Renewable Energy 108 (2017) 569–580

Contents lists available at ScienceDirect

Renewable Energy

journal homepage: www.elsevier.com/locate/renene

Assessment of solar radiation components in Brazil using the BRL model

Leonardo F.L. Lemos [a], Allan R. Starke [a, *], John Boland [b], José M. Cardemil [c], Rubinei D. Machado [a, d], Sergio Colle [a]

[a] LEPTEN-Laboratory of Energy Conversion Engineering and Energy Technology, Department of Mechanical Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil
[b] Centre for Industrial and Applied Mathematics, University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA, 5095, Australia
[c] Mechanical Engineering Department, Universidad de Chile, Beauchef 851, Santiago, Chile
[d] Department of Atmospheric Science, University of Sao Paulo (USP), São Paulo, Brazil

ARTICLE INFO

Article history:
Received 3 November 2016
Received in revised form
18 February 2017
Accepted 24 February 2017
Available online 27 February 2017

Keywords:
Solar radiation
BRL model
Diffuse solar irradiation
Direct normal solar irradiation
Brazil

ABSTRACT

Quality data regarding direct and diffuse solar irradiation is crucial for the proper design and simulation of solar systems. This information, however, is not available for the entire Brazilian territory. However, hourly measurements of global irradiation for more than seven hundred stations over the territory are available. Several mathematical models have been developed over the past few decades aiming to deliver estimations of solar irradiation components when only measurement of global irradiation is available. In order to provide reliable estimates of diffuse and direct radiation in Brazil, the recently presented Boland–Ridley–Laurent (BRL) model is adjusted to the particular features of Brazilian climate data, developing adjusted BRL models on minute and hourly bases. The model is adjusted using global, diffuse and direct solar irradiation measurements at nine stations, which are maintained by INPE in the frame of the SONDA project. The methodology for processing and analyzing the quality of the data-sets and the procedures to build the adjusted BRL model is thoroughly described. The error indicators show that the adjusted BRL model performs better or similarly to the original one, for both diffuse and DNI estimates calculated for each analyzed Brazilian station. For instance, the original BRL model diffuse fraction estimates have MeAPE errors ranging from 16% to 51%, while the adjusted BRL model gives errors from 9% to 26%. Regarding the comparison between the minute and hourly adjusted models, it can be concluded that both performed similarly, indicating that the logistic behavior of the original BRL model is well suited to make estimates in sub-hourly data sets. Based on the results, the proposed adjusted model can be used to provide reliable estimates of the distribution of direct and diffuse irradiation, and therefore, can help to properly design and reduce the risks associated to solar energy systems.

1. Introduction

The design and simulation of solar energy systems require detailed information about incoming solar radiation at the site in question. However, in many locations, only information about global solar radiation is available resulting in a lack of data for direct normal irradiation (DNI), i.e., the non-scattered and reflected radiation that reaches a surface, which is crucial for the analysis of concentrating systems among other technologies. DNI measurements are expensive since they require complex tracking devices and significant operational efforts. As a result, the data is not always available, and in the cases when it is, the measurements are commonly incomplete due to equipment malfunctions or poor maintenance. One possible approach to overcome the lack of reliable information is to use models to assess DNI. Such models were initially developed in the 1960s [1] and have been improved to fit radiation data at different locations worldwide. Ridley et al. [2] proposed a model that uses global irradiation data to estimate diffuse irradiation and DNI based on a set of predictors, which are based on meteorological information commonly available from weather stations.

Regarding the solar resource information in Brazil, there are three main sources: the database derived from the Solar and Wind Resource Assessment (SWERA) project [3], the Brazilian

* Corresponding author.
E-mail address: allan.starke@lepten.ufsc.br (A.R. Starke).

http://dx.doi.org/10.1016/j.renene.2017.02.077
0960-1481/

Environmental Data Organization System (SONDA) stations [4], and the radiation measurement network from the National Institute of Meteorology (INMET) [5]. The SWERA project, funded by the United Nations Environment Programme (UNEP) and the Global Environmental Facility (GEF), is devoted to building up a reliable database of solar and wind energy resources, aiming to foster the insertion of renewables in the energy matrix of developing countries. The data obtained by the SWERA project is derived by a satellite model and compiled into a variety of useful geographic and socioeconomic information in geographic information system (GIS) format, comprising maps of monthly averages for global and direct solar radiation, seasonal and annual averages for global, diffuse, and direct normal irradiation, and the compilation of 20 Typical Meteorological Years (TMYs) for selected locations [6,7]. The satellite model was validated using data from the SONDA station network, an initiative of the National Institute for Space Research (INPE) consisting of several meteorological monitoring stations deployed along the Brazilian territory. The stations in the network measure global, diffuse, and direct normal irradiation on a minute basis. The spatial distribution of the stations comprising the SONDA network is shown in Fig. 1.

An extensive radiation measurement network in Brazil is operated by the National Institute of Meteorology (INMET), whose spatial distribution is also depicted in Fig. 1. This is the largest network of meteorological stations in Latin America, comprising approximately seven hundred stations distributed all over the country. The stations gauge hourly global solar irradiation and ambient variables but do not measure diffuse and direct components. Despite the high volume of available data, a lack of information still exists regarding the characteristics of solar resource in Brazil and the variability of direct solar radiation. This fact induces high uncertainties for project developers, discouraging the deployment of Concentrating Solar Power (CSP) and Concentrating Photovoltaics (CPV).

Based on this scenario, the present work aims to develop a diffuse fraction model adapted for Brazilian local conditions, using irradiation data from stations of the SONDA network. Thus, this national model can be further used to estimate the hourly diffuse and direct irradiation at each station of the INMET network, collecting useful information that can help to reduce the risks of deploying solar energy systems and improving the design process. Building a model adapted to local conditions is necessary in order to provide reliable estimates of the distribution of direct and diffuse irradiation within the Brazilian territory.

To accomplish this objective, the Boland–Ridley–Laurent (BRL) model [2] is employed. First, the original BRL coefficients are used to estimate the diffuse fraction for selected locations in Brazil, and then compared with available diffuse radiation measurements by the SONDA stations. Thereafter, a new set of coefficients is calculated by considering data sets from several stations, resulting in a national model for Brazil. In order to verify the performance of the model over the Brazilian territory, the national adjusted BRL model is further applied to each station separately, and then compared with the results of the original BRL model using formal error analysis.

The BRL model was originally developed to estimate the diffuse fraction on an hourly basis. On the other hand, the SONDA network database consists of minute measurements of solar irradiation, which make it possible to develop adjusted BRL models for minute and hourly basis. Therefore, it is possible to perform a detailed analysis of the performance of the BRL model when used to estimate the diffuse fraction on a minute basis.

This study presents a description of the implementation process and adaptation of the BRL model with multiple predictors for estimating the hourly diffuse fraction, and is organized as follows. The methodology for processing and analyzing the quality of the

Fig. 1. Spatial distribution of the SONDA and INMET station networks over the Brazilian territory.

data-sets and the procedures to build the adjusted BRL model is thoroughly described. The results and performance of the minute and hourly adjusted model, including the comparison against the original BRL model, are assessed. Both adjusted and original BRL model are compared when applied to estimate diffuse irradiation at selected locations in Brazil. Finally, a detailed discussion of the results of this study is presented in the light of the results reported in the related literature.

2. Diffuse radiation model

Because of the passage of solar irradiation through the atmosphere, it is separated into different components. The beam component of solar irradiation is the part that directly reaches the surface of Earth, without suffering any change in direction, and is usually referred to as direct normal irradiation (DNI). The scattering of the solar irradiation in the atmosphere generates the diffuse component, which has no defined direction and arrives from the entire sky dome on a horizontal surface. If fewer sources of disturbance are present in the atmosphere, lower scattering effects are expected, which result in higher DNI values.

One measure of the presence of disturbances in the atmosphere is the *clearness index* (k_T), defined as the ratio of the global irradiation on a horizontal surface (I_g) to the extraterrestrial irradiation at the top of the atmosphere (I_0), also on a horizontal surface. By plotting the diffuse fraction (d) – defined as the ratio of the diffuse irradiation on a horizontal surface (I_d) to the I_g – against k_T, for a given location, a correlation can be observed in an S-shaped curve. Several models using the clearness index as a predictor to estimate the diffuse fraction have been developed since the 1960s [8–12]. However, by using only one predictor, these models result in a curve that only fits the radiation data and is not able to properly model the spread of diffuse fraction values.

By including other input variables in the model, the spread of data can be modeled. Reindl et al. [11] studied the effects of other meteorological variables in predicting the distribution of the diffuse fraction. From a starting set of twenty eight possible variables, the authors concluded that four of these had a significant effect on diffuse radiation and should be used as predictors in a correlation; in particular, the clearness index itself, the solar altitude angle, the ambient temperature, and the relative humidity have a significant effect. Using data from two meteorological stations in the United States and three in Europe, the authors proposed a piece-wise linear correlation to estimate the diffuse fraction using hourly radiation and environment data. Other authors [11,13–15] have proposed similar multi-variable correlations to estimate both hourly and daily diffuse radiation fractions. These models consist of either linear piece-wise or simple nonlinear functions and have been used in software tools, e.g., TRNSYS [16], when complete diffuse radiation data is not available.

One drawback of these correlations, however, is the fact that they have been developed using locations in Europe and North America; hence, they may not function properly in tropical regions and/or within the southern hemisphere. Moreover, the use of piece-wise correlations, such as [11,15], give rise to discontinuities in the predicted diffuse fraction. Additionally, values of k_T delimiting the clearness intervals on these models vary according to each correlation. Adjusting such models to specific climate conditions requires that the boundary points be determined along with the coefficients that multiply the predictors [17].

An improvement of these piece-wise correlations was proposed by Ridley et al. [2] by using a continuous logistic function model. This correlation intends to use a minimum of measured variables; therefore, ambient temperature and humidity were not used as model predictors. Instead, the authors used the clearness index, solar altitude, and three additional variables, which can easily be calculated by applying solar geometry algorithms and measurements of global irradiation. The diffuse fraction is calculated by the BRL model as follows,

$$\widehat{d} = \frac{1}{1 + e^{(\beta_0+\beta_1 k_T+\beta_2 AST+\beta_3\alpha+\beta_4 K_T+\beta_5\psi)}} \tag{1}$$

$$= \frac{1}{1 + e^{(-5.38+6.63k_T+0.006AST-0.007\alpha+1.75K_T+1.31\psi)}},$$

where k_T is the hourly clearness index, AST is the apparent solar time, α is the solar altitude, K_T is the daily clearness index, and ψ is a persistence factor defined in Ref. [2]. The numeric values of the coefficients $(\beta_1, \beta_2, \ldots, \beta_5)$ listed in Equation (1) were also derived in Ref. [2] using data from three meteorological stations in the southern hemisphere and four in the northern hemisphere. The authors showed that the BRL model performs significantly better in the southern hemisphere than other models, while presenting equivalent errors for sites in the northern hemisphere. In addition, Ridley et al. considered the Bayesian Information Criterion (BIC) to demonstrate that the BRL model presents a better trade-off between performance (errors) and complexity (number of predictors), performing as well as, or better than, other models while using fewer predictors.

Considering the estimated diffuse fraction and the measured global irradiation, the direct normal irradiation is estimated by the following expression,

$$\widehat{I}_{bn} = \frac{\left(1 - \widehat{d}\right) I_g}{\sin\alpha}. \tag{2}$$

Based on the aforementioned features, the BRL model is chosen for the present study because of its simplicity, good performance and the least number of variables in Equation (1) to be considered from existing data. The BRL model is also very flexible since it can easily be adjusted to new locations by recalculating the β_i coefficients, while other widely used models, for instance the Perez model [14], cannot accomplish this adjustment. Moreover, the BRL model presents significantly smaller errors for locations in the southern hemisphere when compared to piece-wise models [2], which impact the proper design and evaluation of solar energy systems.

It is worth mentioning that this paper also proposes verification of the usefulness of the BRL model for diffuse fraction estimation considering minute data. In this context, the same methodology is applied by considering minute-based predictors, rather than hourly ones, where k_T, AST, α and ψ are calculated on a minute basis and K_T is kept as daily clearness index.

The diffuse radiation can also be calculated using theoretical detailed models such as UniSky [18–20], using atmospheric radiative transfer models such as MODTRAN [21] and SMARTS2 [22], or using satellite based models such as the developed by Perez et al. [14]. However, the use of these models in Brazil is troublesome because of the lack of readily available values of turbidity, aerosol optical depth (AOD), cloud covers, among other variables commonly used as inputs in those models. In that sense, using simpler models such as the original BRL is of special interest since they can provide reliable estimates by using a set of easily accessible input variables.

3. Methodology

The following section thoroughly describes the methodology used in the study, presenting the characteristics of the irradiation database, the filtering and quality control and the procedures to build the adjusted BRL model.

3.1. Irradiation data

The SONDA network consists of seventeen stations that measure solar irradiation, which are divided in three categories, as reference stations, advanced stations and basic stations. Detailed information about the instruments and methods are presented by Refs. [4,6,23]. However, only nine stations are considered in this study (depicted by the index number in Fig. 1). Some stations were disregarded because they stayed online for less than a year, while other stations were not considered because only a small amount of data remained after the quality-control procedures, compromising the proper development of the Brazilian BRL model. Table 1 summarizes the station information considered in the present work, while Fig. 2 depicts the monthly data availability for each station from January of 2004 to November of 2016 as a heat-map representation. The availability index is calculated as a ratio between measured minutes and respective total minutes of the month.

3.2. Quality control and outlier removal

In order to guarantee the quality of the database, a validation procedure was performed on the meteorological data from the SONDA stations. First, the minute data from each station was organized in a time-series structure, where the missing data was replaced by a non-numeric value, setting up an organized and continuous data set. Then, for each minute, the solar altitude angle α, apparent solar time *AST*, extraterrestrial radiation I_0, and persistence factor ψ were calculated. For an initial assessment, the data was integrated to allow the determination of irradiation values and clearness indexes on an hourly and daily basis. With these results, a set of filters were applied to the resulting data; in particular, data were considered with $\alpha > 7^\circ$, $0 < \psi < 1$, minute, hourly, and daily clearness index between zero and one $\left(0 < k_{T,minute}, k_{T,hourly}, K_t < 1\right)$, and finally, diffuse fractions of $0 < d < 1$.

Measurements taken at low solar altitudes are prone to high uncertainties, but are of minor importance in design applications; therefore, they were not considered in this study. A solar altitude threshold of 7° was adopted from the literature [24]. The other conditions exclude theoretically impossible values of the diffuse fraction, persistence, and clearness indexes.

After applying the filtering process described above, the data set was submitted to plausibility and consistency checks. These tests were adapted from the criteria defined by the Baseline Surface Radiation Network (BSRN) [25]. These tests compare the measured global, diffuse, and beam irradiation with extraterrestrial irradiation and also with each other to check consistency in all measured data, as depicted in Table 2.

A series of tests were then carried out to detect equipment malfunctions and further remove suspicious irradiation values. The procedure used to detect tracker malfunctions or misalignment of shading rings was the same used in the QCRad quality control routine [25]. Following recommendations from the World Meteorological Organization [26], the extreme variations on consecutive measured values of global irradiation were removed (when the minute I_g varies more than 800 W/m^2 between successive time-steps).

Geiger et al. [27] proposed that daily I_g values considered for diffuse fraction estimation should not be lower than 3% of the daily extraterrestrial irradiation. However, this approach was developed for an hourly resolution, and when applied to sub-hourly data a lot of data can be rejected. Since minute data is used in this work, an approach valid for a sub-hourly resolution was applied, such as the methodology presented by Journée and Bertrand [28], which adapted a lower I_g bound for sub-hourly data sets by stating that the clearness index should satisfy the criterion $I_g/I_0 \geq 10^{-4}(\alpha - 10)$ for $\alpha > 10^\circ$ and that the average of the clearness index within a day, $\mu(I_g/I_0)$, should not be smaller than 0.03.

Table 1
Geographical information of the SONDA stations considered in the present study.

Index	Abbreviation	Location	Lat. (°)	Long. (°)	Height (m)
1	BRB	Brasília	−15.60	−47.71	574
2	CGR	Campo Grande	−20.44	−54.54	677
3	CBA	Cuiabá	−15.56	−56.07	185
4	FLN	Florianópolis	−27.60	−48.52	31
5	ORN	Ourinhos	−22.95	−49.89	446
6	PMA	Palmas	−10.18	−48.36	216
7	PTR	Petrolina	−9.07	−40.32	387
8	SLZ	São Luís	−2.59	−44.21	40
9	SMS	S. Martinho da S.	−29.44	−53.82	489

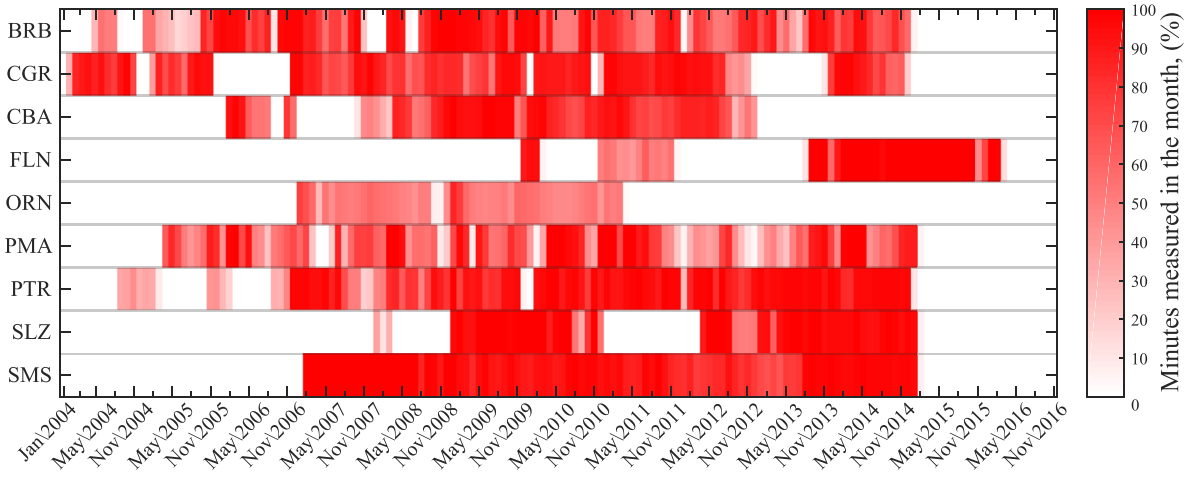
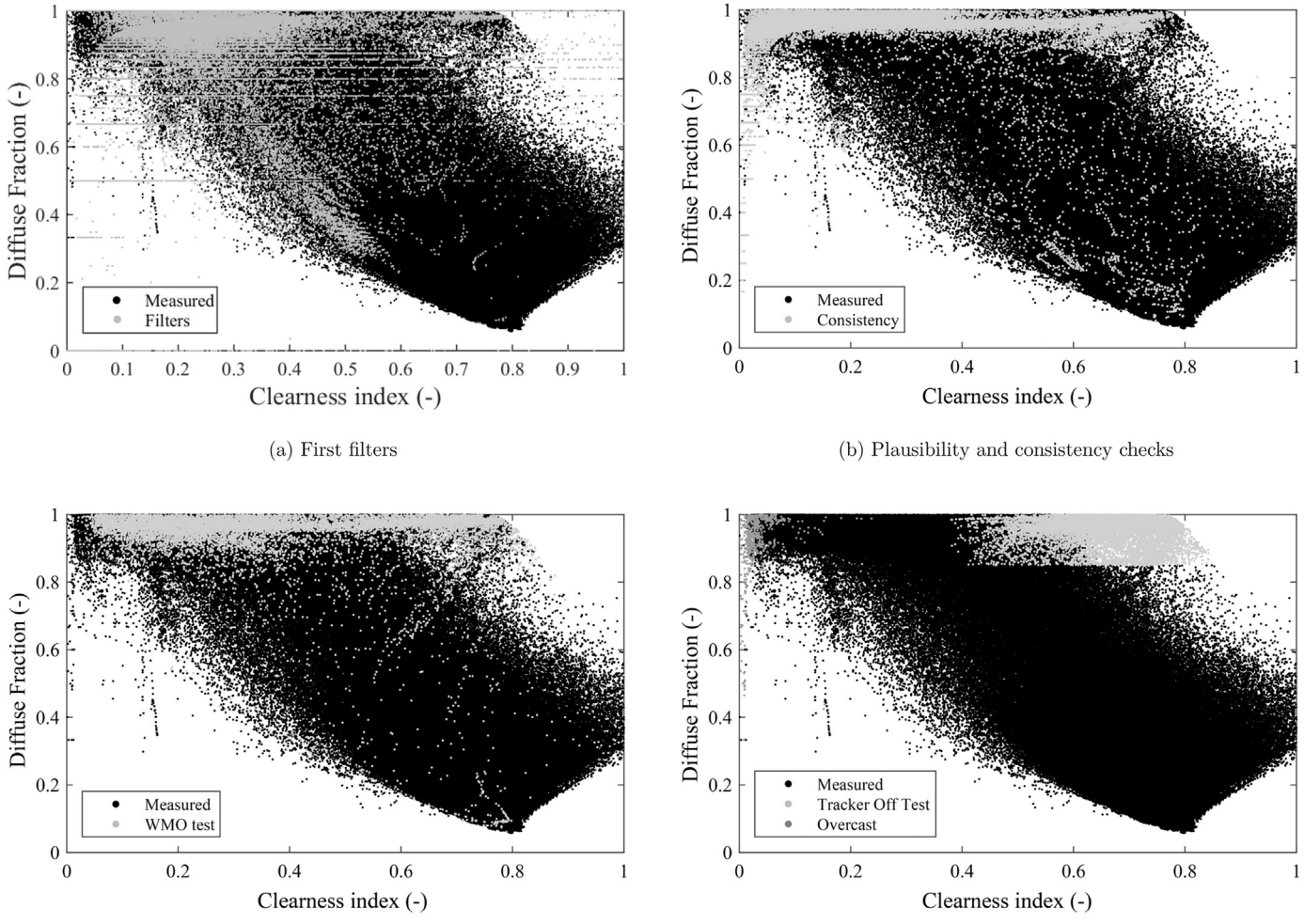


Fig. 2. Heat-map of monthly data availability of the stations – fraction of the measured minutes in a month.

Table 2
Quality tests applied to raw data.

Tests	Description	Criteria
First filters	Solar altitude limit.	$\alpha < 7^\circ$
	Persistence factor plausibility.	$0 < \psi < 1$
	Minute and hourly clearness index limits.	$0 < k_{T,minute}, k_{T,hourly} < 1$
	Daily clearness index limit.	$0 < K_T < 1$
	Diffuse fraction limit.	$0 < d < 1$
Plausibility checks	Physical plausible limits for irradiation adapted	$0 < I_g < 1.5G(\sin\alpha)^{1.2} + 100$
	From Ref. [25]. G is the solar constant adjusted to	$0 < I_d < 0.95G(\sin\alpha)^{1.2} + 50$
	Earth-Sun distance. Irradiation values in W/m^2.	$0 < I_{bn} < G$
Consistency checks	Evaluates if irradiation measurements	$\left\lvert\frac{I_g-(I_d+I_{bn}\sin\alpha)}{I_g}\right\rvert < 0.08$, if $\alpha > 15^\circ$ and $I_d + I_{bn}\sin\alpha > 50$
	are mutually consistent [25].	$\left\lvert\frac{I_g-(I_d+I_{bn}\sin\alpha)}{I_g}\right\rvert < 0.15$, if $\alpha < 15^\circ$ and $I_d + I_{bn}\sin\alpha > 50$
	Irradiation values in W/m^2.	If $I_d + I_{bn}\sin\alpha > 50$, test not possible
Tracker off test	Checks for equipment malfunction [25].	If $\frac{I_g}{I_{gc}} > 0.85$ and $\frac{I_d}{I_g} > 0.85$, measurement rejected
Data variability test	Check for excessive variations between successive time steps [26]. Irradiation values in W/m^2.	$\lvert I_{g,i} - I_{g,i-1}\rvert < 800$ and $\lvert I_{g,i+1} - I_{g,i}\rvert < 800$
Overcast condition test	Lower bound on I_g [28].	$\frac{I_g}{I_0} \geq 10^{-4}(\alpha - 10)$, if $\alpha > 10$
		$\mu\left(\frac{I_g}{I_0}\right) \geq 0.03$, average taken on the whole day
Clear sky Comparison	Removal of measured I_g values larger than clear sky conditions. Solis clear sky model used [32].	$\frac{I_g}{I_{gc}} < 1.1$
Rayleigh limit	Comparison of diffuse irradiation with Rayleigh lower limit [25].	If $I_g > 50$ and $d < 0.8$ and $I_d < R_L - 1$, measurement rejected $R_L = 209.3\sin\alpha - 708.3(\sin\alpha)^2 + 1128.7(\sin\alpha)^3 - 911.2(\sin\alpha)^4 + 287.87(\sin\alpha)^5 + 0.046725(\sin\alpha)P$



(a) First filters

(b) Plausibility and consistency checks

(c) WMO variability check

(d) Tracker malfunction and overcast checks

Fig. 3. Test and filtering data process for the Florianópolis station. Approved data points are shown in gray.

The filtering and test procedures described so far are illustrated in Fig. 3 for Florianópolis. The points removed by each step of the quality control procedure are show in gray.

An additional test was applied to compare global irradiation measurements with clear-sky numeric values. This approach allows for the removal of irradiation values caused by cloud enhancement phenomena, which happens when beam radiation reflected by clouds is detected as diffuse irradiation [29], leading to an upward trend of diffuse fraction values for high k_T values. In sub-hourly data sets, this phenomenon is more frequently detected; however, in hourly data sets, this enhanced diffuse irradiation is not recurrent since the data is averaged; in fact, the models developed for diffuse fraction estimation were made for hourly data and do not consider the cloud enhancement phenomenon, which is the case of the original BRL model.

Another limitation that arises when hourly models are used to make estimations with sub-hourly data is that minute-based data presents higher variability resulting in a larger data spread on the $d - k_T$ chart, which induces larger random errors in the model estimations [30]. It is then necessary to evaluate whether the functional form of the model can capture the greater variability of the minute data. Therefore, cloud enhancement events were removed from the data sets by comparing the global irradiation measurements with clear sky values, as suggested by Engerer [31]. This results in a higher variability data set, without enhanced irradiation values, allowing for the evaluation of the model's ability to simulate greater data variability.

The clear-sky conditions were assessed by adopting the Solis model [32], where the necessary inputs, aerosol optical depth, and atmospheric water-vapor content were taken from the MACC-II project database [33]. Global irradiation measured data was compared with the modeled clear-sky global irradiation, and only data points that satisfied the condition $I_g/I_{gC} < 1.1$ were accepted. The Solis model presents high uncertainties when predicting the diffuse and beam-normal irradiation using inputs from the MACC-II project [34]. Therefore, no comparisons were made between measured and clear-sky values of diffuse and beam irradiation. Instead, to identify a lower bound of diffuse irradiation values, the Rayleigh diffuse limit was applied using the method described in Ref. [25]. This limit is the smallest amount of diffuse irradiation that can reach the surface in ideal clear-sky conditions. Fig. 4 depicts data removal by the clear-sky and Rayleigh tests showing that the clear-sky global irradiation criteria remove a large amount of data, including almost all diffuse fractions, for clear indexes higher than 0.82.

The quality tests stated so far are summarized in Table 2, which gives a brief description and the criteria used to qualify the raw data.

The test previously applied removed several inconsistent measurements; however, the data may still contain inconsistencies, which can be classified as outliers that affect the process of computing the new set of BRL coefficients. These outlier values are difficult to remove by filtering processes. Boland and Ridley [12] used an empirical likelihood to remove outliers. However, this method cannot be applied in the present work because of the large amount of data considered, which leads to an optimization problem with millions of independent variables. Therefore, an outlier identification procedure, similar to that proposed by Younes et al. [24], was applied to address this issue, creating an envelope around plausible data. Younes et al. developed these envelopes by using polynomial functions; however, Boland and Ridley [12] proved that the best curve to model the diffuse fraction versus clearness index is a logistic function with the form $f = 1/(1 + exp(\beta_0 + \beta_0 k_T))$. Therefore, this logistic function was adopted to create the outlier envelope.

The outlier removal procedure was divided into two steps. First, the suspicious data was identified using the value of the absolute residual for the i-th observation as a figure of merit, measuring the vertical distance between the measurements and a fitted curve. Second, the envelope's upper and lower bounds were built. The first step has a significant importance since it removes unlikely data and helps identify data that is not physically plausible. It is important to mention that the first step does not remove any data from the pool; it only disregards data for building the envelope.

In order to calculate the residuals, a logistic function was fitted to the data set that passed the filtering and testing steps. The residual was calculated as the absolute difference between the measured diffuse fraction and the fitted curve, specifically, $R_i = |d_i - f(k_{T,i})|$. Then, a new logistic curve was fitted, $g(k_T)$, disregarding any point where $R_i > 0.3$. This conservative approach helps the logistic curve $g(k_T)$ capture the expected behavior of the data, as observed in Fig. 5. This new resulting curve was used as reference to establish the upper and lower bounds for outlier identification, which were defined by the translation of $g(k_T)$ on the $d - k_T$ chart, as described by the following expressions:

$$UpperBound = g(k_T - a_1) + a_2 \tag{3}$$

$$LowerBound = g(k_T + b_1) - b_2. \tag{4}$$

The coefficients $a_1, a_2, b_1,$ and b_2 were defined for each station to efficiently capture the data shape and remove measurements

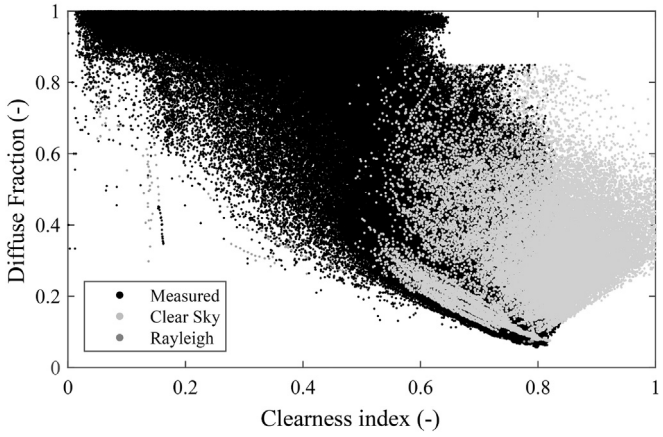


Fig. 4. Clear-Sky and Rayleigh test procedures for the Florianópolis station.

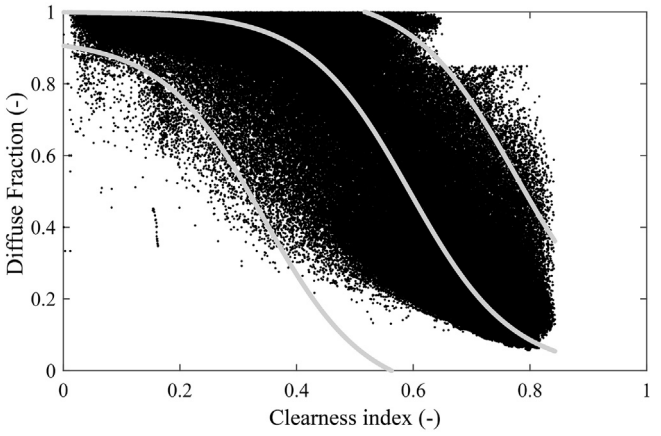


Fig. 5. Outlier envelop for the Florianopólis station.

that were not removed by the filtering process but were visibly suspect. Fig. 5 depicts the logistic curve $g(k_T)$ derived for Florianópolis and the upper and lower bounds used for outlier removal.

To clarify and summarize the quality control and outlier removal routine, Fig. 6 depicts a flowchart of the procedures applied herein. All qualifying processes shown in the flowchart are described in Table 2.

3.3. Hourly averaging of data

The processed minute data was then averaged to obtain hourly values of irradiation. Averaging was only performed when at least 45 min within the hour were valid. In the case where not enough minutes were valid, no averaging was performed, and the hour data point was not considered for further calculations. The approved minutes within this hour were also not considered for the upcoming calculations. This approach aims to avoid too many minute and hourly points being discarded.

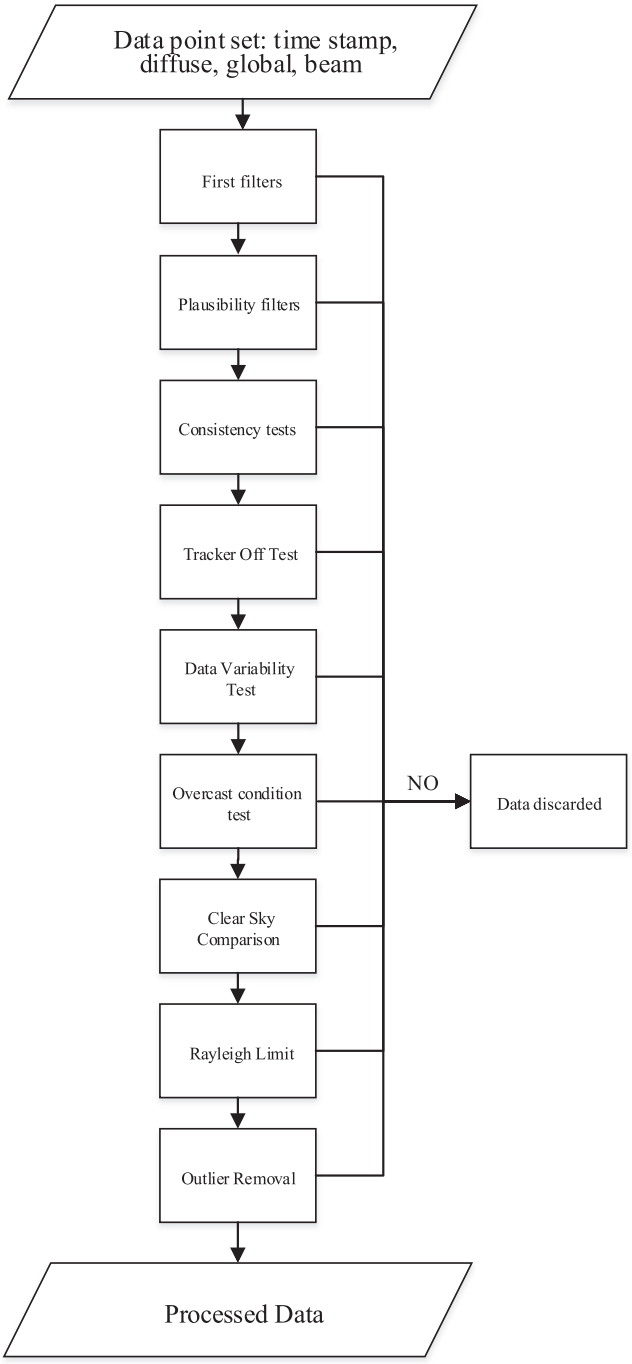


Fig. 6. Flowchart of the quality control and outlier routine.

3.4. Data merge and model fit

After the qualification and outlier removal procedure, the data from individual stations were merged to a single data set. This procedure was done for the minute and hourly sets. However, since each station is operated for a particular period, the amount of measurement points varies from location to location, as noted in Fig. 2. These differences in the size of the data set do not facilitate the simple aggregation into one single data set since stations presenting more points induce a location bias in the regression. To solve this problem, the same amount of data was considered from each station and then subjected to the amalgamation process. Regarding the number of data points, the amount of data of the station that presents the lowest amount of qualified data was chosen; in this case, the station was São Luís - MA. For the other stations, this same amount of data was randomly pulled from their respective data sets.

Minute and hourly merged data sets were individually submitted to a regression in order to adjust the BRL model (Equation (1)). Two thirds of the data points were randomly taken from the set and used for nonlinear least-square regression, which determined the new set of β_i coefficients for the national adjusted correlation. The resulting correlation was then applied to the remaining third of the data to validate the model and calculate its error indicators.

As a formal error analysis, three statistical indicators were considered; in particular, the median absolute percentage error (*MeAPE*), the normalized root mean square error (*NRMSE*), and the Kolmogorov-Smirnov test Integral parameter (*KSI*) were used according to the approach adopted in Refs. [2] and [17]. The *MeAPE*, *NRMSE*, and *KSI* are defined by the following expressions,

$$MeAPE = median\left(\frac{|\widehat{d}_i - d_i|}{d_i} \times 100\right) \tag{5}$$

$$NRMSE = \frac{1}{\overline{d}}\sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{d}_i - d_i\right)^2}{n}} \tag{6}$$

$$KSI = \int_{x_{min}}^{x_{max}} D_n dx, \tag{7}$$

where $d_i = (I_d)_i/(I_g)_i$ is the actual value of the diffuse fraction calculated from measurement i of a set of n measurements, $\widehat{d}_i$ is the estimated value of the diffuse fraction of data point i calculated using the model, and $\overline{d}$ is the mean experimental diffuse fraction of the data set. Finally, x_{max} and x_{min} are the extreme values of the independent variable, while D_n is the difference between the cumulative distribution function (CDF) of the measurements and the estimated values.

The *MeAPE* indicates the size of the errors presented by the correlation, while the *NRMSE* provides a measure of the overall goodness of fit. Finally, the *KSI* was proposed in Ref. [35] as a measure of the similarity of the cumulative distribution of the actual and modeled data, which helps to identify if the model fits

the distribution of the data over the whole range of observed values.

4. Results

A national BRL model was developed for the minute and hourly merged data, and then a graphical comparison between the fit and measured data was constructed to observe the goodness of fit. This process permitted us to observe if the adjusted BRL model can capture the spread and variation present in the minute data set.

The national adjusted model was applied to make estimations for each station individually, while the error indicators were calculated. The original BRL model was also applied to individual stations, which permitted the qualitative evaluation of error indicators offered by the adjusted model, when compared to the original. This procedure also allowed us to assess the correlation when applying the national model for each station separately. Moreover, it demonstrated the improvements achieved by building

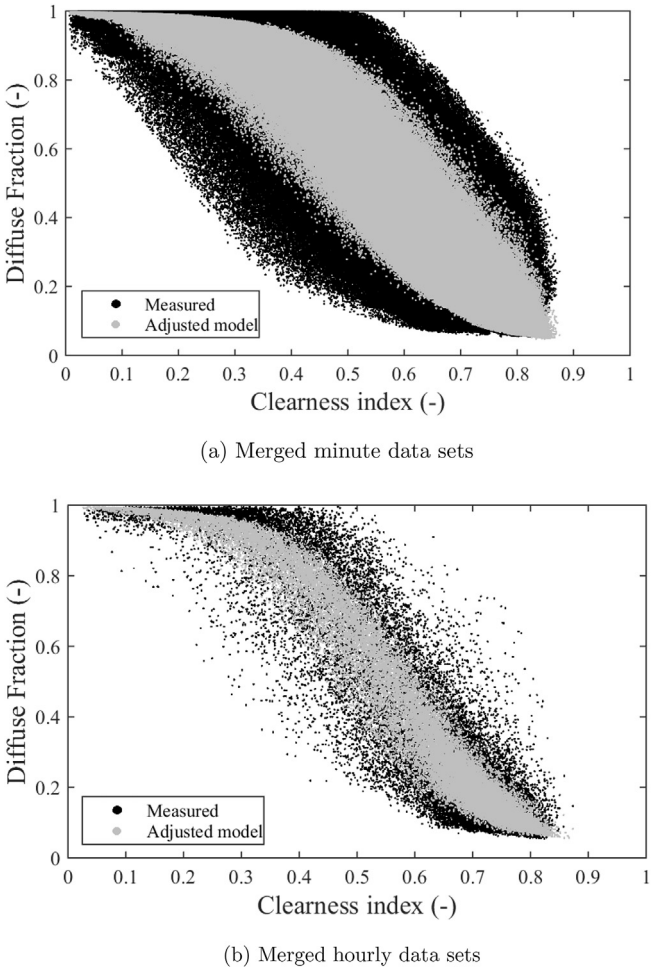


Fig. 7. Adjusted models for minute and hourly merged data sets.

Table 4
Coefficients of the adjusted models. BRL model shown for reference.

	Constant	k_T	AST	α	K_T	ψ
Minute	−6.26	5.97	0.024	−0.00533	2.84	2.41
Hourly	−4.41	7.87	−0.088	−0.00490	1.47	1.10
BRL	−5.38	6.36	0.006	−0.007	1.75	1.31

Table 5
Error measures for diffuse fraction estimates with exchanged coefficients for Florianópolis.

	Minute coefficients on hourly data	Hourly coefficients on minute data
MeAPE (%)	14.860	16.366
nRMSE (%)	21.592	25.416
KSI(−)	0.019	0.051

Table 6
Hourly model results for analyzed stations.

Location	Statistic	Original BRL model		Adjusted BRL model	
		d	DNI	*d*	DNI
All Locations	MeAPE (%)	33.949	17.860	17.530	10.674
	nRMSE (%)	34.856	22.383	25.764	16.078
	KSI(−)	0.077	61.539	0.017	11.328
Brasília	MeAPE (%)	26.016	16.002	13.502	12.189
	nRMSE (%)	26.756	19.815	20.127	15.089
	KSI(−)	0.071	45.893	0.014	15.928
Campo Grande	MeAPE (%)	27.681	15.774	17.984	10.971
	nRMSE (%)	33.968	22.479	29.366	18.589
	KSI(−)	0.046	38.170	0.038	34.810
Cuiabá	MeAPE (%)	38.065	23.458	22.939	15.018
	nRMSE (%)	36.792	29.710	27.669	23.045
	KSI(−)	0.081	64.367	0.031	24.054
Florianópolis	MeAPE (%)	16.774	18.404	12.402	13.539
	nRMSE (%)	22.067	21.358	19.554	17.034
	KSI(−)	0.055	44.448	0.030	23.935
Ourinhos	MeAPE (%)	50.907	15.314	21.364	7.175
	nRMSE (%)	47.750	20.186	34.055	13.260
	KSI(−)	0.105	85.880	0.034	20.902
Palmas	MeAPE (%)	47.893	22.638	23.337	12.581
	nRMSE (%)	46.461	27.034	34.093	18.195
	KSI(−)	0.124	97.572	0.045	29.664
Petrolina	MeAPE (%)	47.328	16.752	25.700	9.091
	nRMSE (%)	54.393	20.975	34.608	13.893
	KSI(−)	0.084	80.585	0.030	21.472
São Luis	MeAPE (%)	16.503	31.997	9.059	22.670
	nRMSE (%)	22.713	41.323	18.128	34.527
	KSI(−)	0.071	46.786	0.019	18.208
São Martinho da Serra	MeAPE (%)	23.249	22.231	11.845	14.997
	nRMSE (%)	25.802	23.586	17.830	16.985
	KSI(−)	0.064	48.793	0.026	23.990

Table 3
Errors of the adjusted models on merged data sets.

Statistic	Adjusted BRL model				Original BRL model			
	Hourly		Minute		Hourly		Minute	
	d	DNI	*d*	DNI	*d*	DNI	*d*	DNI
MeAPE (%)	17.530	10.674	19.941	10.930	33.949	17.860	31.427	15.048
nRMSE (%)	25.764	16.078	27.472	16.656	34.856	22.383	30.178	18.488
KSI (−)	0.017	11.328	0.018	14.686	0.077	61.539	0.052	41.244

the national model.

4.1. Comparison between minute and hourly estimates

A graphical comparison between the adjusted fit and measured data is shown in Fig. 7 for the merged minute and hourly data sets. Notice the excellent fit for the entire data set in both cases. The fitted model captures the shape and spread of the data reasonably well.

Regarding formal error analysis, Table 3 shows the results for both minute and hourly regressions considering the estimations of the diffuse fraction (d) and DNI. When comparing the errors for the minute and hourly estimates, the adjusted model performs slightly better when estimating diffuse fraction for hourly data, with both hourly and minute data presenting diffuse fraction *MeAPE* and *nRMSE* values of around 19% and 26%, respectively. Meanwhile the DNI estimates present *MeAPE* and *nRMSE* measures of around 11% and 16%, respectively.

The *KSI* measure shows that both adjusted models present the same similarity between the CDFs of the actual and modeled diffuse fraction data, presenting values of 0.017 for the hourly and 0.018 for the minute estimated diffuse fractions. For the DNI, the adjusted model has a slightly better KSI value for hourly data than for minute data.

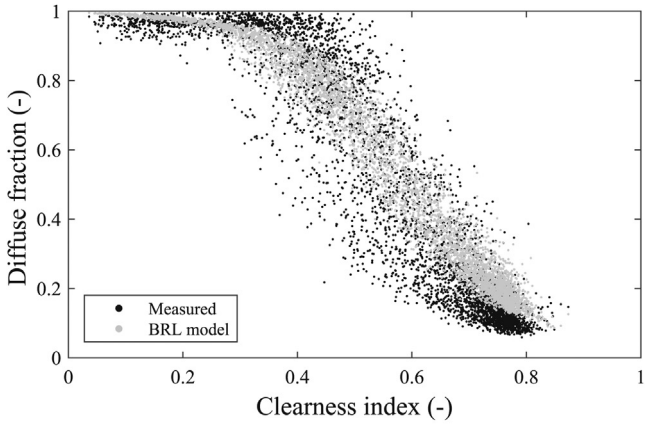
Table 3 shows the error measures of the BRL model estimates of the merged data. By comparing error measures of the adjusted model against the errors of the BRL model, a significant improvement is observed. For the hourly data set, it can be observed that for the diffuse fraction, *MeAPE* reduces from 33.9% to 17.5%, *nRMSE* from 34.8% to 25.7% and the *KSI* is reduced from 0.077 to 0.017. For DNI estimates, *MeAPE* reduces from 17.8% to 10.6%, *nRMSE* from 22.4% to 16%, and the *KSI* is also reduced, from 0.052 to 0.018. For the minute data set, the same behavior was observed, however, the reduction in error measures of the adjusted model is lower.

The results shown in Table 3 demonstrate that the BRL model can be used for both the hourly and minute data sets, presenting similar error measures. However, adjusting the model for local conditions significantly improves the performance of the model.

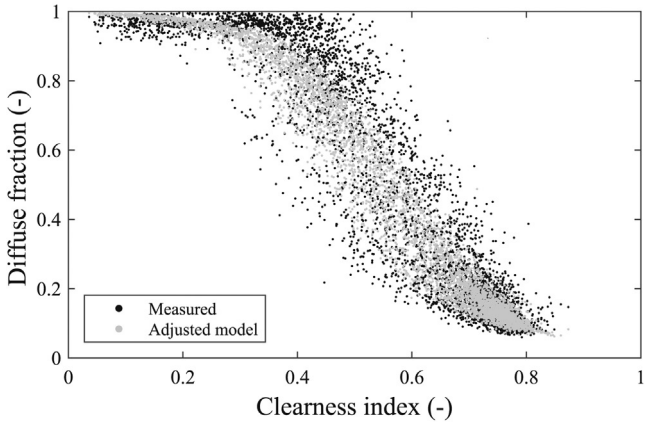
Since the model adjustment resulted in smaller errors when compared to the original BRL model (for both minute and hourly data), it is possible that the BRL model is able to properly describe the larger spread of the minute data. Thus, the model should only be adjusted to include cloud enhancement effects to be used for sub-hourly data sets.

The coefficients of the adjusted models are given in Table 4. The parameters of both adjusted models (hourly and minute) are significantly different than those presented in the original BRL model.

As shown in Table 3, both minute and hourly adjusted models performed similarly. Furthermore, the hourly and minute data sets have similar shapes, despite the larger spread of minute data. It would be interesting to have a single model to be used regardless of the data time resolution, thus eliminating the need for different models for minute and hourly data sets in order to capture the large

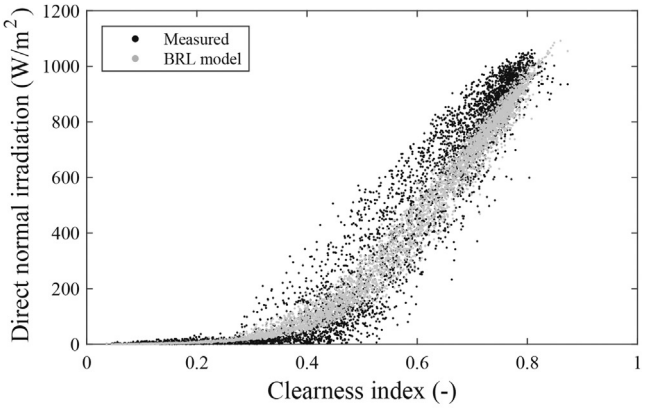


(a) Original BRL model for hourly data

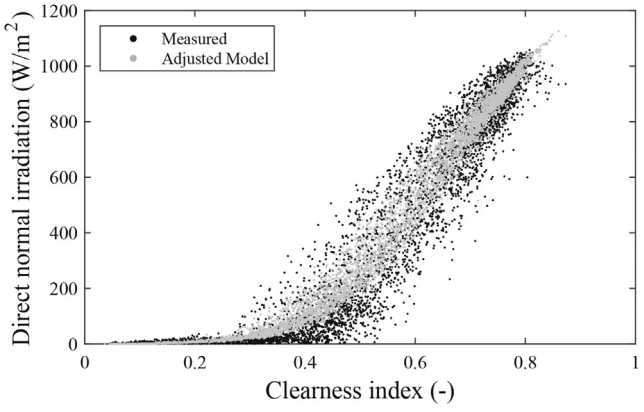


(b) Adjusted BRL model for hourly data

Fig. 8. Diffuse fraction estimates for Florianópolis.



(a) Original BRL model for hourly data



(b) Adjusted BRL model for hourly data

Fig. 9. DNI estimates for Florianópolis.

spread of the minute data. To further investigate this possibility, the correlations were interchanged in relation to the data sets, i.e., the minute correlation was applied to the hourly data set and the hourly correlation was applied to the minute data set. The error measures of the diffuse fractions resulting from this analysis are depicted in Table 5 for the city of Florianópolis. Observe that when the models are interchanged, the estimates present similar *MeAPE* and *nRMSE* values, approximately 15% and 23%, respectively. However, when the hourly coefficients are applied to the minute data, the error measures are slightly higher. Regarding the *KSI* measure, the hourly coefficients applied to the minute data had a *KSI* almost three times the value for the minute coefficients applied to the hourly data. This shows that the hourly coefficients applied to the minute data do not produce a good fit over the whole range of data as the opposite did occur.

The results of Table 5 can be further compared to the error measures of the hourly adjusted model applied to the hourly data for Florianópolis (Table 6). It can be seen that using the minute adjusted model or hourly adjusted model to estimate the diffuse fraction of hourly data results in similar errors; that is, *MeAPE* of 14%, *nRMSE* of 20%, and *KSI* of 0.02.

4.2. Original and adjusted BRL models

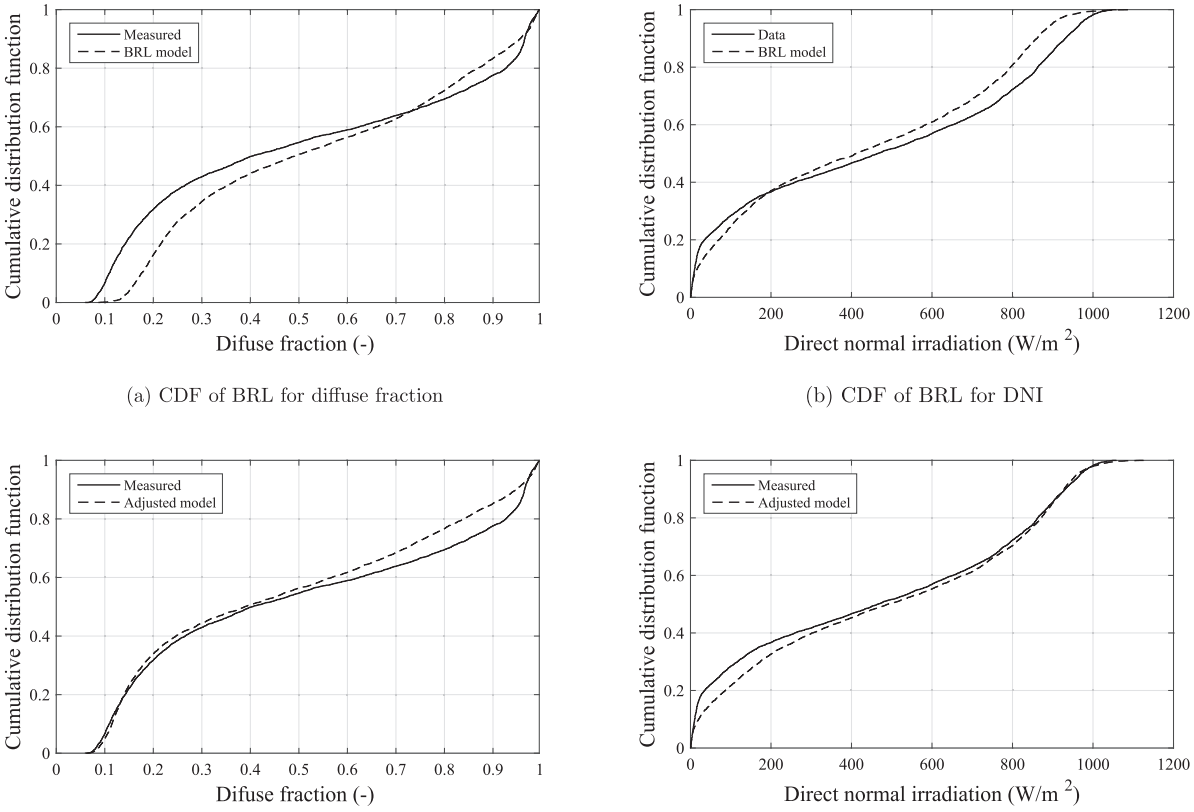
In order to assess the performance of the national adjusted model for each station individually, we considered all qualified data points available from each station. The original BRL model was also applied to each station to quantify the benefits of using the adjusted one.

To graphically illustrate the results, the city of Florianópolis was considered. A comparison of the models' estimations overlaid against the measured data is shown in Figs. 8 and 9 for the diffuse fraction and DNI, respectively. As shown in Fig. 8a, the BRL seems to fit the data reasonably well for all ranges of the clearness index. Meanwhile, the adjusted model (Fig. 8b) presents a very good fit to the data set, adequately capturing the shape of the diffuse fraction for clearness indexes around 0.7 and 0.8. Inspecting the results shown in Fig. 9, the same behavior is observed since the adjusted model presents a better fit for high DNI values.

Fig. 10 allows us to evaluate the goodness of fit over the whole range of data. For the BRL and the adjusted models, Fig. 10a and c shows an overlay of the CDFs of the estimated and measured diffuse fractions. Considering the shape of the CDFs, the adjusted model performs better than the original BRL model for values of the diffuse fraction below 0.6, but the errors increase for larger diffuse fraction values. This lower performance for high *d* values is not enough to counter better estimates at lower *d* values, which result in a lower *KSI* value for the adjusted BRL model, as shown in Table 6.

Similarly, Fig. 10b and d shows the CDFs of the DNI values obtained using both models. In this case, the adjusted BRL model gives slightly worse estimations for DNI values lower than 300 W/m^2, but considerably better results for larger DNI values. This behavior also results in lower *KSI* values for the adjusted model, as shown in Table 6.

These results are highlighted in Fig. 11, which depicts the differences between the CDFs of the model estimates and measured data for the diffuse fraction and DNI. Notice in Fig. 11a that the adjusted BRL model presents a significant improvement since the differences are reduced to almost zero for a diffuse fraction lower than 0.6 when compared to the original BRL model. On the other hand, Fig. 11b depicts the same analysis for DNI, showing that the adjusted BRL model has significantly lower differences, i.e., a better



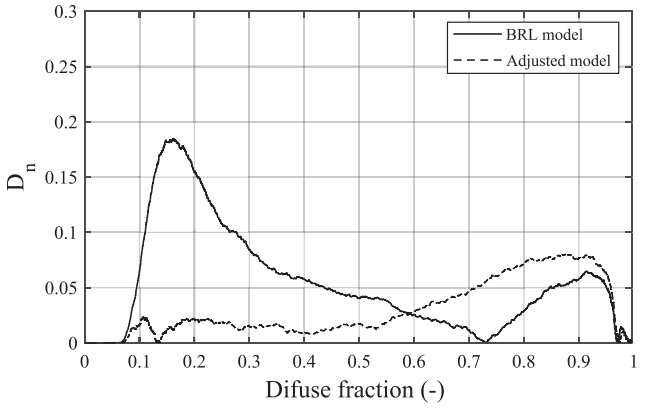
(a) CDF of BRL for diffuse fraction

(b) CDF of BRL for DNI

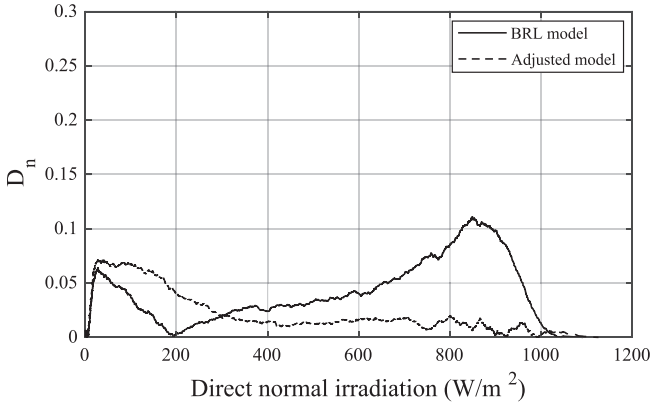
(c) CDF of adjusted model for diffuse fraction

(d) CDF of adjusted model for DNI

Fig. 10. CDF of the hourly diffuse fraction and DNI for Florianópolis.



(a) Differences in diffuse fraction



(b) Differences in DNI

Fig. 11. Differences D_n between CDFs of the measured and estimated hourly diffuse fraction and DNI for Florianópolis.

fit over the range of data above 300 W/m^2, and larger differences below that value.

The results for all the stations are summarized in Table 6. Notice that the adjusted model performs better than the BRL one for all of the locations and for both the diffuse fraction and DNI estimations. The adjustment of the original BRL model to Brazilian stations resulted in overall improvements in terms of the percentage error of the diffuse fraction and DNI estimations. For all locations analyzed, the adjusted model resulted in a better fit to the measured data and better similarity between the CDFs of the estimates and the data, as evidenced by the decrease in the *nRMSE* and *KSI* values given in Table 6. Therefore, it is possible to affirm that the new adjusted model is a useful tool for estimating the diffuse fraction and DNI values in Brazilian locations.

4.3. Discussion

The diffuse fraction *MeAPE* indicators calculated for the adjusted BRL model in Table 6 were between 9% and 26%, while for the original BRL model the range was between 16% and 50%. For the DNI estimates, the *MeAPE* for the adjusted model ranged from 7% to 23%, while the range for the original BRL model was from 15% to 31%. Therefore, adjusting the model coefficients resulted in improved model performance.

The MeAPE range for the adjusted model in Brazil is similar to the MeAPE range of the original BRL model presented for the seven stations used in Ridley et al. [2], where the BRL model had MeAPE ranging from 5% to 21%. Thus, the adjusted Brazilian model performed in Brazil similarly to how the original BRL model performed for the stations it was adjusted from.

The adjusted BRL model performed equally well when adjusted to both minute and hourly data sets, as shown in Table 3. This indicates that the logistic behavior of the original BRL model is well suited to make estimates in sub-hourly data sets, as previously demonstrated for hourly data in Ref. [12]. A further development of the logistic form model was the equation proposed by Ref. [31], which used a logistic function along with an additional clear sky model, to estimate the diffuse fraction for minute data. A review by Gueymard and Ruiz-Arias [36] analyzing several stations worldwide showed that the model presented by Engerer [31] delivered the best DNI estimates in tropical, arid, and temperate climates. This superior performance was found to be a result of considering clear-sky models to simulate the radiation enhancement phenomenon.

However, assessment of the diffuse fraction in Brazil using models that take clear-sky conditions into account is troublesome given the lack of readily available ground-measured values of turbidity and aerosol optical depth (AOD), among other variables commonly used as inputs to clear-sky models. In this sense, using simpler models, such as the original BRL model, is of interest if they can provide reliable estimates by using a set of easily obtainable input variables.

5. Conclusions

Solar resource in Brazil have yet to be characterized for the whole territory, thus, compromising the deployment of solar power facilities throughout the country. To address this issue, the BRL model was adjusted to Brazilian data to provide better irradiation estimates in Brazilian locations. To adjust the model, minute data from nine stations spread across the country was processed using a methodology based on testing procedures presented in the literature.

Previous studies [30,36] have assessed if solar radiation models developed with hourly data, such as the BRL model, are able to deliver estimates using sub-hourly data sets. A significant increase of random errors was found when the models were applied to minute data sets, as expected from the higher variability of minute data. It was also shown that hourly models are not able to properly describe cloud enhancement phenomena, which are more frequent in sub-hourly data.

In light of these facts, the methodology presented in this work aimed to adapt the coefficients of the BRL model to reduce random errors when minute data is considered, in order to achieve better estimations, and simulate the higher spread of minute data. However, further modifications still need to be included in the BRL model to improve its performance under radiation enhancement events.

Hourly integration of the processed data was considered so that the model could also be adjusted to irradiation measurements on an hourly basis. The processed data was merged, and a regression of the BRL model equation over the merged data set resulted in an adjusted BRL model for Brazil. The estimates given by the new model were compared with the estimates from the original one to evaluate the model performance improvements resulting from the adjustments.

The results showed that the original BRL model is able to properly describe the larger spread of minute data. On an hourly basis, the adjustments of the BRL model to Brazilian stations resulted in overall improvements in the estimation error of both diffuse fraction and DNI. In all of the locations analyzed, the

adjusted BRL model presented a better fit to the measured data and better similarity between the cumulative distributions.

The adjusted model represents a significant contribution to estimate the components of solar radiation (diffuse and DNI) over the entire Brazilian territory. Therefore, can be used in the extensive global irradiation measurement networks of Brazil to assess solar resources and help to properly design and reduce the risks associated to solar energy systems.

Acknowledgement

The authors would like to thank ANEEL/Petrobras, which sponsor this research under contract # 0050.0080593.12.9 through the Department of Mechanical Engineering at the Federal University of Santa Catarina, Florianópolis, Brazil.

References

[1] J.A. Duffie, W.A. Beckman, Solar Engineering of Thermal Processes, fourth ed., John Wiley & Sons, 2013.
[2] B. Ridley, J. Boland, P. Lauret, Modelling of diffuse solar fraction with multiple predictors, Renew. Energy 35 (2010) 478–483.
[3] SWERA, Solar and Wind Energy Resource Assessment, 2017. http://en.openei.org/wiki/SWERA, (Accessed 11 January 2017).
[4] INPE, SONDA Network, 2017. http://sonda.ccst.inpe.br/, (Accessed 11 January 2017).
[5] INMET, INMET Solar Stations, 2017. http://www.inmet.gov.br/portal/index.php?r=estacoes/estacoesAutomaticas, (Accessed 11 January 2017).
[6] E.B. Pereira, F.R. Martins, S.L.D. Abreu, R. Rüther, Atlas Brasileiro de Energia Solar, 2006.
[7] F.R. Martins, E.B. Pereira, S.A.B. Silva, S.L. Abreu, S. Colle, Solar energy scenarios in Brazil, Part one: resource assessment, Energy Policy 36 (2008) 2843–2854.
[8] B.Y. Liu, R.C. Jordan, The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, Sol. Energy 4 (1960) 1–19.
[9] J.F. Orgill, K.G.T. Hollands, Correlation equation for hourly diffuse radiation on a horizontal surface, Sol. Energy 19 (1977) 357–359.
[10] D. Erbs, S. Klein, J. Duffie, Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation, Sol. Energy 28 (1982) 293–302.
[11] D. Reindl, W. Beckman, J. Duffie, Diffuse fraction correlations, Sol. Energy 45 (1990) 1–7.
[12] J. Boland, B. Ridley, Models of diffuse solar fraction, Renew. Energy 33 (2008) 575–584.
[13] E.L. Maxwell, A Quasi-physical Model for Converting Hourly Global to Direct Normal Insolation, 1987.
[14] R. Perez, P. Ineichen, K. Moore, M. Kmiecik, C. Chain, R. George, F. Vignola, A new operational model for satellite-derived irradiances: description and validation, Sol. Energy 73 (2002) 307–317.
[15] A. Skartveit, J.A. Olseth, M.E. Tuft, An hourly diffuse fraction model with correction for variability and surface albedo, Sol. Energy 63 (1998) 173–183.
[16] S.A. Klein, TRNSYS: a Transient Systems Simulation Program, vol. 17, 2010.
[17] J. Boland, J. Huang, B. Ridley, Decomposing global solar radiation into its direct and diffuse components, Renew. Sustain. Energy Rev. 28 (2013) 749–756.
[18] M. Kocifaj, Angular distribution of scattered radiation under broken cloud arrays: an approximation of successive orders of scattering, Sol. Energy 86 (2012) 3575–3586.
[19] M. Kocifaj, Unified model of radiance patterns under arbitrary sky conditions, Sol. Energy 115 (2015) 40–51.
[20] M. Kocifaj, L. Kómar, Modeling diffuse irradiance under arbitrary and homogeneous skies: comparison and validation, Appl. Energy 166 (2016) 117–127.
[21] A. Berk, P. Conforti, R. Kennett, T. Perkins, F. Hawes, J. van den Bosch, MODTRAN6: a major upgrade of the MODTRAN radiative transfer code, p. 90880H.
[22] C.A. Gueymard, Parameterized transmittance model for direct beam and circumsolar spectral irradiance, Sol. Energy 71 (2001) 325–346.
[23] T. C. Martins, F. R., Guarnieri R. A., Chagas R. C., Neto S. L. M., Pereira E. B., Andrade E., Projeto Sonda – Rede Nacional De Estaçoes Para Coleta De Dados Meteorológicos Aplicados Ao Setor De Energia, in: Congresso Brasileiro de Energia Solar - CBENS I, p. 9.
[24] S. Younes, R. Claywell, T. Muneer, Quality control of solar radiation data: present status and proposed new approaches, Energy 30 (2005) 1533–1549.
[25] C.N. Long, Y. Shi, The QCRad value added product: surface radiation measurement quality control testing, including climatology configurable limits, Tech. Rep. (2006). ARM TR-074, Atmospheric Radiation Measurement Program.
[26] WMO, Guide to the Global Observing System, Report No. 488, WMO-No. 488, 2007, p. 170.
[27] M. Geiger, L. Diabaté, L. Ménard, L. Wald, A web service for controlling the quality of measurements of global solar irradiation, Sol. Energy 73 (2002) 475–480.
[28] M. Journée, C. Bertrand, Quality control of solar radiation data within the RMIB solar measurements network, Sol. Energy 85 (2011) 72–86.
[29] R.H. Inman, Y. Chu, C.F.M. Coimbra, Cloud enhancement of global horizontal irradiance in California and Hawaii, Sol. Energy 130 (2016) 128–138.
[30] C. A. Gueymard, J. A. Ruiz-Arias, Performance of separation models to predict direct irradiance at high frequency: validation over arid areas, in: EuroSun 2014 ISES Conference Proceedings, pp. 79–12.
[31] N.A. Engerer, Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia, Sol. Energy 116 (2015) 215–237.
[32] P. Ineichen, A broadband simplified version of the Solis clear sky model, Sol. Energy 82 (2008) 758–762.
[33] J. Kaiser, V.-H. Peuch, A. Benedetti, O. Boucher, R. Engelen, T. Holzer-Popp, J.-J. Morcrette, M. Wooster, The MACC-II Management Board: The pre-operational GMES Atmospheric Service in MACC-II and its potential usage of Sentinel-3 observations, in: Proceedings of the 3rd MERIS/(A)ATSR and OCLI-SLSTR (Sentinel-3) Preparatory Workshop, Frascati, Italy.
[34] P. Ineichen, Validation of models that estimate the clear sky global and beam solar irradiance, Sol. Energy 132 (2016) 332–344.
[35] B. Espinar, L. Ramírez, A. Drews, H.G. Beyer, L.F. Zarzalejo, J. Polo, L. Martín, Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations, Sol. Energy 83 (2009) 118–125.
[36] C.A. Gueymard, J.A. Ruiz-Arias, Extensive worldwide validation and climate sensitivity analysis of direct irradiance predictions from 1-min global irradiance, Sol. Energy 128 (2016) 1–30.