Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

# Assessment of employing different cloud cover data sources to model the Brazilian solar energy potentiality

## Avaliação do emprego de diferentes fontes de dados de cobertura de nuvens para modelar a potencialidade da energia solar brasileira

*Emerson D. Oliveira*[1], *Thomás Rocha Ferreira*[2], *Carlos Denyson da Silva Azevedo*[3], *Mario de Miranda Vilas Boas Ramos Leitão*[1], *Maria Luciene Melo*[2]

## ABSTRACT

A simplified atmospheric transmittance model based on the Beer-Lambert law was utilized to analyze surface solar radiation (SSR) variability based on different sources of cloud cover datasets (CMIP6, ERA5, NCEP, ISCCP, and EUMETSAT). This study evaluated the performance of various modeled SSR datasets against observed data from the Brazilian Daily Weather Gridded Data (BR-DWGD) over the period from 1983 to 2009. Contour plots of annual average SSR from the five modeled datasets were compared with BR-DWGD observations, revealing spatial agreements and discrepancies. The highest SSR values were consistently observed in the Brazilian semi-arid Northeast, while the Amazon region exhibited the lowest values. In the analysis of annual averages, the International Satellite Cloud Climatology Project (ISCCP) demonstrated the closest agreement with BR-DWGD, while the National Center for Environmental Prediction (NCEP) showed the most significant deviations. Root mean square error (RMSE) analysis highlighted seasonal variability in model performance, with the Coupled Model Intercomparison Project Phase 6 (CMIP6) and the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) performing best during equinoxes, and ISCCP showing the lowest annual RMSE (16.9 $Wm^{-2}$). Hierarchical clustering further grouped EUMETSAT and CMIP6 as the most similar and accurate datasets, while NCEP remained the least consistent. Global horizontal irradiance maps corroborated SSR patterns, with higher values in the Northeast and lower values in the Amazon and Southern regions. These findings underscored the importance of dataset selection for accurate SSR modeling in Brazil, with ISCCP, EUMETSAT, and CMIP6 emerging as the most reliable options.

**Keywords:** cloud cover; solar radiation; reanalysis data; satellite data; CMIP6 models.

## RESUMO

Um modelo simplificado de transmitância atmosférica baseado na lei de Beer-Lambert foi utilizado para analisar a variabilidade da radiação solar de superfície (SSR) com base em diferentes fontes de dados de cobertura de nuvens (CMIP6, ERA5, NCEP, ISCCP e EUMETSAT). Este estudo avaliou o desempenho de diversos conjuntos de dados modelados de SSR em comparação com dados observados do Brazilian Daily Weather Gridded Data (BR-DWGD) no período de 1983 a 2009. Gráficos de contorno da média anual de SSR dos cinco conjuntos de dados modelados foram comparados com as observações do BR-DWGD, revelando concordâncias e discrepâncias espaciais. Os maiores valores de SSR foram consistentemente observados no semiárido do nordeste brasileiro, enquanto a região amazônica apresentou os menores valores. Na análise das médias anuais, o Projeto Internacional de Climatologia de Nuvens por Satélite (ISCCP) demonstrou a maior concordância com o BR-DWGD, enquanto o Centro Nacional de Previsão Ambiental (NCEP) apresentou os desvios mais significativos. A análise do erro quadrático médio (RMSE) destacou a variabilidade sazonal no desempenho dos modelos, com o Projeto de Intercomparação de Modelos Acoplados Fase 6 (CMIP6) e a Organização Europeia para a Exploração de Satélites Meteorológicos (EUMETSAT) apresentando os melhores resultados durante os equinócios, e o ISCCP exibindo o menor RMSE anual (16,9 $Wm^{-2}$). O agrupamento hierárquico em seguida classificou o EUMETSAT e o CMIP6 como os conjuntos de dados mais semelhantes e precisos, enquanto o NCEP permaneceu como o menos consistente. Mapas de irradiância global horizontal corroboraram os padrões de SSR, com valores mais altos no nordeste e mais baixos nas regiões da Amazônia e do sul. Esses resultados reforçam a importância da seleção do conjunto de dados para uma modelagem precisa da SSR no Brasil, com o ISCCP, EUMETSAT e CMIP6 emergindo como as opções mais confiáveis.

**Palavras-chave:** cobertura de nuvens; radiação solar; dados de reanálise; dados de satélite; modelos do CMIP6.

[1]Universidade Federal do Vale do São Francisco – Juazeiro (BA), Brazil.
[2]Universidade Federal de Alagoas – Maceió (AL), Brazil.
[3]Instituto Nacional de Pesquisas Espaciais – Cachoeira Paulista (SP), Brazil.
Corresponding author: Emerson D. Oliveira – Avenida Antônio Carlos Magalhães, 510, Santo Antônio – CEP: 48902-300 – Juazeiro (BA), Brazil.
E-mail. emerson.oliveira@univasf.edu.br
Funding: none.
Conflicts of interest: the authors declare no conflicts of interest.
Received on: 02/07/2025. Accepted on: 05/03/2025.
https://doi.org/10.5327/Z2176-94782451

## Introduction

In recent years, some researchers investigated the cloud cover amount disagreements from different data sources (e.g., Free et al., 2016; Chakraborty and Lee, 2021; Wu et al., 2023). For instance, according to Free et al. (2016), reanalysis products generally capture the variability and trends of cloud cover but tend to underestimate cloud amounts when compared to ground-based observations. Likewise, Wu et al. (2023) compared monthly total and high cloud cover from four atmospheric reanalysis datasets (ERA-Interim, ERA5, MERRA-2, and NCEP) with the Moderate Resolution Imaging Spectroradiometer (MODIS) retrievals during 2001–2018, and found generally consistent spatial and temporal distributions with MODIS. However, discrepancies exist, especially over polar regions and the Tibet Plateau (Wu et al., 2023). Furthermore, Chakraborty and Lee (2021) analyzed the differences in solar radiation considering reanalysis data, satellite products, and simulations; in general, their study revealed systematic positive biases in total shortwave radiation, which were largely attributed to the intermodel variability from cloud cover simulations.

In this context, some studies using the Coupled Model Intercomparison Project (CMIP) scenarios to estimate how climate changes could impact the photovoltaic (PV) potential were performed using the ensemble technique (Wild et al., 2015; Danso et al., 2022; Ha et al., 2023; Isaza et al., 2023; Nwkolo et al., 2023). For example, according to Danso et al. (2022), the CMIP6 scenarios suggest a decrease in solar PV potential in West Africa, which is attributed primarily to a reduction in surface solar radiation (SSR); this, in turn, is due to changes in cloud cover and aerosol depth. In addition, Ha et al. (2023) study highlights the importance of climate change mitigation to support the sustainable development of solar energy, where the authors compare the future potential of solar power output under different climate attributes, technology types, climate projection models (CMIP5 *vs.* CMIP6), and emission scenarios (Shared Socioeconomic Pathways—SSP).

Nowadays, the most common technologies for solar panels are monocrystalline silicon (Mono-Si), polycrystalline silicon (Poly-Si), and thin-film solar panels (Vunnam et al., 2021). Inevitably, there is pressure toward a transition to clean and sustainable energy production. Consequently, many countries worldwide are investing in renewable energy sources to reduce fossil fuel consumption, which is an effort to decrease pollution and climate change effects. Considering this, Silva et al. (2021) analyzed the economic feasibility of expanding a previously installed PV system in a Brazilian high school institution with multicampus. In this study, the authors explored aspects like the benefit-to-cost ratio and time payback from different installation scenarios, where installation scenarios 1 and 2 demanded four and eight years as time payback, respectively. To learn about these scenarios, read Silva et al. (2021).

Therefore, accurate estimation of surface shortwave irradiance is essential for climate and weather studies, as well as for the operation of PV power plants. Simplified atmospheric transmittance models, such as those based on the Beer-Lambert law (Bird and Hulstrom, 1981; Bird and Riordan, 1986), can be particularly useful in reducing uncertainties in this process. Previous works (Oliveira et al., 2019; Oliveira and Fernandez, 2020) have demonstrated the potential of using simplified parameterizations to account for cloud cover effects, thereby improving the representation of downward shortwave radiation.

Beyond the role of clouds and the ozonosphere, aerosols also affect the solar radiation that reaches the ground. As revealed by Wild et al. (2005), between the 1950s and 1980s, there was a global decline of roughly 4–6% in SSR attributable to aerosol pollution. Isaza et al. (2023) stated that, depending on their size, microphysical and optical characteristics, aerosols can perform as scatterers (e.g., sulfates and sea salt, producing a cooling effect), absorbers (e.g., black carbon, producing a warming effect), or moderate absorbers (e.g., dust and organic matter). Undoubtedly, we have to consider that forecasting future climate conditions, particularly for the long term (2100, for example), is influenced by numerous sources of uncertainty in the model structure, such as aerosol-cloud interaction, cloud physics, convection schemes, and ocean-atmosphere interaction (Collins et al., 2013). According to Wang et al. (2021), SSP3-7.0, a pathway with weak global warming mitigation, aligns better with current carbon emission trends and political realities. However, the authors also clarify that due to inaccuracies in observed data, the CMIP6 models underestimated the recent decline in anthropogenic aerosol over China, affecting the temperature response simulated for the next few decades.

Furthermore, Isaza et al. (2023) assessed the cost of future solar power plants, revealing substantial disparities between the SSP1-2.6 and SSP3-7.0 scenarios. The study found that cost savings in the SSP1-2.6 scenario compared to the SSP3-7.0 scenario could amount to 12.4 billion US dollars in one year or 372 billion US dollars over the PV plants' 30-year lifetimes. Indeed, climate change forces a quick adaptation process, affecting many different areas, such as agronomy. For instance, according to Souza et al. (2021), many studies investigating the microclimate within protected environment agriculture, like net houses, have been developed to improve agricultural productivity in global warming scenarios.

Dubey et al. (2013) reported that, in the context of global warming, solar panel efficiency is expected to diminish with rising temperatures, including increases in average temperature and occurrences of extremely high temperatures. Indeed, it represents an enormous challenge for future solar energy plants, considering the current technology constraints. If we consider a pessimistic scenario, it will probably encourage the development of new technologies over the following decades, which will be crucial to maintaining renewable energy transition goals.

In light of the aforementioned concerns, this study aimed to analyze the sensitivity of modeling solar radiation at the surface level by

considering different gridded cloud cover data sources. As described previously, before solar radiation reaches the ground, it interacts with the atmosphere (aerosols, ozone, total cloud amount, water vapor, etc.). Inevitably, due to the interactions, there are uncertainties in estimating how much solar radiation is absorbed or reflected into space, which is modeled independently in atmospheric models. To overcome this issue, we modeled the SSR using a simplified approach, investigating how the total cloud amount affects it. Therefore, we assumed different cloud cover data sources, such as reanalysis, satellite products, and the CMIP6 ensemble dataset.

Additionally, we intended to ensure that the estimates of SSR were computed using the same methodology, allowing us to individually evaluate the impact of each cloud cover dataset on solar energy power estimations. Our study area covered the Brazilian territory, providing error estimations from each cloud dataset. The findings presented herein are significant for the scientific community, the solar energy market, and the sustainable energy transition efforts.

## Methods

### Solar radiation model

Our empirical model for estimating solar radiation on the ground was inspired by some classical papers (Berger, 1978; Bird and Hulstrom, 1981; Bird and Riordan, 1986). We adopted four steps to model SSR, mixing clear sky and cloudiness conditions, as presented in Figure 1. Thus, the first computes the solar radiation at the top of the atmosphere (SRTOA). The second, the third, and the fourth estimates the solar radiation after passing the ozone layer, the atmospheric water content, and the possible existing clouds, respectively. The SRTOA was computed using the Daily INSOlation (DINSOL) software application (Oliveira, 2023), while the ozone and precipitable water data were obtained from the National Center for Environmental Prediction (NCEP).

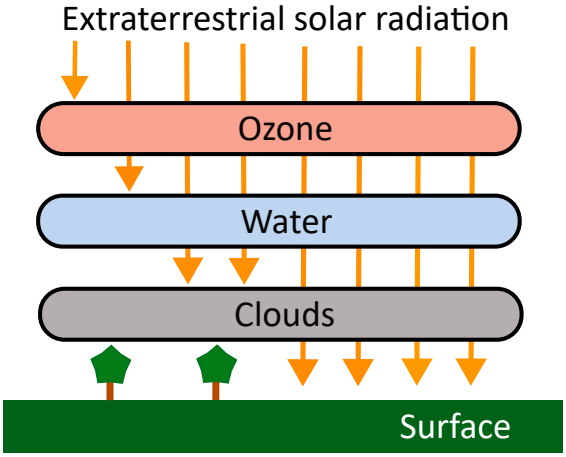


**Figure 1 – Representation of the simplified solar radiation model considering the transmittance in three different layers.**

Essentially, we are using a simplified Beer-Lambert law approach to estimate the transmittance of the ozone and water content, which is a typical clear sky model (see Equations 1 and 2). On the other hand, considering the transmittance under cloudiness conditions, we adapted the parameterization employed by Oliveira et al. (2019) and Oliveira and Fernandez (2020) (Equation 3). Therefore, our shortwave solar radiation model suitably emulates net surface irradiance, presenting a good fit for this study's aims.

$$\tau_o = e^{(-k_o m)} \quad (1)$$

Where:
$\tau_o$ = ozone transmittance;
$k_o$ = absorption coefficient; and
$m$ = thickness of the ozone.

$$\tau_w = e^{(-k_w m)} \quad (2)$$

Where:
$\tau_w$ = water transmittance; and
$k_w$ = absorption coefficient.
$m$ = precipitable water thickness.

Note that the ozone and precipitable water data were obtained from the NCEP reanalysis.

In Equation 3, we have the parameterization to estimate the cloud transmittance:

$$\tau_{cld} = (1 - CLD\alpha_{cld}) \quad (3)$$

Where:
$\tau_{cld}$ = cloud transmittance;
$CLD$ = cloud cover amount; and
$\alpha_{cld}$ = average cloud albedo parameter.

As the last step, we employed Equation 4 to compute the surface solar irradiance, which is estimated by multiplying the SRTOA by the continuum attenuation of the solar radiation, in other words, the transmittance product.

$$I = I_0\ (\tau_o\ \tau_w \tau_{cld}) \quad (4)$$

$I$ = surface solar irradiance;
$I_0$ = solar radiation at the top of the atmosphere;
$\tau_o$ = ozone transmittance;
$\tau_w$ = water transmittance; and
$\tau_{cld}$ = cloud transmittance.

Table 1 shows the constant values employed in Equations 1–3. Notice that accurate models adopt different coefficients for each wavelength, while our simplified model assumes a more general approach.

To conclude this subsection, Figure 2 shows contour plots of the four variables adopted to estimate SSR from our method. Note that Figure 2A is the SRTOA, Figure 2B represents the ozone layer (NCEP) in Dobson units (DU), Figure 2C is the precipitable water (NCEP) layer, and Figure 2D is the percentage of cloud amount obtained from the International Satellite Cloud Climatology Project (ISCCP).

## Cloud cover samples

As mentioned in the introduction, this research investigated the impact of employing different cloud data sources to estimate SSR in the context of solar energy potentiality. Hence, we chose five cloud datasets: CMIP6 models, ERA5 and NCEP reanalysis, and European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) and ISCCP satellites. It is worth mentioning that the ISCCP was part of the World Climate Research Project (WCRP), a pioneering initiative aimed at inferring cloud properties, global distribution, and seasonal and interannual variations. The ISCCP data were obtained from the following webpage: https://isccp.giss.nasa.gov, last access 04/11/2024.

**Table 1 –Values of the ozone and water absorption coefficients and the average cloud albedo parameter employed in Equations 1–3.**

| $k_o$ | $k_w$ | $\alpha_{cld}$ |
|---|---|---|
| 0.35 $cm^{-1}$ | 0.01 $cm^{-1}$ | 0.70 |

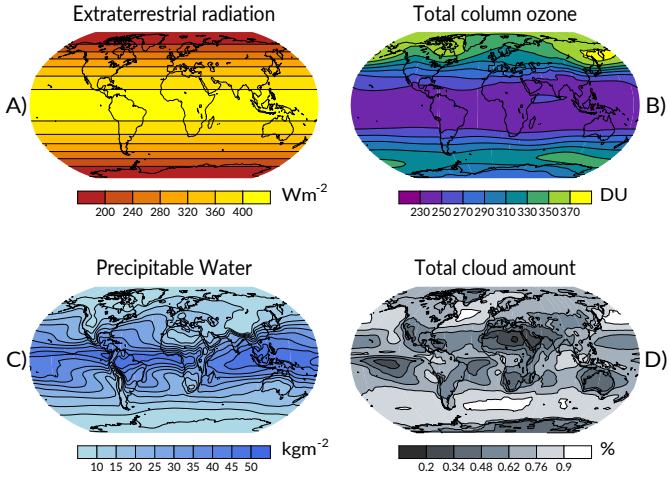


**Figure 2 – Contour plots representing the annual mean of each variable (averaged) adopted in our surface solar radiation model: (A) Solar radiation at the top of the atmosphere computed using the Daily INSOlation program; (B) Total column ozone in Dobson units obtained from National Center for Environmental Prediction; (C) Precipitable water obtained from National Center for Environmental Prediction; and (D) Total cloud amount obtained from the International Satellite Cloud Climatology Project.**

The European Centre for Medium-Range Weather Forecasts (ECMWF) reanalysis (ERA5) and the EUMETSAT data were obtained online from the Climate Data Store platform (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=form, last access: 04/11/2024). Moreover, the NCEP data was obtained from the webpage: https://psl.noaa.gov/data/gridded/data.ncep.reanalysis2.html, last access 04/11/2024. Notice that all CMIP6 data were obtained from the Earth System Grid Federation (https://esgf-data.dkrz.de/search/cmip6-dkrz/, last access: 04/11/2024).

In this study, we adopted eight CMIP6 models: ACCESS-ESM-1-5, CAN-ESM5, CAS-FGOALS-g3, INM-CM5-0, IPSL-CM6A-LR, MIROC6, MPI-ESM1-2, and NCAR-CESM2. Likewise, these data represent the experiment id "hist," that is, the Earth's climate historical reconstruction (1850–2014). All data have the variant label "r4i1p1f1", representing the same realization, initialization, physical, and forcing conditions. We employed the ensemble technique for all models, assuming just one average data product for this study.

## Time series and spatial resolution

We assumed a time series from 1983 to 2009, according to the ISCCP dataset availability. Although the other data had a more extensive time range, we defined the ISCCP time range to perform the assessment, in other words, establishing the same conditions during the evaluation. Moreover, all the data obtained were initially monthly, but for this study, we calculated the annual and seasonal averages based on the climatology of this period (1983–2009). Regarding the spatial resolution, we interpolated each one of them to work with a standard grid. To perform data interpolations, we adopted the Climate Data Operator (CDO). For example, the bilinear interpolation was considered mostly with the CMIP6 models, assuming a 144×72 gridded setup at a global scale. Ultimately, the final gridded data covers Brazil's longitudinal and latitudinal ranges: 74.85°W–34.85°W and 33.85°S–5.35°N. Notice that the spatial resolution is about 0.1°, which is given from 391 points of longitude and 393 points of latitude.

## Model evaluation

To analyze which cloud cover dataset provides the best adjustment for our surface radiation model, we assumed the gridded data provided by Xavier et al. (2022), the Brazilian Daily Weather Gridded Data (BR-DWGD) for validation. These datasets are based on observational data registered in Brazilian weather stations managed by the National Institute of Meteorology (INMET, *Instituto Nacional de Meteorologia*) and the National Water and Sanitation Agency (ANA, *Agência Nacional de Águas e Saneamento Básico*). Xavier et al. (2022) interpolated these data to create regular gridded time series data, including air temperature, relative humidity, wind speed, precipitation, evapotranspiration, and solar radiation. Due to the high accuracy of these datasets, they are ideal for model assess-

ments. Thus, we adjusted our modeled SSR data to match the spatial resolution of the BR-DWGD.

For proper evaluation, we used two strategies. The first was plotting the contour fields and comparing the modeled and observed data. The second was computing the root mean square error (RMSE) to estimate the impact of using different cloud cover data sources to model the SSR. Hence, Equation 5 presents the RMSE Equation (Oliveira, 2023):

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(S_i - O_i)^2} \quad (5)$$

Where:
$N$ = number of gridded elements;
$S_i$ = simulated data; and
$O_i$ = observational data.

Additionally, we applied hierarchical clustering to assess the similarities among the different cloud datasets. This was achieved by computing the Euclidean distance matrix between all RMSE values, employing the "complete" linkage method for clustering. The dendrogram was constructed by defining two distinct groups, which allowed us to visualize the relationships and similarities between the datasets.

## Solar energy potentiality

We computed a simple approach to estimate the Brazilian Photovoltaic Power Potential (PPP) for all datasets used in this study. Notice that the evaluation process assumed the annual average given in $Wm^{-2}$. Consequently, to compute the solar energy potentiality, we only need to multiply it by 24 hours and divide by 1,000 to obtain the result in $kWh/m^2$. It is worth mentioning that the panel's efficiency is strongly variable, where the panel area significantly impacts this computation as well as the kind of technology involved (e.g., Mono-Si, Poly-Si, and Thin-Film). Therefore, to reduce uncertainties, it is interesting to consider different Brazilian PPP estimations. Nevertheless, even though the BR-DWGD data can be considered accurate in most Brazilian regions, the data on the North of Brazil is less reliable than in other areas due to the lower weather station density. Inevitably, the modeled data might be considered a reasonable option for Brazilian PPP estimations.

## Results

Figure 3 illustrates the contour plots of modeled SSR (Figures 3A–3E) and the observed SSR based on BR-DWGD (Figure 3F), both representing the annual average from 1983 to 2009. Figure 3A exhibits the SSR data modeled using an ensemble dataset from CMIP6 data clouds, where the contour lines presented important agreements compared to the BR-DWGD (Figure 3F). The highest values, as expected (Pereira et al., 2017), were found in the Brazilian semi-arid region, mainly in the Northeast of the country.

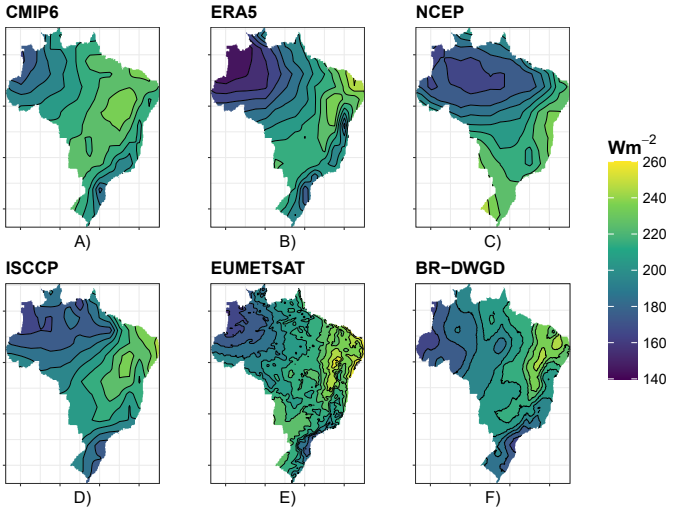


**Figure 3 – Annual contour plots of the surface solar radiation from modeled (A–E) and observed dataset (F).**

Analyzing the ERA5 plot (Figure 3B), we realized another important agreement over the Amazon region, where we had low values of SSR. It is worth mentioning that the ERA5, compared to other modeled SSRs, presented the lowest values. Alternatively, the NCEP (Figure 3C) presented the most notable contour line differences to BR-DWGD, taking into consideration all the modeled SSRs. Like CMIP6 and ERA5, the ISCCP data (Figure 3D) showed important agreements with BR-DWGD. For example, we could mention the coastline of the Brazilian South, where low SSR values were found in both plots. In addition, the modeled SSR was analyzed using EUMETSAT (Figure 3E) cloud data. This map clearly evidenced more disturbed contour lines, but spatially, the fields agreed well with BR-DWGD. This disturbance is due to the EUMETSAT data being provided in high resolution, about 0.5°. Overall, the EUMETSAT revealed the highest modeled values compared to other fields and showed high SSR values over the Brazilian Northeast coastline.

Figure 4 exhibits the differences between the modeled and observed SSR by showing the great and minor discrepancies for each cloud dataset. For example, in Figure 4A, we observe CMIP6 overestimations in most Brazilian regions, where the only exception lay in the Northeast region. Regarding the ERA5 (Figure 4B), the differences were mostly close to zero, except for the Amazon region (in the North), where we had more prominent underestimations. On the other hand, our model indicated a more significant disparity when using the NCEP data (Figure 4C), which was properly discussed in the next section. Likewise, considering all cloud datasets, the ISCCP (Figure 4D) presented the lowest differences, suggesting it is a good fit to model the SSR. The more prominent difference was strictly in the North, which is negative close to the Amazon basin and positive over the Acre state

(extreme West of the country). Thus, in the last plot, we analyzed the differences by considering the EUMETSAT (Figure 4E). The EUMETSAT, such as the ISCCP and ERA5, predominantly presented differences closer to zero, except for the coastline of the Northeast and over the South region, both with overestimations.

We employed the RMSE (Figure 5) to assess error estimations, which provides a robust metric for evaluating differences between modeled and observed datasets. In our study, the annual RMSE values revealed that the highest value was associated with the NCEP cloud dataset, approximately 22.5 $Wm^{-2}$ (Figure 5E), whereas the lowest RMSE value was obtained using the ISCCP dataset, approximately 16.9 $Wm^{-2}$.

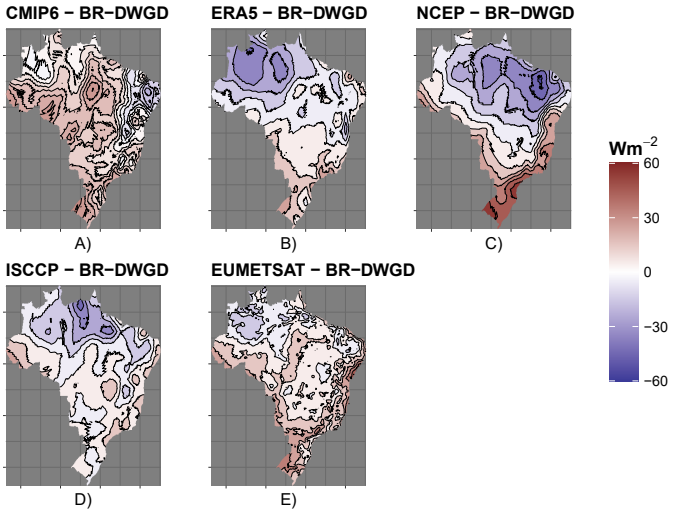


**Figure 4 – Annual contour plots of the surface solar radiation differences considering each modeled (A–E) data minus the Brazilian Daily Weather Gridded Data.**
RMSE: root mean square error; DJF: December-January-February (summer); MAM: March-April-May (autumn); JJA: June-July-August (winter); SON: September-October-November (spring).

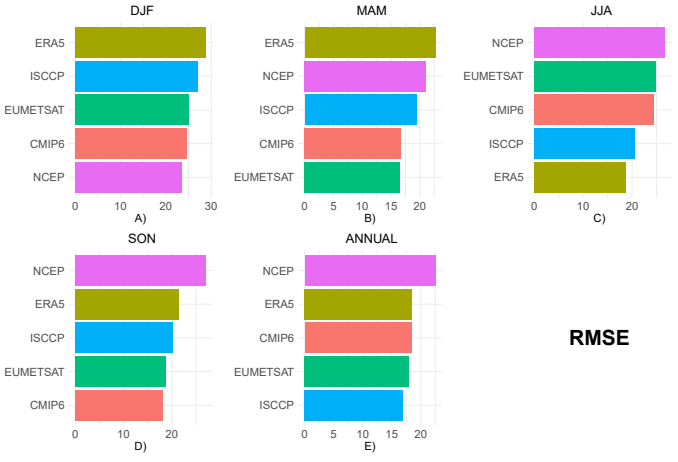


**Figure 5 – Horizontal bar charts showing the hierarchical organization of root mean square error values for different cloud data types, grouped by seasons.**
Seasonal Root Mean Square Error Distribution by Dataset

EUMETSAT demonstrated the second-best performance, followed closely by CMIP6 and ERA5.

Analysis of seasonal variations revealed distinct patterns. During the austral summer (December-January-February; DJF), error estimations differed significantly from the annual pattern, with NCEP exhibiting the lowest RMSE value of approximately 23.5 $Wm^{-2}$, while ERA5 showed the highest value of roughly 28.8 $Wm^{-2}$.

For the austral autumn (March-April-May; MAM; Figure 5B), ERA5 continued to demonstrate the poorest model accuracy, whereas EUMETSAT and CMIP6 exhibited superior performance with lower RMSE values of approximately 16.6 and 16.7 $Wm^{-2}$, respectively. During the austral winter (June-July-August; JJA; Figure 5C), ERA5 achieved the lowest RMSE of approximately 18.7 $Wm^{-2}$, while NCEP showed the poorest accuracy with approximately 26.7 $Wm^{-2}$. In the austral spring (September-October-November; SON; Figure 5D), NCEP displayed higher RMSE values again, while CMIP6 demonstrated the best accuracy among all datasets.

By analyzing the RMSE values from the heatmap (Figure 6), it became evident that CMIP6 and EUMETSAT exhibited similar patterns, demonstrating their best performance during the equinoxes and a slightly reduced performance during the solstices. In contrast, the reanalysis datasets revealed significant seasonal variability, with more pronounced fluctuations between seasons.

Overall, Figure 6 clearly shows that the EUMETSAT, CMIP6, and ISCCP datasets exhibited the best transmittance model performance during the MAM season. To further analyze this, Figure 7 presents the spatial distribution of differences between the modeled and observed datasets for MAM.

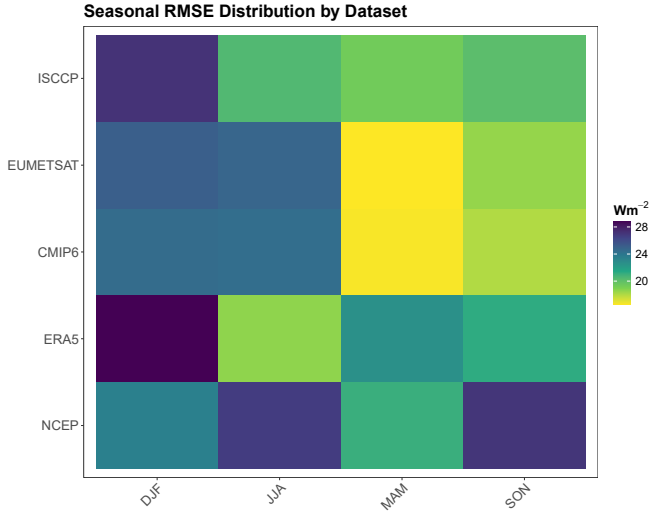


**Figure 6 – Heatmap showing the seasonal variation of root mean square error for different cloud data types, with averages calculated for the seasons.**
ISCCP: International Satellite Cloud Climatology Project; EUMETSAT: European Organisation for the Exploitation of Meteorological Satellites; CMIP6: Coupled Model Intercomparison Project Phase 6; ERA5: ECMWF Reanalysis 5; NCEP: National Center for Environmental Prediction; DJF: December-January-February (summer); MAM: March-April-May (autumn); JJA: June-July-August (winter); SON: September-October-November (spring).

The results indicated an underestimation over the Northern region when using ERA5 (Figure 7B), EUMETSAT (Figure 7E), and ISCCP (Figure 7D). Conversely, overestimations were predominantly observed in the Southern region, particularly in NCEP (Figure 7C), CMIP6 (Figure 7A), and EUMETSAT (Figure 7E). Moreover, when considering the spatial patterns of other seasons, we observed, in general, underestimations in DJF, overestimations in JJA, and behavior in SON similar to that of MAM.

As a final analysis, we employed hierarchical clustering (Figure 8) to evaluate the similarity among datasets based on seasonal and annual RMSE values. The resulting dendrogram revealed distinct groupings, highlighting the relationships between the datasets. Specifically, the most accurate and closely related datasets were EUMETSAT and CMIP6, which exhibited the highest degree of similarity. The intermediate group consisted of ERA5 and ISCCP, with ISCCP demonstrating slightly better and more consistent estimations. In contrast, NCEP showed the least similarity with the other datasets, displaying only marginal agreement with ERA5.

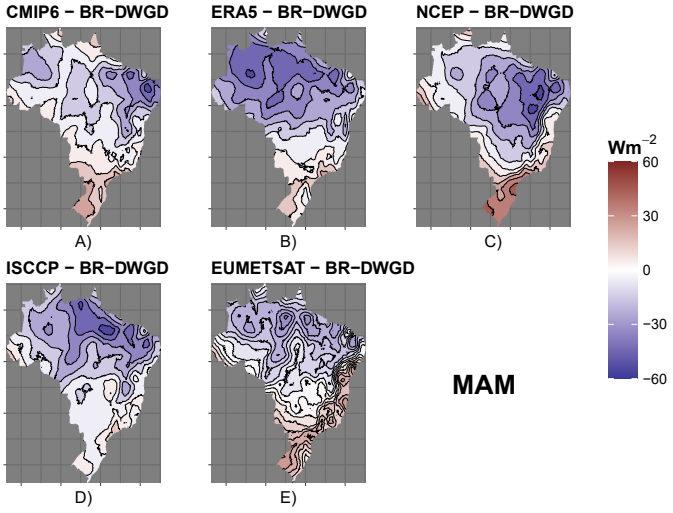


**Figure 7 – Contour plots of surface solar radiation differences, calculated as the difference between each modeled dataset (A–E) and BR-DWGD for the austral autumn (MAM).**

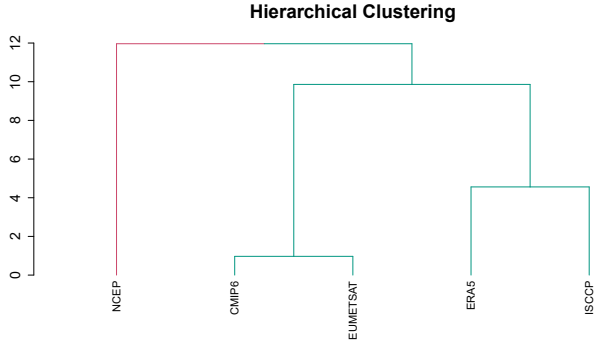


**Figure 8 –Hierarchical clustering dendrogram based on root mean square error values from seasonal and annual means for different cloud datasets.**

## Discussion

After evaluating the datasets, we computed the Global Horizontal Irradiance (GHI) for all datasets, as illustrated in Figure 9. The GHI maps exhibited a direct relationship with SSR, revealing higher GHI values in the Brazilian Northeast region and lower values in the Amazon (North) and the South regions. These findings align with previous studies by Martins et al. (2012) and Pereira et al. (2017).

These results are crucial for estimating the PV potential of solar energy systems. However, it should be emphasized that additional factors, such as air temperature and panel technology, must also be considered to accurately estimate PPP. Our findings provide valuable insights to support future solar power plant projects across Brazil. Specifically, the modeled GHI can help reduce uncertainties in the North region, particularly in the Amazon basin, where the sparse distribution of weather stations poses significant challenges, as Xavier et al. (2022) noted. Based on our analysis, we strongly recommend prioritizing satellite-derived cloud cover data, complemented by CMIP6 reconstructions, while strictly adhering to our methodology—including the choice of models, ensemble techniques, and experimental design—to achieve more accurate and reliable GHI modeling.

Furthermore, our results indicate that cloud data obtained from the NCEP reanalysis produce less reliable modeled GHI, consistent with the findings of Free et al. (2016) and Wu et al. (2023). It is crucial to address a common misunderstanding in the literature regarding the distinction between total cloud amount and total cloud cover. According to the World Meteorological Organization, the total cloud amount is estimated from superposed cloud layers, whereas total cloud cover refers only to the visible clouds in the sky (https://cloudatlas.wmo.int/en/total-cloud-cover-and-cloud-amount.html, last access: 05/08/2024). Consequently, the NCEP reanalysis cloud data, which represent only visible clouds, tend to exhibit greater variability, making these reanalysis data unsuitable for accurate GHI modeling.

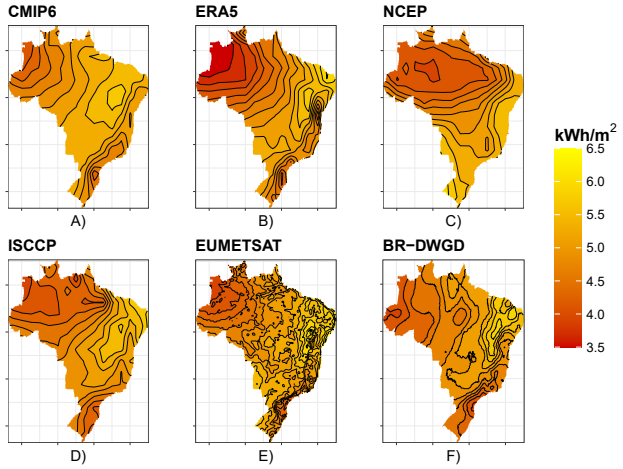


**Figure 9 – Contour plots of Brazilian global horizontal irradiance from modeled datasets (A–E) and observed dataset (F) for the annual average.**

## Conclusion

This study highlights the critical role of cloud cover data sources in modeling SSR and estimating PPP across Brazil. On an annual scale, the ISCCP dataset demonstrated the highest accuracy in SSR modeling, with the lowest RMSE of approximately 16.9 $Wm^{-2}$, followed closely by EUMETSAT and ERA5. In contrast, the NCEP reanalysis data exhibited significant discrepancies, underestimating SSR in the North and Northeast regions while overestimating it in the South. These biases stem primarily from the conceptual differences between total cloud cover and total cloud amount in the dataset. Although the CMIP6 ensemble tended to overestimate SSR, it provided intermediate performance, particularly in transitional seasons. These findings underscore the importance of using satellite-derived cloud data, such as ISCCP and EUMETSAT, to enhance SSR modeling accuracy, especially in regions with sparse observational data, such as the Amazon basin.

Seasonal analysis revealed distinct patterns in model performance. During the austral autumn (MAM), CMIP6 and EUMETSAT exhibited the lowest RMSE values (16.6 and 16.7 $Wm^{-2}$, respectively), indicating superior accuracy. In contrast, ERA5 performed poorly in DJF, with the highest RMSE (28.8 $Wm^{-2}$), while NCEP showed the least accuracy in JJA (26.7 $Wm^{-2}$). These seasonal variations highlight the influence of cloud dynamics on SSR modeling and emphasize the need for tailored approaches depending on the time of year. Overall, the study demonstrates that SSR modeling is a reliable alternative for estimating GHI and PPP, particularly in data-scarce regions. By prioritizing satellite-derived cloud data and adhering to robust methodologies, future solar energy projects in Brazil can achieve more accurate and reliable energy potential assessments, supporting the country's transition to sustainable energy.

## Code and data availability

The datasets and computational scripts supporting this study are publicly available at Zenodo: https://doi.org/10.5281/zenodo.15168357

## Authors' Contributions

**Oliveira**, E.D.: conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; project administration, supervision, writing – original draft; writing – review & editing. **Ferreira**, T.R.: conceptualization; data curation; formal analysis; investigation; methodology; validation; visualization; writing – review & editing. **Azevedo,** C.D.S.: conceptualization, formal analysis, investigation; methodology; visualization; writing – review & editing. **Leitão**. M.M.V.B.R.: investigation; methodology; visualization; writing – review & editing. **Melo**, M.L.D.: investigation; methodology; visualization; writing – review & editing.

## References


Berger, A., 1978. Long-term variations of daily insolation and Quaternary. Journal of Atmospheric Sciences, v. 35 (12), 2362-2367. https://doi.org/10.1175/1520-0469(1978)035<2362:LTVODI>2.0.CO;2.

Bird, R.; Riordan, C., 1986. Simple solar spectral model for direct and diffuse irradiance on horizontal and tilted planes at the earth's surface for cloudless atmospheres. Journal of Applied Meteorology and Climatology, v. 25 (1), 87-97. https://doi.org/10.1175/1520-0450(1986)025<0087:SSSMFD>2.0.CO;2.

Bird, R.E.; Hulstrom, R.L., 1981. A Simplified Clear Sky Model for Direct and Diffuse Insolation on Horizontal Surfaces. Tr-642-761. Solar Energy Research Institute (SERI)—US Department of Energy, Golden, CO. http://dx.doi.org/10.2172/6510849.

Chakraborty, T.; X. Lee., 2021. Large differences in diffuse solar radiation among current-generation reanalysis and satellite-derived products. Journal of Climate, v. 34 (16), 6635-6650. https://doi.org/10.1175/JCLI-D-20-0979.1.

Collins, M.R.; Knutti, J.; Arblaster, J.-L.; Dufresne, T.; Fichefet, P.; Friedlingstein, X.; Gao, W.J.; Gutowski, T.; Johns, G.; Krinner, M.; Shongwe, C.; Tebaldi, A.J.; Weaver, A. J.; Wehner, M., 2013. Long-term Climate Change: Projections, Commitments and Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F.; D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA. https://doi.org/10.1017/CBO9781107415324.024.

Danso, D.K.; Anquetin, S.; Diedhiou, A.; Lavaysse, C.; Hingray, B.; Raynaud, D.; Kobea, A.T., 2022. A CMIP6 assessment of the potential climate change impacts on solar photovoltaic energy and its atmospheric drivers in West Africa. Environmental Research Letters, v. 17. https://doi.org/10.1088/1748-9326/ac5a67.

Dubey, S.; Sarvaiya, J.N.; Seshadri, B., 2013. Temperature Dependent Photovoltaic (PV) efficiency and its effect on PV production in the world – a review. Energy Procedia, v. 33, 311-321. https://doi.org/10.1016/j.egypro.2013.05.072.

Free, M.; Sun, B.; Yoo, H.L., 2016. Comparison between total cloud cover in four reanalysis products and cloud measured by visual observations at U.S. weather stations. Journal of Climate, v. 29 (6), 2015-2021. https://doi.org/10.1175/JCLI-D-15-0637.1.

Ha, S.; Zhou, Z.; Im, E-S.; Lee, Y-M., 2023. Comparative assessment of future solar power potential based on CMIP5 and CMIP6 multi-model ensembles. Renewable Energy, v. 206, 324-335. https://doi.org/10.1016/j.renene.2023.02.039.

Isaza, A.; Kay, M.; Evans, J.P.; Prasad, A,; Bremner.; S., 2023. Maximizing photovoltaic potential and minimizing costs in a future warmer climate: the role of atmospheric aerosols and greenhouse gas emissions. Renewable Energy, v. 219 (Part 2), 119561. https://doi.org/10.1016/j.renene.2023.119561.

Martins, F.R.; Abreu, S.L.; Pereira, E.B., 2012. Scenarios for solar thermal energy applications in Brazil, Energy Policy. v. 48, 640-649. https://doi.org/10.1016/j.enpol.2012.05.082.

Nwokolo, S.C.; Ogbulezie, J.C.; Ushie, O.J., 2023. A multi-model ensemble-based CMIP6 assessment of future solar radiation and PV potential under various climate warming scenarios. Optik, v. 285, 170956. https://doi.org/10.1016/j.ijleo.2023.170956.

Oliveira, E.D., 2023. Daily INSOLation (DINSOL-v1.0): an intuitive tool for classrooms and specifying solar radiation boundary conditions. Geoscientific Model Development, v. 16 (9), 2371-2390. https://doi.org/10.5194/gmd-16-2371-2023.

Oliveira, E.D.; Fernandez, J.H., 2020. Um novo modelo climático simplificado: global resolved energy balance with galactic cosmic rays (GREB-GCR) theory. Revista Brasileira de Geografia Física, v. 13 (2), 842-854. https://doi.org/10.26848/rbgf.v13.2.p842-854.

Oliveira, E.D.; Fernandez, J.H.; Mendes, D.; Spyrides, M.H.C.; Gonçalves, W.A., 2019. Validação de um modelo climático simplificado adaptado para simular os efeitos do aumento da concentração de CO2 associados à teoria dos raios cósmicos galácticos (Validation of a simplified climate model adapted to simulate the effects of increased CO2 concentration associated with the galactic cosmic rays theory). Revista Brasileira de Geografia Física, v. 12 (3), 768-778. https://doi.org/10.26848/rbgf.v12.3.p768-778.

Pereira, E.B.; Martins, F.R.; Gonçalves, A.R.; Costa, R.S.; Lima, F.L.; Rüther, R.; Abreu, S.L de.; Tiepolo, G.M.; Pereira, S.V.; Souza, J.G., 2017. Atlas Brasileiro de Energia Solar. 2. ed. INPE, São José dos Campos, 80 p. https://doi.org/10.34024/978851700089.

Silva, O.A.V.O.L.; Moita Neto, J.M.; Lira, M.A.T.; Morais, F.H.M., 2021. Expansion of photovoltaic systems in multicampi higher education institutions: evaluation and guidelines. Revista Brasileira de Ciências Ambientais, v. 56 (4), 697-709. https://doi.org/10.5327/Z217694781009.

Souza, R.M.A.; Leitão, M.V.B.R.; Oliveira, G.M.; Oliveira, E.D.; Correia, L.T., 2021. Demanda hídrica e adubação orgânica no cultivo protegido de pimentão na região norte da Bahia. Irriga, v. 26 (4), 787-800. https://doi.org/10.15809/irriga.2021v26n4p787-800.

Vunnam, S.; VanithaSri, M.; RamaKoteswaraRao, A., 2021. Performance analysis of mono crystalline, poly crystalline and thin film material based 6 × 6 T-C-T PV array under different partial shading situations. Optik, v. 248, 168055. https://doi.org/10.1016/j.ijleo.2021.168055.

Wang, Z.; Lin, L.; Xu,Y.; Che, H.; Zhang, X.; Zhang, H.; Dong, W.; Wang, C.; Gui, K.; Xie, B., 2021. Incorrect Asian aerosols affecting the attribution and projection of regional climate change in CMIP6 models. npj Climate and Atmospheric Science. v. 4 (2). https://doi.org/10.1038/s41612-020-00159-2.

Wild, M.; Folini, D.; Henschel, F.; Fischer, N.; Müller, B., 2015. Projections of long-term changes in solar radiation based on CMIP5 climate models and their influence on energy yields of photovoltaic systems, Solar Energy, v. 116, 12-24. https://doi.org/10.1016/j.solener.2015.03.039.

Wild, M.; Gilgen, H.; Roesch, A.; Ohmura, A.; Long, C.N.; Dutton, E.G.; Forgan, B.; Kallis, A.; Russak, V.; Tsvetkov, A., 2005. From Dimming to Brightening: Decadal Changes in Solar Radiation at Earth's Surface. Science, v. 308 (5723), 847-850. https://www.science.org/doi/10.1126/science.1103215.

Wu, H.; Xu, X.; Luo, T.; Yang, Y.; Xiong, Z.; Wang, Y., 2023. Variation and comparison of cloud cover in MODIS and four reanalysis datasets of ERA-interim, ERA5, MERRA-2 and NCEP, Atmospheric Research, v. 281, 106477. https://doi.org/10.1016/j.atmosres.2022.106477.

Xavier, A.C.; Scanlon, B.R.; King, C.W.; Alves, A.I., 2022. New improved Brazilian daily weather gridded data (1961–2020). International Journal of Climatology, v. 42 (16), 8390-8404. https://doi.org/10.1002/joc.7731.