Renewable Energy 177 (2021) 700–714

Contents lists available at ScienceDirect

Renewable Energy

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

Assessing one-minute diffuse fraction models based on worldwide climate features

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

[a] LEPTEN - Laboratory of Energy Conversion Engineering and Energy Technology, Department of Mechanical Engineering, Federal University of Santa Catarina (UFSC), Florianopolis, Brazil
[b] Department of Mechanical and Metallurgical Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile
[c] Industrial AI, Centre for Industrial and Applied Mathematics, University of South Australia, Mawson Lakes Boulevard, Mawson Lakes, SA, 5095, Australia

ARTICLE INFO

Article history:
Received 5 January 2021
Received in revised form
16 May 2021
Accepted 18 May 2021
Available online 2 June 2021

Keywords:
Diffuse fraction model
One-minute data
Cloud enhancement
Köppen-Geiger climate classification
BSRN
Robust statistics

ABSTRACT

Despite the variety of irradiance separation models available in the literature, there is not yet a model that performs equally well for every location worldwide. Nevertheless, separation models still represent an interesting approach when only global irradiance is measured. This work presents further developments of the recently proposed Boland-Ridley-Lauret minute model and proposes 1-min diffuse fraction models for each climate zone in the globe. To accomplish this, 1-min worldwide irradiance data from 51 the Baseline Surface Rresearch Network stations and one from Australian Bureau of Meteorology were used. The Köppen-Geiger classification was used to determine the climate zone of each station: both the simple climate classification (A, B, C, D, E) and the complete classification (Aw, Bsh, etc.). Furthermore, a robust nonlinear regression method was used to build the separation models without removing measurement outliers in advance. The climate-specific models proposed herein present better performance than other models from the literature, such as the BRL, Perez and Skartveit models (developed using hourly irradiance data) and the Engerer model (developed using minute data). Generally, the models proposed for climate zones in the complete Köppen classification presented better similitude with the measured data than the models for the simple classification, especially for stations within climate zones B and E.

1. Introduction

Accurate values of the three irradiance components, namely the global horizontal irradiance (GHI), diffuse horizontal irradiance (DIF) and direct normal irradiance (DNI), are fundamental for the correct design and performance assessment of solar energy systems. Designing solar systems (thermal or PV) requires the value of the global tilted irradiance (GTI), which is estimated from the three irradiance components, while the assessment of concentrating systems requires information about direct normal irradiance.

There are several ways to obtain data regarding all three components of solar radiation. One option is the direct ground level measurement of all three components of solar radiation. This task, however, requires complex tracking devices and significant operational efforts, which lead to high costs. As a result, there are few stations providing complete, reliable measurements of the three components, while several stations measure only global irradiance due to lower operational efforts and costs.

Another option is the use of satellite-based methods, that have been employed since the 1980s to estimate hourly global horizontal irradiance using albedo and cloud cover indexes calculated from satellite images [1]. Since then, such methods have been improved to use satellite images combined with clear sky models [2] and with ground measured bright sunshine hours [3–5]. The use of ground data together with satellite images is an interesting approach for high quality estimates of solar irradiance components, by combining empirical or semiempirical data (measured or statistically modeled ground level irradiance) with data derived from physical models of the atmosphere (which is often the case in satellite-based methods). A common agreement has been established in the scientific community that the combination of physical and semiempirical approaches, in other words, the inclusion of

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

https://doi.org/10.1016/j.renene.2021.05.108
0960-1481/

physical descriptions into semiempirical models, has a key role in the characterization of the solar resource [6].

This novel approach also allows for the calibration of solar resource models for specific locations. Since the solar resource models have an empirical or semi-empirical aspect, one can gather local solar datasets by a short-term local measurement campaign and use it to conduct a calibration of the model, thus substantially improving its accuracy. For this site adaptation process to have good results, the uncertainty associated with the semiempirical estimation models represents a key issue.

It is in this context, diffuse fraction models, also named separation models, can be useful. Such models can estimate both DIF and DNI from GHI observations, and have been used not only in cases where only ground-measured GHI is available, but also in combination with satellite derived GHI. These models provide an easy-to-implement semiempirical approach to be used in solar resource assessment.

Since the initial work by Liu and Jordan [7], several separation models have been developed using measured data from stations in various climatic conditions, using various temporal scales (*e.g.* monthly, daily, hourly, minute), as well as a large variety of methods to calculate the diffuse fraction of solar radiation.

Some of the earlier developed models used linear piece-wise equations with one predictor [8,9], a strategy that was further improved by the addition of new predictors and better selection of breakpoints for the piece-wise equation [8,10]. Other proposed methods included polynomial piece-wise equations [11], quasi-physical models [12], logistical relationships [13–16], artificial neural networks [17], or even a more complex approach, using machine learning techniques to combine different separations models [18].

However, many of the separation models presented in the literature have been developed using hourly irradiance data, and do not properly describe fast transient episodes in solar irradiance, which happen on time scales much smaller than an hour. As a result, these models are not adequate to meet the current demands from the industry [19–21]. For instance, it has been shown that the non-linear and transient effects on CSP plants are only properly assessed in simulations with temporal resolution of the solar data of 10-min or less [22,23]. For PV applications, even higher temporal resolution of solar data is required, because of the impact of cloud enhancement events (CEE) on PV systems yield. The performance of such systems is currently being studied using solar data with a timestep of 3-s [20,24,25].

In their extensive validation study, Gueymard and Ruiz-Arias [21] reviewed 140 separation models, most of them developed using hourly data, in order to assess how those models would perform in estimating diffuse radiation at 1-min resolution. The 140 selected models were tested using high-quality 1-min data from 54 research-class stations from the Baseline Surface Radiation Network (BSRN). The authors reported that cloud enhancement events and high-albedo effects intensify the errors in the irradiance estimates of the selected separation models. Based on new criteria to evaluate the robustness of a separation model and for comparing each model's performance within different climate zones, the authors recommended the development of separation models for each specific climatic region. The authors also concluded that models that consider irradiance variability and clear-sky irradiance as predictors tend to perform better, and thus recommended a total of nine models as candidates for providing accurate and consistent estimations worldwide. The models recommended were the ones developed by Engerer [16], Perez et al. [26], Boland et al. [27], Hollands and Chra [28], Perez et al. [29], Perez et al. [30], Skartveit et al. [31] and Yao et al. [32]. Of the listed models, the one developed by Engerer presented the best generalization skill and can be considered a "quasi-universal' 1-min separation model.

Engerer [16] is the first study reporting the development of a separation model using 1-min irradiance data. The author proposed a separation model based in a generalized logistic function considering clear-sky irradiance (CSI) as a predictor and used a linear correction to account for the cloud enhancement events.

Since the work by Engerer, other studies have assessed the use of separation models specifically developed for 1-min data. Lemos et al. [33] demonstrated that the logistical equation [13] is able to properly describe the larger spread of minute data. Hofmann and Seckmeyer [34] developed another model with 1-min irradiance data from 28 locations worldwide, taken from the BSRN. The model was structured in three parts, which are independently calculated, and then combined to determine the diffuse fraction, depending on the clearness index. Instead of using a function, the author used a probability matrix that models the correlation between diffuse fraction and clearness index (first part) and models the correlation between the changes in the diffuse fraction and the changes in the clearness index (second part). The third part of the model used a posynomial equation to account for the diffuse fraction characteristic of clear sky days. The posynomial equation is based on a geometrical approach and considers clearness index, variability of clearness index, air mass, aerosol optical depth, water vapor, and the up/down time.

Starke et al. [35] presented a logistical model for 1-min irradiance based on the BRL model. The proposed model, named the BRL-minute model, uses a piecewise function, defined by two sub-domains, one for CEE and the other for non-CEE. The same logistical equation is used in both sub-domains, but with different coefficients. In addition, clear-sky irradiance was also added as a predictor. The authors proposed a model for Brazil and another for Australia, the former using 1-min data from four BSRN stations, and the latter using 1-min data from four stations of the Australian Government Bureau of Meteorology (BOM). The new model performed better than other separation models previously published, such as the original BRL, Engerer, Skartveit and Perez models.

Paulescu and Blaga [10] proposed a linear piecewise model to estimate diffuse fraction using only the clearness index as a predictor. The model was adjusted to data from the BSRN station au Palaiseau, France, and compared to the models of Engerer [16], Starke [35] and Hofmann and Seckmeyer [34], using data from several BSRN stations other than the Palaiseau station. Despite its simplicity and even though this model was adjusted to data from only one location, the authors showed that their proposed model performed well when compared to the other, more complex models.

In order to contribute with the development of separation models for 1-min diffuse radiation, one of the goals of this study is to further improve the BRL-minute model proposed by Starke et al. [35], by adding a new predictor to improve model performance. To support the inclusion of all predictors in the final model, the partial F-test was used, allowing assessment of whether the addition of CSI and the new predictor improved the model, and also to evaluate if all predictors are needed for modeling the CEE.

Another suggestion from the extensive review by Gueymard and Ruiz-Arias [21] was to develop diffuse fraction models for specific climate zones. Some recent works have considered the type of climate when proposing new separation models [36–39], however such models were made for hourly, daily or monthly irradiance data.

Therefore, the main goal of this study is to develop 1-min climate-specific separation models, by using 1-min worldwide irradiance data from several BSRN stations. The Köppen-Geiger climate classification was used to determine the climate zone of each station, which is represented by two or three letters, the first

of which indicates the broader climate zone (tropical, dry, temperate, continental and polar), and the following letters indicate the climate subtype. The separation models were implemented for each of the five broader climate zones in the Köppen classification (the first letter). As a second step, fourteen separation models were developed clustering the climate zones by its subtypes, i.e. three-letter classification. To support the concept of developing models for each climate zone, a cross-test was performed, where each broader climate model was applied for all stations, comparing the performance of each model for each station. Finally, to assess the performance of the new model, a formal error analysis was performed, by direct comparison between the model described herein and the models of Ridley et al. [14], Engerer [16], Skartveit et al. [31] and Perez et al. [26].

This paper is organized as follows: after the introduction, the BRL minute model and proposed modes are presented. Section three presents the methodology developed, including the description of the used data. Section four presents the 1-min diffuse fraction models results for each BSRN station considered. Section five presents the diffuse fraction climate models developed for each climate zone. The conclusions are presented on section six, followed by the acknowledgments, supplementary material appendix and data availability.

2. BRL minute model

Recently, Starke et al. [35] proposed a piecewise logistical model that can predict the fast transient phenomena observed in 1-min data. The BRL-minute model proposed by the authors is given by,

$$\widehat{d} = \begin{cases} \dfrac{1}{1 + e^{\left(\beta_0 + \beta_1 k_T + \beta_2 AST + \beta_3 \alpha + \beta_4 \overline{K_T} + \beta_5 \psi + \beta_6 CSI\right)}}, & K_{CSI} < 1.05 \\ \dfrac{1}{1 + e^{\left(\beta_7 + \beta_8 k_T + \beta_9 AST + \beta_{10} \alpha + \beta_{11} \overline{K_T} + \beta_{12} \psi + \beta_{13} CSI\right)}}, & K_{CSI} \geq 1.05 \text{ and } k_T > 0.65 \end{cases} \quad (1)$$

where $\widehat{d}$ is the estimated diffuse fraction, k_T is the clearness index at minute basis, defined as the ratio of the global irradiance on a horizontal surface (I_g) to the horizontal extra-terrestrial irradiance at the top of the atmosphere (I_0). AST is the apparent solar time in hours, α is the solar altitude in degrees, $\overline{K_T}$is the daily clearness index and ψ is a persistence factor defined by Ridley et al. [14]. CSI is the clear-sky irradiance in MJ/(hr.m^2) and K_{CSI} is the ratio of the measured global horizontal irradiance and the modeled global clear-sky irradiance. The numeric values of the coefficients (β_0, β_1, …, β_{13}) in Eq. (1) were reported in Starke et al. [35], where a set of coefficients was calculated for Brazil and another for Australia.

Eq. (1) represents a two-part piecewise model. This division aims to separate diffuse fraction values that correspond to cloud enhancement from those that do not. To set the domain breakpoint, i.e. to identify whether a data point is CEE or not, the variable K_{CSI} was adopted, which is defined as the ratio between the GHI and the CSI.

Most of the predictors of the BRL-minute model can be easily calculated through the measurement time and location by applying well-known functions of solar geometry [40,41]. Measurements of global irradiance are used to calculate the clearness index, daily clearness index and persistence factor. The BRL-minute model also requires additional measured variables such as aerosol optical depth and atmosphere water vapor content, which are not commonly available. To overcome this problem, the values provided by the CAMS Reanalysis [42] dataset were used.

Fig. 1a shows the relationship between diffuse fraction and clearness index (k_T) for 1-min data. As depicted in the figure, four regions can be identified: overcast, partly cloudy, clear-sky, and cloud enhancements events. Fig. 1b depicts the relationship between the diffuse fraction and the clearness index for Florianopolis (Brazil). The measured diffuse fractions are shown in black, and the diffuse fraction estimates given by the BRL-minute model are shown in gray. It is observed that the BRL-minute model efficiently captures both the shape of the measured data and the cloud enhancement phenomena.

Although Starke et al. have presented significant advances on the performance of the logistic model for 1-min data, the authors did not present a statistical justification for using CSI as a predictor in the model nor for the use of six predictors on the CEE part of equation. Furthermore, it can be argued that, for 1-min models, an intuitive approach is to use hourly clearness index (K_T) instead of daily clearness index ($\overline{K}_T$). Therefore, the present work evaluates the improvement of the BRL-minute model by including CSI as a model predictor, as well as including other predictors on the CEE equation, such as the hourly clearness index. The partial F-test was selected for assessing the improvement obtained from the addition or removal of predictors, as in the following expression [43],

$$F = \frac{\frac{SSE_{Reduced\ Model} - SSE_{Full\ model}}{k_{Full\ model} - k_{Reduced\ model}}}{\frac{SSE_{Full\ model}}{df_{Full\ model}}} \quad (2)$$

where the subscript *"Full model"* denotes the model that includes all the variables of interest, *"Reduced Model"* denotes a model that includes all the variables of interest except those whose statistical significance is tested. *df* is the degree of freedom, calculated as $df = n - k - 1$, being n the number of observations (i.e. sample size) and k is the number of variables of interest (predictors). Finally, *SSE* is the sum of squared errors, calculated as,

$$SSE = \sum_{i=1}^{n} \left(d - \widehat{d}\right) \quad (3)$$

where d and $\widehat{d}$ are the observed and modeled diffuse fraction, respectively. Table 1 summarizes the partial F-test calculations for the Florianopolis station, where the effect of the number of predictors is assessed for both sections of the piecewise equation. The BRL-minute model expressed by Eq. (1) is used as a reference, and the tests of adding and/or removing predictors in the model is reported accordingly. It is observed in Table 1 that the influence of *CSI* and K_T are highly significant in both sections of the model. That effect is observed since removing *CSI* as a predictor from the model increases the SSE, while including K_T reduces the SSE. Therefore, these two variables should be considered in the model. Regarding

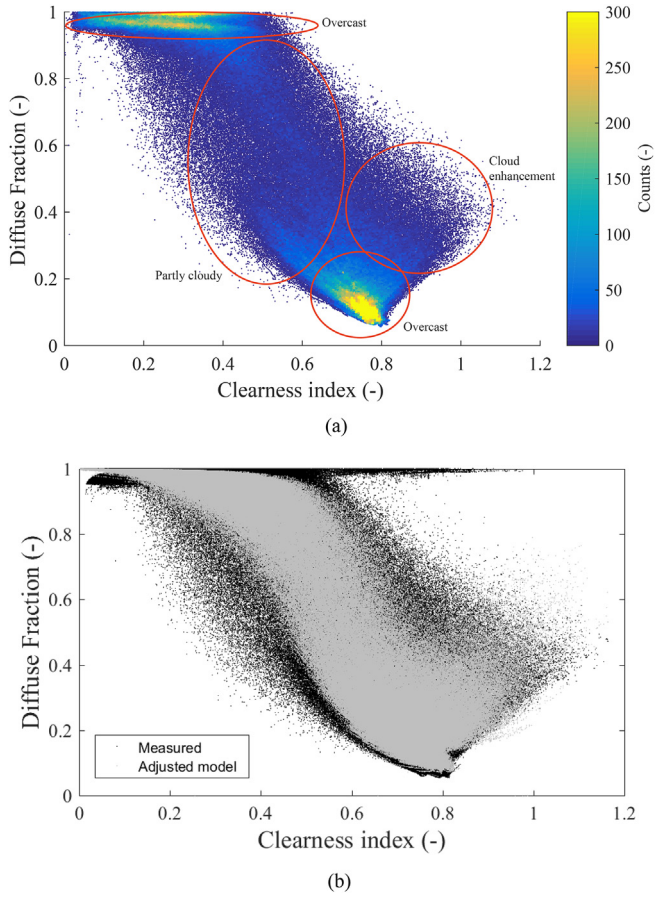


Fig. 1. Observed d-kT envelop of 1-min irradiance data from Florianopolis/Brazil; (a) the colormap denotes the density of points (counts number). Four regions are highlighted: overcast, partly cloudy, clear-sky, and cloud enhancements events; (b) measured data overlaid with BRL-minute model estimates.

the other predictors, $\overline{K}_T$ is highly significant for the Non-CEE equation and removing it negatively affects the model's performance. For the CEE equation, almost all predictors are observed as significant and removing any of them might increase the SSE, therefore all those predictors should be considered in the model. The exception is the *AST*, which appears not to be significant for the CEE equation; however, removing it does not provide a significant reduction on the SSE, and can be included in the model for sake of generality.

Based on the Partial F-test, the proposed model is given by,

$$\hat{d} = \begin{cases} \dfrac{1}{1+e^{\left(\beta_0+\beta_1 k_T+\beta_2 AST+\beta_3\alpha+\beta_4\overline{K}_T+\beta_5\psi+\beta_6 CSI+\beta_7 K_T\right)}}, & K_{CSI} \geq 1.05 \text{ and } k_T > 0.75 \\ \dfrac{1}{1+e^{\left(\beta_8+\beta_9 k_T+\beta_{10} AST+\beta_{11}\alpha+\beta_{12}\overline{K}_T+\beta_{13}\psi+\beta_{14} CSI+\beta_{15} K_T\right)}}, & \text{Otherwise} \end{cases} \quad (4)$$

where the values of the breakpoint were defined as constants, as explained in section 3.5. The subdomain intervals were reorganized in a clearer form, where the first equation models the cloud enhancements events and the additional equations model the other events. Unlike Eq. (1), in Eq. (4) the CSI inputs for the equation are given in W/m^2.

Table 1
Summary of partial F-test for Florianopolis station (n = 737,731).

	Variable	F	SSE	k
Non-CEE eq.	Reference (BRL-minute, Eq. (1))	-	2590	14
	Remove *CSI*	48,985	2762	13
	Add K_T	53,630	2415	15
	Remove $\overline{K}_T$	172,599	3196	13
CEE eq.	Add K_T	3554	2578	15
	Remove *CSI*	6497	2613	13
	Remove ψ	3430	2602	13
	Remove $\overline{K}_T$	6282	2612	13
	Remove α	1395	2595	13
	Remove *AST*	−22	2590	13

3. Methodology

Like most separation models proposed in the literature, the model proposed herein is derived from irradiance measurements [21], and the model coefficients are estimated through a regression fit to the measured data. High-quality 1-min irradiance data is required to build the separation model. The quality of the model depends on the quality of the irradiance measurements, which depend on radiometer performance and calibration, station maintenance, and instrument cleaning [21]. Having this in mind, only data from research-class stations were considered.

3.1. Irradiance data

The 1-min database contains the three irradiance components (GHI, DIF, DNI) measured by thermopile radiometers. Most of the stations (51) belong to the Baseline Surface Radiation Network (BSRN) [44,45], a project of the Data and Assessments Panel from the Global Energy and Water Cycle Experiment (GEWEX) under the umbrella of the World Climate Research Programme (WCRP). One station belongs to the Australian Government Bureau of Meteorology (BOM) [46]. A summary of all stations considered herein is presented in Table 2, and their geographical localization is depicted in Fig. 3. Only data after the year 2000 is being considered, as most stations do not have 1-min irradiance measurements before that.

3.2. Clear-sky model

The clear-sky irradiance (CSI) values used in this work were calculated from the broadband simplified analytical version of the Solis model [47]. This model estimates CSI at the evaluated site by multiplying the extraterrestrial irradiance by a correction factor that is a function of site altitude, aerosol optical depth (AOD) and atmosphere water vapor content data taken from the CAMS

Table 2
General information on the 53 stations whose data were available for this study, including, in order of appearance, three leter code (the ones used by BSRN, for stations that belong to that network), station complete name, Latitude and Longitude in degrees, elevation in meters above sea level, acronyms for the data source, climate of the station region according to the Köppen-Geiger classification and the number of valid data points (after quality control).

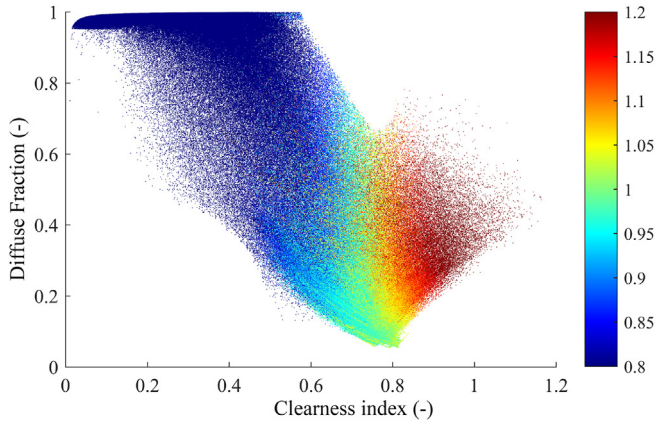
	Code	Station	Latitude	Longitude	Elevation	Source	Climate	Data points
1	ADL	Adelaide	−34.929	138.601	61.7	BOM	Csb	444 557
2	ALE	Alert	82.490	−62.420	127	BSRN	ET	860 461
3	ASP	Alice Springs	−23.798	133.888	547	BSRN	BSh	3 413 934
4	BAR	Barrow	71.323	−156.607	8	BSRN	ET	571 991
5	BER	Bermuda	32.267	−64.667	8	BSRN	Cfa	727 743
6	BIL	Billings	36.605	−97.516	317	BSRN	Cfa	1 326 227
7	BOS	Boulder	40.125	−105.237	1689	BSRN	Cfb	443 931
8	BOU	Boulder	40.050	−105.007	1577	BSRN	Cfb	685 653
9	BRB	Brasilia	−15.601	−47.713	1023	BSRN	Aw	960 528
10	CAB	Cabauw	51.971	4.927	0	BSRN	Cfb	2 101 944
11	CAM	Camborne	50.217	−5.317	88	BSRN	Cfb	1 700 471
12	CAR	Carpentras	44.083	5.059	100	BSRN	Cfb	3 229 538
13	CLH	Chesapeake Light	36.905	−75.713	37	BSRN	Cfa	2 579 937
14	CNR	Cener	42.816	−1.601	471	BSRN	Cfb	1 362 342
15	COC	Cocos Island	−12.193	96.835	6	BSRN	Aw	1 795 416
16	DAA	De Aar	−30.667	23.993	1287	BSRN	BSk	1 046 037
17	DAR	Darwin	−12.425	130.891	30	BSRN	Aw	2 221 084
18	DOM	Concordia Station	−75.100	123.383	3233	BSRN	EF	1 134 106
19	DWN	Darwin Met Office	−12.424	130.893	32	BSRN	Aw	1 797 269
20	E13	Southern Great Plains	36.605	−97.485	318	BSRN	Cfa	1 839 139
21	EUR	Eureka	79.989	−85.940	85	BSRN	ET	601 728
22	FLO	Florianopolis	−27.605	−48.523	11	BSRN	Cfa	737 731
23	FUA	Fukuoka	33.582	130.376	3	BSRN	Cfa	1 246 885
24	GCR	Goodwin Creek	34.255	−89.873	98	BSRN	Cfa	303 879
25	GOB	Gobabeb	−23.561	15.042	407	BSRN	BWk	1 226 622
26	GVN	Georg von Neumayer	−70.650	−8.250	42	BSRN	EF	1 819 251
27	ISH	Ishigakijima	24.337	124.164	5.7	BSRN	Cfa	1 369 443
28	IZA	Izaña	28.309	−16.499	2372.9	BSRN	Csb	1 736 320
29	KWA	Kwajalein	8.720	167.731	10	BSRN	Af	581 092
30	LAU	Lauder	−45.045	169.689	350	BSRN	Cfb	2 479 186
31	LER	Lerwick	60.139	−1.185	80	BSRN	Cfc	1 285 996
32	LIN	Lindenberg	52.210	14.122	125	BSRN	Cfb	2 191 167
33	LRC	Langley Research Center	37.104	−76.387	3	BSRN	Cfa	579 451
34	MAN	Momote	−2.058	147.425	6	BSRN	Af	2 121 521
35	MNM	Minamitorishima	24.288	153.983	7.1	BSRN	Aw	1 654 244
36	NAU	Nauru Island	−0.521	166.917	7	BSRN	Af	1 602 555
37	NYA	Ny-Ålesund	78.925	11.930	11	BSRN	ET	1 786 803
38	PAL	Palaiseau	48.713	2.208	156	BSRN	Cfb	1 917 902
39	PAY	Payerne	46.815	6.944	491	BSRN	Cfb	1 567 868
40	PSU	Rock Springs	40.720	−77.933	376	BSRN	Dfb	363 090
41	PTR	Petrolina	−9.068	−40.319	387	BSRN	BSh	879 163
42	REG	Regina	50.205	−104.713	578	BSRN	Dfb	1 875 990
43	SAP	Sapporo	43.060	141.329	17.2	BSRN	Dfb	1 283 889
44	SBO	Sede Boqer	30.860	34.779	500	BSRN	BWh	788 147
45	SMS	São Martinho da Serra	−29.443	−53.823	489	BSRN	Cfa	1 040 359
46	SON	Sonnblick	47.054	12.958	3108.9	BSRN	Dfc	431 236
47	SOV	Solar Village	24.910	46.410	650	BSRN	BWh	473 286
48	SPO	South Pole	−89.983	−24.799	2800	BSRN	EF	1 125 927
49	SYO	Syowa	−69.005	39.589	18	BSRN	EF	1 656 354
50	TAM	Tamanrasset	22.790	5.529	1385	BSRN	BWh	1 594 857
51	TAT	Tateno	36.058	140.126	25	BSRN	Cfa	3 220 379
52	TOR	Toravere	58.254	26.462	70	BSRN	Dfc	1 894 060

Reanalysis [42] dataset, which provides atmospheric composition data from 2003 up to 2017. However, for stations with data extending beyond 2017, the information of the last year is used as an approximation to the atmospheric composition of the following years. As reported in Starke et al. [35] this approximation has a minor impact on model performance.

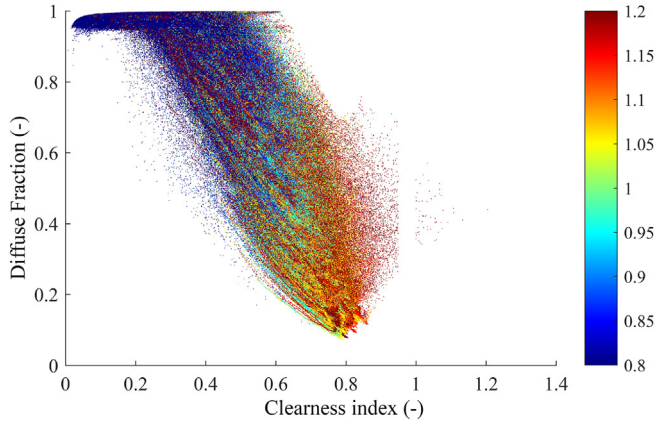
Some of the stations used in this work are placed in sites whose elevation is much higher than the average pixel elevation for the CAMS dataset, which might cause inconsistencies when computing the CSI estimations. In order to address this issue, an altitude correction was implemented, as described by Gueymard and Thevenard [48],

$$AOD(h) = AOD(h_0)e^{((h-h_0)/H_a)} \quad (5)$$

where $AOD(h)$ is the AOD estimate at the station altitude h, informed by BSRN, $\mathrm{AOD}(h_0)$ is the AOD pixel estimate with reference altitude h_0, obtained using the geopotential measurement for the considered pixel, obtained from the CAMS dataset. Finally, H_a is the scale height, adopted as a constant 2100 m as suggested by Gueymard and Thevenard for coastal sites. This adoption was implemented because most stations that presented anomalous behavior regarding the CSI estimate are located close to the coast, but at high elevations, i.e. mountains near the sea. The same procedure was used for correcting the atmospheric water vapor content, by using the water vapor W instead of AOD in the equation with the same scale height.



(a) Florianopolis/Brazil



(b) Xianghe/China

Fig. 2. d-kt plots for (a) Florianopolis/Brazil and (b) Xianghe/China (not considered). The colormap indicates the K_{CSI}. The Clear-sky region should have K_{CSI} values ranging from 0.95 to 1.05, with higher K_{CSI} in the cloud enhancement region.

3.3. Quality control

Even in high quality stations, there may exist some variation in equipment performance and operational difficulties may occur. Such limitations should be acknowledged, and the raw data should be initially filtered and qualified, to ensure that inconsistent and suspicious data is removed.

The first quality checks applied to the solar data were the ones used by Gueymard and Ruiz-Arias [21] in their extensive review. These checks ensure that the measured irradiance value is physically possible and that GHI, DIF, and DNI measurements are mutually consistent. Further quality checks were implemented, based on the quality control methodology described by Lemos et al. [33], such as the “Rayleigh limit test”, guaranteeing that the measured diffuse irradiance is not below a minimum physically possible value. The “overcast test” was applied as a lower allowable limit to GHI.

The number of qualified data points for each station is presented in Table 2. Stations with less than 262,800 qualified data points (roughly sunlight minutes within one year) were removed from the pool. This criterion was used so that only stations with at least one year of valid data were considered for building the models.

As the last quality check, the CSI values calculated were compared with the actual measurements and a d-kt scatter plot was made (see Fig. 2), where all points plotted were colored using the K_{CSI} values, defined as the ratio between the GHI and CSI. Almost all stations performed visually well in this test, and the couple stations for which the clear-sky model clearly did not represent a faithful description, even with the altitude correction of the AOD and water vapor content data, were also removed from the pool. The discrepancy of those stations was probably caused by some bad AOD and/or atmospheric water-vapor content data for those places, so using them would produce unreliable results.

3.4. Climate classification

In order to develop separation models for specific climate zones, the Köppen-Geiger climate classification was used to classify the different stations considered in the present work. To do so, data provided on the website developed by Chen [49], which is based on the work by Chen and Chen [50], were used. The authors used global temperature and precipitation observations over the period of 1901–2010 to build the Köppen classification dataset on the interannual, interdecadal, and 30-year time scales. The result is a classification indicated by a letter code, where the first letter indicates the main climate type (A for tropical, B for dry, C for temperate, D for continental and E for polar), and the following letters denote the specific climate type. The reader is referred to the work by Chen and Chen [50] for a more detailed description of the letters in the Köppen classification scheme.

In the present work the 30-year time scale was used, with data from 1981 to 2010, comprising most of the measurement period of the stations considered. The Köppen climate of each station is presented in Table 2, while Fig. 3 depicts the geographical location of each station with the markers and the color legend indicating the climate group.

3.5. Regression method

Boland and Ridley [13] established the procedure for the construction of the logistical separation model, while Ridley et al. [14] presented the method for building the BRL model using the nonlinear least-squares (NLS) method. However, the presence of outliers in the data can severely affect the results when the separation model is adjusted to the measured data through a least-squares fit.

Lemos et al. [33] presented a method for removing possible outliers in solar data sets which is similar to the one proposed by Younes et al. [51], and consists of creating an envelope around plausible data, using specified functions. Although this method is effective, it requires building envelope curves for each individual dataset (i.e. station), a process carried out by a thorough analysis of the data. Building an outlier removal procedure for a heterogeneous database such as the BSRN is an infeasible task, requiring the individual inspection of each station's data and site-specific knowledge of the climate and irradiation behavior.

One solution for building the separation models without removing the outliers in advance is to use a robust nonlinear regression method. Robust statistical methods began to emerge in the nineteenth century, and nowadays are considered a well-established method for estimating location, scale and regression parameters [52,53]. These methods were derived to correct the least squares indicators in cases where classical assumptions, such as the homoscedasticity of the data, that is, the data is distributed normally, do not apply. Therefore, the methods considering this robust approach can be employed to find reliable parameters estimations when the data follow an arbitrary non-normal distribution. As highlighted by Riazoshams et al. [54], in real-life data the homoscedastic assumption might not be correct, because of the natural behavior of the data; or because the data is affected by some

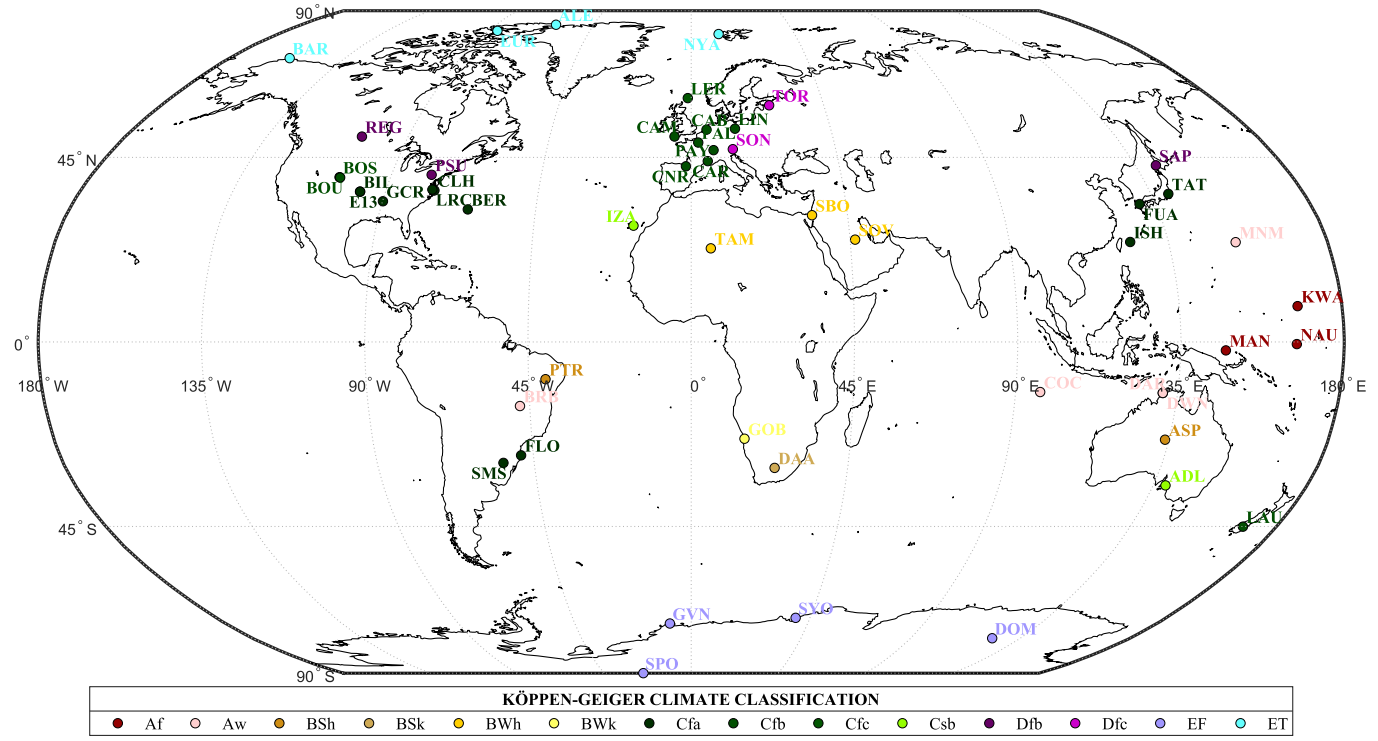


Fig. 3. World Map showing the Köppen-Geiger climate classification for each station used in this work.

discrepancy in the observations – i.e., the outliers. Although the homoscedasticity[1] hypothesis is not considered for deriving the NLS method, when this hypothesis is not true, the NLS standard errors are no longer valid for constructing confidence intervals, t statistics, and F statistics [55].

In the case of the data from the station at Florianopolis/Brazil, Fig. 4a shows that the residuals of the adjusted model using a NLS regression method deviates significantly from a normal distribution. On the other hand, Fig. 4b shows that, using a robust least-square method, this discrepancy becomes considerably smaller, guaranteeing that the statistical indicators will be close enough to the exact ones. A brief discussion on how the regression algorithm affects model performance is presented at the end of section 4.

Based on the results observed in Fig. 4, a robust nonlinear regression method, available in Mathworks [56], was adopted to build the diffuse fraction models of the present study. An iterative reweighted least squares algorithm [57,58] is used by the software, which recalculates the weights at each iteration based on the residual of the observations of the last iteration. This approach reduces the influence of the outliers on the fit at each iteration, where the process continues until the convergence. A weight function and a tuning constant must be defined for the robust fitting. After testing different equations for the weight's estimation, it was decided to use the following logistic function,

$$w_i = \frac{\tanh r_i}{r_i} \tag{6}$$

where w_i is the robust weight for the residual of the regression r_i applied to the measurement i. The tuning constant used was the default given by the MathWorks function. Fig. 5 shows the estimated weights for the Florianopolis/Brazil dataset. As observed in the figure, the majority of high weight values are in a region comparable to the envelopes proposed by Younes et al. [51], which is the reason for the robust method to work without the need of an outlier removal procedure. Even though this method does not completely remove the influence of obvious outliers, it assures that even the improbable, but properly measured, values are considered on the diffuse fraction model.

One feature resulting of the proposed methodology is that the robust nonlinear regression method could provide a simple method for removing outliers, not requiring, a priori, a dataset with low quantity of outliers or with low quantity of erroneous data, as in the approaches proposed by Younes et al. [51] and Lemos et al. [33]. By performing a robust nonlinear regression – in specific using an iterative reweighted least squares algorithm - the weights generated on the procedure can be used to remove outliers, being necessary to define a limit value. This approach could be used as one step of the quality control procedures.

Moreover, as the model proposed in this study is based on a piecewise function, the breakpoint value must also be defined. This process was carried out by testing many combinations of breakpoint values for the k_T and K_{CSI}, which are, respectively, the clearness index at minute basis and the ratio of the measured global horizontal irradiance and the modeled global clear-sky irradiance. The datapoints being regressed were split in two classifications, Cloud Enhancement Event (CEE) or non-CEE, and a regression was made for each of the two new datasets, returning the coefficients for each part of the final model and the weights considered for each datapoint.

3.6. Building local and climate-specific models

After the qualification procedure, the regression method was applied to the data from each individual station to create a local optimization model, valid for that station. On the other hand, to

[1] Residuals of the adjusted model do not need to have a normal distribution for its resulting parameters to be the best linear unbiased estimator (BLUE).

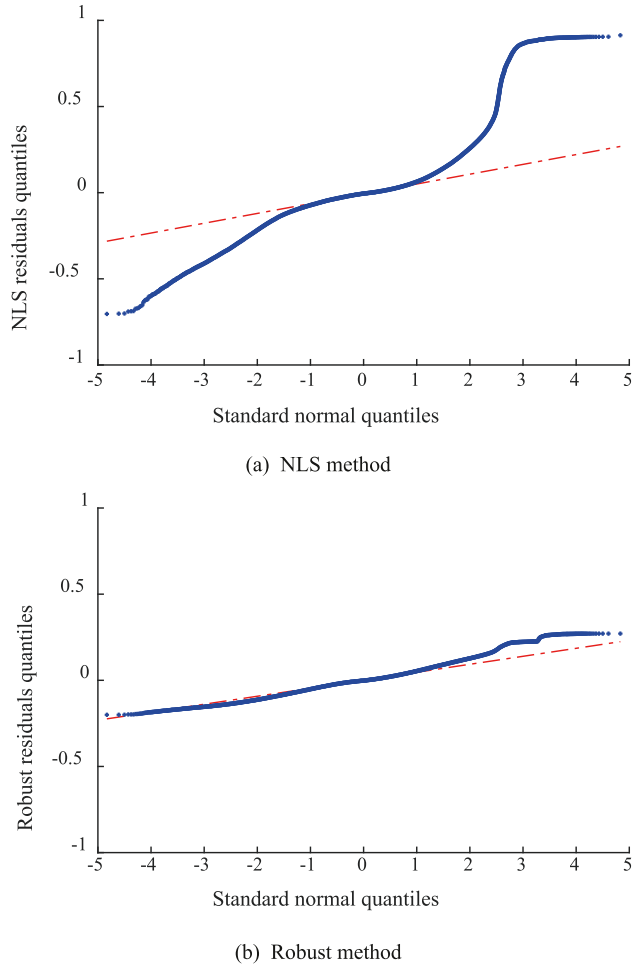


Fig. 4. Q-Q plots of the residuals of the regressions for Florianopolis/Brazil using (a) a NLS regression method and (b) a robust nonlinear regression method; The red dashed lines represent a normal distribution for each regression, while the blue markers represent the distribution for the residuals of each regression.

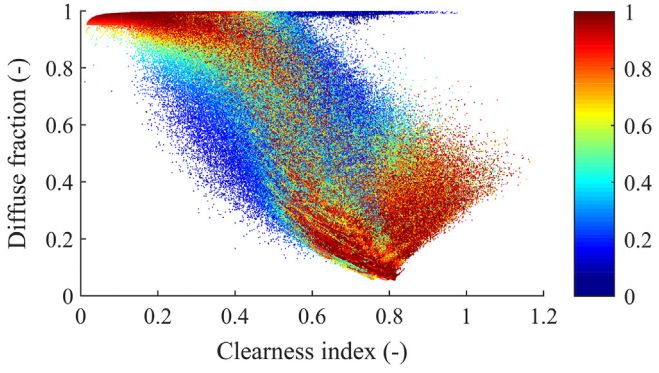


Fig. 5. Observed clearness index (k_T) and diffuse fraction (d) correlation of 1-min irradiance data from Florianopolis/Brazil; The colormap denotes the estimated weight w_i for each point.

create a model valid for a region (climate, country or "universal"), the data from individual stations need to be merged to a single data set, and then analyzed with the defined regression method to create a regional model.

The first step to build the climate models was merging the data from stations belonging to the same broad climate zones in the

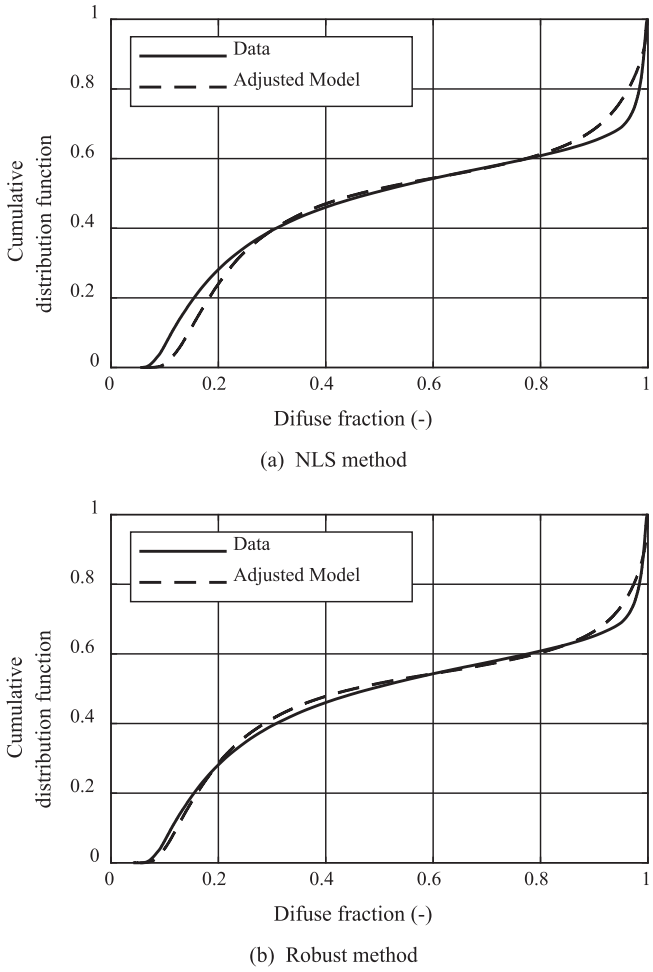


Fig. 6. CDFs for the diffuse fraction in Florianopolis/Brazil. The continuous line represents the actual dataset, while the dashed line represents the results from the model generated using a NLS (a) and a robust regression methods (b).

Köppen classification (A, B, C, D and E). This was done by randomly selecting the same number of datapoints for each station of that climate, so each station has equal contributions to the climate dataset. For each of these datasets a separation model was estimated. The next step was to merge the data from stations belonging to the same three-letter Köppen climate zone. However, some zones were disregarded for having no available stations with qualified data.

3.7. Statistical indicators of model performance

Gueymard [59] presented a complete review of performance indicators that can be used in radiation models for validations purposes. Among those, four statistical indicators were considered: normalized MAD, normalized RMSE, normalized MBE, and the Kolmogorov–Smirnov test Integral (KSI). In the present work, these indicators have been slightly modified to be able to use weighted residuals (r_i), which is defined as,

$$r_i = \sqrt{w_i}\left(\widehat{d}_i - d_i\right) \tag{10}$$

where d_i is the actual value of the diffuse fraction calculated, $\widehat{d}_i$ is the estimated value of the diffuse fraction for the point i, and w_i is the weight estimated using the robust method for the point i.

Fig. 7. Comparison between KSI (a), nRMSE (b), and nMBE (c) error indicators for each station when using a NLS or a robust regression method. The black markers represent the robust method, and the red markers represent the NLS method. The stations are grouped by climate. The lines between each marker are presented just to ease the visualization.

In the robust regression model, the weighted residuals are used to calculate the goodness of fit of the regressed model. Such an approach allows assessing the ability of the model for predicting the main trend of the dataset that originated the model, reducing the penalty for not fitting the outliers. However, when assessing the performance of different models, or applying it for different datasets, the weights are not available. In this case, the weight for each data point is set equal to the unity, $w_i = 1$, which reduces Eq. (10) to the standard residuals of the NLS method,

$$r_i = \left(\widehat{d}_i - d_i\right) \tag{11}$$

3.8. Other separation models

In order to assess the performance of this new model an error analysis was performed, based on the comparisons between the model described herein and the models of Ridley et al. [14], Engerer [16], Skartveit et al. [31] and Perez et al. [26]. The models selected for comparison are well known and are often considered as reference models in the literature. Indeed, these models are among the ones highlighted in the review by Gueyamrd and Ruiz-Arias [21] as the best performing separation models.

Table 3
Set of coefficients generated for each climate model, A - tropical, B - dry, C - temperate, D - continental and E – polar.

Coefficients	Climate model				
	A	B	C	D	E
β_0	0.29566	−1.7463	−0.083	0.67867	0.51643
β_1	−3.64571	−2.20055	−3.14711	−3.79515	−5.32887
β_2	−0.00353	0.01182	0.00176	−0.00176	−0.00196
β_3	−0.01721	−0.03489	−0.03354	−0.03487	−0.07346
β_4	1.7119	2.46116	1.40264	1.33611	1.6064
β_5	0.79448	0.70287	0.81353	0.76322	0.74681
β_6	0.00271	0.00329	0.00343	0.00353	0.00543
β_7	1.38097	2.30316	1.95109	1.82346	3.53205
β_8	−7.00586	−6.53133	−7.28853	−7.90856	−11.70755
β_9	6.35348	6.63995	7.15225	7.63779	10.8476
β_{10}	−0.00087	0.01318	0.00384	0.00145	0.00759
β_{11}	0.00308	−0.01043	0.02535	0.10784	0.53397
β_{12}	2.89595	1.73562	2.35926	2.00908	1.76082
β_{13}	1.13655	0.85521	0.83439	1.12723	0.41495
β_{14}	−0.0013	−0.0003	−0.00327	−0.00889	−0.03513
β_{15}	2.75815	2.63141	3.19723	3.72947	6.04835

4. Locally adjusted models

In order to assess the improvements of the model when regressed using a consistent dataset, two regressions were made for each station in Table 2: one using the NLS method, and another using the robust method. Both regressions produce similar results, but the robust method suffers less influence from the points that do not follow the main trend, like occasional outliers, resulting in a model with better similitude. The results for Florianopolis/Brazil are presented in Fig. 6, where it is observed that the robust method generates a model that predicts more accurately points in the clear sky region. The resulting models for all stations are available in the Supplementary Material.

The goodness of fit of each model was also calculated using the error indicators previously described, with which it is proven that the robust method is consistently better than the NLS method in two of the three indicators used, as shown in Fig. 7. The only case that the robust method does not perform consistently better than the NLS method is when comparing the bias between each model. This can be explained because the weight function (Eq. (6)) used on the robust method is not perfectly suited for the data, such a function is just an estimate of the variance in the data subsets, which introduces some bias to the model.

The results shown in Figs. 4, 6 and 7 illustrate how the use of the robust regression method results in models that are better adjusted to the data cumulative distribution and have smaller nRMSEs. For that reason, it was decided that all models to be proposed and discussed in the following sections will come from a robust regression of the measured data.

5. Climate zones models

The data for each major Köppen climate was merged as described in section 3.5.1, and submitted to the robust regression method to determine the model coefficients to each major zone. These coefficients are presented in Table 3.

After that, each climate model was reapplied for each station that belonged to that climate, but considering all qualified data (just a portion of the data of each station was used in the regression). Fig. 8 shows a comparison between the estimation errors from the proposed climate zone model and those from other models taken from the literature. As depicted in Fig. 8, the model proposed in this work produces errors comparable to the models selected by Gueymard and Ruiz-Arias [21], and even surpassing those models on most occasions.

Regarding the KSI indicator, it can be noticed that the climate models show values close to the obtained on the local optimization, which represents the lowest possible error that the proposed model can reach for each station. There are some exceptions, like “PTR”, “SOV” and “TAM” in climate B, and “FLO”, “IZA” and “ADL” in climate C, and “SON” in climate D. Those anomalies probably happen because the data reported by the stations do not behave as expected for the climate group to which it was assigned, indicating that a local climate may have been neglected, or the existence of a more appropriate methodology for assigning a model to a climate. In addition, for climate E, the new model is able to capture the behavior of the data much better than any of the other models to which it was compared. Comparing to the other models, the climate models perform significantly better; for instance, the Engerer model performs well only for climate C, presenting worse similitude for the other climate zones.

In terms of the nRMSE, Fig. 8b depicts that the climate models proposed herein present lower values of the nRMSE for all stations, compared to the other models analyzed. In fact, using the proposed climate models instead of the other models reduces the nRMSE by at least 5%. As expected, the local optimization presents the lowest errors, between 10 and 15%. It is worth mentioning that, for the local optimization, the weighted residual was used to calculate the nRMSE and nMBE, disregarding any influence of the outliers.

Fig. 8c presents the performance of the proposed models in terms of the nMBE. As expected, the local optimization models present bias close to zero. The climate models perform reasonably well, with errors close to zero for most stations, but in some cases, there are large deviations, for instance, “PTR”, “SBO”, “SOV” and “TAM” of climate B, and “IZA” on climate C. The Engerer model also presents good performance on climate C, but it shows large deviations on the other climates.

Another important feature that should be noted on Fig. 8 is that comparing the climate models against the other models (Engerer, BRL, Perez, Skartveit), the model proposed herein presents better statistical error indicators in almost every dataset and for all the metrics analyzed.

Furthermore, to check the hypotheses that some station does not behave like the major climate zones, models were developed for each three-letter Köppen climate zones to which there are data available. The same methodology used to create the models for the five major climate zones was used. The model coefficients of the 14 climate models are available in the Supplementary Material, along with the formal error analysis for each station.

Fig. 9 depicts the performance of the climate models clustered using the three-letter classification, namely Climate model - Full, in terms of the KSI error. Fig. 9 also presents the KSI errors for the major climate zones (Climate model – Simple), and for the model created for each station (Local model).

As observed in Fig. 9, there are some reduction on the KSI errors developed by the Climate model - Full, showing the benefits of subdividing the major climate zones, especially for climates B, D and E. For example, “PTR” and “TAM” present large values of KSI when using the Climate model for the B climate. However, when using the BSh climate model for “PTR”, and the BWh climate model for “TAM” a significant reduction on the error values is noticeable. Same effect is observed for climates EF and ET, where the climate models present errors significantly lower than the Climate mode - Simple, with exception of “BAR”.

Generally, the Climate models – Full present better similitude than the Climate models – Simple, especially for stations with climates B and E. The drawback of this approach is the number of

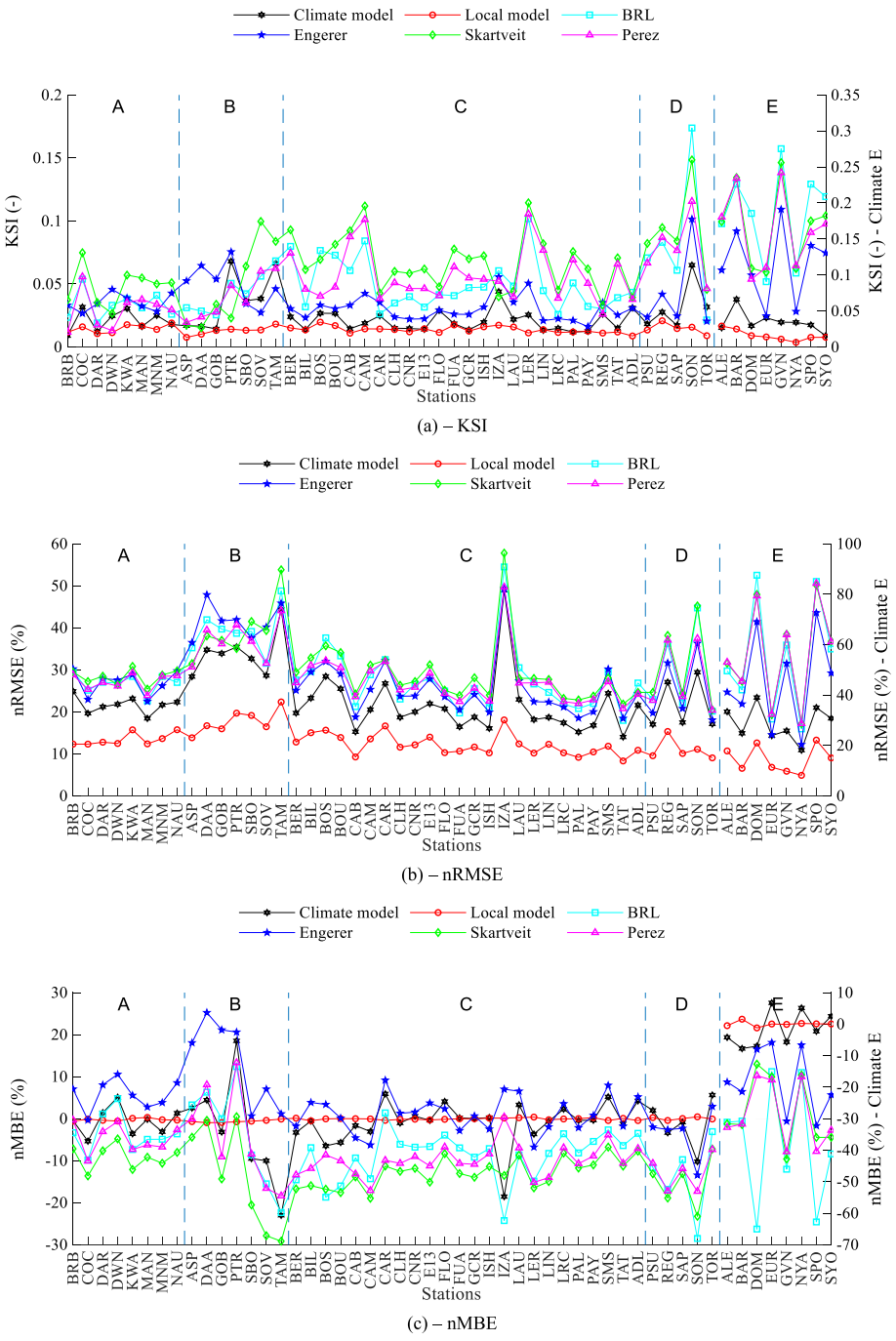


Fig. 8. Comparison between the KSI (a), nRMSE (b), and nMBE (c) error indicators for the climate (black), locally optimized (red), Engerer (darker blue), BRL (cyan) Perez (magenta) models. The stations are grouped by climate. The right axis represents the values for climate E only. The lines between each marker are presented just to ease the visualization.

models developed. Using the major climate classification five models were created, while using the three-letter classification 14 models were developed. Moreover, some of these models were created using data from only one station, which is not recommended, but it is a feasible solution considering the finite number of stations with reliable data available. The results of this analysis may indicate that applying a methodology such as cluster analysis could result in a better classification and a smaller number of climate groups.

5.1. Cross test

An additional "cross test" was carried out, where the models generated for a specific climate were applied to all stations, including the stations that did not belong to that specific climate. This is shown in Fig. 10, where, for instance, the "Climate A model" curve shows the results of climate A model applied for all the stations, in terms of the KSI, nRMSE and nMBE errors.

Fig. 10 shows that using the "wrong" climate model on a specific

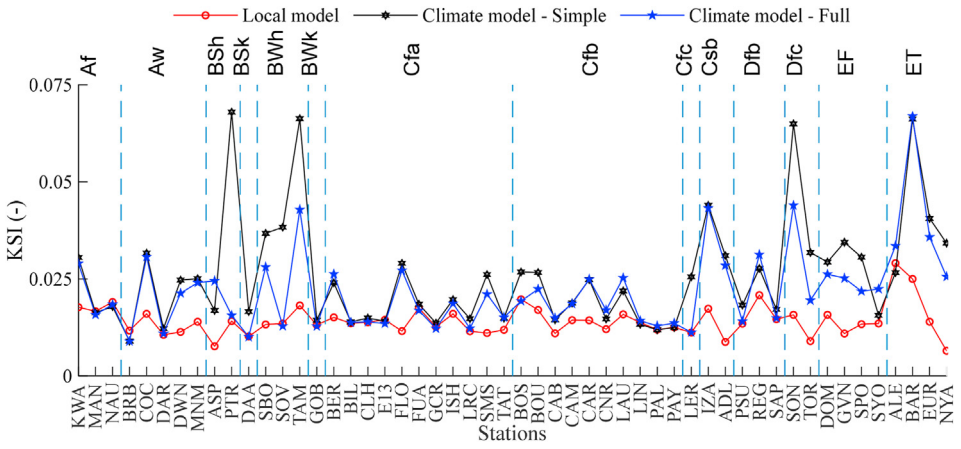


Fig. 9. KSI error indicator for Climate models - Simple (black) – created for the major climate zone, Climate models - Full (blue) – created for the three-letter classification and for the Local models (red) – created for each station. The lines between each marker are presented just to ease the visualization.

station results in significantly larger errors, showing the importance of developing climate models instead of a "universal" model. That conclusion is evident when analyzing the KSI and nMBE errors in Fig. 10a and c, which present large deviations on the cross test. Naturally, climate E (polar) has a unique behavior and performs completely different than other stations. Similarly, climate A and climate B models perform poorly for stations of climate C. On the other hand, some stations of climate B, "ASP" and "DAA" perform better when using climate A model, presenting lower KSI values. Same behavior is observed for station "CAR", "SMS" and "ADL" which are classified as climate C. Similar observations can be made when applying the climate D model for some station classified as climate C. For example, stations "IZA" and "LER", both present significantly lower KSI when the diffuse fraction is modeled using the climate D model. These observations corroborate the hypothesis that some stations do not behave as its climate classification, which suggest that other clustering method could produce better results.

In terms of nRMSE, Fig. 10b depicts that the climate E model should be only used for stations classified as climate E: When used for stations with other climate, it results in diffuse fraction with large nRMSE errors. Same observation is valid for the other climate models applied to stations with climate E. However, for the other cases as using climate model B for stations classified as climate C, it cannot be observed larger values of nRMSE. Although there are no differences on the nRMSE, the estimates of the diffuse fraction show larger bias (larger nMBE) and lower similitude (larger KSI).

Based on the results described above, it can be concluded that the Climate models present better results than the other "universal" models, delivering better diffuse fraction estimates. Furthermore, each climate model is unique and not interchangeable, being recommended that each one should be used only for its specific climate.

6. Conclusions

The objective of this study is to develop a separation model based on a logistical function using 1-min data that is reliable and accurate for different locations worldwide. To do so, an alternative version of the BRL-min model was proposed, introducing the hourly clearness index (K_T) as a new predictor, and analyzing the merit of the presence of all the predictors from the original BRL model in this new one. Instead of generating one set of coefficients for a "universal" model, a different set of coefficients is proposed for each climate in the Köppen climate classification. To develop the models, we have introduced the use of a robust nonlinear regression method in order to adjust the coefficients to the irradiance data, reducing the effects of outliers in the dataset. Compared to models developed with nonlinear least squares, the models developed using robust regression were better adjusted to the data cumulative distribution and had smaller nRMSEs. Therefore, the separation models proposed in this work were developed using robust regression instead of nonlinear least squares regression.

The climate specific separation models proposed herein present better performance than other models in the literature (BRL, Perez, Skartveit), including the Engerer model that was developed for minute data. Indeed, the climate specific models reduces the nRMSE by at least 5%, and a larger reduction is expected (between 10 and 15%) if local optimization of the separation model is carried out. The proposed models show better statistical error indicators in almost every dataset and for all metrics considered: KSI, nRMSE and nMBE. In general, the Climate models – Full present better similitude than the Climate models – Simple, especially for stations with climates B and E with a reduction of 0.025 in the KSI. However, the former approach implies in a larger number of models from which the user must choose. Based on the results, the proposed climate models provide significantly better results than the "universal" models.

It is worth mentioning that the climate models proposed herein are not interchangeable. For instance: applying the Climate A model in stations classified as Climate B or C mostly results in larger KSI and nMBE. However, there are some few locations that do not follow this trend, like some stations classified as climate C where the climate A model performs better than the climate C model itself. These results suggest that, in future works, classification methodologies other than the three-letter Köppen classification may be used, possibly resulting in a smaller number, and more accurate, separation models.

Data availability

Datasets related to this article can be found at https://doi.org/10.17632/b5ymcdcrrp.1, an open-source online data repository hosted at Mendeley Data (https://data.mendeley.com/datasets/b5ymcdcrrp/1).

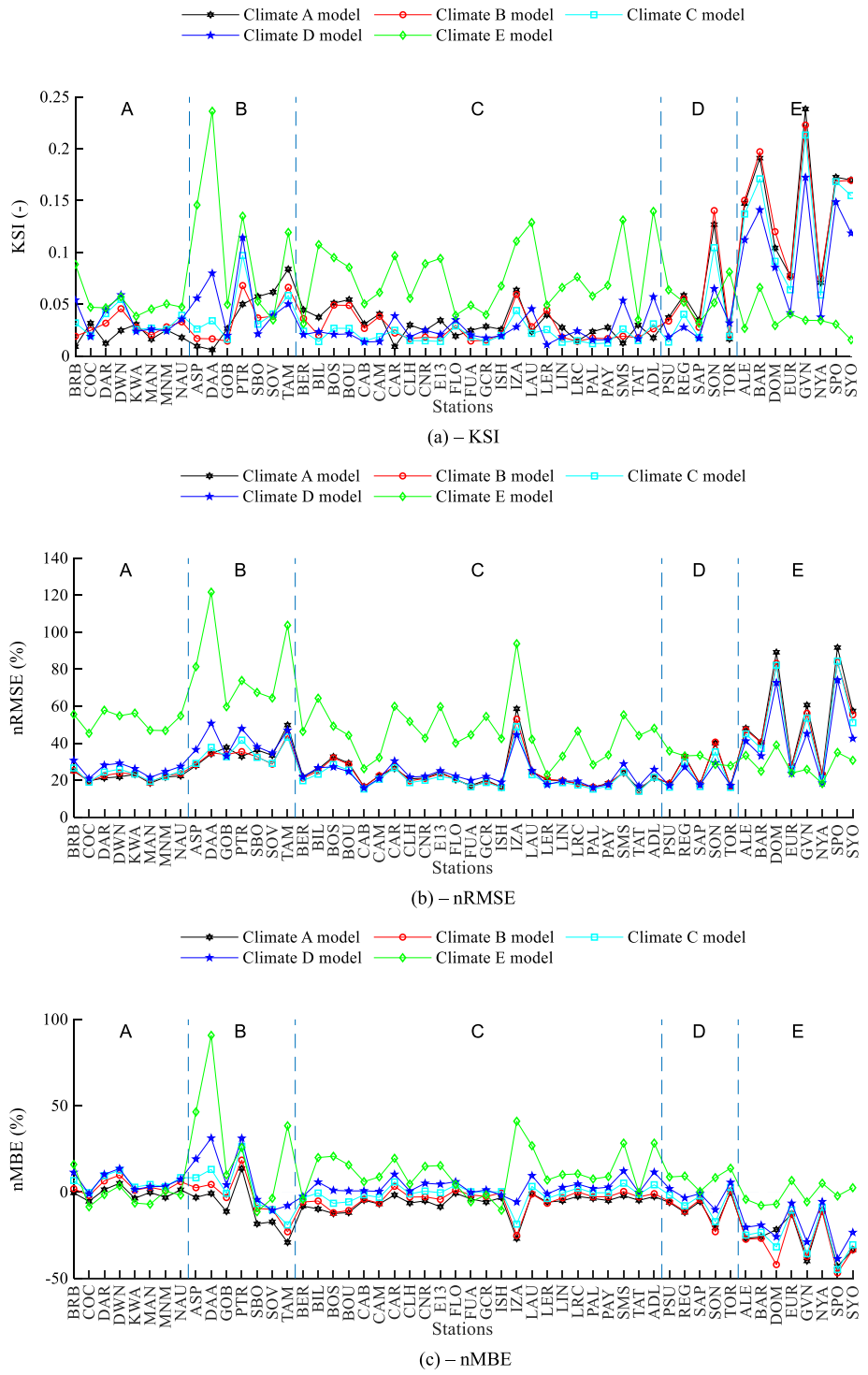


Fig. 10. Error analysis of the cross-test in terms of the KSI (a), nRMSE (b), and nMBE (c) error indicators for the five climate models proposed. The lines between each marker are presented just to ease the visualization.

CRediT authorship contribution statement

Allan R. Starke: Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Supervision, Writing – original draft, Writing – review & editing. **Leonardo F.L. Lemos:** Conceptualization, Methodology, Software, Formal analysis, Investigation, Resources, Writing – original draft, Writing – review & editing. **Cristian M. Barni:** Investigation, Methodology, Software, Resources, Visualization, Writing – original draft. **Rubinei D. Machado:** Methodology, Investigation, Resources, Writing – original draft. **José M. Cardemil:** Methodology, Writing – original draft, Writing – review & editing. **John Boland:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Sergio Colle:** Funding acquisition, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We would like to express our gratitude to all BSRN scientists and technicians working to keep the high quality of BSRN standards up and making the data available. Thanks as well to all the institutions supporting the BSRN project. We also thankfully appreciate the Australian Government Bureau of Meteorology for kindly providing irradiance data for the Adelaide station. We are also grateful for the atmospheric composition data provided by the Copernicus Atmosphere Monitoring Service (CAMS) project, implemented by the European Centre for Medium-Range Weather Forecasts (ECMWF). Furthermore, we wish to express our gratitude to the institutions that sponsored this work. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001 through scholarship for Leonardo Lemos. The authors appreciate the support from the project ANID/FONDAP/15110019 "Solar Energy Research Center"-SERC-Chile. And finally, the authors would like to thank ANEEL/Petrobras, which sponsored this research under contract PD-00553-0042/2016 and ANP/Petrobras, which sponsored this research under contract PT-151.01.12675, through the Department of Mechanical Engineering at the Federal University of Santa Catarina, Florianopolis, Brazil.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.renene.2021.05.108.

Nomenclature

AST	Apparent Solar Time (h)
CPI	Combined Performance Index (%)
d	Diffuse fraction calculated with measured data (−)
$\widehat{d}$	Diffuse fraction calculated with model estimated data (−)
I_0	Extraterrestrial irradiance at the top of the atmosphere (W/m^2)
Ig	Global irradiance on a horizontal surface (W/m^2)
K_{CSI}	Ratio between GHI and CSI (−)
k_T	Clearness index (−)
K_T	Daily clearness index (−)
KSI	Kolmogorov-Smirnoff test Integral (%)
MAD	Mean Absolute Difference (%)
MBE	Mean Bias Error (%)
OVER	Ratio of the
RMSE	Root Mean Square Error (%)

Acronyms

BOM	Australian Government Bureau of Meteorology
BRL	Boland - Ridley - Lauret
BSRN	Baseline Surface Radiation Network
CDF	Cumulative Distribuction Function
CEE	Cloud Enhancement Event
CPV	Concentrating Photovoltaics
CSI	Clear-sky Irradiance
CSP	Concentrating Solar Power
DIF	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
GHI	Global Horizontal Irradiance
GTI	Global Tilted Irradiance
INMET	National Institute of Meteorology
MACC II	Monitoring Atmospheric Composition and Climate Interim Implementation
PV	Photovoltaics
SONDA	Brazilian Environmental Data Organization System
TMY	Typical Meteorological Year

Greek

α	Solar altitude angle (deg)
β	Coefficients of separation model (−)
ψ	Persistence factor (−)

References

[1] D. Cano, J.M. Monget, M. Albuisson, H. Guillard, N. Regas, L. Wald, A method for the determination of the global solar radiation from meteorological satellite data, Sol. Energy 37 (1986) 31–39, https://doi.org/10.1016/0038-092X(86)90104-0.

[2] S.E. Rusen, Modeling and analysis of global and diffuse solar irradiation components using the satellite estimation method of HELIOSAT, Comput. Model. Eng. Sci. 115 (2018) 327–343, https://doi.org/10.3970/cmes.2018.00159.

[3] S. Ener Rusen, A. Hammer, B.G. Akinoglu, Estimation of daily global solar irradiation by coupling ground measurements of bright sunshine hours to satellite imagery, Energy 58 (2013) 417–425, https://doi.org/10.1016/j.energy.2013.05.062.

[4] S.E. Rusen, A. Hammer, B.G. Akinoglu, Coupling satellite images with surface measurements of bright sunshine hours to estimate daily solar irradiation on horizontal surface, Renew. Energy 55 (2013) 212–219, https://doi.org/10.1016/j.renene.2012.12.019.

[5] S. Ener Rusen, A. Konuralp, Quality control of diffuse solar radiation component with satellite-based estimation methods, Renew. Energy 145 (2020) 1772–1779, https://doi.org/10.1016/j.renene.2019.07.085.

[6] J. Polo, S. Wilbert, J.A. Ruiz-Arias, R. Meyer, C. Gueymard, M. Súri, L. Martín, T. Mieslinger, P. Blanc, I. Grant, J. Boland, P. Ineichen, J. Remund, R. Escobar, A. Troccoli, M. Sengupta, K.P. Nielsen, D. Renne, N. Geuder, T. Cebecauer, Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets, Sol. Energy 132 (2016) 25–37, https://doi.org/10.1016/j.solener.2016.03.001.

[7] B.Y.H. Liu, R.C. Jordan, The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, Sol. Energy 4 (1960) 1–19, https://doi.org/10.1016/0038-092X(60)90062-1.

[8] D.T. Reindl, W.A. Beckman, J.A. Duffie, Diffuse fraction corrections, Sol. Energy 45 (1990) 1–7.

[9] J.F. Orgill, K.G.T. Hollands, Correlation equation for hourly diffuse radiation on a horizontal surface, Sol. Energy 19 (1977) 357–359, https://doi.org/10.1016/0038-092X(77)90006-8.

[10] E. Paulescu, R. Blaga, A simple and reliable empirical model with two predictors for estimating 1-minute diffuse fraction, Sol. Energy 180 (2019) 75–84, https://doi.org/10.1016/j.solener.2019.01.029.

[11] D.G. Erbs, S.A. Klein, J.A. Duffie, Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation, Sol. Energy 28 (1982) 293–302, https://doi.org/10.1016/0038-092X(82)90302-4.

[12] E.L. Maxwell, A quasi-physical model for converting hourly global to direct normal insolation. http://rredc.nrel.gov/solar/pubs/PDFs/TR-215-3087.pdf, 1987, 35-46.

[13] J. Boland, B. Ridley, Models of diffuse solar fraction, Model. Sol. Radiat. Earth's Surf. Recent Adv. (2008) 193–219, https://doi.org/10.1007/978-3-540-77455-6_8.

[14] B. Ridley, J. Boland, P. Lauret, Modelling of diffuse solar fraction with multiple predictors, Renew. Energy 35 (2010) 478–483, https://doi.org/10.1016/j.renene.2009.07.018.

[15] J. Boland, J. Huang, B. Ridley, Decomposing global solar radiation into its direct and diffuse components, Renew. Sustain. Energy Rev. 28 (2013) 749–756, https://doi.org/10.1016/j.rser.2013.08.023.

[16] N.A. Engerer, Minute resolution estimates of the diffuse fraction of global irradiance for southeastern Australia, Sol. Energy 116 (2015) 215–237, https://doi.org/10.1016/j.solener.2015.04.012.

[17] H.K. Elminir, Y.A. Azzam, F.I. Younes, Prediction of hourly and daily diffuse fraction using neural network, as compared to linear regression models, Energy 32 (2007) 1513–1523, https://doi.org/10.1016/j.energy.2006.10.010.

[18] R. Aler, I.M. Galván, J.A. Ruiz-Arias, C.A. Gueymard, Improving the separation of direct and diffuse solar radiation components using machine learning by gradient boosting, Sol. Energy 150 (2017) 558–569, https://doi.org/10.1016/j.solener.2017.05.018.

[19] C.A. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode

radiometers. Part 1: impacts on global horizontal irradiance, Sol. Energy (2017) 1–14, https://doi.org/10.1016/j.solener.2017.05.004.
[20] C.A. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 2: performance of separation and transposition models for global tilted irradiance, Sol. Energy 153 (2017) 766–779, https://doi.org/10.1016/j.solener.2017.04.068.
[21] 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, https://doi.org/10.1016/j.solener.2015.10.010.
[22] Tobias Hirsch, Heiko Schenk, Dynamics of oil-based parabolic trough plants - a detailed transient simulation model, in: Proceedings of the SolarPACES 2010 conference. SolarPACES. SolarPACES 2010, 21-24 Sep 2010, Perpignan, France, 2010.
[23] Tobias Hirsch, Heiko Schenk, Norbert Schmidt, Richard Meyer, Dynamics of oil-based parabolic trough plants - impact of transient behaviour on energy yields, in: Proceedings of the SolarPACES 2010 conference. SolarPACES. SolarPACES 2010, 21-24 Sep 2010, Perpignan, France, 2010.
[24] J. Luoma, J. Kleissl, K. Murray, Optimal inverter sizing considering cloud enhancement, Sol. Energy 86 (2012) 421–429, https://doi.org/10.1016/j.solener.2011.10.012.
[25] 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, https://doi.org/10.1016/j.solener.2016.02.011.
[26] 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, https://doi.org/10.1016/S0038-092X(02)00122-6.
[27] J. Boland, J. Huang, B. Ridley, Decomposing global solar radiation into its direct and diffuse components, Renew. Sustain. Energy Rev. 28 (2013) 749–756, https://doi.org/10.1016/j.rser.2013.08.023.
[28] K.G.T. Hollands, S.J. Crha, An improved model for diffuse radiation: correction for atmospheric back-scattering, Sol. Energy 38 (1987) 233–236, https://doi.org/10.1016/0038-092X(87)90044-2.
[29] R.R. Perez, P. Ineichen, E.L. Maxwell, R.D. Seals, A. Zelenka, Dynamic global-to-direct irradiance conversion models, Build. Eng. (1992) 354–369.
[30] R. Perez, R. Seals, A. Zelenka, P. Ineichen, Climatic evaluation of models that predict hourly direct irradiance from hourly global irradiance: prospects for performance improvements, Sol. Energy 44 (1990) 99–108, https://doi.org/10.1016/0038-092X(90)90071-J.
[31] 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, https://doi.org/10.1016/S0038-092X(98)00067-X.
[32] W. Yao, Z. Li, Y. Lu, F. Jiang, C. Li, New models for separating hourly diffuse and direct components of global solar radiation, in: Proc. 8th Int. Symp. Heating, Vent. Air Cond., 2014, pp. 653–663.
[33] L.F.L. Lemos, A.R. Starke, J. Boland, J.M. Cardemil, R.D. Machado, S. Colle, Assessment of solar radiation components in Brazil using the BRL model, Renew. Energy 108 (2017) 569–580, https://doi.org/10.1016/j.renene.2017.02.077.
[34] M. Hofmann, G. Seckmeyer, A new model for estimating the diffuse fraction of solar irradiance for photovoltaic system simulations, Energies 10 (2017), https://doi.org/10.3390/en10020248.
[35] A.R. Starke, L.F.L. Lemos, J. Boland, J.M. Cardemil, S. Colle, Resolution of the cloud enhancement problem for one-minute diffuse radiation prediction, Renew. Energy 125 (2018) 472–484, https://doi.org/10.1016/j.renene.2018.02.107.
[36] Y. Zhou, D. Wang, Y. Liu, J. Liu, Diffuse solar radiation models for different climate zones in China: model evaluation and general model development, Energy Convers. Manag. 185 (2019) 518–536, https://doi.org/10.1016/j.enconman.2019.02.013.
[37] Y. Zhou, Y. Liu, Y. Chen, D. Wang, General models for estimating daily diffuse solar radiation in China: diffuse fraction and diffuse coefficient models, Energy Procedia 158 (2019) 351–356, https://doi.org/10.1016/j.egypro.2019.01.101.
[38] F. Li, Y. Lin, J. Guo, Y. Wang, L. Mao, Y. Cui, Y. Bai, Novel models to estimate hourly diffuse radiation fraction for global radiation based on weather type classification, Renew. Energy 157 (2020) 1222–1232, https://doi.org/10.1016/j.renene.2020.05.080.
[39] J. Fan, L. Wu, F. Zhang, H. Cai, X. Ma, H. Bai, Evaluation and development of empirical models for estimating daily and monthly mean daily diffuse horizontal solar radiation for different climatic regions of China, Renew. Sustain. Energy Rev. 105 (2019) 168–186, https://doi.org/10.1016/j.rser.2019.01.040.
[40] J.J. Michalsky, The astronomical almanac's algorithm for approximate solar position (1950-2050), Sol. Energy 40 (1988) 227–235, https://doi.org/10.1016/0038-092X(88)90045-X.
[41] I. Reda, A. Andreas, Solar Position Algorithm for Solar Radiation Applications (Revised), 2008.
[42] ECMWF, CAMS Reanalysis of Global Atmospheric Composition (2003 - 2018, 2019.
[43] M.R. Mahmoudi, M. Mahmoudi, E. Nahavandi, Testing the difference between two independent regression models, Commun. Stat. Theor. Methods 45 (2016) 6284–6289, https://doi.org/10.1080/03610926.2014.960584.
[44] Wrmc-BSRN, World Radiation Monitoring Center (WRMC) - Baseline Surface Radiation Network (BSRN), WRMC-BSRN, 2018. https://bsrn.awi.de/.
[45] A. Driemel, J. Augustine, K. Behrens, S. Colle, C. Cox, E. Cuevas-Agulló, F.M. Denn, T. Duprat, M. Fukuda, H. Grobe, M. Haeffelin, N. Hyett, O. Ijima, A. Kallis, W. Knap, V. Kustov, C.N. Long, D. Longenecker, A. Lupi, M. Maturilli, M. Mimouni, L. Ntsangwane, H. Ogihara, X. Olano, M. Olefs, M. Omori, L. Passamani, E.B. Pereira, H. Schmithüsen, S. Schumacher, R. Sieger, J. Tamlyn, R. Vogt, L. Vuilleumier, X. Xia, A. Ohmura, G. König-Langlo, Baseline Surface Radiation Network (BSRN): structure and data description (1992–2017), Earth Syst. Sci. Data Discuss. (2018) 1–17, https://doi.org/10.5194/essd-2018-8.
[46] BOM, Australian Government Bureau of Meteorology, 2021. http://www.bom.gov.au/. (Accessed 5 May 2021).
[47] P. Ineichen, A broadband simplified version of the Solis clear sky model, Sol. Energy 82 (2008) 758–762, https://doi.org/10.1016/j.solener.2008.02.009.
[48] C.A. Gueymard, D. Thevenard, Monthly average clear-sky broadband irradiance database for worldwide solar heat gain and building cooling load calculations, Sol. Energy 83 (11) (2009) 1998–2018, https://doi.org/10.1016/j.solener.2009.07.011.
[49] H.W. Chen, Köppen climate classification. http://hanschen.org/koppen/, 2018. (Accessed 10 July 2019).
[50] D. Chen, H.W. Chen, Using the Köppen classification to quantify climate variation and change: an example for 1901–2010, Environ. Dev. 6 (2013) 69–79, https://doi.org/10.1016/j.envdev.2013.03.007.
[51] S. Younes, R. Claywell, T. Muneer, Quality control of solar radiation data: present status and proposed new approaches, Energy 30 (2005) 1533–1549, https://doi.org/10.1016/j.energy.2004.04.031.
[52] P.J. Huber, E.M. Ronchetti, Robust Statistics, John Wiley and Sons Ltd, 2009.
[53] R.A. Maronna, D.R. Martin, V.J. Yohai, Robust Statistics: Theory and Methods, John Wiley & Sons, 2006.
[54] H. Riazoshams, H. Midi, G. Ghilagaber, Robust Nonlinear Regression: with Applications Using R, John Wiley & Sons, 2018.
[55] J.M. Wooldridge, Introductory Econometrics: A Modern Approach, fifth ed., South Western Educational Publishing, 2012.
[56] MathWorks, MATLAB R2018a. https://www.mathworks.com/help/stats/fitnlm.html, 2018. (Accessed 22 March 2019).
[57] W. Dumouchel, F. O'Brien, in: A. Buja, P.A. Tukey (Eds.), Computing and Graphics in Statistics, Springer-Verlag New York, Inc., New York, NY, USA, 1991, pp. 41–48.
[58] P.W. Holland, R.E. Welsch, Robust regression using iteratively reweighted least-squares, Commun. Stat. Theor. Methods 6 (1977) 813–827, https://doi.org/10.1080/03610927708827533.
[59] C.A. Gueymard, A review of validation methodologies and statistical performance indicators for modeled solar radiation data: towards a better bankability of solar projects, Renew. Sustain. Energy Rev. 39 (2014) 1024–1034, https://doi.org/10.1016/j.rser.2014.07.117.