MINISTÉRIO DA CIÊNCIA, TECNOLOGIA, INOVAÇÕES E COMUNICAÇÕES
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

ANALYSIS OF CLOUD COVER VARIABILITY USING GEOSTATIONARY SATELLITE DATA

Eduardo Weide Luiz

Doctoral Thesis of the Graduate Course in Earth System Science, supervised by Drs. Enio Bueno Pereira and Fernando Ramos Martins.

INPE
São José dos Campos
2018

AGRADECIMENTOS

Gostaria de agradecer, primeiramente, a meus pais, Vanderlei e Rosane, e à minha irmã, Simone, pelo amor incondicional e apoio constante, pela compreensão de minha ausência e, principalmente, pelos esforços empreendidos com o objetivo de ajudar-me a superar as dificuldades que encontrei durante o transcurso do Doutorado.

Agradeço sinceramente ao Dr. Enio Bueno Pereira e ao Dr. Fernando Ramos Martins, meus orientadores, por todo o apoio e orientação necessários para que esta Tese estivesse concluída.

Agradeço também o apoio dos colegas do Laboratório de Modelagem e Estudos de Recursos Renováveis de Energia (LABREN), pela companhia diária e pela ajuda no desenvolvimento deste estudo. Um agradecimento especial ao Rodrigo e ao André que me apoiaram e aguentaram todas as minhas dúvidas e divagações a respeito do conteúdo da tese.

Aos meus amigos que me apoiaram em todos os momentos. À minha namorada, Ielena, que aguentou junto comigo todos os momentos bons e conturbados que passei.

Ao Instituto Nacional de Pesquisas Espaciais (INPE), ao Centro de Ciência do Sistema Terrestre (CCST) e à Coordenação de Aperfeiçoamento Pessoal de Nível Superior – CAPES, pela oportunidade de estudo, disponibilização de infraestrutura e suporte financeiro.

Por fim, a todos que de alguma forma contribuíram para o cumprimento de mais esta etapa.

ABSTRACT

One of the main barriers to increase the solar energy share is its intermittency. Solar energy has a large variability in different time-scales driven by the solar astronomical cycles and by weather. Ground-based measurements are important to evaluate the variability at high resolutions, but they are only representative of small areas close to the measurement sites. Satellite observations come as a solution for the analysis over large areas, however they have coarse temporal and spatial resolutions. The main objective of this thesis is to develop a methodology for the characterization of the variability of the solar resource, focusing on the cloud effects. This simple methodology will allow to evaluate the variability of the solar power generation over large areas, using only data of geostationary satellite images, with no need of ground data. First, we compared the cloud cover fraction obtained through a satellite-based methodology with site-specific data from all-sky cameras. This comparison presented a Pearson correlation of 0.9. In addition, we evaluated the similarity between the cumulative distributions functions of both datasets using the Kolmogorov-Smirnov test and the results pointed out for statistically significant similarity between them, even though their time resolutions were different. Then, we examined the variability of the global horizontal irradiance ramp rates from ground-based radiometers and compared it with the satellite cloud cover variability in 3 different Brazilian climate regimes. The results showed that the driest periods have lower solar irradiance variability. However, this result is not necessarily valid for different climate regimes. For instance, Petrolina, the driest place, exhibited the higher variability for shorter timescales, probably due to the rapid passage of small clouds shadowing the sun. When comparing the variability of the satellite cloud cover with that of the solar irradiance, the Pearson correlation reached up to 0.93, depending on the site, for the same time resolution (30 minutes). However, considering smaller time steps for solar irradiance ramps, the correlation decreased to values lower than 0.66 in all sites. The proposed methodology has broad application in the planning and management of solar power generation in countries with large territorial extension, such as Brazil.

Key Words: Cloudiness, Satellite, Solar Variability

ANÁLISE DA VARIABILIDADE DA COBERTURA DE NUVENS EMPREGANDO DADOS DE SATÉLITE GEOESTACIONÁRIO

RESUMO

Uma das principais barreiras para aumentar a participação de tecnologias solares na matriz energética é sua intermitência. A energia solar apresenta grande variabilidade em diferentes escalas de tempo impulsionadas pelos ciclos naturais do sol e pelo clima. Medidas de superfície são importantes para avaliar a variabilidade em altas resoluções, mas são apenas representativas de pequenas áreas próximas aos locais de medição. Observações por satélite são uma solução para a análise em grandes áreas, no entanto com resoluções espaciais e temporais mais grosseiras. O objetivo principal desta tese foi desenvolver uma metodologia para a caracterização da variabilidade do recurso solar no território brasileiro, com foco nos efeitos das nuvens. Esta tese apresenta o desenvolvimento de uma metodologia simples, para avaliar a variabilidade da produção solar em grandes áreas utilizando apenas dados de satélite geoestacionário, sem a necessidade de modelar a irradiação solar em superfície. Inicialmente, investigou-se a relevância estatística de uma metodologia baseada em dados de satélite geoestacionário para medir a fração de cobertura de nuvens, comparando-a com os resultados de câmeras imageadoras do céu. A comparação apresentou boa concordância, com índice de correlação Pearson de 0,9. Além disso, as funções de distribuição cumulativa de ambas bases de dados foram analisadas através do teste de Kolmogorov-Smirnov demonstrando que as duas bases de dados possuem distribuições estatisticamente similares, mesmo quando apresentam resoluções temporais diferentes. Além disso, a variabilidade das taxas de variação da irradiância global em superfície foi comparada com a variabilidade da cobertura da nuvem calculada através de dados de satélite em 3 diferentes regimes climáticos brasileiros. A investigação confirmou que os períodos mais secos apresentam menor variabilidade, devido à menor presença nebulosidade. No entanto, isso não é válido para diferentes climas, porque Petrolina, o local de clima mais seco, apresentou maior variabilidade em escalas de tempo mais curtas, associada provavelmente com a rápida passagem de pequenas nuvens no local. Ao

comparar a variabilidade da cobertura de nuvens obtida através de dados de satélite com a irradiância solar global horizontal, a correlação foi de até 0,93, para a mesma resolução de tempo (30 minutos). No entanto, para taxas de variação em intervalos menores a correlação diminuiu até valores menores que 0,66. Este método tem ampla aplicação no planejamento e gerenciamento da geração de energia com tecnologia solar em países com grande extensão territorial, como o Brasil.

Palavras-chave: Nebulosidade, Satélite, Variabilidade Solar

LIST OF FIGURES

Pg.

LIST OF TABLES

Pág.

LIST OF ABREVIATIONS

CCF – Cloud Cover Fraction

CDF – Cumulative Distribution Function

CPA – *Cachoeira Paulista*

FAR – False Alarm Rate

GOES - Geostationary Operational Environmental Satellite

IP – Internet Protocol

KS - Kolmogorov-Smirnov

MD - Mean Deviation

NOAA – National Oceanic and Atmospheric Administration

POD – Probability of Detection

PTR – *Petrolina*

RGB – Red Green Blue

RR – Ramp Rates

SD – Standard Deviation

SMS – *São Martinho da Serra*

TCP – Transmission Control Protocol

VS – Variability Score

CONTENT

1 INTRODUCTION

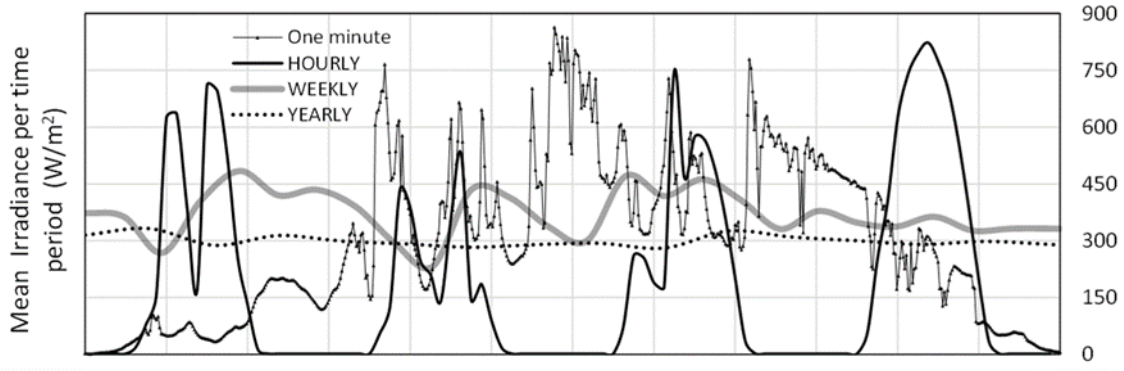
The world energy consumption is expected to increase at least twice by the middle of this century, mostly due to the population growth and economic development (LEWIS and NOCERA, 2006). To tackle with this growth in energy demand ensuring energy security with minimal impact on the climate system, the use of alternative energy sources becomes an important commitment. In this context, the solar energy comes as one of the leading options, especially in the last decade, due to the declining costs of solar power technologies and the increase of its efficiency (MARTÍNEZ-CHICO; BATLLES; BOSCH, 2011).

One of the major barriers to increase the solar energy share in the energy matrix is its intermittency. The solar resource has a large variability in different time-scales driven mostly by the solar natural cycles (the diurnal and the annual seasonality) and by weather. While the natural cycles can be precisely estimated by calculating the apparent motion of the Sun in the sky, the variability caused by weather and atmospheric conditions depends highly on clouds and weather systems dynamics, which makes it much harder to predict (PEREZ et al., 2016; WATANABE; OISHI; NAKAJIMA, 2016). Clouds passing in front of the Sun can cause drastic fluctuations in the surface solar irradiance and this variability has a huge impact in the power output of PV or concentrated solar power plants (ARI; BAGHZOUZ, 2011; LAVE; RENO; BRODERICK, 2015; PEREZ et al., 2016). Also, the solar radiation variability ends up producing transients that are incompatible with the required standards for the electricity distribution system, including voltage variability and frequency disturbances caused by imbalance between power generation and electricity demand (KLEISSL, 2013). In addition, the solar energy intermittency can cause rapid changes in the receiver temperature and it may lead to thermal stress of the devices, increasing the maintenance costs (ARI; BAGHZOUZ, 2011; KAZANTZIDIS et al., 2012).

To understand and manage the output variability of a solar power plant, it is important to examine it in an appropriate spatial and temporal context. For example, when expanding the area of the solar plant (footprint), the intermittency is attenuated, because some photovoltaic panels may be shaded while others may be exposed to clear sky. Furthermore, when the solar irradiation is integrated

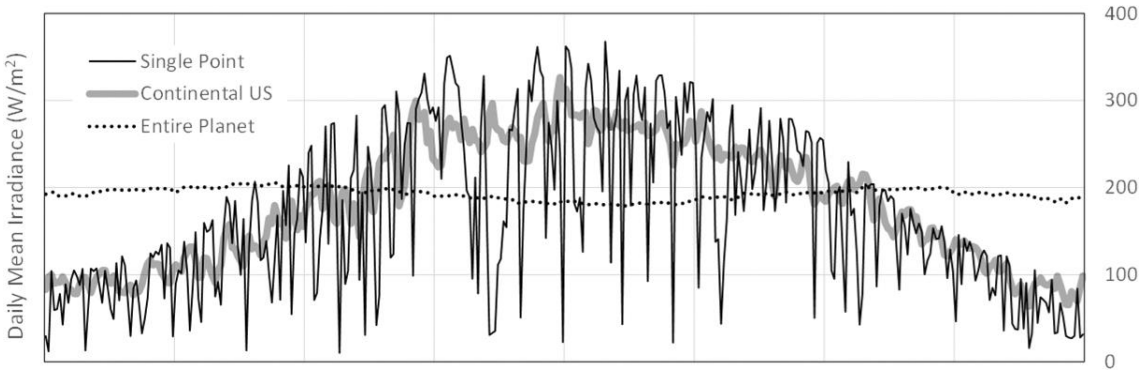
over different time intervals, minutes, hours or days, the high-frequency variability will be attenuated and become insignificant for the longer integration times. Figures 1.1 and 1.2 show the effect of different integration times and footprints on the variability of the solar resource. The smoothing effect on solar irradiance is strongly related to the distance between the solar panels, the time interval of interest, the speed of the cloud and its size (KLEISSL, 2013; PEREZ et al., 2016).

Figure 1.1 **-** The variability of global irradiance time series in a North-America location as a function of the integration time. The Figure includes 1 day of minute data, 4 days of hourly data, 26 weeks of weekly data and 16 years of yearly data.



Source: Perez et al. (2016).

Figure 1.2 **-** The variability of daily global irradiance time series observed along one year for different footprints.



Source: Perez et al. (2016).

These two factors, footprint and time scale, are of primary concern for grid operators and lead to different load management challenges. Increasing the solar generation footprint from a single location to a resource dispersed over an entire region will reduce intermittency considerably. Thus, shorter-term variability matters for smaller spatial scales. One-minute fluctuations are relevant for grid operators of single distribution systems and large centralized plants, due to voltage control issues. In other hand, for grid balancing of distribution systems,

the variability below 30 minutes is not relevant, while hourly and above variability remains of higher concern (PEREZ et al., 2016).

This thesis aims to deepen the knowledge about the variability of the cloud coverage in order to minimize its effect on solar generation. To pursue this goal the main task of this thesis is to develop a methodology for the characterization of the variability of the solar resource in the Brazilian territory, focusing on the effects of the clouds. This methodology should be applicable throughout the country and not only for a specific site, so the existing methodology based only on ground measurements has not the required spatial coverage. The use of geostationary satellite imagery comes as an alternative tool for large areas, however with coarse temporal and spatial resolutions (LAVE; BRODERICK; RENO, 2017; WATANABE; OISHI; NAKAJIMA, 2016). One of the main challenges faced in the development of this thesis was to define the limitations of the satellite methodology and to develop a reliable tool to deal with this limitation.

The first step was to investigate the statistical relevance of the satellite-based methodology when compared with the methodology employing ground-based cloud cover data acquired by an all-sky camera. This comparison aimed to analyze the reliability and the statistical relevance of the satellite results, taking into account the differences in temporal and spatial resolutions. After that, we investigated if the cloud-cover variability derived by the satellite-based methodology was statistically correlated with the variability of global solar irradiance in locations presenting distinct typical climate and atmospheric characteristics.

An immediate application of this new methodology will be the identification of hotspots more appropriate for the solar technologies in Brazil, concerning not only the total amount of the solar resource, but also its temporal variability. This information is critical not only for the generation and distribution of photovoltaic solar energy, but also for the strong impact on the penetration of the concentrated solar energy technologies (CSP) in the Brazilian market.

This document is based on two published papers. Both were produced within the PhD research program in the Center for Earth System Science at the National Institute for Space Research (INPE), Brazil. Also, part of the contents of these

two papers were developed during a period of studies in the Institut für Meteorologie und Klimatologie at the Leibniz Universität Hannover, Germany. A brief description of the thesis structure is presented below:

Chapter 2 includes the area of study of this thesis and the general part of the methodology used in both papers. Data from the Cachoeira Paulista site were used in both papers, however the data from São Martinho da Serra and Petrolina were used only in the second one, because Cachoeira Paulista is the only site with data from an all-sky camera available.

Chapter 3 includes a comparison between results of cloud cover fraction estimated using imagery from GOES 13 geostationary satellite with those estimated by using site-specific images from All-Sky camera. The study assumes as reference the results obtained with the All-Sky camera method and analyzes the reliability and the statistical relevance of the satellite results, taking into account the differences in temporal and spatial resolutions.

Chapter 4 presents a characterization of the global solar irradiance variability for different timescales and different climate regimes. Additionally, the chapter includes a comparison between the cloudiness variability obtained from a geostationary satellite method and those of ground measurements of solar irradiance in the same situations. The study described here aimed at evaluating the relationship between the satellite cloud cover variability and the surface solar irradiance variability in different climate regimes and timescales. The Chapter presents a methodology for the variability analysis of the solar resource using only satellite imagery.

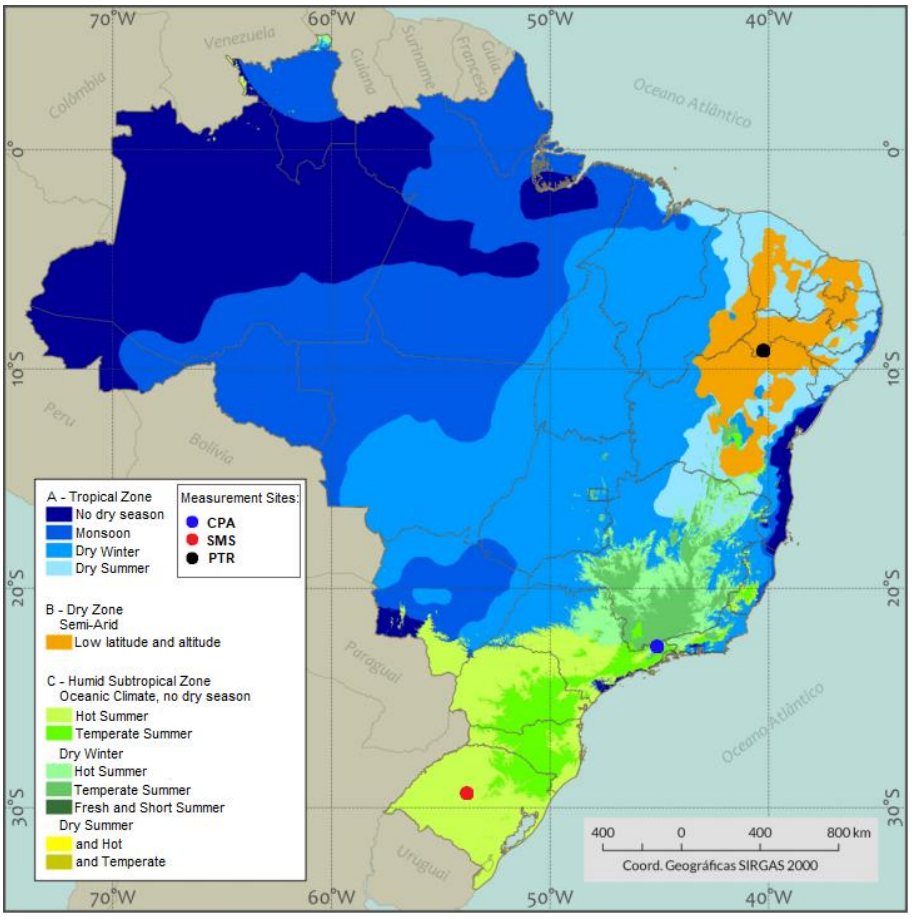
Chapter 5 presents the major findings and the general conclusions of the research. The Chapter also includes some proposals and recommendations for future research, applications of the results and knowledge acquired during this doctorate research.

2 STUDY AREA AND GENERAL METHODOLOGY

2.1. Study Area

Brazil has several distinct climate regimes mostly because of its large territorial extension and the atmospheric circulation. As a consequence, the rainfall and nebulosity regimes are quite different throughout the Brazilian territory. Figure 2.1 shows the location of measurements sites superimposed on a map presenting the major Brazilian climate regimes. The climate in the central area of the Northeastern region (in orange color) is the semi-arid characterized by very dry weather all year long (PEREIRA et al., 2017). The Petrolina (PTR) measurement site designed to acquire meteorological and solarimetric data is located in this area. The other two measurement sites, Cachoeira Paulista (CPA) and São Martinho da Serra (SMS), are located in humid subtropical climate zones.

Figure 2.1. Brazilian climate zones and the measurement sites analyzed during this work.



Source: Adapted from Alvares et al. (2013).

2.1.1. Cachoeira Paulista (CPA)

The Brazilian Institute for Space Research (INPE) operates the ground measurement site located in Cachoeira Paulista (CPA) (22° 41’ 23’’S; 45° 00’ 22’’W). The annual rainfall in the region is around 1360 mm and it presents two distinct climate seasons – a wet season (from October to March) when most of the precipitation occurs (~176 mm/month) and a dry season (from April to September) with low rainfall (~51 mm/month). This also influences the mean cloud cover fraction in the region, which has a mensal mean around 80% in the wet season and 60% in the dry season, according to visual observations (INMET, 2018). During the winter (dry season) the occurrence of cold fronts is common, bringing most of the cloudiness for this season. In summer (wet season), the region is affected mostly by the South Atlantic Convergence Zone (ZCAS) and natural convection (NUNES; VICENTE; CANDIDO, 2009).

2.1.2. São Martinho da Serra (SMS)

The ground measurement site São Martinho da Serra (SMS) is located in at the Southern Space Observatory *campus* (29° 26’ 34’’S; 53° 49’ 23’’W) and managed by INPE. The annual rainfall in the region is around 1800 mm presenting monthly averages ranging from 111 mm to 173 mm. This low seasonal variability in the rainfall also reflects the mean cloud cover fraction in the region, which has a monthly mean around 50% (INMET, 2018). The occurrence of cold fronts during the austral winter and the occurrence of convective systems and local convection in summer explain the low rainfall variability over the year (PEREIRA et al., 2017).

2.1.3. Petrolina (PTR)

The Petrolina (PTR) measurement site (09° 04’ 08’’S; 40° 19’ 11’’W) is located in the driest area of the Northeastern region of Brazil. The annual rainfall is around 580 mm with a wet season from November to April when most of the precipitation occurs. The monthly averages of precipitation range from approximately 44 mm to 114 mm during the wet season (November to April) and from approximately 1 mm to 13 mm during the dry season (from May to October). The cloud cover fraction has a monthly mean around 60% in the wet season and 50% in the dry season, according to visual observations (INMET, 2018). The dry climate of the

region is due to influence of the Walker's circulation, which is only broken in the months of February to April. This wet period accounts for ~60% of the total precipitation, because the Intertropical Convergence Zone (ITCZ) is in its southernmost position (CAVALCANTI et al., 2009; STRANG, 1972).

2.2. Effective cloud cover coefficient

Geostationary satellite imagery has become an important tool for cloud monitoring. The Geostationary Operational Environmental Satellite GOES-13 satellite, operated by the National Oceanic and Atmospheric Administration (NOAA) provided images with 30-min time resolution for South America during the 2016-2017 in five different spectral intervals. The visible spectral band (0.55 to 0.75 μm) allows us to observe clouds, fog, and pollutants in the atmosphere during the day.

The effective cloud cover coefficient (Ceff) is defined for each pixel of the satellite image, in terms of the radiance (L_r) measured by the satellite in the visible channel, the clear-sky radiance (L_{clr}) and the radiance for overcast condition (L_{cld}), as described by Equation 2.1. The L_{clr} and L_{cld} values can be determined by composing clear and overcast images from all satellite images acquired during one-month period. In this step, we divided the reflectance by the cosine of the sun zenith angle to avoid the influence of the illuminance geometry in the radiance data observed by the satellite. According to Martins et al. (2008), Ceff is an dimensionless coefficient related to the cloud optical depth in each image pixel. The Ceff ranges in the 0 to 1 interval from very clear sky conditions until completely overcast sky with no direct solar irradiance reaching the Earth's surface.

$$\text{Ceff} = (L_r - L_{clr})/(L_{cld} - L_{clr}) \tag{2.1}$$

Usually, the clear sky radiance estimation assumes the lowest visible radiance measured by the satellite during the one-month period. However, there may be instances where a cloudless pixel is shaded by clouds from adjacent pixels thus providing an incorrect cloud cover value. In order to minimize this issue, we used a procedure based on the percentile 5 to estimate L_{clr} values.

Similarly, the overcast radiance is usually determined for each pixel as the highest visible radiance observed by the satellite during the period of one month. However, this technique becomes flawed when the completely overcast condition does not occur in a pixel during this sampling period. In fact, this condition occurs frequently in very dry arid areas of the Brazilian Northeastern region. To account for this condition, we assume that similar cloud types are found all over the target area and the radiance observed by the satellite will be quite the same after correction for illuminance geometry. Therefore, we established a fixed L_{cld} value for the overcast sky condition equals to the L_{cld} radiance observed in areas with more frequent cloud cover.

3 COMPARISON OF METHODOLOGIES FOR CLOUD COVER ESTIMATION IN BRAZIL[1]

3.1. Introduction

The understanding of the several factors influencing the Earth's energy balance is fundamental for assessment of the Earth's climate and its variability. Clouds are the major modulator of the shortwave and longwave radiation components of the Earth's energy balance and, as such, help regulate the planet's temperature. In general, high clouds act as greenhouse gases, increasing the longwave radiation (LW) at the surface and warm up the atmosphere, while low clouds have a cooling effect by reflecting the solar radiation back to space (LIOU, 2002; MALEK, 1997). Clouds can even enhance the solar radiation at surface, sometimes to values higher than the ones observed at the top of the atmosphere. This effect happens due the reflection by cloud edges and/or forward scattering of the radiation by the clouds nearby, when the Sun is not obstructed by them (ANTÓN et al., 2011; CALBÓ, 2005; TZOUMANIKAS et al., 2016).

In this context, clouds have attracted increasing interest in the solar energy sector. The first method for assessment of cloud coverage was the visual observations made by operators of meteorological stations, and it is still used today. The method classifies clouds according to visual analysis of shape and appearance, dividing the sky into eight parts (octas) (ROBAA, 2008; WERKMEISTER et al., 2015). Because of the high subjectivity of the method, nowadays, several authors reported different ways for estimating the amount of clouds in the sky in a more objective way. Some authors report methods using downward longwave radiation, along with other meteorological parameters acquired at the surface (DÜRR; PHILIPONA, 2004; MALEK, 1997; MARTY; PHILIPONA, 2000). Others investigate the cloud cover conditions based on all-sky camera images (KAZANTZIDIS et al., 2012; LONG et al., 2006; NETO et al., 2010) and/or on satellite data (ESCRIG et al., 2013; LIANG; YUAN, 2016).

[1] This chapter is based on the paper: LUIZ, E.W.; MARTINS, F.R.; COSTA, R.S.; PEREIRA, E.B.; Comparison of methodologies for cloud cover estimation in Brazil – A case study, as a part of the requirements for obtaining the Ph.D. title in the Earth System Science graduate program. Published at: **Energy for Sustainable Development**, v. 43, p. 15-22, 2018. DOI: 10.1016/j.esd.2017.12.001

There are several difficulties in combining satellite with ground-based data when related to measures of cloudiness. Differences in both spatial and temporal resolutions can be cited. In addition, satellites feature instantaneous measurements at small solid angles, while ground measurements are made at large solid angles from the sky dome (ESPINAR et al., 2009). Also, multiple cloud layers, may lead to misclassifications; e.g. high clouds appear sooner at the all-sky camera images than low clouds at the same distance, while height has no effect for the satellite analysis (ESCRIG et al., 2013).

Many authors have worked comparing different methods for cloud analysis. Wacker et al. (2015) compared different methods for the estimation of the total cloud cover over Switzerland. The methods used all-sky cameras, downward longwave radiation, visual observations, the Meteosat Second Generation (MSG) satellite and ceilometers. With direct comparisons of the results in octas, the results indicated that the automatic methods underestimated the nebulosity estimated by the observer; however, the method using the all-sky camera obtained the closest results. In the comparison between the automatic methods, the data with the better agreement were those obtained with the camera and the MSG satellite. In 52% of cases, the two methods obtained the same result, while concordance within ± 1 or ± 2 octas were 72% and 84%, respectively.

Escrig et al. (2013) compared different cloud situations over Almería, Spain, using the MSG satellite and an all-sky camera. Their satellite algorithm always detected clouds when the camera classified a condition as overcast (over 7 octas) and never classified as overcast when the camera classified as cloudless (below 1 octa), both with > 90% of agreement. For partially cloudy situations the method had approximately 75% of agreement. Werkmeister et al. (2015) made the same comparison for Hannover (Germany), but they classified the situations differently: the overcast condition was stipulated for cloud fractions over 5 octas, the cloudless condition for fractions below 3 octas and broken sky from 3 until 5 octas. For overcast situations, the satellite probability of detection presented very good skill (94%), for cloudless the skill was good (72%), but for intermediate fractions the skill was unsatisfactory (12%).

This study aims at comparing two approaches to get cloud cover information in Cachoeira Paulista (SP), located at the Southeastern region of Brazil. The

methodologies use geostationary satellite data and all-sky camera images. The comparative evaluation assumed the all-sky camera method as the reference methodology, due to better data resolution. The specific methodology is described in Section 3.2. Section 3.3 describes the relationship between both methodologies. The specific conclusions of Chapter 3 are presented in Section 3.4.

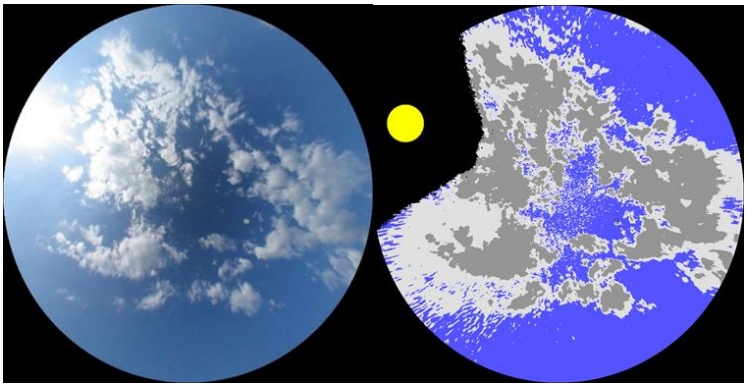
3.2. Methodology

3.2.1. Determination of cloud cover fraction using an *All-Sky* camera

The all-sky camera SRF-02 (EKO Instruments) is a digital camera with fish eye lens and 180° field of view. The camera is encased in a weatherproof housing with a heater system for temperature stability. The user can remotely set up the image acquisition parameters using a desktop computer through TCP/IP connection. The instrument is operating at the roof of the Laboratory of Meteorological Instrumentation (LIM) from INPE. The image acquisition was from July 4th, 2016 to June 30th, 2017, in a 10-min interval during the year of 2016 and 5-min in 2017. The camera takes two sky images with different light exposure: one normal exposed and one underexposed. The EKO instrument provides the Cloud Cover Fraction, hereafter called CCF_{Cam}, using the company's software package to identify clouds and calculate the cloud cover fraction, for each acquired image, using the method proposed by Ghonima et al. (2012). In reason of the hazy sky and the presence of some obstacles close to the horizon line, the image pixels with zenith angles larger than 70° were discarded.

To classify the images, it is necessary to previously manually select clear sky library of images. Then, the software compares each image with the library image with the closest solar zenith angle and classifies each pixel as clear sky, thin cloud or opaque cloud, based in thresholds, as shown in Figure 2.1. It can be noted that the classification of thin clouds presented incorrect results and, because of that, the opaque cloud cover fraction was used as the total fraction. In addition, the software does not classify very bright areas near the sun, masking them in black.

Figure 3.1. Picture from the sky without the elevation angle of 30° (left) and evaluation made by the software (right).

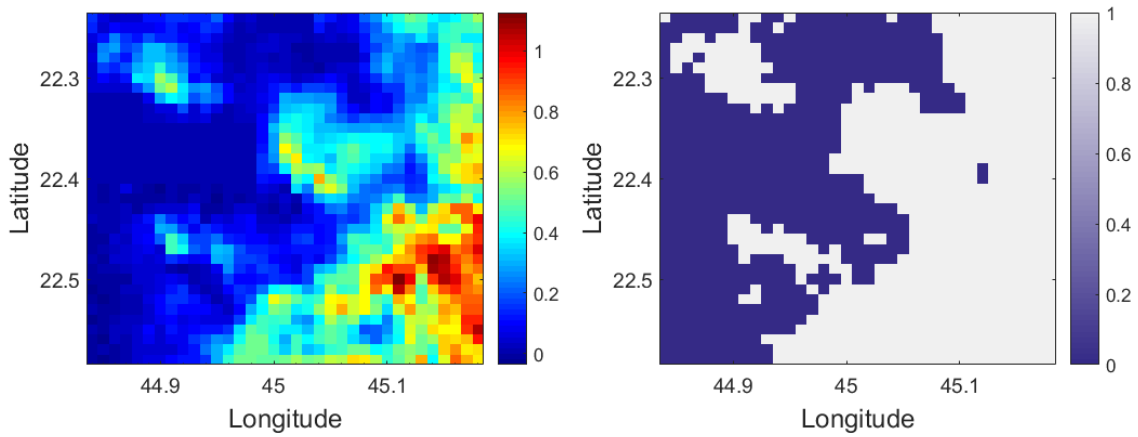


3.2.2. Determination of cloud cover fraction using geostationary satellite imagery

To compare the satellite data with the cloud cover fraction results obtained with the All-Sky Camera, it was necessary to adapt the Ceff methodology. Since Ceff refers only to the cloud coverage of each pixel, a Ceff value threshold was used to classify each pixel as clear sky or cloudy. The threshold used was 0.14, because of best correlation with the camera results. Thus, the fraction of cloud cover estimated by the satellite (CCF_{Sat}) was defined as the number of pixels classified as cloudy, divided by the total in the analyzed area. The image acquisition was done with a 30-minutes time resolution (with exception to the periods of fast scanning mode), from July/2016 to June/2017.

The satellite area necessary to compare with the camera is not easy to determine, because clouds with different heights appear differently in the camera image, e.g. higher clouds appear sooner in the camera horizon. So, with the discard of zenith angles larger than 70º of the camera image, we started with a mean cloud height of 3 km, which gives an area of ~290 km², and then changed the area until the best correlation with the camera results were found. The best correlation were found using an area of 23x23 pixels (~530 km²), which refers to a mean cloud height of ~4 km. Figure 2.3 presents the Ceff of an area of 35x35 pixels over CPA on the left and the mask made using the threshold on the right.

Figure 3.2. Ceff of an area of 35x35 pixels over CPA on the left and the mask made using the threshold of 0.14 on the right (white pixels refers to cloud contaminated and blue to clear-sky).



3.2.3. Statistical analysis

Three statistical indexes were used to compare the methodologies to estimate cloud cover fraction. The method using all-sky camera was the reference methodology. The first statistical index used was the Mean Deviation described in Equation 3.1 where n is the number of match ups between the two data sets and x_m is the difference between them. The Mean Deviation (MD) is positive if the all-sky camera method provided lower cloud cover fraction values than the satellite methods.

$$MD = \frac{1}{n}\sum_{m=1}^{n} x_m \tag{3.1}$$

The Standard Deviation (SD) is defined by Equation 3.2. The expected value for MD is zero if both methods provide equal estimations for cloud cover fractions. The SD value provides information on the dispersion of the discrepancies around the MD value.

$$SD = \sqrt{\frac{1}{n}\sum_{m=1}^{n}(x_m - MD)^2} \tag{3.2}$$

The last statistical index used for comparisons between two datasets was the Correlation Coefficient R as defined by Equation 3.3.

$$R(y,z) = \frac{\sum_{m=1}^{n}(z_m-\bar{z})(y_m-\bar{y})}{\sqrt{\sum_{m=1}^{n}(z_m-\bar{z})^2 \cdot \sum_{m=1}^{n}(y_m-\bar{y})^2}} \tag{3.3}$$

where z_m and y_m represent the different data sets and $\bar{z}$ and $\bar{y}$ are their average value.

In addition to these statistical indexes, we calculated the probability of detection (POD) of similar estimations of cloud cover fractions provided by the two methodologies. Also, a false alarm rate (FAR) was calculated to indicate the probability of finding dissimilar estimation of cloud cover fractions. This analysis was based on the scheme proposed by Werkmeister et al. (2015).

The POD and FAR were calculated using the CCF estimated in octas, as explained by the World Meteorological Organization for weather observations. The CCF values, ranging from 0 to 1, were transformed into octas multiplying by 8 and then rounding them for the nearest integer between 0 (completely clear) and 8 (completely overcast).

For this analysis, the sky conditions were arranged composing three different scenarios: cloud free (CCF ≤ 2 octas), broken clouds (3 octas ≤ CCF ≤ 5 octas) and completely cloudy sky (CCF ≥ 6 octas). The contingency table is presented in Table 3.1, according to Reuter et al. (2009). The table can be used to identify the number of cases in each of the cloud coverage conditions, considering the CCF_{Cam} as reference. The POD and FAR are defined in Equations 3.4 to 3.9.

$$POD_{clear} = \frac{a}{a+d+g} \tag{3.4}$$

$$POD_{broken} = \frac{e}{b+e+h} \tag{3.5}$$

$$POD_{cloudy} = \frac{i}{c+f+i} \tag{3.6}$$

$$FAR_{clear} = \frac{b+c}{a+b+c} \tag{3.7}$$

$$FAR_{broken} = \frac{d+f}{d+e+f} \tag{3.8}$$

$$FAR_{cloudy} = \frac{g+h}{g+h+i} \tag{3.9}$$

Table 3.1. Contingence Table

		Reference Data		
	Scenario	**Clear**	**Broken**	**Cloudy**
	Clear	a	b	c
	Broken	d	e	f
	Cloudy	g	h	i

Source: Reuter et al. (2009)

Also, for a better analysis of the distributions of the generated data, the Kolmogorov-Smirnov (KS) statistical test was used to determine if they differ significantly. The test is based on the cumulative distribution function, where if the maximum difference between the distributions is less than a threshold, the two databases are similar and can be considered statistically the same. According to Espinar et al. (2009), the test is done by converting the data to an unbiased estimator S(xi) of the cumulative distribution function (CDF), at i = 1...N, where N is the population size. For the calculation, it is necessary to find the maximum value of the absolute difference (D) between the two CDFs, according to Equation 3.10.

$$D = max|S(x_i) - R(x_i)| \tag{3.10}$$

where R (xi) is the CDF of the reference database.

In this way, the null hypothesis is formulated so that if D is less than the limit value (Vc), the two sets of data have a very similar distribution and, statistically, could be considered the same. The critical value depends on N and is calculated according to Equation 3.11 with 0.01 significance level (ESPINAR et al., 2009; MASSEY, 1951).

$$V_c = \frac{1{,}63}{\sqrt{N}}\ , N \geq 35 \tag{3.11}$$

In this part of the analysis the comparison between both distributions was done with different time resolutions; 30 minutes for the satellite data and 5-10 minutes for the camera data. The difference on the resolutions may imply that even with coarser time resolution, the satellite still has similar distribution, and its variability may be representative of higher timespans.

3.3. Results

3.3.1. Comparison among CCF_{Cam} and CCF_{Sat} values

Table 3.2 presents the frequency distribution for the cloud cover fractions in octas produced by the satellite and the all-sky camera images. Figure 3.3 shows the histogram of the deviations between CCF_{Sat} and CCF_{Cam} values. Both instruments calculated the same CCF in 59% of the measurements, while in 23% (10%) of them, the differences were ± 1 (± 2) octas.

The Mean Deviation (MD) of -0.02 octas between CCF data provided by CCF_{Sat} and CCF_{Cam} methods was low (0.5% of the mean value of CCF_{Cam}). Furthermore, the Standard Deviation of 2.2 octas and the Correlation Coefficient R of 91.6% denote a good agreement among both methods. Table 3.3 presents the contingency table for the three cloud coverage scenarios based on the CCF_{Sat} and CCF_{Cam} results.

Table 3.2. Density of occurrences between CCF_{Sat} and CCF_{Cam}. The highlighted numbers refer to the cases in which both methods obtained the same result.

		CCF_{Cam}								
	Octas	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
	0	**1450**	182	59	21	11	2	4	3	1
	1	123	**123**	80	42	20	17	5	5	1
	2	47	56	**45**	47	26	18	7	6	1
	3	19	39	35	**31**	29	22	10	6	4
	4	11	25	20	29	**34**	32	27	16	10
	5	6	8	17	18	29	**40**	34	26	16
	6	5	8	17	22	31	35	**59**	45	52
	7	5	1	7	14	25	42	59	**98**	238
	8	2	3	10	10	13	28	58	113	**1148**
	%	**32%**	**9%**	**6%**	**5%**	**4%**	**5%**	**5%**	**6%**	**29%**

Figure 3.3. Frequency histogram for the deviations between cloud cover fraction provided by CCF_{Sat} and CCF_{Cam} in Octas. Positive deviations indicate the CCF_{Sat} values were larger than CCF_{Cam} values.

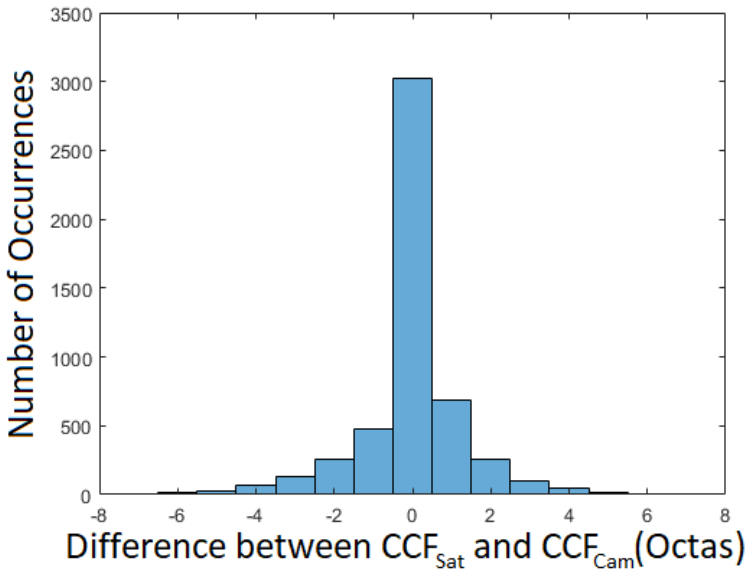


Table 3.3. Contingency table for three cloud coverage scenarios provided by satellite and all-sky camera methodologies.

		CCF_{Cam}		
	Scenario	**Clear**	**Broken**	**Cloudy**
	Clear	42.1%	3.5%	1.1%
	Broken	4%	5.1%	4.3%
	Cloudy	0.6%	2.9%	36.4%

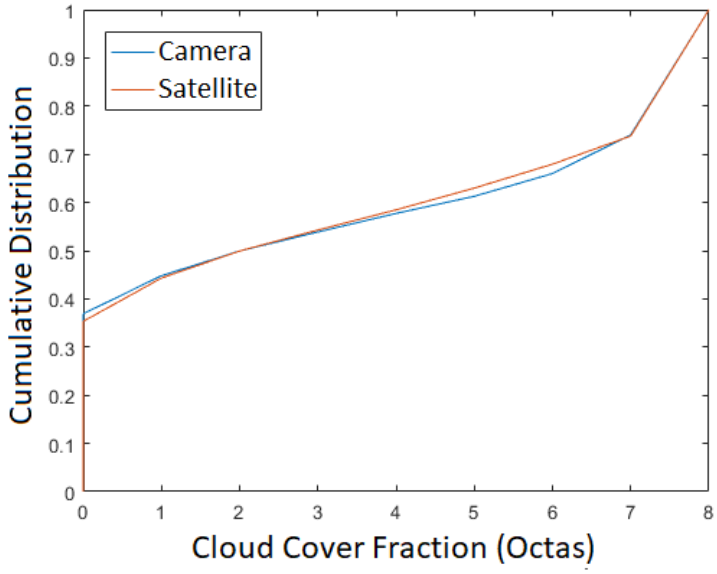
Table 3.4 shows the POD and FAR values for the three cloud scenarios produced by satellite and the all-sky camera methods. It can be noted that both methods present good agreement for cloudless and overcast sky conditions. The satellite method has identified the cloudless scenario in > 91% of the cloudless cases indicated by the all-sky camera method. Furthermore, FAR_{broken} around 61% means that both methods provided data presenting larger discrepancies for partially cloudy scenarios.

Table 3.4. PODs and FARs between CCF_{Sat} e CCF_{Cam}.

POD_{clear}	POD_{broken}	POD_{cloudy}	FAR_{clear}	FAR_{broken}	FAR_{cloudy}
91.1%	44.5%	87.1%	9.9%	61.6%	8.9%

In addition, the cumulative distribution functions (CDF) of the two databases were analyzed through the Kolmogorov-Smirnov (KS) test. Figure 3.4 shows the two functions, where it can be observed that the two distributions have very similar behavior. Also, with respect to the test, using CCF_{Cam} as a reference, the null hypothesis was accepted, with a threshold value of 0.023 and maximum difference between the CDFs of 0.019. This result denotes that despite the differences in the partially cloudy scenarios, the two datasets can be considered statistically the same.

Figure 3.4. Cumulative Distribution Functions of CCF_{Sat} and CCF_{Cam}.



In summary, the larger discrepancies between the cloud cover fractions provided by the satellite and by the all-sky camera methods occur for partially cloudy scenarios. The disparities can be related to three sources of uncertainty: differences on viewing geometry, the spatial resolution and algorithm deficiencies (ESCRIG et al., 2013; WERKMEISTER et al., 2015).

The spatial resolution is also a major source of differences among the CCF estimates by the two methods. Convective clouds with areas smaller than 1 km^2 can be misidentified by the visible satellite imagery. Besides that, the presence of stratocumulus clouds surrounding small clear sky areas can be a source of error. These cloudless areas can be identified by the all-sky camera,

but are usually incorrectly classified as totally cloudy by the satellite due to its low spatial resolution (WERKMEISTER et al., 2015).

Concerning the viewing geometry, the parallax error can contribute to the CCF deviations. Parallax is a displacement or difference in the apparent position of an object viewed along two different lines of sight. For this study, the parallax error depends on the relative position of the cloud and the image sensors. Consequently, clouds can be positioned in different locations by satellite and all-sky camera. The parallax error depends only on the cloud location, because the satellite is in a fixed position relative to the surface.

3.4. Discussion and Conclusions

Two methodologies for estimation of CCF in Cachoeira Paulista, Brazil, were compared. The satellite method presented good agreement with the all-sky camera for clear sky and overcast conditions, with probabilities of detection of 91.1% and 87.1%. The major problem occurs in the broken cloud scenarios, with probabilities of detection of 44.5%. The R correlation between the methods is around 92%.

Our results presented lower probability of detection for overcast situations, but improved results for clear and broken cloud scenarios when comparing it with Werkmeister et al. (2015), which obtained probabilities of detection of 94%, 72% and 12% for overcast, clear sky and broken clouds, respectively. Additionally, when comparing the uncertainties of the methodologies with Wacker et al. (2015), the method proposed here presented 59% of agreement, while for them 52%.

Additionally, when comparing the results presented here with Luiz et al., (2018), the probabilities of detection for overcast and clear sky were similar (93.3% against 91.1% and 87.6% against 87.1%, respectively), but for broken clouds scenarios the probability presented here of 44.5% were superior against the 38% obtained on the paper. It is important to highlight that in Luiz et al., (2018) we analyzed only three months of data during the dry season, while here we analyzed one full year of data.

Regarding the satellite method, the most important issue is the spatial resolution, which has the major impact on the broken-clouds sky scenarios. Because of the

better agreement for cloudy and clear sky conditions, it’s possible to say that the cloud masking works properly for large clouds with, at least, the size comparable to the satellite image pixel. The apparent larger error may be related to the existence of only three scenarios in these analyzes, since small errors at the ends of the scenarios can be considered as completely wrong because they are in different scenarios.

Furthermore, the cumulative distributions of the two data sets of data are very similar, even with different time resolutions and they can be considered statistically the same by the KS test. This result implies that the satellite variability is representative of higher time resolutions, at least related with the data distribution.

The method using the all-sky camera needs also to be improved, as it presented some deficiencies. For example, the very bright areas around the sun are not analyzed by the software. Such issues can be at least partly overcome by using some different techniques like sky whiteness corrections in the area around the sun and should be a topic for further studies.

4 ANALYSIS OF THE SOLAR IRRADIANCE INTRA-DAY VARIABILITY IN DIFFERENT BRAZILIAN CLIMATE ZONES[2]

4.1. Introduction

The knowledge on the solar irradiance variability is fundamental and necessary for understanding the output variability of a solar power plant. So, it is important to find out the best method to evaluate such variability. Different physical variables are relevant for different technologies, similarly, different timescales are important for different technologies and areas of interest (PEREZ et al., 2016).

Grid operators are interested in the power output, which reflects the underlying variability of solar irradiance. For this, two fundamental parameters can be used: global horizontal irradiance (relevant for PV technologies) and direct normal irradiance (relevant for concentrating solar technologies). However, these two parameters embed both the astronomical cycle of the sun and cloud weather effects. To focus on the second it is helpful to use a parameter that normalizes the irradiance by the sun geometry, conserving the cloud variability (PEREZ et al., 2016). For PV technologies, the clearness index, Kt (ratio between global and the extraterrestrial irradiances) and the clear sky index, Kt* (ratio of the measured and the clear sky modeled global irradiances) may be good indexes to meet this criterion (LAVE; BRODERICK; RENO, 2017; PEREZ et al., 2016; WATANABE; OISHI; NAKAJIMA, 2016).

After the definition of the physical quantity, it is necessary to define how the variability should be measured. Watanabe et al. (2016) used the mean, standard deviation and sample entropy to evaluate the variability of the surface solar irradiance over Japan. Perez et al. (2016) proposed the nominal variability of Kt* ramp rates (RR) as a metric to study the solar power variability. The RR, defined as the change in magnitude over a timeframe of interest, was also used by Lave

[2] This chapter is based on the paper: LUIZ, E.W.; MARTINS, F.R.; GONÇALVES, A.R.; PEREIRA, E.B.; Analysis of intra-day solar irradiance variability in different Brazilian climate zones, as a part of the requirements for obtaining the Ph.D. title in the Earth System Science graduate program. Published at **Solar Energy**, v. 167, p. 210-219, 2018. DOI: 10.1016/j.solener.2018.04.005.

et al. (2015) to quantify the local high-frequency variability based on their cumulative distribution functions.

Besides that, assuming that cloud cover is the major driver of the surface solar irradiance, it is useful to examine the cloud variability in an appropriate spatial and temporal perspective, in order to get a better comprehension of the intermittency issues and to develop mitigation solutions. In a specific location, during a partly cloudy day, the one-minute incoming solar irradiance will experience very large variability, because of the clouds shadowing the Sun. But, the daily total of the incoming solar radiation will experience much less inter-day variability. Furthermore, in spatial terms, for larger the areas of interest (footprint), lower is the variability of the incoming solar radiation (PEREZ et al., 2016).

Combining ground-based and satellite measurements is a good way to investigate solar irradiance variability over large areas. However, the poor temporal and spatial resolutions make it necessary to use temporal downscaling and/or spatial interpolation (PEREZ et al., 2016). Many authors have used different spatial interpolations to estimate the solar irradiance variability at some distance from the measurement sites (ARIAS-CASTRO; KLEISSL; LAVE, 2014; ELSINGA; VAN SARK, 2015; PEREZ et al., 2012; YANG et al., 2014). Ngoko *et al.* (2014) and Wegener *et al.* (2012) used different downscaling methods to synthetize data with higher frequency than the available, while Stein *et al.* (2011) synthetized high-frequency data from satellite data in the nearby areas. Hummon *et al.* (2012) and Lave *et al.* (2017) found associations between 1 hour satellite data and higher-frequency ground data for locations in USA and Watanabe *et al.* (2016) characterized the variability in Japan using cloud properties.

In this work, we propose a methodology to evaluate the variability of the incoming surface global solar irradiance using ramp rates of the effective cloud cover coefficient estimated from the visible GOES-13 satellite imagery data, using one year of data (July/2016 until June/2017). The study investigates the variability of ground-based Kt ramp rates in different timescales (1-, 5- and 30-minutes) and compared them with ramp rates of the cloud cover coefficient obtained from 30-minutes time resolution satellite data. The novelty of this work is to use a simple method to analyze solar variability using only visible satellite imagery, instead of estimating the irradiance on surface. The large extension of the Brazilian territory

makes it possible to evaluate the applicability and the performance of the method for different climate characteristics, and so, we analyzed data from three different climate zones (CPA, SMS and PTR). The methodology is described in Section 4.2. Section 4.3.1 encompasses the evaluation of the ground-based irradiance variability at the 3 ground sites and Section 4.3.2 describes the relationship between the satellite cloud ramp rates and the Kt ground-based ramp rates. The specific conclusions of Chapter 4 are presented in Section 4.4.

4.2. Methodology

4.2.1. Ground-based solar irradiance index

For this analysis, we used the clearness index (Kt) for the three observation sites calculated using the global irradiance data recorded with a one-minute temporal resolution, from July/2016 to June/2017, and dividing it by the irradiance at the top of the atmosphere at each moment. As mentioned before, the use of Kt allows to remove both the diurnal and seasonal cycles to the Earth movements. Additionally, the empirical correction, proposed by Perez et al. (1990) was applied to remove the effect of the air mass in large solar zenith angles. Equation 4.1 shows the corrected value (Kt'), where "am" is the relative air mass. The irradiance is more affected by the air mass in moments

$$Kt' = Kt/(1.031\exp(-1.4/(0.9 + 9.4/am)) + 0.1) \quad (4.1)$$

4.2.2. Ramp rates

Several metrics can be applied to understand the solar irradiance variability at a specific location. The difference between two sequential values of the surface incoming solar irradiance were used in several studies (PEREZ et al., 2016; WATANABE; OISHI; NAKAJIMA, 2016). However, Kleissl (2013) proposed a moving average, as described in Equation 4.2, to avoid the influence of different time steps used for the ramp rates evaluation. The ramp rates of Kt', designated from now on as $RR_{\Delta t}^{Kt'}$, were used in this study.

$$RR_{\Delta t}^{Kt'} = \frac{1}{\Delta t}\left(\sum_{t}^{t+\Delta t} Kt' - \sum_{t-\Delta t}^{t} Kt'\right) \quad (4.2)$$

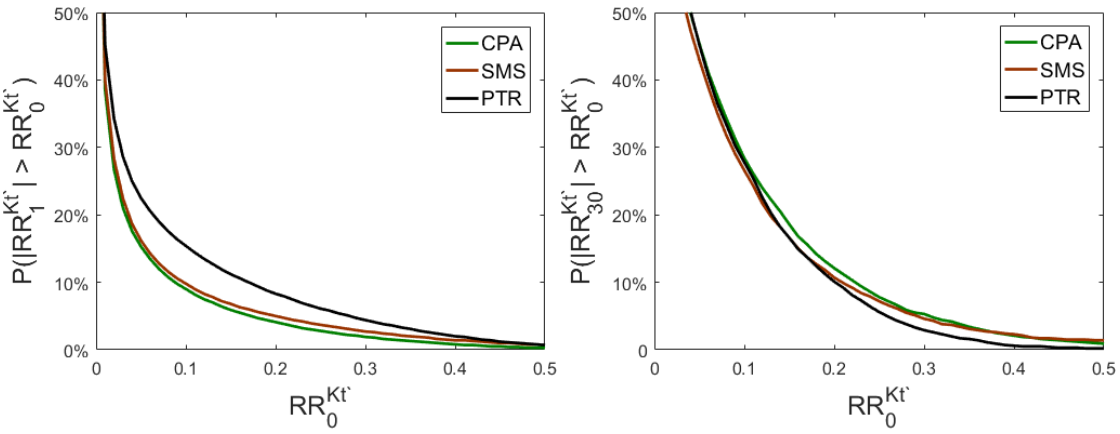
where Δt is the timescale of interest.

In this work, we used Δt equals to 30, 5 and 1 minutes, hereafter called $RR_{30}^{Kt'}$, $RR_{5}^{Kt'}$and $RR_{1}^{Kt'}$, respectively. We also proposed to use the ramp rates based on the Ceff values (RR_{30}^{Ceff}) as the difference of Ceff values in two successive images for each image pixel. The index "30" in the symbol stands for the 30-min temporal resolution of the satellite images.

4.2.3. Variability Score

The cumulative probability distribution is useful to understand the behavior of the ramp rates, $RR_{\Delta t}^{Kt}$, in different timescales, although it is hard to exactly compare the variability from one place to another, as can be seen in Figure 4.1. It is important to highlight that at the vertical axis $P(|RR_{\Delta t}^{Kt'}| > RR_{0}^{Kt'})$ is the complementary probability of the typical cumulative distribution functions (CDFs), representing the probability (on the vertical axis) of having in the dataset values larger than the values of the horizontal axis.

Figure 4.1. The annual cumulative distributions $(P(|RR_{\Delta t}^{Kt'}| > RR_{0}^{Kt'}))$ for one-minute ramp rates ($RR_{1}^{Kt'}$) on the left side and for 30-minutes ramp rates ($RR_{30}^{Kt'}$) on the right side. The probability and Kt' are non-dimensional.



To solve the comparison issue and taking into consideration the high impact of large values of $RR_{\Delta t}^{Kt}$, Lave et al. (2015) proposed a metric, called Variability Score (VS). The VS is defined as the maximum value of the product between the $RR_{\Delta t}^{Kt}$ and its probability, scaled by 100 as shown in Equation 4.3.

$$VS(\Delta t) = 100.\max[RR_{0}^{Kt'}.P\left(\left|RR_{\Delta t}^{Kt'}\right| > RR_{0}^{Kt'}\right)] \quad (4.3)$$

where $RR_0^{Kt'}$ is the $RR_{\Delta t}^{Kt'}$ magnitude and $P(RR_{\Delta t}^{Kt'} > RR_0^{Kt'})$ is the probability of finding values higher than $RR_0^{Kt'}$ on the data set.

According Lave et al. 2015, usually the VS is generally well correlated with number of tap changes on a solar electric plant, however variables like tracking setup, cloud velocity and geographic smoothing may lead to some erroneous evaluations and need to be better analyzed. The VS(Δt) value ranges from 0 (no variability) to 100 (all ramp rates with the maximum value). Larger VS(Δt) indicates more variability.

4.2.4. Statistical analysis

Three statistical indexes were used in this Chapter to compare the monthly VS's estimated using the irradiance ground-data and the Ceff data. The first statistical index used was the Mean Deviation described in Equation 4.4 where n is the number of match ups between the two data sets and x_m is the difference between them. The Mean Deviation (MD) is positive if the irradiance method provided lower VS values than the satellite methods.

$$MD = \frac{1}{n}\sum_{m=1}^{n} x_m \tag{4.4}$$

The Root Mean Square Deviation (RMSD) is defined in Equation 4.5, where z_m and y_m represent the different data sets.

$$RMSD = \sqrt{\frac{1}{n}\sum_{m=1}^{n}(z_m - y_m)^2} \tag{4.5}$$

The last statistical index used for comparisons between two datasets was the Pearson Correlation Coefficient R as defined by Equation 4.6.

$$R(y,z) = \frac{\sum_{m=1}^{n}(z_m-\bar{z})(y_m-\bar{y})}{\sqrt{\sum_{m=1}^{n}(z_m-\bar{z})^2 \cdot \sum_{m=1}^{n}(y_m-\bar{y})^2}} \tag{4.6}$$

where z_m and y_m represent the different data sets and $\bar{z}$ and $\bar{y}$ are their average value.

4.3. Results

4.3.1. Variability comparison at the three sites

As mentioned before, the three ground sites are located in areas with very distinct climates. While CPA and SMS have very similar annual rainfall, PTR has a much dryer climate, close to one third of the precipitation of the first two sites. On the other hand, CPA and PTR present two distinct seasons (wet and dry), while SMS has an almost constant precipitation all over the year.

Figure 4.2 shows the ramp rate magnitudes ($RR_0^{Kt\prime}$) times the annual cumulative probabilities $P(\left|RR_{\Delta t}^{Kt'}\right| > RR_0^{Kt'})$ scaled by 100 for 1-min ($RR_1^{Kt\prime}$) and 30-min time steps ($RR_{30}^{Kt\prime}$) for the three ground measurement sites. We omitted $RR_5^{Kt\prime}$ from the comparison, because the plot has very similar behavior with $RR_1^{Kt\prime}$. Therefore, the VS in PTR is much higher than in CPA and SMS for 1-min and 5-min timescales. This could be related to the presence of a larger number of small clouds in Petrolina, even in dry conditions. Furthermore, for the 30-min timescale, the VS values are very similar in the three sites, which reinforces the smoothing effect attenuating the variability when analyzing larger timespans. Besides that, it is important to highlight that the VS(Δt) for all the situations occurs for $RR_0^{Kt\prime}$ between of 0.1 and 0.2, which shows they have the larger impact.

Figure 4.2. Annual cumulative probabilities $P(\left|RR_{\Delta t}^{Kt'}\right| > RR_0^{Kt'})$ multiplied by the ramp rate magnitudes $RR_0^{Kt\prime}$ values scaled by 100 for time-steps of 1-minute ($RR_1^{Kt\prime}$) on the left and 30-minutes ($RR_{30}^{Kt\prime}$) on the right for the three ground

measurement sites. The dots at the maximum values of the curves are the VS(Δt) values for each site. The probability and Kt' are non-dimensional.

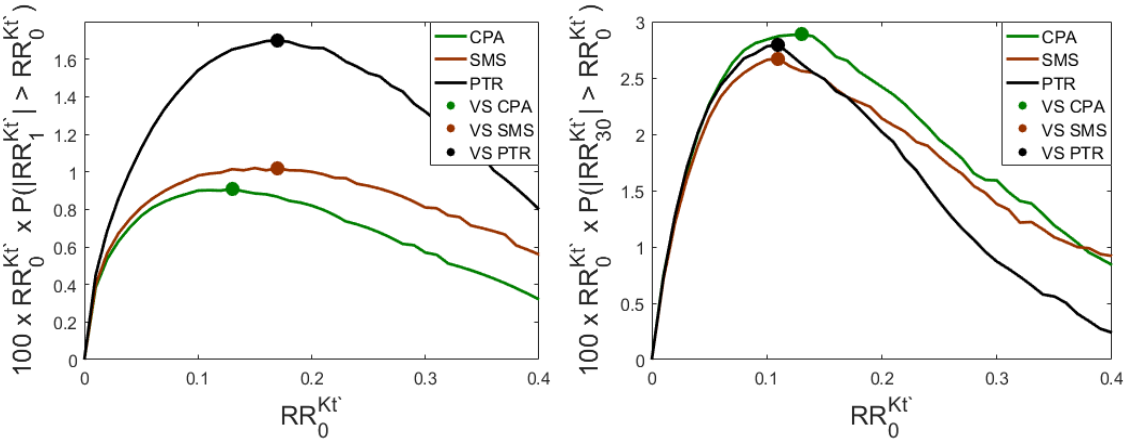


Figure 4.3 presents a comparison of the annual VS(Δt) index for different timescales (Δt) at the three ground measurement sites. We can note that the PTR has a much higher VS(Δt) index from 1 to 15 minutes, while from 20 until 30 the three sites have more similar VS values. It is also important to highlight that CPA and SMS have very similar VS(Δt) for Δt from 1 to 15, but for longer Δt, the VS values for CPA gets closer to PTR faster than SMS. This issue could be related to climate similarities between CPA and PTR like the two distinct wet and dry seasons.

Moreover, it is expected that the $RR_{\Delta t}^{Kt'}$ values increase with Δt, due to the longer time interval over which Kt' can variate from the previous value (LAVE; KLEISSL; ARIAS-CASTRO, 2012). So, since VS(Δt) is based on the $RR_{\Delta t}^{Kt'}$ probability distribution, it is also expected to increase for larger Δt. Nevertheless, it is not observed in PTR. Figure 4.3 shows that the VS(Δt) has similar values from 5 to 30 minutes Δt in PTR, which reinforces the very high short-term variability there.

Figure 4.3. Annual Variability Score (VS) for different time steps (Δt) used in the ramp rates evaluation at the three ground sites. VS is non-dimensional.

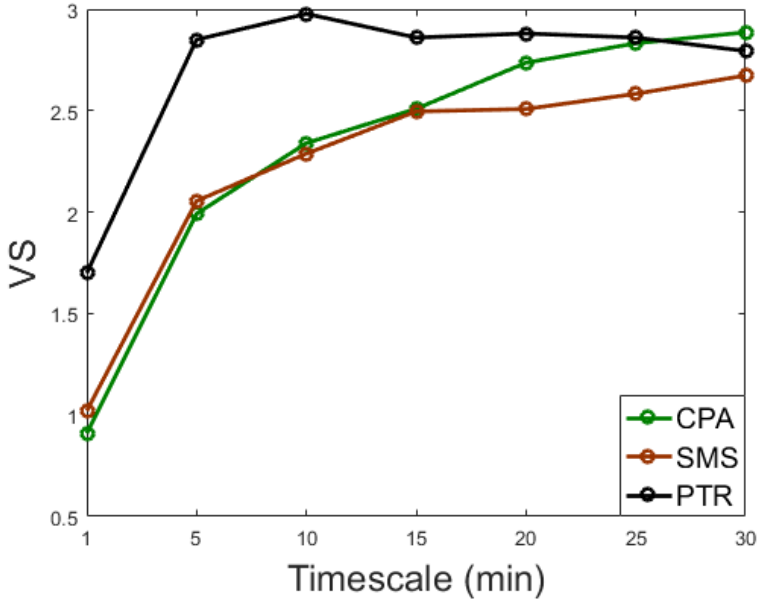
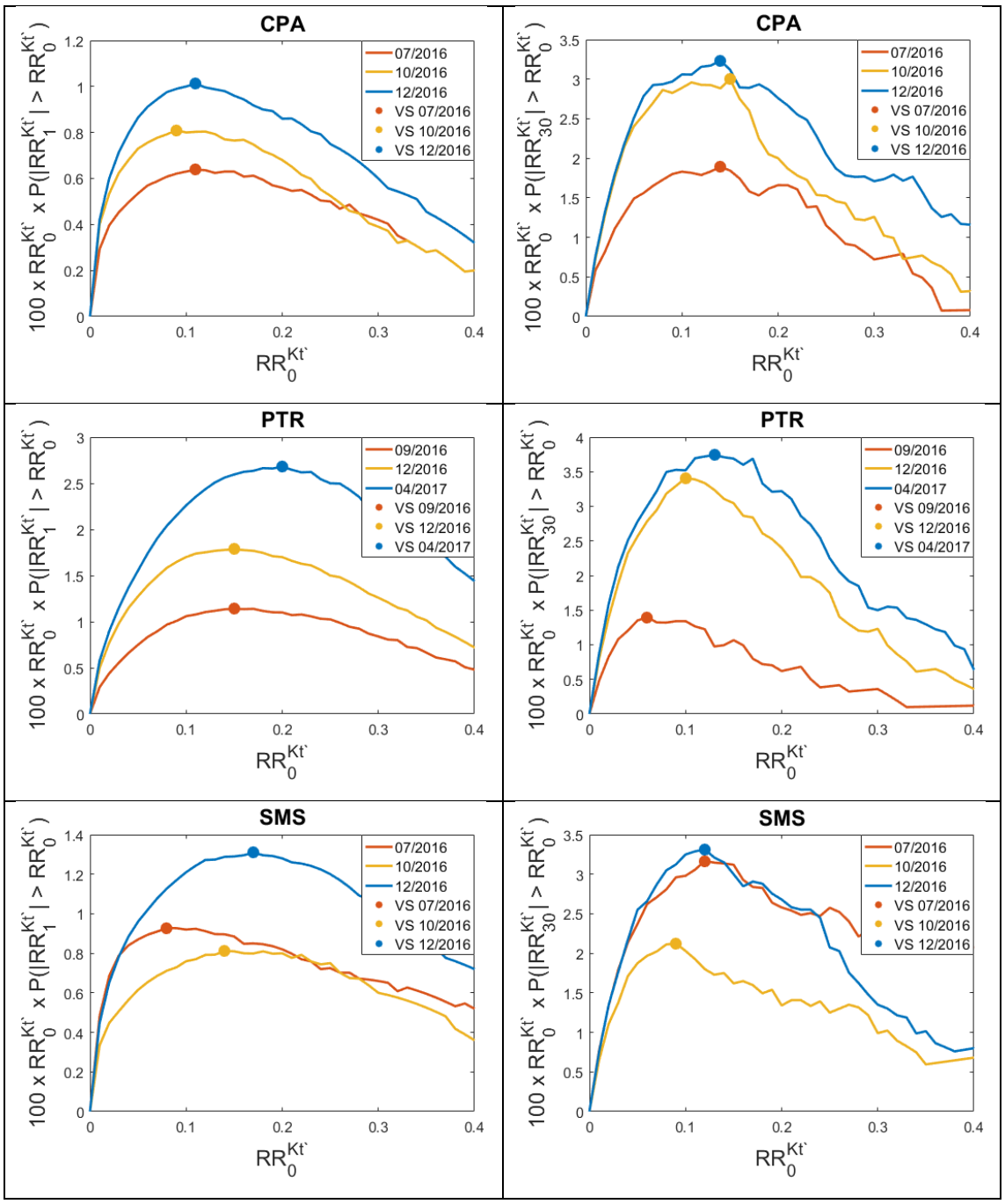


Figure 4.4 presents similar plots as presented in Figure 3.3, but using only one-month periods at the three ground sites. In CPA and PTR, the red line stands for a month in the dry season, the blue line stands for a month in the wet season and the yellow line for a month in between. It's possible to see that the wet season showed the highest Kt' variability, which is related to more convection activity in this season.

Figure 4.4. Monthly cumulative probabilities $P(\left|RR_{\Delta t}^{Kt'}\right| > RR_0^{Kt'})$ multiplied by the ramp rate magnitudes $RR_0^{Kt'}$ values scaled by 100 for time-steps of 1-minute ($RR_1^{Kt'}$) on the left and 30-minutes ($RR_{30}^{Kt'}$) on the right, in CPA (first line), PTR (second line) and SMS (third line). The dots at the maximum values of the curves are the VS(Δt) value for each month. In CPA and PTR, the red line stands for a month in the dry season, the blue line stands for a month in the wet season and the yellow line for a month in between. SMS has no distinct wet and dry seasons. The probability and Kt' are non-dimensional.



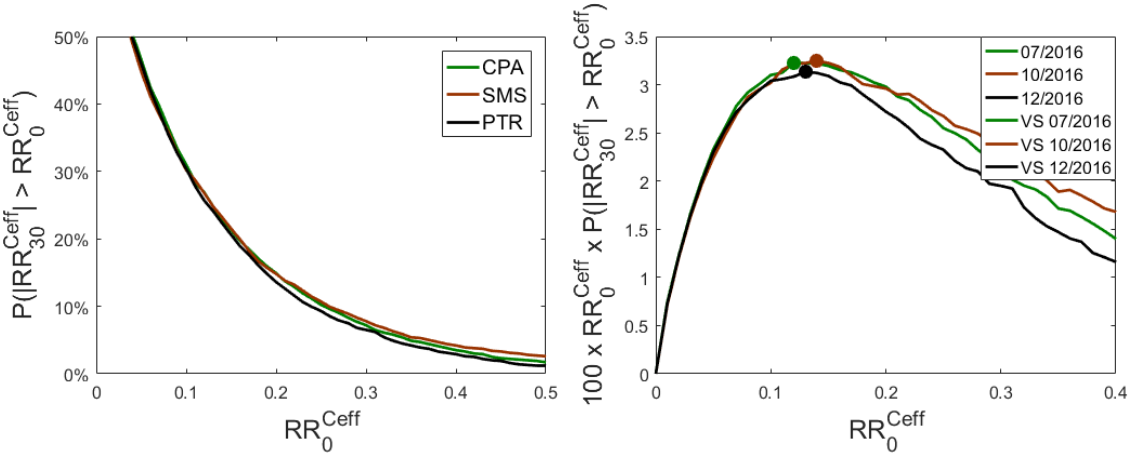
On the other hand, for SMS, with no distinct wet and dry seasons, the difference between the months can be related to different cloud types. In austral winter (red line), most of the clouds are related to the passage of cold fronts, while in summer

(blue line) most of the clouds are related to local convection. The local convection creates broken clouds fields and, therefore, larger $RR_1^{Kt'}$variability. Small clouds have less influence on the $RR_{30}^{Kt'}$ distribution probabilities, thus, resulting in a similar VS(30) in both summer and winter months. The yellow line stands for a transition season (spring) presenting fewer passages of cold fronts and less intense local convection.

4.3.2. Relationship between satellite and ground based variability

After the analysis of the solar irradiance variability using ground measurements, we investigated their relationship with the observed variability of RR_{30}^{Ceff} obtained from Ceff satellite data. Figure 4.5 shows the complementary probability distribution for RR_{30}^{Ceff} (left) and the plot for RR_{30}^{Ceff} times its probability scaled by 100 (right) for the three ground sites.

Figure 4.5. Ceff ramp rates annual cumulative distributions ($P(|RR_{\Delta t}^{Ceff}| > RR_0^{Ceff})$) on the left and the same cumulative distribution multiplied by the Ceff ramp rate magnitudes (RR_0^{Ceff}) scaled by 100 on the right, at the three sites. The dots at the maximum values of the curves are the VS for each ground site. The probability and Ceff are non-dimensional.



The first step was to compare the monthly and annual VS(Δt) values obtained using $RR_{30}^{Ceff}$and $RR_{30}^{Kt'}$. Figure 4.6 shows the scatter plots of VS(30) values between $RR_{30}^{Kt'}$ and RR_{30}^{Ceff}at the three ground sites. It can be noticed that there is an almost linear relationship for all sites, with a slight overestimation from RR_{30}^{Ceff}. Table 4.1 presents the Pearson correlation, the Bias deviation and the Root Mean Square deviation (RMSD) for this comparison. PTR had the largest Pearson correlation (0.93) and the lowest RMSD. The lowest correlation (0.67)

was obtained using data from CPA, but the Bias deviation was the lowest there. The low correlation in CPA needs to be better investigated, but could be related to weather characteristics or the high number of burning biomass events close to the site during the dry season, which may interfere in the satellite analysis and in the solar irradiance at surface. Also, January may be an outlier month in CPA and if we take the month out of the analysis, the Pearson correlation increases to 0.84, similar to the ones obtained at the other two sites.

Figure 4.7 presents a comparison between the monthly and annual VS(Δt) values for $RR_5^{Kt'}$ and $RR_1^{Kt'}$. Table 4.1 also presents the Pearson correlation, the Bias deviation and the Root Mean Square Deviation (RMSD) for these comparisons. The PTR site presented the smaller deviations and the largest correlation. There is a high decrease in the correlation for 1-min time step. This is an expected result, because the satellite cloud cover data have 30-min time resolution, which increases the smoothing effect and may be too poor to describe high frequency changes caused by clouds.

Besides that, the deviations are supposed to increase for the $RR_{\Delta t}^{Kt'}$ calculated for lower time steps (Δt), because VS is larger for longer time steps. As we can notice in Table 4.1 and Figure 4.7, this behavior is observed CPA and SMS, but not at PTR, which has very low deviation and high correlation in the comparison between the VS for RR_{30}^{Ceff} and $RR_5^{Kt'}$. This fact reinforces the earlier remarks from Figure 4.3, where we can observe similar VS in PTR for time steps longer than 5 minutes, but now for the monthly analysis.

Table 4.1. Statistical comparison between the VS values obtained from the Cloud Cover coefficient (Ceff) provided by satellite images and the ground-based Kt' values at the three ground sites.

	30 minutes			**5 minutes**			**1 minute**		
	R	**MD**	**RMSD**	**R**	**MD**	**RMSD**	**R**	**MD**	**RMSD**
CPA	0.67	0.33	0.63	0.59	1.30	1.41	0.56	2.39	2.46
SMS	0.89	0.62	0.74	0.77	1.30	1.41	0.64	2.32	2.42
PTR	0.93	0.40	0.51	0.8	0.33	0.62	0.66	1.48	1.62

Figure 4.6. Scatter plot between the VS(30) of the cloud cover coefficient (Ceff) and the VS of 30-minutes ramp rates of Kt' at the three sites. The blue dots are the monthly VS and the red dots are the annual VS. VS is non-dimensional.

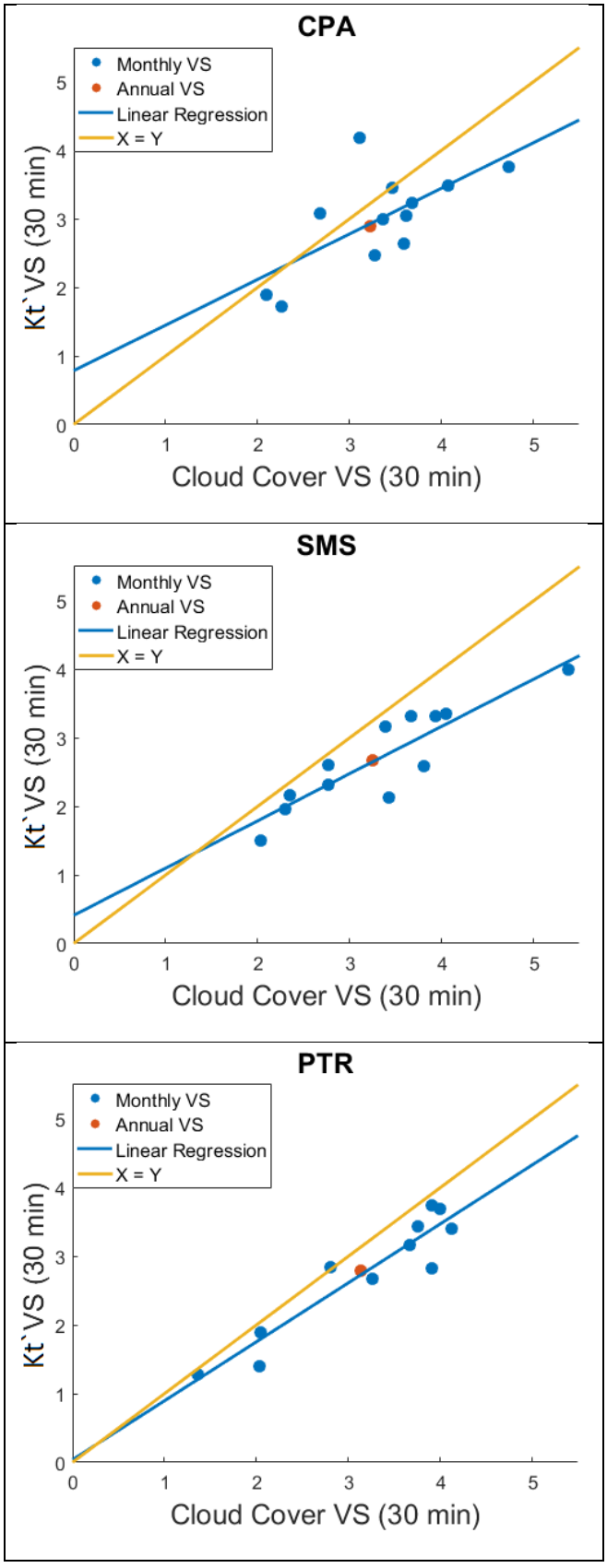
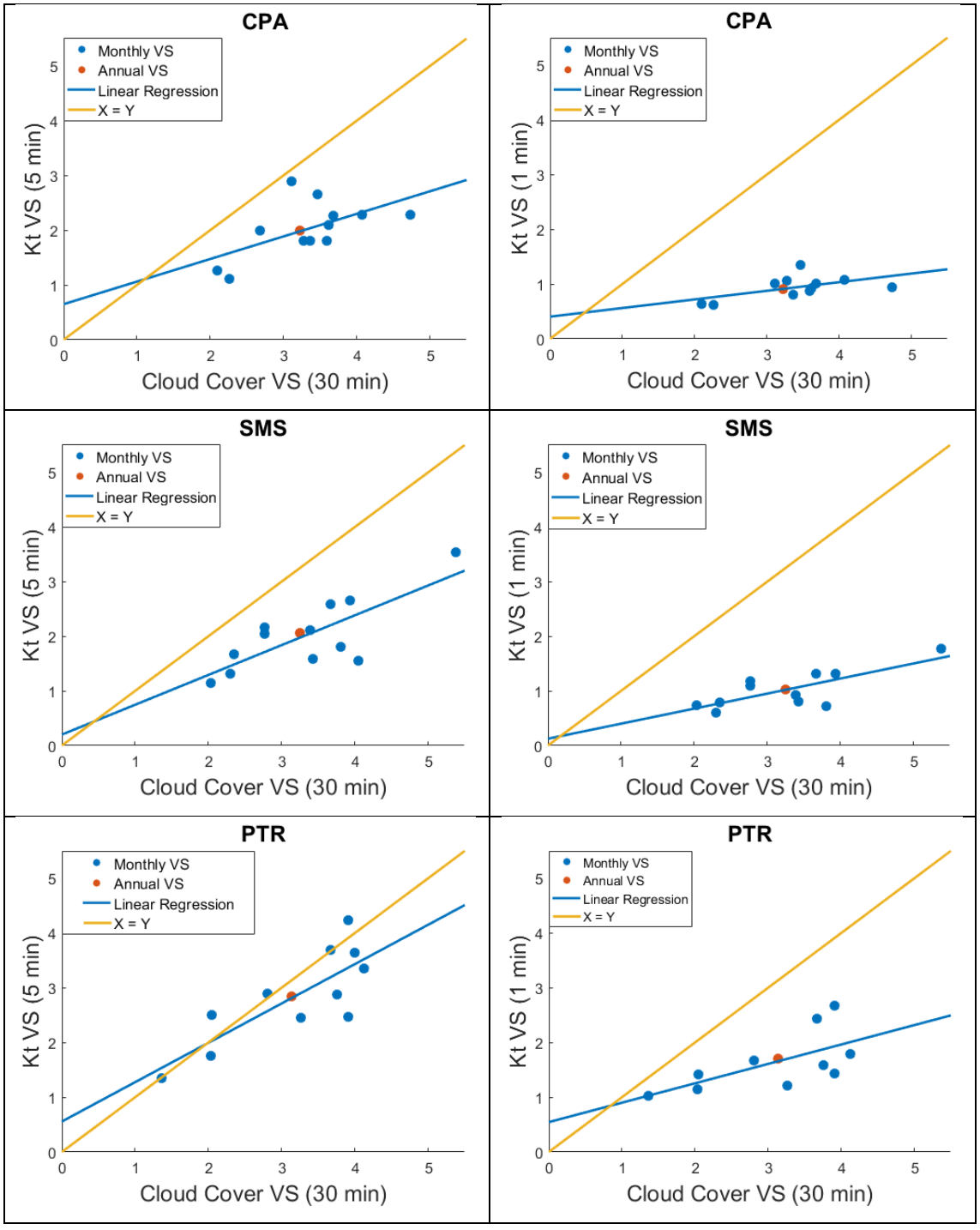


Figure 4.7. Scatter plot between the VS(30) for cloud cover coefficient (Ceff) and the 5-minutes (left column) and 1-minute (right column) VS for Kt', at CPA (first line), SMS (second line) and PTR (third line). The blue dots are the monthly indexes and the red dots are the annual indexes. VS is non-dimensional.



4.4. Discussion and conclusions

The solar irradiance variability over three different regions of Brazil were investigated using the Variability Score (VS), obtained using local ground-irradiance data, and the effective cloud cover obtained from GOES-13 satellite.

One year of data was used to encompass the different weather characteristics on those places.

As expected, the humid months have the highest variability because of the larger number of clouds in the sky. However, the driest location evaluated in this study, Petrolina (PTR), presented larger ramp rates variability than the other two ground sites (Cachoeira Paulista - CPA and São Martinho da Serra - SMS) for time steps shorter than 15-minutes. This fact can be of high concern for the Brazilian energy planning, because PTR is located in the region pointed out as the best for solar power generation (PEREIRA et al., 2017). This large variability may be related to the presence of a large number of clouds in the region, as we can see in Figure 4.8. The Figure presents a picture with a very common sky condition in the Caicó (RN) site, which have similar climate characteristics with PTR. Anyhow, the annual 30–min ramp rates variability is very similar for all three sites due to smoothing effects when analyzing longer time steps.

Figure 4.8. Picture of the Sky of Caicó (RN) located in the same region of Petrolina (PTR).

Furthermore, the VS obtained for CPA ground site is very similar to the VS for the SMS site for the ramp rates in shorter time steps (≤ 15 min). By the contrary, for longer time steps (in the 20- to 30-min range), the VS values in CPA site are closer to the ones obtained in PTR. This fact may be related with the weather

similarity between CPA and SMS in the wet season and CPA and PTR in the dry season.

With respect to the comparison of solar variability for 30-min time steps, using the VS provided by the ground-based measurements and the satellite, the results presented good agreement. Our findings indicate that the variability of the cloud coverage in a region is representative of the Kt' variability (and hence solar production), without the need for costly and time demanding field measurements or modeling the solar irradiance. The CPA site has the lowest Pearson correlation (0.67) between the monthly VS values, but the smaller Bias. The reason to such low correlation needs to be better investigated, but may be related to local weather characteristics or data acquisition issues in January 2017.

Using the same database, we investigated the relationship between the VS of Ceff ramp rates in 30-min time step and Kt' ramp rates at lower time steps. For the 5-min Kt' ramp rates, there was a high Pearson correlation (up to 0.80). An interesting outcome was that the magnitude of the VS index values in PTR was very similar for 5 and 30 minutes. It was not intuitive because the index is proportional to the ramp rate values, usually higher for 30-min time steps. This fact reinforces the high variability in the PTR site for smaller timescales.

Concerning the ramp rates for 1-min time steps, the agreement with the VS obtained using satellite 30-min resolution data was not satisfactory. The best Pearson correlation was equals to 0.66 and RMSD up to 2.46. The correlation and the deviations are supposed to be worse for large differences between the satellite and ground data timescales, because the low time resolution presented by satellite data increases the smoothing effect and may be not enough to describe high frequency variability caused by clouds. A solution for this problem may be to analyze some neighbor pixels around the sites. This larger area of analysis, together with cloud motion information, can provide information on clouds about to cover the site and thus, some predictive information. However, this is subject for future work. Besides that, downscaling methods to synthetize higher frequency cloud cover data with similar characteristics may be an important tool to a more precise characterization of the solar irradiance variability.

The most important result achieved was the development of a simple methodology to evaluate the surface solar irradiance variability using cloud cover obtained from only visible satellite imagery in locations where there is no ground data available. We proved to be possible to get reliable information on solar irradiance variability using the proposed methodology instead of using radiative transfer models that demand larger computational resources and can present large uncertainties.

5 FINAL REMARKS

The focus of this thesis was the development of a methodology for the analysis of the cloud cover variability and, as consequence, of the solar resource variability. The importance of this methodology is the capacity to find hotspots for using solar power technologies, related also with the variability of the resource. The method proposed here uses geostationary satellite imagery and so, is able to analyze the entire Brazilian territory. To do so, first we analyzed the representativeness of the cloud cover fraction obtained with satellite imagery related with ground-based All-Sky camera results, which has better spatial and time resolutions. After that, we compared the satellite cloud cover variability with ground-based global solar irradiance variability, in different Brazilian climate regimes, besides of an analysis of the irradiance variability on the same sites.

This final chapter synthetizes the major findings of the thesis and discusses their importance. Besides that, proposals and recommendations for the future work are presented here to deepen the knowledge on the solar resource variability over Brazil. The chapter is divided in 2 sections, section 5.1 present the major findings of each chapter and section 5.2 discusses the future research needs.

5.1. Major finds

In Chapter 3, we analyzed the confidence of the cloud cover fraction provided by satellite analysis in an area of ~500 km² to represent the cloud cover fraction provided by All-Sky camera. The satellite method presented good agreement with the all-sky camera method, with Pearson R correlation around 92%. The best agreement was found for the clear sky and overcast conditions, with probabilities of detection equals to 91.1% and 87.1%, respectively. The major discrepancy occurred in the broken cloud conditions when the probability of detection was 44.5%.

In addition to the direct comparison between methods, the most important result presented in Chapter 3 was the statistical analysis of the cumulative distribution functions of both datasets. The cumulative distributions showed that the two data sets are very similar, even for different time resolutions. The KS test demonstrated that they can be considered statistically the same, i.e. the cloud

cover variability obtained from satellite images can provide representative information of cloud cover variability observed from the ground in finer time resolutions.

The most important issue related to the satellite method is its low spatial resolution, which has the major impact on the broken-clouds sky scenarios and implies that the cloud masking works better for large clouds with, at least, the size comparable to the satellite image pixel. Despite the low resolution, the satellite method to evaluate the cloud cover may provide reliable information about its variability.

Differently from Chapter 3, Chapter 4 describes a statistical evaluation only in the pixel over the measurement site. The analysis was done using the VS index based on the cumulative distribution functions of the ramp rates of cloud cover and represent its variability.

First, we analyzed the VS of the ground-based global irradiance ramp rates in 3 different Brazilian climate regimes. The results confirmed that the driest periods have lower solar irradiance variability, due to the lower nebulosity. However, this fact is not necessarily valid for different climates. As we found out, Petrolina, the driest place, has the higher variability for smaller time steps, probably due to the high occurrence of small clouds passing in front of the Sun. This fact can be of high concern for the Brazilian energy planning, because the site is located in the best region for power generation using solar technologies (PEREIRA et al., 2017). Anyhow, for 30–minutes time step, the ramp rates variability is very similar in all three sites, what shows the smoothing impact when analyzing larger timescales.

After that, we compared the VS index for the satellite cloud cover variability with the variability of the solar resource in surface. The results presented good relationship, with Pearson R correlation between 0.67 and 0.93, depending on the site, when both datasets have the 30-minutes time resolution (time resolution of the satellite images). However, for lower time steps of the solar irradiance ramp rates the correlation decreased to values smaller than 0.66 in all sites. This decrease was already expected, because of the smoothing effect when analyzing

the 30-minutes time resolution of the satellite data, which ends up masking the variability of the smaller time steps.

The most important result presented in Chapter 3 was the development of a simple methodology to evaluate the surface solar irradiance variability using cloud cover obtained from the satellite images, without the necessity of modeling the solar irradiance at surface. Only the visible geostationary imagery proved itself to be enough to characterize the solar energy variability on surface with sufficient accuracy.

In summary, the main achievement of this thesis was to provide a simple methodology for analyzing the variability of the solar resource using only geostationary satellite data, what can be of great importance for the future energy planning in Brazil and other countries presenting large territories.

5.2. Proposals and Recommendations for Future Research

This study was the first step to increase the comprehension on the influence of typical cloud cover on the surface solar irradiance, in some Brazilian climate zones. It is important to continue advancing the investigation in other climate zones than the ones that have already been analyzed in this research. One important example to be cited would be the Midwest region of Brazil, where a lot of biomass burning events happen every year during the dry season releasing high concentration of aerosols to the atmosphere and affecting the cloud cover determination by satellite images. The Brazilian Northern region is another important area to be investigated due to the presence of the Amazon rainforest, and a very hot and humid tropical climate.

The key limitation of the methodology proposed here is the satellite time resolution of 30 minutes. Even though this timespan provides the required information for grid balancing of distribution systems, it does not provide the required precision necessary for controlling and solving voltage control issues by operators of single or centralized plants. To overcome this issue, downscaling methods to synthetize data with similar characteristics may be an important tool to a more precise characterization of a region.

Additionally, an important fact to be highlighted is the recent launch and activation of the GOES-16 satellite, managed by NOAA. The new satellite provides images in the two visible bands with spatial resolution of 500m and 15–min time resolution. The new satellite data will contribute in the improvement of the methodology, bringing more representativeness and accuracy to the cloud cover coefficient data.

The future research may lead to an improvement of the Brazilian Atlas for Solar Energy (PEREIRA et al., 2017) presenting also the intermittency of the solar energy based on the cloud cover influence. This information will be of great importance for the energy planning in Brazil with higher level of solar energy penetration on the electricity grid.

REFERENCES

ALVARES, C. A. et al. Köppen's climate classification map for Brazil. **Meteorologische Zeitschrift**, v. 22, n. 6, p. 711–728, 2013.

ANTÓN, M. et al. Short-term variability of experimental ultraviolet and total solar irradiance in Southeastern Spain. **Atmospheric Environment**, v. 45, n. 28, p. 4815–4821, 2011.

ARI, G. K.; BAGHZOUZ, Y. **Impact of high PV penetration on voltage regulation in electrical distribution systems**. In: International Conference on Clean Electrical Power (ICCEP), 2011 Disponível em: <http://ieeexplore.ieee.org/document/6036386/>.

ARIAS-CASTRO, E.; KLEISSL, J.; LAVE, M. A Poisson model for anisotropic solar ramp rate correlations. **Solar Energy**, v. 101, p. 192–202, 2014.

CALBÓ, J. Empirical studies of cloud effects on UV radiation: a review. **Reviews of Geophysics**, v. 43, n. 2, p. RG2002, 2005.

CAVALCANTI, I. F. A. et al. **Tempo e clima no Brasil**. São Paulo (SP), Brazil: Oficina de Textos, 2009.

DÜRR, B.; PHILIPONA, R. Automatic cloud amount detection by surface longwave downward radiation measurements. **Journal of Geophysical Research**, v. 109, n. D5, p. 1–9, 2004.

ELSINGA, B.; VAN SARK, W. Spatial power fluctuation correlations in urban rooftop photovoltaic systems. **Progress in Photovoltaics: Research and Applications**, v. 23, n. 10, p. 1390–1397, 2015.

ESCRIG, H. et al. Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast. **Energy**, v. 55, p. 853–859, 2013.

ESPINAR, B. et al. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. **Solar Energy**, v. 83, n. 1, p. 118–125, 2009.

GHONIMA, M. S. et al. A method for cloud detection and opacity classification based on ground based sky imagery. **Atmospheric Measurement Techniques**, v. 5, n. 11, p. 2881–2892, 2012.

HUMMON, M. et al. **Sub-hour solar data for power systems modeling from static spatial variability analysis**. Lisbon, Portugal: [S.n.]. Disponível em: <http://www.osti.gov/bridge>.

INMET. **Normais Climatológicas do Brasil 1981-2010**. Disponível em: <http://www.inmet.gov.br/portal/index.php?r=clima/normaisClimatologicas>.

KAZANTZIDIS, A. et al. Cloud detection and classification with the use of whole-sky ground-based images. **Atmospheric Research**, v. 113, p. 80–88, 2012.

KLEISSL, J. **Solar energy forecasting and resource assessment**. [S.l.]: Elsevier, 2013.

LAVE, M.; BRODERICK, R. J.; RENO, M. J. Solar variability zones: satellite-derived zones that represent high-frequency ground variability. **Solar Energy**, v. 151, p. 119–128, 2017.

LAVE, M.; KLEISSL, J.; ARIAS-CASTRO, E. High-frequency irradiance fluctuations and geographic smoothing. **Solar Energy**, v. 86, n. 8, p. 2190–2199, 2012.

LAVE, M.; RENO, M. J.; BRODERICK, R. J. Characterizing local high-frequency solar variability and its impact to distribution studies. **Solar Energy**, v. 118, p. 327–337, 2015.

LEWIS, N. S.; NOCERA, D. G. Powering the planet: chemical challenges in solar energy utilization. **Proceedings of the National Academy of Sciences**, v. 103, n. 43, p. 15729–15735, 2006.

LIANG, X.; YUAN, C. Derivation of 3D cloud animation from geostationary satellite images. **Multimedia Tools and Applications**, v. 75, n. 14, p. 8217–8237, 2016.

LIOU, K. N. **An introduction to atmospheric radiation**. 2. ed. Amsterdam: Academic Press, 2002

LONG, C. N. et al. Retrieving cloud characteristics from ground-based daytime color all-sky images. **Journal of Atmospheric and Oceanic Technology**, v. 23, n. 5, p. 633–652, 2006.

LUIZ, E. W. et al. Comparison of methodologies for cloud cover estimation in Brazil - A case study. **Energy for Sustainable Development**, v. 43, p. 15–22, 2018.

MALEK, E. Evaluation of effective atmospheric emissivity and parameterization of cloud at local scale. **Atmospheric Research**, v. 45, p. 41–54, 1997.

MARTÍNEZ-CHICO, M.; BATLLES, F. J.; BOSCH, J. L. Cloud classification in a mediterranean location using radiation data and sky images. **Energy**, v. 36, n. 7, p. 4055–4062, 2011.

MARTINS, F. R. et al. The influence of cloud cover index on the accuracy of solar irradiance model estimates. **Meteorology and Atmospheric Physics**, v. 99, n. 3–4, p. 169–180, 2008.

MARTY, C. A.; PHILIPONA, R. The clear-sky index to seperate clear-sky from cloud-sky situations in climate research. **Geophysical Research Letters**, v. 27, n. 17, p. 2649–2652, 2000.

MASSEY, F. J. The Kolmogorov-Smirnov test for goodness of fit. **Journal of the American Statistical Association**, v. 46, n. 253, p. 68, 1951.

NETO, S. L. M. et al. The use of Euclidean geometric distance on RGB color space for the classification of sky and cloud patterns. **Journal of Atmospheric and Oceanic Technology**, v. 27, n. 9, p. 1504–1517, 2010.

NGOKO, B. O.; SUGIHARA, H.; FUNAKI, T. Synthetic generation of high temporal resolution solar radiation data using Markov models. **Solar Energy**, v. 103, p. 160–170, 2014.

NUNES, L. H.; VICENTE, A. K.; CANDIDO, D. H. Clima da região sudeste do brasil. In: CAVALCANTI, I. F. A. et al. (Eds.). **Tempo e clima no brasil**. Oficina de Textos, 2009. p. 243–258.

PEREIRA, E. B. et al. **Atlas brasileiro de energia solar**. 2. ed. São José dos Campos (SP), Brazil: INPE, 2017.

PEREZ, R. et al. Making full use of the clearness index for parameterizing hourly insolation conditions. **Solar Energy**, v. 45, n. 2, p. 111–114, 1990.

PEREZ, R. et al. Short-term irradiance variability: preliminary estimation of station pair correlation as a function of distance. **Solar Energy**, v. 86, n. 8, p. 2170–2176, 2012.

PEREZ, R. et al. Spatial and temporal variability of solar energy. **Foundations and Trends® in Renewable Energy**, v. 1, n. 1, p. 1–44, 2016.

REUTER, M. et al. The CM-SAF and FUB cloud detection schemes for SEVIRI:

validation with synoptic data and initial comparison with MODIS and CALIPSO. **Journal of Applied Meteorology and Climatology**, v. 48, n. 2, p. 301–316, 2009.

ROBAA, S. Evaluation of sunshine duration from cloud data in Egypt. **Energy**, v. 33, n. 5, p. 785–795, 2008.

STEIN, J. S. et al. **Estimating annual synchronized 1-min power output profiles from utility-scale PV plants at 10 Locations in Nevada for a solar grid integration study**. USA: NNSA, 2011.

STRANG, D. M. G. D. **Análise climatológica das normais pluviométricas do Nordeste brasileiro**. São José dos Campos (SP), Brazil: CTA/IAE, 1972.

TZOUMANIKAS, P. et al. The effect of clouds on surface solar irradiance, based on data from an all-sky imaging system. **Renewable Energy**, v. 95, p. 314–322, 2016.

WACKER, S. et al. Cloud observations in Switzerland using hemispherical sky cameras. **Journal of Geophysical Research: Atmospheres**, v. 120, n. 2, p. 695–707, 2015.

WATANABE, T.; OISHI, Y.; NAKAJIMA, T. Y. Characterization of surface solar-irradiance variability using cloud properties based on satellite observations. **Solar Energy**, v. 140, p. 83–92, 2016.

WEGENER, J. et al. **Temporal downscaling of irradiance data via Hidden Markov models on wavelet coefficients: application to california solar intiative data**. San Diego: University of California, 2012.

WERKMEISTER, A. et al. Comparing satellite- to ground-based automated and manual cloud coverage observations - a case study. **Atmospheric Measurement Techniques**, v. 8, n. 5, p. 2001–2015, 2015.

YANG, D. et al. Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage. **Solar Energy**, v. 103, p. 550–562, 2014.