introduction

The geographic variation of solar radiation and sun light availability at the surface depends on the sun position and radiometric properties of the atmosphere. The position of the sun is easily estimated from zenith and azimuth solar angles, while the actual state of the atmosphere is much more difficult to obtain. One way to assess the radiometric properties of the atmosphere, which will be explored in this work, is by the sky condition.

In general, sky conditions are classified in three categories: clear sky, partially cloudy and cloudy. Under clear sky condition, the intensity of solar irradiance at the surface is a function of the content of ozone, water vapor and aerosol in the atmosphere. For cloudy conditions, the major depletion is caused by clouds. They act according to type, number of layers and thickness, forming complex structures that reduce the solar radiation at the surface (Iqbal, 1983). Daily sky condition can be easily identified using clearness index or relative sunshine. Clearness index is the ratio of incoming

global solar irradiance at the surface to extraterrestrial solar irradiance and indicates the degree of atmospheric transparency to solar radiation. Relative sunshine is the ratio of sunshine duration at the surface to daytime period and, at practical terms, indicates the fraction of daytime covered by clouds.

It is available in the literature several works relating sky condition to different ranges of hourly and daily values of clearness index (Orgill and Hollands, 1977; Skartveit and Olseth, 1987; Erbs et al., 1982; Reindl et al., 1990; Zangvil and Lamb, 1997; Li and Lam, 2001; Calbó et al., 2001). However, there are no unique set of clearness index values that is generally accepted to define hourly or daily sky conditions in terms of clear sky, partially cloudy and cloudy sky. The main reason for this discrepancy is because the definition of sky condition in terms of clearness index values depends on local climate and the available method of solar radiation measurements.

Historically, solar irradiance as well as other meteorological parameters is reported in terms of hourly, daily and monthly values. Although hourly values are common and appropriated for general proposes, short term characterization of sky condition is important to estimate direct and diffuse solar radiations at the surface; particularly to correct effects caused by shadow-band and shadow-disk devices over pyranometers due to the anisotropy of solar radiation field (LeBaron et al., 1990; Batlles et al., 1995). It is important, also, in the evaluation of indoor light efficiency (Robledo and Soler, 2000, 2001; Littlefair, 2001) and shading effects in photoelectric control systems (Hiller et al., 2000).

Gansler et al. (1995) verified that cumulative distribution function of solar radiation quantities measured during one minute-interval (instantaneous values) differs considerably from ones measured during one hour-interval. Suehrcke and McCormick (1988) observed that frequency distribution for 5 minutes-averaged solar irradiance values is similar to frequency distribution of instantaneous values and concluded that hourly and daily values of solar irradiance do not represent appropriately the statistical properties of instantaneous values. This discrepancy is particularly important in applications where apparatus deploying solar energy respond almost instantaneously to fluctuations caused by clouds and other atmospheric effects.

The main goal of this work is to describe a new algorithm developed to identify, *objectively (independent of climate and measurement methods)*, instantaneous sky condition in intervals of 5 min using four categories of sun exposition: apparent sun with cloud reflection effects, apparent sun without cloud effects, sun partially concealed by clouds and sun totally concealed by clouds. The algorithm is based only on 5 minutes-averaged values of clearness index and relative optical air mass. Therefore, it requires measurements of global solar radiation at the surface every 5 min. This algorithm allows also classifying hourly and daily sky conditions in terms of clear sky, partially cloudy and cloudy, as traditionally carried out in most of the works in the literature. The algorithm identifies instantaneous, hourly and daily sky conditions within a confidence interval of 95%.

It will be shown that the *shape* of frequency distributions of 5 minutes-averaged values of clearness index in Botucatu, located in a countryside rural area of São Paulo State, Southeastern of Brazil, using 6 years of continuously observed values of global solar radiation, have a universal behavior for clear sky and cloudy days; consequently their statistical properties can be extend and applied to other places, as long as the parameters used to define them are evaluated according to the methodology presented here. Thus, the frequency distribution curves of clearness index obtained for Botucatu will be used to generate a set of cumulative frequency distribution curves in terms of relative optical air mass, which in turn will be used to develop the new algorithm. It will be also shown that new algorithm can be used to estimate sunshine duration and relative sunshine.

The data used in this work is described in Sect. 2. Details about the modeling of frequency distribution of clearness index are presented in Sect. 3. The new algorithm is described in Sect. 4 and applied to determine the seasonal evolution of the predominant daily sky condition and sunshine duration in the city of Botucatu in Sect. 5. The results of this research are summarized in the Sect. 6.

2. Data set

This work is based on 5 minutes-averaged values of global (G) and direct (G_n) solar irradiance, and

daily values of sunshine duration measured in the Radiometric Station at the University of the State of São Paulo, located in the rural area of Botucatu city, State of São Paulo, Southeast of Brazil (22°51' S; 48°26' W).

Botucatu, a city with 110 thousand habitants, is located in the countryside of Brazil, at 786 m above the mean sea level, and approximately 221 km far from the Atlantic Ocean. It is characterized by mild cold and dry winter (June to September) and by warm and wet summer (December to March). The averaged air temperature (relative humidity) varies from a minimum of 17.5 °C (55%) in the winter to a maximum of 22.5 °C (80%) in the summer. The minimum precipitation occurs in August (38 mm) and the maximum in January (260 mm).

Global solar irradiance was measured with a spectral precision pyranometer model PSP, manufactured by Eppley Inc, and direct solar irradiance was measured by a pyrliometer coupled to a sun tracker model ST-3, both manufactured by Eppley Inc. All irradiance values were sampled with frequency of 0.2 Hz using the data acquisition system model CR23X, manufactured by Campbell Scientific Inc. Relative sunshine (S) was obtained from sunshine duration (N) measurements carried out using a Campbell-Stocks recorder located near to the radiometers in the radiometric station of Botucatu. All solar radiation values used in this work are referred to true solar time. The pyranometer and pyrliometer were periodically checked using as reference a secondary pyranometer and an absolute pyrliometer according to WMO (1983). All radiation measurements used in this work were carried out continuously during 6 years, from 1996 to 2001.

The algorithm proposed here is based on values of solar irradiance for two distinct sky conditions: clear sky and cloudy (Suehrcke and McCormick, 1988; Jurado et al., 1995). The original data set was divided in two mutually excluding subsets. The first subset corresponds to clear sky condition, defined by days with $S \geq 0.9$ and by a symmetric diurnal evolution of 5 minutes-averaged values of global solar radiation at the surface. Days when the time evolutions in the morning and in the afternoon follow the extraterrestrial solar radiation diurnal evolution are considered symmetric (Skartveit et al., 1998). The second one corresponds to cloudy condition, defined by days with $S \leq 0.1$ and $G_n < 120 \text{ W m}^{-2}$.

This lower limit of G_n is considered as the minimum solar radiation value required to burn the heliograph record (Gueymard, 1993; Skartveit et al., 1998; Suehrcke, 2000). These two mutually excluding data sets are used to estimate the frequency distributions of 5 minutes-averaged clearness indexes for clear sky and cloudy conditions used here to develop the new algorithm.

3. Modeling frequency distribution

The algorithm is based on the fact that short-term variations existing in the frequency distributions of clearness index are exclusively due to changes in the relative optical air mass (Suehrcke and McCormick, 1988; Skartveit and Olseth, 1992; Jurado et al., 1995; Tovar et al., 1998; Assunção et al., 2003).

The clearness index is estimated, for each relative optical air mass, as indicated below:

$$k_t = \frac{G}{G_0} \quad (1)$$

where G_0 is 5 minutes-averaged value of solar irradiance at the top of the atmosphere, expressed in W m^{-2} and evaluated according to Iqbal (1983).

The relative optical air mass is estimated neglecting the Earth curvature effects and assuming a homogeneous and not refractive atmosphere so that:

$$m_a = \sec \theta_Z \cdot \exp(-0.0001184z) \quad (2)$$

where θ_Z is the sun zenith angle, expressed in degrees and estimated according to Iqbal (1983) and z is height above the surface expressed in meters.

Table 1. Number of 5 minutes-averaged values of global solar radiation for six classes of relative optical optic mass used to construct the frequency distribution of k_t for clear sky and cloudy conditions

Class	m_a		Number of k_t values	
	Range	Central value	Clear sky condition	Cloudy condition
1	$0.9 \leq m_a \leq 1.1$	1	6436	6422
2	$1.9 \leq m_a \leq 2.1$	2	1142	918
3	$2.9 \leq m_a \leq 3.1$	3	418	384
4	$3.9 \leq m_a \leq 4.1$	4	213	233
5	$4.9 \leq m_a \leq 5.1$	5	152	137
6	$5.9 \leq m_a \leq 6.1$	6	95	96

To estimate the variations of the frequency distributions of k_t in terms of m_a , each of the two mutually excluding data sets are divided in 6 subsets representing 6 classes of relative optical air mass values described in Table 1, for clear sky and cloudy conditions (Suehrcke and McCormick, 1988). Each subset, represented by a unique value of m_a , is divided in 50 intervals of clearness index and used to estimate the frequency distribution of k_t in terms of m_a (Tovar et al., 1998; Assunção et al., 2003).

3.1 Clear sky condition

Under clear sky condition the observed frequency distributions of k_t show a symmetric pattern for all relative optical air mass classes (Fig. 1a). According to Tovar et al. (1998), this type of

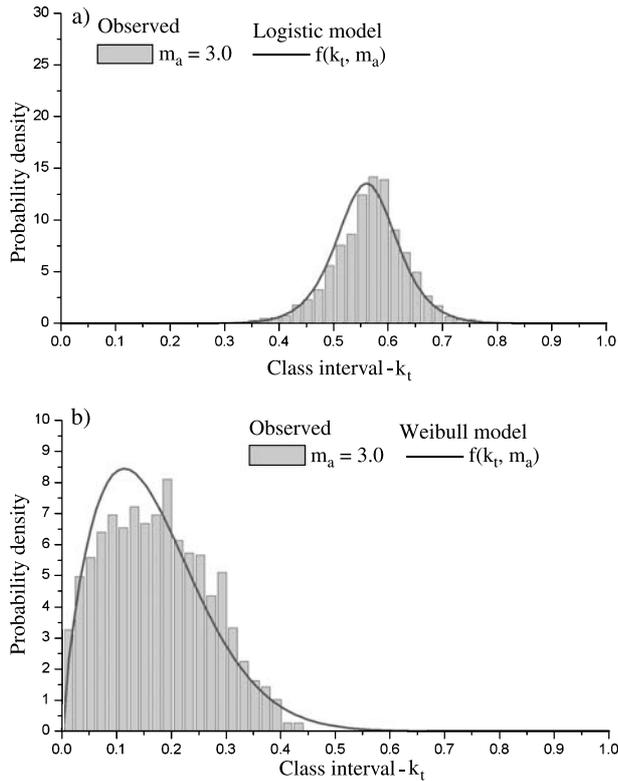


Fig. 1. Frequency distribution of clearness index for a) clear sky and b) cloudy conditions observed in Botucatu from 1996 to 2001. The continuous line in a) corresponds to fitted logistic probability density function (Appendix A) for $\mu = 0.16175 + 0.71628 \exp(-m_a/5.74442)$ and $\delta = 0.04736 - 0.04128 \exp(-m_a/3.82816)$. The continuous line in b) corresponds to a fitted Weibull probability density function (Appendix B) for $\alpha = 1.30649 + 0.74763 \exp(-m_a/2.58206)$ and $\delta = 0.09919 + 0.14525 \exp(-m_a/3.04931)$

frequency distribution is satisfactorily adjusted by the following family of logistic curves:

$$f(k_t, m_a) = \frac{\exp\{-[k_t - \mu(m_a)]/\delta(m_a)\}}{\delta(m_a)\{1 + \exp\{-[k_t - \mu(m_a)]/\delta(m_a)\}\}^2} \quad (3)$$

where $f(k_t, m_a)$ is the probability density function of k_t for a given relative optical air mass.

The corresponding cumulative distribution function is expressed as:

$$F(k_t, m_a) = \frac{1}{1 + \exp\{-[k_t - \mu(m_a)]/\delta(m_a)\}} \quad (4)$$

Conceptually, $F(k_t, m_a)$ is the probability of finding values of clearness index equal or smaller than k_t for a given relative optical air mass. The parameters $\mu(m_a)$ and $\delta(m_a)$ are respectively, the center and the size of the modeled frequency distributions of k_t . Table 2 shows values of these parameters considering the mean (*avg*) and standard deviation (*sd*) of observed distribution of k_t as indicated in the Appendix A. Values of R^2 very close to 1 indicated that logistic curve (3) adjusted very well the observed distribution of k_t for all relative optical air mass classes.

Table 2. Clearness index parameters as a function of relative optical air mass, estimated from 5 minutes-averaged values of solar radiation for clear sky conditions

m_a	μ	δ	<i>avg</i>	<i>sd</i>	R^2
1	0.76569	0.01588	0.76569	0.02880	0.99977
2	0.66632	0.02215	0.66632	0.04018	0.99990
3	0.57842	0.02867	0.57842	0.05200	0.99992
4	0.52498	0.03309	0.52498	0.06002	0.99956
5	0.46838	0.03660	0.46838	0.06639	0.99878
6	0.40815	0.03832	0.40815	0.06950	0.99964

3.2 Cloudy condition

Under cloudy condition the observed frequency distributions of k_t are asymmetrical (Fig. 1b), as consequence they are properly adjusted by Weibull probability density function, described by:

$$f(k_t, m_a) = \alpha(m_a)\beta(m_a)^{-\alpha(m_a)}k_t^{\alpha(m_a)-1} \exp\{-[k_t/\beta(m_a)]^{\alpha(m_a)}\} \quad (5)$$

and the corresponding cumulative distribution function is expressed as:

$$F(k_t, m_a) = 1 - \exp\{-[k_t/\beta(m_a)]^{\alpha(m_a)}\} \quad (6)$$

where the parameters $\alpha(m_a)$ and $\beta(m_a)$ are estimated from the observed mean (*avg*) and standard deviation (*sd*) values of k_t distributions (Table 3) as indicated in Appendix B.

Table 3. Observed probability distribution parameters for clearness index (k_t) as a function of relative optical air mass (m_a). Cloudy days

m_a	α	β	<i>avg</i>	<i>sd</i>	R^2
1	1.80633	0.20551	0.18272	0.10471	0.99870
2	1.67323	0.16910	0.15105	0.09279	0.99870
3	1.53682	0.15884	0.14299	0.09494	0.99843
4	1.45340	0.13781	0.12491	0.08732	0.99924
5	1.42314	0.13045	0.11860	0.08454	0.99402
6	1.38736	0.11881	0.10844	0.07915	0.99258

3.3 Parameterization of $\mu(m_a)$, $\delta(m_a)$, $\alpha(m_a)$ and $\beta(m_a)$

The derivation of $f(k_t, m_a)$, in the previous Sects. 3.1 and 3.2, has indicated that the magnitude of all relevant parameters varies continuously as a function of the relative optical air mass. An analysis of how the discrete variation $\mu(m_a)$, $\delta(m_a)$, $\alpha(m_a)$ and $\beta(m_a)$ alters each one of the frequency distribution curves and, consequently, the amplitude of k_t (Tables 2 and 3) has indicated that they can be parameterized in terms of the relative optical air mass according to the following generic expression:

$$g(m_a) = a + b \exp(-m_a/c), \text{ for } 0.9 \leq m_a \leq 6.1 \tag{7}$$

where the coefficients a , b and c are indicated in Table 4. The R^2 very close to 1 indicated that these coefficients are very well estimated by non-

Table 4. Coefficients of adjusting parameters μ , δ , α and β as a function of relative optical air mass (Eq. (7))

$g(m_a)$	a	b	c	R^2	Probability density function
$\mu(m_a)$	0.16175	0.71628	5.74442	0.99781	Logistic
$\delta(m_a)$	0.04736	-0.04128	3.82816	0.99718	
$\alpha(m_a)$	1.30649	0.74763	2.58206	0.99257	Weibull
$\beta(m_a)$	0.09919	0.14525	3.04931	0.99254	

linear fitting of (7) through the observed discrete values of $\mu(m_a)$, $\delta(m_a)$, $\alpha(m_a)$ and $\beta(m_a)$ indicated in Tables 2 and 3.

Figure 3 indicate a set of probability functions of k_t for clear sky and cloudy conditions obtained using (7).

4. Algorithm

According to Iqbal (1983) the amount of solar irradiance that reaches the surface of the Earth depends on the concentration of ozone, water vapor, aerosol and clouds. They are the major atmospheric components responsible by the atmospheric transmittance. Due to several causes, the concentrations of these components are affected by short-term random variations, causing frequent changes in the state of the atmosphere. However, a continuous monitoring of solar irradiance for a long period of time will certainly generate a data set that contains statistical information about a significant large range of atmospheric transmittance values. Therefore, by knowing the models that describe the amplitude of solar irradiance based on frequency distribution functions derived from long-term observation, it is possible to identify the sky condition. To eliminate latitude effects the sky condition is evaluated in terms of the clearness index (Li and Lam, 2001; Calbó et al., 2001).

Although instantaneous values of clearness index portray quite well sky condition, it is not technically practical to classify the sky conditions in terms of traditional way (clear sky, partially cloudy and cloudy) taking into consideration only one measurement of global solar irradiance at the surface without a simultaneous visual inspection of sky. On the other hand, working with a series of global solar irradiance measurements (5 minutes-averaged for instance) taken during the day and considering the sky condition associated to each measurement in terms of the degree of sun exposition (or sun concealing) the traditional classification becomes feasible.

Thus, based on these two considerations the new algorithm will be developed to classify instantaneous sky condition in intervals of 5 minutes using four categories of sun exposition (Table 5, Fig. 2).

Table 5. Characterization of instantaneous sky condition (sky_{ij})

Category j	Instantaneous sky condition (sky_{ij})
1	Sun apparent with cloud reflection effects
2	Sun apparent without cloud effects
3	Sun partially concealed by clouds
4	Sun totally concealed by clouds

4.1 Boundary values

According to Jurado et al. (1995), exist, for each of these four categories, very well defined boundary values of clearness indexes that can be justified mathematically, by considering the clearness index as a combination of several factors such as relative optical air mass and the content of ozone, water vapor and aerosol.

Thus, it is possible to estimate boundary values of k_t and to set ranges of sky conditions with confidence interval $CI = 0.950$. This confidence interval is distributed symmetrically around the mean value of k_t so that: $P(k_t \leq k_{tmin}^s) = 1 - CI/2 = 0.025$ and $P(k_t \leq k_{tmax}^s) = 1 - (1 - CI/2) = 0.975$. Here, k_{tmin}^s and k_{tmax}^s are, respectively, the minimum and maximum expected values of clearness index for a specific relative optical air mass under clear sky condition.

Inverting (4) and replacing $F(k_t, m_a)$ by $P(k_t \leq k_{tmin}^s)$ for the lower limit, and substituting $F(k_t, m_a)$ by $P(k_t \leq k_{tmax}^s)$ for the upper limit, one obtains the following interval of k_t values for clear sky condition:

$$k_{tmin}^s = \mu(m_a) - \delta(m_a) \ln \left[\frac{1}{0.025} - 1 \right] \quad (8)$$

$$k_{tmax}^s = \mu(m_a) - \delta(m_a) \ln \left[\frac{1}{0.975} - 1 \right] \quad (9)$$

Similarly, inverting (6), assuming $P(k_t \leq k_{tmin}^c) = 0$, $P(k_t \leq k_{tmax}^c) = CI$ and substituting $F(k_t, m_a)$ by CI value, one obtains for cloudy sky condition:

$$k_{tmax}^c = \beta(m_a) \ln \left(\frac{1}{0.05} \right)^{\frac{1}{\alpha(m_a)}} \quad (10)$$

where $P(k_t \leq k_{tmin}^s)$, $P(k_t \leq k_{tmax}^s)$, $P(k_t \leq k_{tmin}^c)$ and $P(k_t \leq k_{tmax}^c)$ are the probabilities associated to the confidence intervals. The superscripts c and s indicate, respectively, presence (cloudy condition) and absence (clear sky condition) of clouds between the sun and the pyranometer.

The meaning of these boundary values are illustrated in Fig. 2. There it is possible identify the relation between the actual appearance of the sky, sky condition and boundary values for four categories of sun exposition by using photography.

4.2 Instantaneous sky condition

Taking into consideration the level of solar exposition, the instantaneous sky condition can be identified in terms of the clearness index value. Since daytime variation of relative optical air mass depends on latitude and time of the year, the way the relative optical air mass varies during the day affects the distribution of global solar radiation and clearness index.

Thus, the classification of instantaneous sky condition to: clear sky, partially cloudy and cloudy (Table 6) is given by:

$$SKY = \begin{bmatrix} sky_{11} & sky_{12} & sky_{13} & sky_{14} \\ sky_{21} & sky_{22} & sky_{23} & sky_{24} \\ \vdots & \vdots & \vdots & \vdots \\ sky_{n1} & sky_{n2} & sky_{n3} & sky_{n4} \end{bmatrix} \quad (11)$$

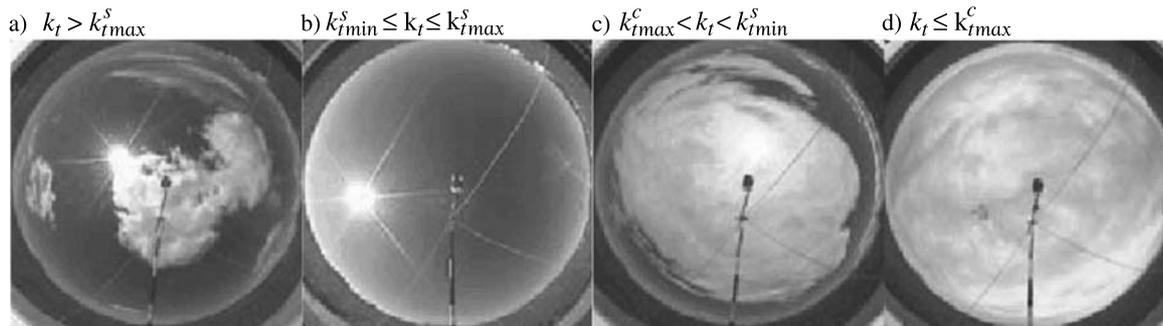


Fig. 2. Photographic examples of: **a)** Sun apparent with cloud reflection; **b)** Sun apparent without cloud; **c)** Sun partially concealed by clouds; **d)** Sun totally concealed by clouds, in Botucatu

The matrix SKY has n lines and four columns, corresponding respectively to the total number of clearness indexes observations during a given day and total number of categories of sun exposition considered (Table 5).

Each element of SKY obeys the following rules:

$$\begin{aligned}
 sky_{i1} &= \begin{cases} 1 & \text{if } k_t > k_{tmax}^s \\ 0 & \text{otherwise} \end{cases} \\
 sky_{i2} &= \begin{cases} 1 & \text{if } k_{tmin}^s \leq k_t \leq k_{tmax}^s \\ 0 & \text{otherwise} \end{cases} \\
 sky_{i3} &= \begin{cases} 1 & \text{if } k_{tmax}^c < k_t < k_{tmin}^s \\ 0 & \text{otherwise} \end{cases} \\
 sky_{i4} &= \begin{cases} 1 & \text{if } k_t \leq k_{tmax}^c \\ 0 & \text{otherwise} \end{cases}
 \end{aligned} \tag{12}$$

The instantaneous sky condition is obtained by the category of nonzero element of each line of SKY .

Table 6. Classification of sky condition based on effective relative sunshine (f_j^d)

Sky condition	Effective relative sunshine (f_j^d)
Clear sky	$f_2^d \geq 0.9$
Partially cloudy	$f_2^d < 0.9$ and $f_4^d < 0.9$
Cloudy	$f_4^d \geq 0.9$

4.3 Effective relative sunshine and daily sky condition

The matrix SKY can also be used to estimate the effective daytime duration (D_L). In the case reported here D_L is approximated by:

$$D_L = \frac{5}{60} \cdot \sum_{j=1}^4 \sum_{i=1}^n sky_{ij} \tag{13}$$

Here, because k_t is available in 5 minutes-averaged values, the ratio 5/60 transforms minute into hour.

Because the algorithm works only for relative optical air mass <6.1 , the values of D_L do not represent the exact daytime duration, but an approximated value of N_0 .

The effective relative sunshine for category j of sun exposition is evaluated as:

$$f_j^d = \frac{1}{D_L} \sum_{i=1}^n sky_{ij} \tag{14}$$

Thus, the classification of sky condition in any specific day is given by the effective relative sunshine of the dominant category according to Table 6.

Iqbal (1983) considered the possibility of small cloud to maintain the sun concealed as it moves through the sky slowly, or the possibility of small among clouds openings to maintain the sun shining during long periods of time. Even considering these possibilities, it seems very unlikely for the sun to remain exposed (or concealed) more than 90% of the time without the day being considered totally clear (or totally covered by clouds).

According to Li and Lam (2001) there are two types of cloudiness, one due to shallow and bright clouds (Fig. 2c) and another caused by thick and dark clouds (Fig. 2d). It is plausible to assume that shallow clouds transmit enough direct solar radiation to burn the sunshine recorder card of heliograph.

Hence, relative sunshine can be expressed considering the prevailing irradiance values in terms of categories 1, 2 and 3:

$$S = \sum_{j=1}^3 f_j^d \tag{15}$$

Adding up the fraction of time that the solar disk is not totally covered by clouds has an advantage over direct measurements carried out with heliographs, because it does not depend on the observer abilities and wetness of the sunshine record card.

5. Application of the new algorithm for Botucatu

The algorithm developed above was applied to the entire data set to identify instantaneous sky condition, to classify the predominant sky condition during the entire daytime (daily sky condition) and to estimate the relative sunshine for Botucatu.

To facilitate the classification of the sky condition, all boundary values will be expressed in terms of the maximum (G_{max}^s) and minimum (G_{min}^s) expected values of global solar irradiance for apparent sun and maximum (G_{max}^c) expected value of global solar irradiance for concealed sun, given in terms of the extraterrestrial solar radiation by:

$$G_{max}^s = k_{tmax}^s G_0 \tag{16}$$

$$G_{\min}^s = k_{\min}^s G_0 \quad (17)$$

$$G_{\max}^c = k_{\max}^c G_0 \quad (18)$$

5.1 Sun apparent with cloud reflection effects (sky_{i1})

In this category the global solar irradiance is intensified by multiple reflections caused by clouds or by ground and clouds (Fig. 2a). These reflections occur when the sky is partially cloudy and can generate clearness index values greater than 1.

Figure 4a and b show four instantaneous sky conditions found in two different days: May 28, 1996 and December 5, 2000. They exemplified events with anomalous high values of global solar irradiance, well above the boundary values for clear sky (G_{\max}^s). According to Suehrcke and McCormick (1988) and Orgill and Hollands (1977) in all these cases clouds reflect solar radiation towards the pyranometer, increasing diffuse radiation without affecting the direct component of solar radiation at the surface. In the case of Botucatu the frequency distribution of this category is, in average, smaller than 2% (Table 7).

Suehrcke and McCormick (1988) pointed out that the cloud reflection do not affect the daily values of solar radiation because they are short time phenomena that are compensated by the attenuation of solar radiation by clouds during the period of one day.

Table 7. Seasonal distribution of fractions of sun exposition in the sky based on apparent sun with cloud reflection effects; apparent sun without cloud effects; sun partially concealed by clouds; sun totally concealed by clouds

Instantaneous sky conditions	Frequency (%)			
	Spring	Summer	Autumn	Winter
Sun apparent with cloud reflection effects	1.2	1.6	1.2	0.8
Sun apparent without cloud effects	45.5	37.9	63.4	66.2
Sun partially concealed by clouds	9.5	9.4	5.9	6.9
Sun totally concealed by clouds	43.8	51.1	29.5	26.1

5.2 Sun apparent without cloud effects (sky_{i2})

Clearness index in this category is controlled by direct solar irradiance in which the transmittance is regulated mainly by atmospheric turbidity and relative optical air mass (Fig. 2b). It can be verified in Fig. 3 that the k_t distribution curves for clear sky undergo a displacement as the relative optical air mass increases. This behavior confirms the strong association between k_t and relative optical air mass, already verified by Suehrcke and McCormick (1988), Tovar et al. (1998) and Assunção et al. (2003).

Here, the diurnal evolution of global solar irradiance at the surface is characterized by a smooth curve (Fig. 4d). Almost undetectable oscillations in global solar irradiance are due to variation in the atmospheric transmittance. In some cases global solar irradiance is affected by high and shallow clouds like *cirrus* (Fig. 4c). Thus, even in sunny days these types of clouds are common and difficult to be visualized. In average, during summer in Botucatu the sun remains uncovered in 38% of time, while in the winter it can reach about 66% (Table 7).

5.3 Sun partially concealed by clouds (sky_{i3})

This category represents the transition between sun apparent and concealed, in which the radiation field is transmitted diffusely throughout the clouds (Fig. 2c) but the direct solar irradiance remains above 120 W m^{-2} .

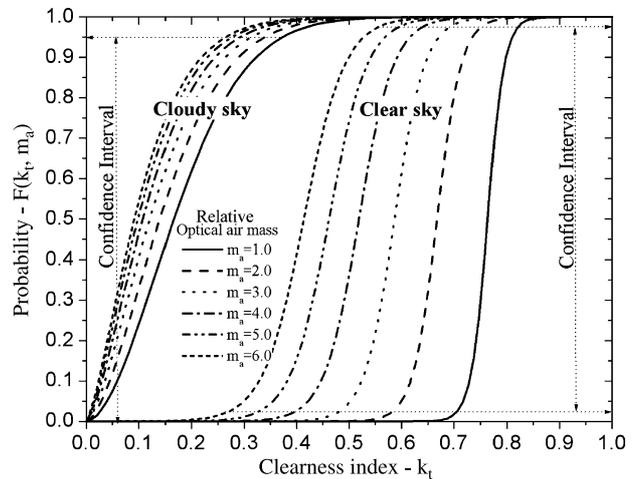


Fig. 3. Cumulative probability function of k_t for clear and cloudy sky conditions defined by an interval of confidence of 95%

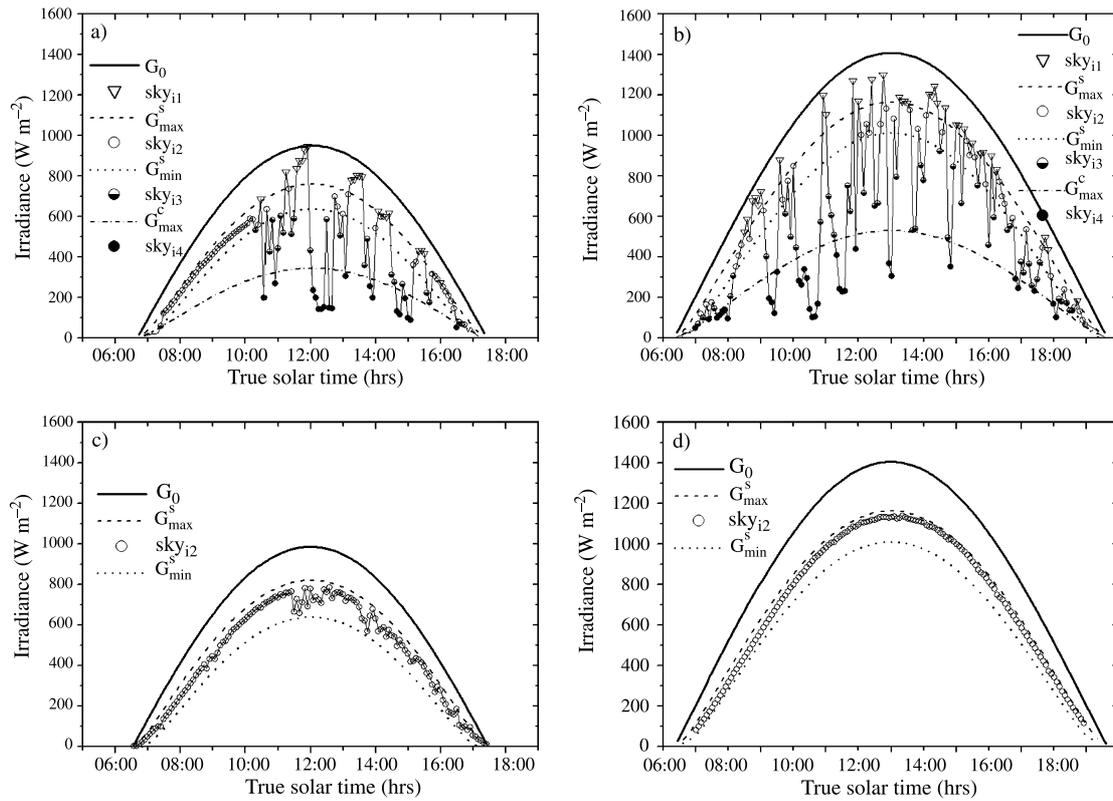


Fig. 4. Diurnal evolution of extraterrestrial solar irradiance (solid line), measured global solar irradiance (symbols) and expected global solar irradiance (dash line, dot line, dash-dot line) at the surface during (a) May 28, 1996; (b) December 5, 2000; (c) July 29, 2001 and (d) December 2, 2001. Symbols indicate 5 minutes-averaged sky condition

According to Suehrcke and McCormick (1988) this phenomena is characteristic of fast variations of solar radiation in the boundaries of thick clouds. Figure 5a shows a typical day representing this category, when the variability in the transmittance caused by clouds is clearly smoothed out by the process of averaging the data with interval of 5 min.

Figure 5b shows a day with sky apparently clear, however the observed values of global irradiance are below the critical level expected for the category of apparent sun, and well above the category of concealed sun. Although Suehrcke and McCormick (1988) claimed that this type of phenomenon is due to uniform and thin cloud layers such as *cirrus-stratus*, we believed that the excess of attenuation observed in this day is due to the ozone, water vapor and clear sky turbidity caused by biomass burning at the end of the sugar-cane harvest season, a widespread activity in Southeast of Brazil in September. According to Skartveit and Olseth (1992) the attenuations in the direct solar irradiance due to aerosol and

shallow clouds are counterbalanced by an increase of the diffuse solar irradiance, reducing the effects of aerosol and clouds on the global solar irradiance.

Table 7 indicates that this category occurs in less than 10% of the cases in Botucatu. This result agrees with the one obtained by Suehrcke (2000) for instantaneous values of clearness index between 0.48 and 0.73, which were attributed to the thin clouds.

5.4 Sun totally concealed by clouds (sky_{i4})

In this category prevails diffuse component of the global solar irradiance, with very little or totally absent direct component (Fig. 2d). A visual inspection of Fig. 3 indicates that the probability of a particular value of k_t to occur increases with the relative optical air mass for cloudy condition. According to Tovar et al. (1998) this behavior may be associated to the fact that the area covered by clouds at the surface increases with the relative optical air mass. Moreover, cloud thick-

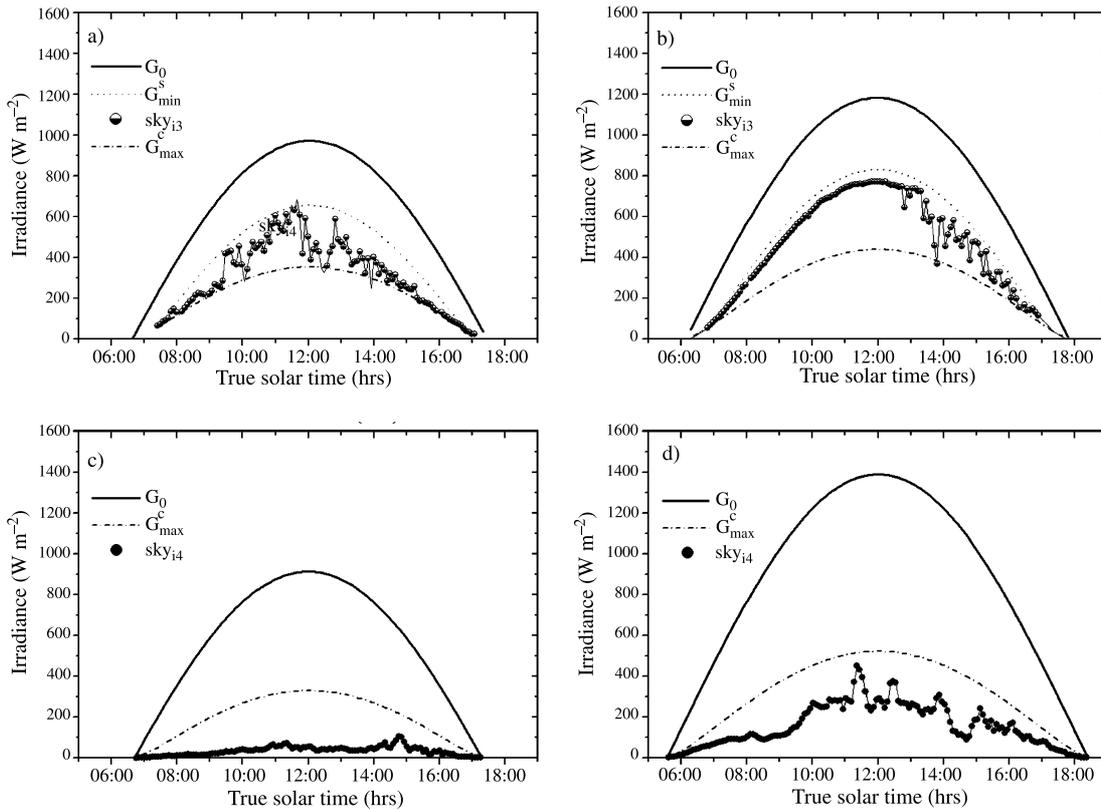


Fig. 5. Diurnal evolution of extraterrestrial solar irradiance (solid line), measured global solar irradiance (symbols) and expected global solar irradiance (dash line, dash-dot line) at the surface during: (a) May 21, 1995; (b) September 9, 1997; (c) June 20, 1999 and (d) February 10, 2001. Symbols indicate 5 minutes-averaged sky condition

ness and horizontal extension do also increase with relative optical air mass as the clouds become less transparent at large values of relative optical air mass. Therefore, the blocking effect caused by clouds on the sun is stronger for large values of relative optical air mass.

Figure 5c and d show that the distribution of global solar irradiance, during a cloudy day, is not as symmetric as suggested by the maximum expected value of global solar irradiance (G_{max}^c). Besides, clearness index depends more on the physical properties of clouds than on the relative optical air mass. The smaller spreading of frequency distribution of k_t curves for cloudy conditions confirms this conjecture (Fig. 3). According the Iqbal (1983), when the sky is entirely covered by clouds the diffuse solar irradiance at the surface becomes isotropic and the clearness index shows a very weak dependence on the relative optical air mass. Table 7 shows that instantaneous cloudy condition is more frequent during summer (51%) and less frequent in the winter (26%) in Botucatu.

5.5 Relative sunshine

The frequency of clear sky, partially cloudy and cloudy days by month in Botucatu are displayed in Fig. 6. These frequencies were estimated using (14) and the definitions in Table 6. The number of clear days shows a maximum in the winter

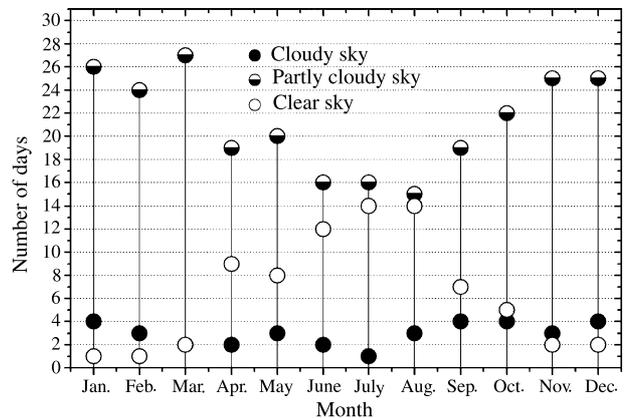


Fig. 6. Frequency distribution of average number of days with clear, partially cloudy and cloudy conditions in Botucatu

(Southern Hemisphere) and partially cloudy days are the prevailing condition during all months of the year. Considering the entire year, the percentage of clear, partially clear and cloudy days are 21, 69 and 10%, respectively.

Figure 7 shows the linear correlation between relative sunshine from a Campbell-Stocks recorder and estimated by the new algorithm. The determination coefficient equal to 0.9564 indicates a good matching between these two methods. Statistical analyses indicate that the new algorithm overestimates the relative sunshine obtained by the Campbell-Stocks recorder in about 6.4%, with a *RMSE* of 18.2%.

Benson et al. (1984) compared relative sunshine measured with Campbell-Stocks recorder and pyrhelimeter, found discrepancies of the order of $\pm 15\%$ for hourly values, $\pm 8\%$ for daily values and $\pm 4\%$ for monthly values. They also found that Campbell-Stocks recorder underestimate relative sunshine for high values of relative optical air mass and underestimate for small ones during periods of intense brightness. Besides, they concluded that the measurements taking with Campbell-Stocks recorder are affected by errors related to the atmospheric moisture content and misreading of the record.

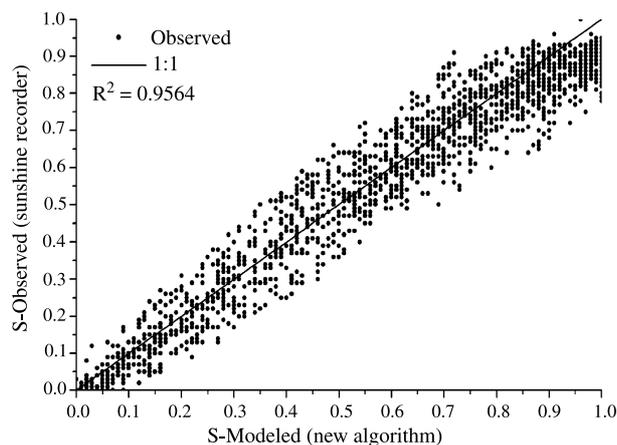


Fig. 7. Dispersion diagram of daily values of relative sunshine obtained from new method and using Campbell-Stocks sunshine recorder for Botucatu

Figure 8 shows the seasonal evolution of monthly-averaged values of daily relative sunshine estimated from Campbell-Stokes recorder and using the new algorithm. The new algorithm overestimated measured relative sunshine in about 4.3%, with *RMSE* of 5.7% for monthly

time scale. The good agreement between new method and observed values of relative sunshine with Campbell Stocks is present in all months of the year and corroborates the observations carried out by Benson et al. (1984). Udo (2000) did also find a high correlation between monthly values of clearness index and relative sunshine. Comparatively to the results obtained by others researchers, the relative sunshine duration obtained by the new algorithm show a slightly larger discrepancy, probably due to differences in the criteria of burning threshold. For instance, Benson et al. (1984) and Li and Lam (2001) considered 210 W m^{-2} as burning threshold, while Skartveit et al. (1998) and Suehrcke (2000) adopted the WMO standard value of 120 W m^{-2} .

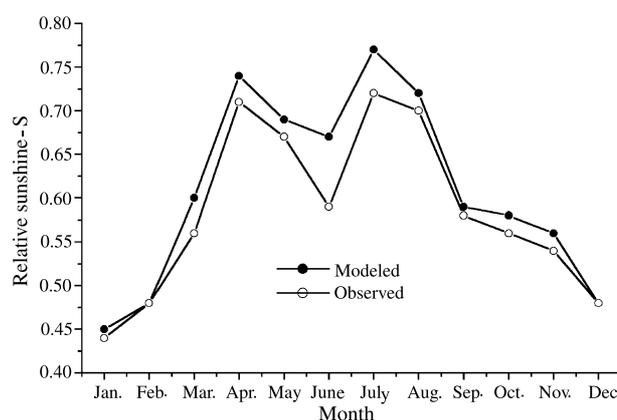


Fig. 8. Seasonal evolution of monthly values of relative sunshine obtained from new method and using Campbell-Stocks sunshine recorder for Botucatu

6. Conclusion

A new algorithm to characterize sky condition in intervals of 5 min using four categories of sun exposition: apparent sun with cloud reflection effects; apparent sun without cloud effects; sun partially concealed by clouds; and sun totally concealed by clouds, objectively, within an interval of confidence of 95%, was developed and tested using solar radiation data measured continuously between 1996 and 2001 in Botucatu, Southeast of Brazil.

This algorithm is based on the observational fact that short-term variation of frequency distribution of clearness index is due to relative optical air mass variations. The algorithm is built on observed frequency distributions of 5 minutes-averaged values of clearness index, block-averaged

in terms of relative optical air mass, that are used to derive a set of logistic cumulative probability functions for clear sky conditions and Weibull cumulative probability function for cloudy sky conditions.

The algorithm can be applied to estimate hourly and daily sky condition in terms of the traditional three categories: clear, partially cloudy and cloudy day, minimizing local climate and measurement methods effects. The new algorithm can also be used to estimate relative sunshine.

The application of the new algorithm to Botucatu indicated that, in daily basis, clear sky condition occurs with smallest frequency in January and February and largest in July and August. Partially cloudy condition occurs with relatively constant frequency all year long. In Botucatu, clear sky conditions (apparent sun categories) are more frequent during winter and less in summer.

Correlation between daily values of relative sunshine evaluated by the new algorithm and using a Campbell-Stocks recorder show a determination coefficient equal to 0.9564, indicating a very good agreement. Comparatively, the statistical errors for daily values of relative sunshine are 6.4% for daily values and 4.3% for monthly values, indicating that new algorithm overestimate the values of relative sunshine obtained with Campbell-Stocks record.

Although several researchers have used different methods to characterize the sky conditions for different places with promising results, the new algorithm proposed here seems to be more general, because it does not depend on climate conditions and on the method of solar radiation measurements. To be applied to other locations is necessary to consider the following points: (a) data set from radiometric stations should be based on 5 minutes-averaged values; (b) the coefficients a , b and c of (7) have to be generated for the new location considering time series of global solar radiation measurements carried out during at least 5 years; (c) the new algorithm should be compared with other methods (for instance, Zangvil and Lamb, 1997; Skartveit and Olseth, 1992; Calbó et al., 2001) to evaluate its performance. It is expected that the application of this new method to other places will help to homogenize the information about sky condition, facilitating comparison at regional and larger spatial scales.

List of symbols

<i>Symbols</i>	<i>Description</i>
a, b, c	Coefficients
Avg, \bar{x}	Mean value of the observed frequency distribution
CI	Confidence interval
D_L	Effective sunshine duration time
f_j^d	Effective relative sunshine for category j
$\delta(m_a), \delta$	Scale parameter for logistic distribution
f_i	Observed frequency of k_t in the i th class interval
x_i	i th class interval of k_t
$f(k_t, m_a), f(x)$	Probability density function
$F(k_t, m_a), F(x)$	Cumulative distribution function
$g(m_a)$	Generic function representation
G	Global solar irradiance at the surface ($W m^{-2}$)
G_{max}^s	Maximum global solar irradiance expected for clear sky condition ($W m^{-2}$) – boundary value
G_{min}^s	Minimum global solar irradiance expected for clear sky condition ($W m^{-2}$) – boundary value
G_{max}^c	Maximum global solar irradiance expected for cloudy condition ($W m^{-2}$) – boundary value
G_n	Direct solar irradiance at the surface ($W m^{-2}$)
G_0	Extraterrestrial solar irradiance ($W m^{-2}$)
k_t, x	Clearness index
k_{tmax}^c	Maximum clearness index value for cloudy condition – boundary value
k_{tmax}^s	Maximum clearness index value for clear sky condition – boundary value
k_{tmin}^s	Minimum clearness index value for clear sky condition – boundary value
m_a	Relative optical air mass
$\mu(m_a), \mu$	Position parameter of the logistic distribution of probability
N	Sunshine duration (hour)
n	Number of events between sun rise and sun set
N_0	Daytime duration (hour)
$P(k_t \leq k_{tmin}^c)$	Probability to find values of k_t smaller than and equal to k_{tmax}^c
R	Linear correlation coefficient
R^2	Determination coefficient
$RMSE$	Root Means Square Error
S	Relative sunshine
SKY	Matrix of daily mean sky condition
sky_{ij}	Element of SKY
sd, s	Standard deviation of observed frequency distribution of
z	Height above the surface (m)
$\alpha(m_a), \alpha$	Shape parameter for the Weibull distribution of probability
$\beta(m_a), \beta$	Scale parameter for the Weibull distribution of probability
θ_z	Sun zenith angle (degree)

γ	Generic variable of gamma function
π	3.1415
$\Gamma(\gamma)$	Gamma function

Acknowledgments

The authors thank to FAPESP, CAPES and CNPq for the financial support for this research, and to *Alexandre Dal Pai* and *Flávio Rodrigues Soares* for the technical assistance.

Appendix A

The frequency distribution of clearness index for clear sky condition (Fig. 1a) can be fitted by a logistic probability density function:

$$f(x) = \frac{e^{-(x-\mu)/\delta}}{\delta[1 + e^{-(x-\mu)/\delta}]^2} \quad (A1)$$

The associated probability function is given by:

$$F(x) = \frac{1}{1 + e^{-(x-\mu)/\delta}} \quad (A2)$$

and its inverse by:

$$x = \mu - \delta \ln \left[\frac{1}{F(x)} - 1 \right] \quad (A3)$$

The logistic probability density function yields a symmetric curve around the average value, μ , where the frequency $f(x)$ is a maximum, calculated as it follows:

$$\mu = \bar{x} = \frac{\sum_{i=1}^n f_i x_i}{\sum_{i=1}^n f_i} \quad (A4)$$

where f_i is the observed frequency of k_t in the i th class interval x_i .

The parameter δ is the scale parameter of the logistic distribution. It establishes the variation rate of the logistic probability density function and can be calculated by:

$$\delta = \frac{s}{\pi} \sqrt{3} \quad (A5)$$

where s is the standard deviation of the observed frequency distribution, evaluated as:

$$s = \sqrt{\frac{\sum_{i=1}^n f_i x_i^2 - \bar{x}^2 \sum_{i=1}^n f_i}{\sum_{i=1}^n f_i - 1}} \quad (A6)$$

where \bar{x} is the average clearness index considering the entire observed frequency distribution.

Appendix B

The frequency distribution of clearness index for cloudy conditions (Fig. 1b) can be adjusted by a Weibull probability density function:

$$f(x) = \alpha \beta^{-\alpha} x^{\alpha-1} e^{-(x/\beta)^\alpha} \quad x \in [0, \infty) \quad (B1)$$

Expression (B1) can be analytically integrated, so that the associated probability function is given by:

$$F(x) = 1 - e^{-(x/\beta)^\alpha} \quad (B2)$$

and its inverse by:

$$x = \beta \ln \left(\frac{1}{1 - F(x)} \right)^{\frac{1}{\alpha}} \quad (B3)$$

The parameters α and β indicate, respectively the shape and scale of the frequency distribution of clearness index. They can be estimated by an iterative numerical method based on mean (\bar{x}) and standard deviation (s) values, both derived from the observed frequency distribution of clearness index and by considering the following relations:

$$\bar{x} = \beta \Gamma(1 + \alpha^{-1}) \quad (B4)$$

and

$$s = \beta \sqrt{\Gamma(1 + 2\alpha^{-1}) - \Gamma^2(1 + \alpha^{-1})} \quad (B5)$$

where $\Gamma(\gamma)$ is a gamma function, calculated by:

$$\Gamma(\gamma) = \sqrt{\frac{2\pi}{\gamma}} e^{\gamma} \left[\ln(\gamma) - \left(1 - \frac{1}{12\gamma^2} + \frac{1}{360\gamma^4} - \frac{1}{1260\gamma^6} \right) \right] \quad (B6)$$

where γ is a generic variable and π is set equal to 3.1415.

The values of \bar{x} and s are estimated according to A4 and A6 in the Appendix A.

References

Assunção HF, Escobedo JF, Oliveira AP (2003) Modelling frequency distributions of 5 minute-averaged solar radiation indexes using Beta probability functions. *Theor Appl Climatol* 75(3–4): 213–224

Battles FJ, Olmo FJ, Alados-Arboledas L (1995) On shadowband correction methods for diffuse irradiance measurements. *Sol Energy* 54: 105–114

Benson RB, Paris MV, Sherry JE, Justus CG (1984) Estimation of daily and monthly direct, diffuse and global solar radiation measurements. *Sol Energy* 32(4): 523–535

Calbó J, González JA, Pagés D (2001) A method for sky-condition classification from ground-based solar radiation measurements. *J Appl Meteor* 40: 2193–2199

Erbs DG, Klein SA, Duffie JA (1982) Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation. *Sol Energy* 28(4): 293–302

Gansler RA, Klein SA, Beckman WA (1995) Investigation of minute solar radiation data. *Sol Energy* 55: 21–27

Gueymard C (1993) Analysis of monthly average solar radiation and bright sunshine for different thresholds at Cape Canaveral, Florida. *Sol Energy* 51: 139–145

Hiller MDE, Beckman WA, Mitchell JW (2000) TRNSHD – a program for shading and insolation calculations. *Build Environ* 35: 633–644

Iqbal M (1983) An introduction to solar radiation. Toronto: Academic Press, 390 pp

Jurado M, Caridad JM, Ruiz V (1995) Statistical distribution of the clearness index with radiation data integrated over five minute intervals. *Sol Energy* 55: 469–473

LeBaron BA, Michalsky JJ, Perez R (1990) A simple procedure for correcting shadowband data for all sky conditions. *Sol Energy* 44: 249–256

Li DHW, Lam JC (2001) An analysis of climatic parameters and sky condition classification. *Build Environ* 36: 435–445

- Littlefair P (2001) Daylight, sunlight and solar gain in the urban environment. *Sol Energy* 70(3): 177–185
- Orgill JF, Hollands KG (1977) Correlation equation for hourly diffuse radiation on a horizontal surface. *Sol Energy* 19: 357–359
- Reindl DT, Beckman WA, Duffie JA (1990) Diffuse fraction correlations. *Sol Energy* 45: 1–7
- Robledo L, Soler A (2000) Luminous efficacy of global solar radiation for clear skies. *Energ Convers Manage* 41: 1769–1779
- Robledo L, Soler A (2001) On the luminous efficacy of diffuse solar radiation. *Energ Convers Manage* 41: 1181–1190
- Skartveit A, Olseth JA (1987) A model for the diffuse fraction hourly global radiation. *Sol Energy* 28: 293–302
- Skartveit A, Olseth JA (1992) The probability density and autocorrelation of short-term global and beam irradiance. *Sol Energy* 49: 477–487
- Skartveit A, Olseth JA, Tuft ME (1998) An hourly diffuse fraction model with correction for variability and surface albedo. *Sol Energy* 63: 173–183
- Suehrcke H (2000) On the relationship between duration of sunshine and solar radiation on the Earth's surface: Ångström's equation revisited. *Sol Energy* 68: 417–425
- Suehrcke H, McCormick PG (1988) The frequency distribution of instantaneous insolation values. *Sol Energy* 40: 413–422
- Tovar J, Olmo FJ, Alados-Arboledas L (1998) One-minute global irradiance probability density distributions conditioned to the optical air mass. *Sol Energy* 62: 387–393
- Udo SO (2000) Sky conditions at Ilorin as characterized by clearness index and relative sunshine. *Sol Energy* 69: 45–53
- WMO (1983) Guide to Meteorological instruments and methods of observation, World Meteorological Organization, Geneva, Switzerland, 289 pp
- Zangvil A, Lamb PJ (1997) Characterization of sky conditions by the use of solar radiation data. *Sol Energy* 61: 17–22

Authors' addresses: Hildeu F. Assunção, Department of Geography, Federal University of Goiás, Campus of Jataí, Jataí, Goiás, Brazil; João F. Escobedo, Department of Natural Resources, School of Agronomic Science, University of State of São Paulo, Botucatu, São Paulo, Brazil; Amauri P. Oliveira (e-mail: apdolive@usp.br), Department of Atmospheric Sciences, Institute of Astronomy, Geophysics and Atmospheric Sciences, University of São Paulo, São Paulo, Brazil.