Cloud variability and cloudless percentage days calculation using cluster analysis

Ricardo Almeida de Siqueira[1], Vinícius Roggério da Rocha[1], Rodrigo Santos Costa[1] and André Rodrigues Gonçalves[1]

[1] Division of Impacts, Adaptations, and Vulnerability, National Institute for Space Research, Av. dos Astronautas, 1758, São José dos Campos (Brazil)

Abstract

Cloudiness variability across the Brazilian territory was analyzed using cluster analysis based on daily values of the *Ceff* and *DARR* indices. Cloud reflectance data with a 10-minute resolution, collected between 2020 and 2024, was employed to compute these daily metrics. A subset of the data exhibiting the lowest values for both *Ceff* and *DARR* was selected to calculate the Cloudless Percentage Days (*CPD*) index, which quantifies the most favorable locations for solar energy utilization across the region. Northwestern Minas Gerais and western Bahia emerged as the top-performing areas for solar energy generation during the study period (*CPD* > 0.65). Additionally, the Central region of Brazil recorded moderate to high CPD values during the dry season.

Keywords: Solar resource, Cloud variability, Cloudless days, Cluster analysis

1. Introduction

The variability of the solar resource on an hourly scale is mainly associated with atmospheric phenomena, such as cloud cover, rain, and aerosol concentration (Da Rocha et al, 2022). Photovoltaic (*PV*) and Concentrated Solar Power (*CSP)* systems need to be installed in regions with low variability to maximize their gains. High-intensity variations are usually responsible for the reduction of solar irradiance due to cloud shading, but increases can also occur due to lateral reflection from clouds (cloud enhancement).

Several works in the literature describe indices for quantifying the variability in the global horizontal irradiance (*GHI*). The index *DARR*, *Daily Aggregate Ramp Rate*, was proposed in van Haaren et al. (2014). It calculates the daily variability of the sum of changes in *GHI* normalized by a constant *C* equal to 1000 Wm^{-2}:

$$DARR = \sum_{n=1}^{N-1} \left| GHI(t) - GHI(t-1) \right| / C \qquad \text{(eq. 1)}$$

where *t* represents the time instant and *N* the time interval. High *DARR* values indicate high variability in solar irradiance and vice versa. For a better interpretation of the variability indices, the work of Stein et al. (2012) presented the so-called *Arrow Head Plot*, where the variability index is plotted as a function of the clearness index. In this way, regions with different types of variability can be easily identified, avoiding ambiguities in their classifications.

2. Objective

The objective of this work is to present the results regarding the analysis of cloudiness variability in Brazilian territory. A cluster analysis was performed with cloud cover index (*Ceff*) and variability index *DARR* data to identify the types of variability in solar irradiance. In other words, the graph of *Ceff* as a function of *DARR* was used as an *Arrow Head Plot,* and the cluster analysis was used to delimit the different regions. The quantification of the best locations was performed by calculating the percentage index of *Cloudless Percentage Days* (*CPD*). Previous work of Da Rocha et al, (2022) used the *GOES-13* reflectance

10.18086/swc.2025.04.12 Available at http://proceedings.ises.org

data with 30 minute time resolution to calculate *Ceff* values. This study employs 10-minute *GOES-16* satellite data to compute the same set of variables. The enhanced temporal resolution is anticipated to more accurately capture daily fluctuations in *GHI*, which predominantly influence the *DARR* values.

3. Methodology

2.1. Cloud cover index calculation

Geostationary satellite *GOES-16* reflectance data in the visible spectral range (channel 2 – 0.64 µm) of the Advanced Baseline Imager (*ABI*), Cloud and Moisture Imagery Product (*CMIP*) level 2, was acquired with 10-minute time resolution for the 2020-2024 period. The data was used to calculate the cloud cover index *Ceff*, a parameter commonly used by models for estimating the surface *GHI* in the parameterization of clouds radiative processes (Siqueira et al., 2025). *Ceff* concerns the cloud optical depth and the percentage of the pixel area effectively covered by them. The *Ceff* was calculated by the following normalized ratio:

$$C_{eff} = \left(\rho - \rho_{clear}\right) / \left(\rho_{cloud} - \rho_{clear}\right) \qquad \text{(eq. 2)}$$

where ρ is the visible reflectance observed by the satellite in a given pixel, ρ_{clear} and ρ_{cloud} are the reflectances for clear and overcast sky conditions, respectively. The ρ_{clear} for each pixel was the second percentile (*P*2) of the data series acquired in a preceding 30-day period on a one-hour interval basis, while for ρ_{cloud}, the value of 0.8 was used following the work of Siqueira et al.,(2025).

2.2. Cloud variability index calculation

To quantify the variability of the solar resource due to the effect of clouds, the *Ceff* values were used to calculate the *DARR* index. Although the original expression was proposed with *GHI* values in mind, it is easily modifiable for the use of *Ceff* because both are highly correlated with cloud variability. Mathematically:

$$DARR = \sum_{n=1}^{N-1} \left| C_{eff}(t) - C_{eff}(t-1) \right| \qquad \text{(eq. 3)}$$

where *t* represents the time instant, *N* the time interval, and Δt (*t – (t-1)*) of 10 minutes. Recalling that as C_{eff} varies from 0 to 1, the value of *C* was considered as unitary. The advantage of using C_{eff} instead of *GHI* lies in the fact that it is a parameter that can be calculated directly via remote sensing data without the need to estimate or use a model to calculate *GHI*. Also, the *DARR* index was chosen for its robustness in identifying periods with high variability, for its mathematical simplicity, and for not depending on a clear sky model (Da Rocha et al., 2022).

2.3. Cloudless percentage days index calculation

Daily averaged and normalized data of *Ceff* and *DARR* were used in the cluster analysis (*k-means*) aiming at the identification of the different regions in the dimensional space of *Ceff* as a function of *DARR*. For simplicity, the number of 3 clusters was adopted. The cluster region with lower values of *Ceff* and *DARR* (*bestcluster*) was used in the calculation of *CPD*. In this context, *CPD* is the fraction of the number of days that a pixel presented the values at the *bestcluster* region (*Nbestcluster*) divided by the total number of days considered *N*:

$$CPD = N_{bestcluster} / N \qquad \text{(eq. 4)}$$

4. Results and discussions

Figure 1 presents the result of the cluster analysis through the graph of *Ceff* as a function of *DARR*. The same graph shows the location of the centroids for the 3 identified regions. With the centroids, it is possible to notice that the distribution of points in each region is not uniform and has distinct point densities. The green region presents the sample space with the best conditions for solar energy generation, as it has the lowest values for *Ceff* (0.2) and *DARR (6),* according to the position of the centroid. In the same figure, the map of Brazil is presented with the location of the pixels used to compose the green cluster. It is possible to note that the central region of Brazil was largely favored in the selection of pixels, although other regions such as the west of São Paulo, west of Minas Gerais, and west of Bahia also presented selected pixels. It is worth mentioning that since the cluster analysis used average data for the entire period 2020–2024, this does not mean that these regions in their entirety will be suitable for solar energy generation, as it does not consider the frequency of these conditions throughout the year. Therefore, the calculation of *CPD* is necessary.

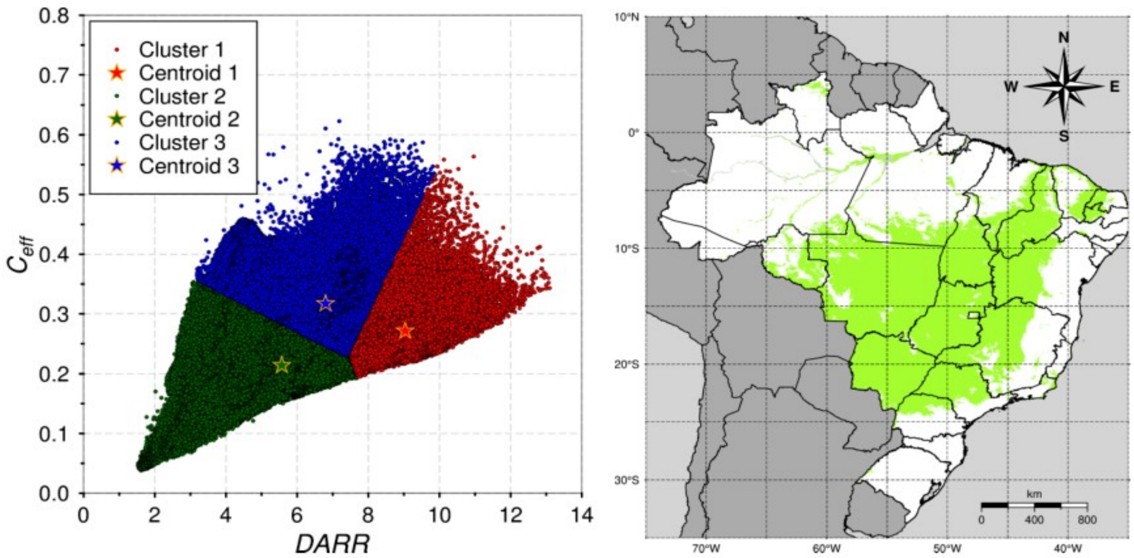

Fig. 1: In the left, plot of C_{eff} vs *DARR* where each color represents a cluster. In the right, Brazil map with the selected pixels in the *bestcluster* region (green color).

Figure 2 presents the average daily *Ceff* and *DARR* data for Brazil from 2022 to 2024. *DARR* values were lowest in central Brazil, while the highest values were obtained in the northern part of the country, concentrated in the states of Amapá, northern Pará, and northern Maranhão, and in the coastal and eastern regions of the Brazilian Northeast. *Ceff* values were lowest in central Brazil, such as *DARR*, but also in the northern part of the Brazilian Northeast. The highest values were obtained in the southern region of Brazil, concentrated in the central and eastern parts of the states of Rio Grande do Sul and Santa Catarina.

The average *CPD* values are presented in Figure 3. It is possible to note that the regions that presented the best conditions for solar energy generation throughout the year, were those in northwestern Minas Gerais and western Bahia (*CPD* > 0.65). Although the central region also presents relatively high *CPD* values (between 0.5 and 0.65), they are mainly concentrated in the dry season (see Figure 4). During the autumn and winter period, there is a reduction in the availability of moisture and active convective systems, hindering cloud formation. This dry period from May to September is characterized by low variability in cloud cover (low *DARR*), while the coverage is sparse (low values of C_{eff}). With the arrival of spring and summer, there is an intensification in moisture availability, due to moisture transport by low-level jets originating from the North of Brazil, and greater activity of the intertropical convergence zone (*ITCZ*) and the South Atlantic (*SACZ*), increasing both the coverage fraction and the variability of the clouds. The North, the eastern part and coast of the Northeast and Southeast regions, as well as the South region, were the least favored regions (values less than 0.3).

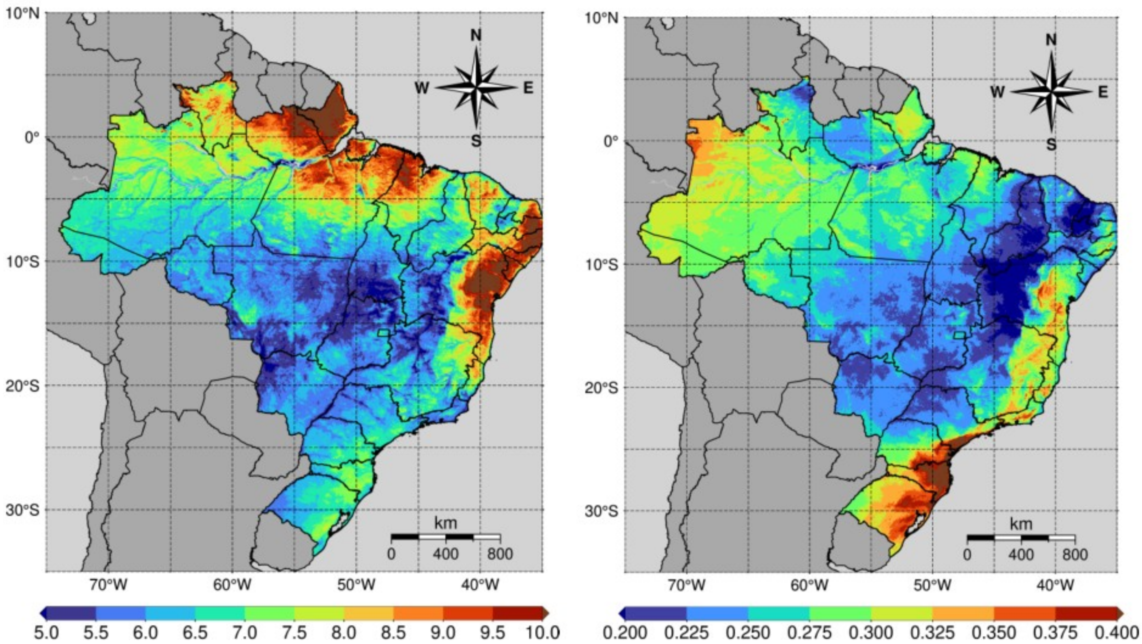

Fig. 2: Maps of Brazil showing the average values of *DARR* (left) and Ceff (right) for the 2020–2024 period.

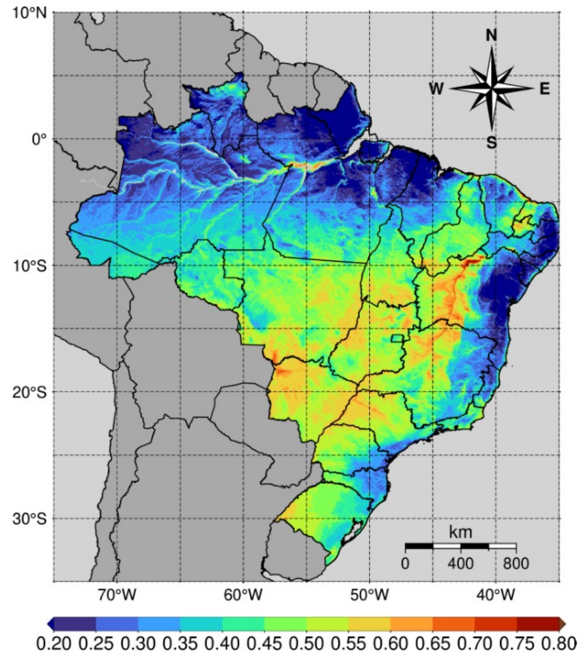

Fig. 3: Brazil map with *CPD* values for the 2020-2024 period.

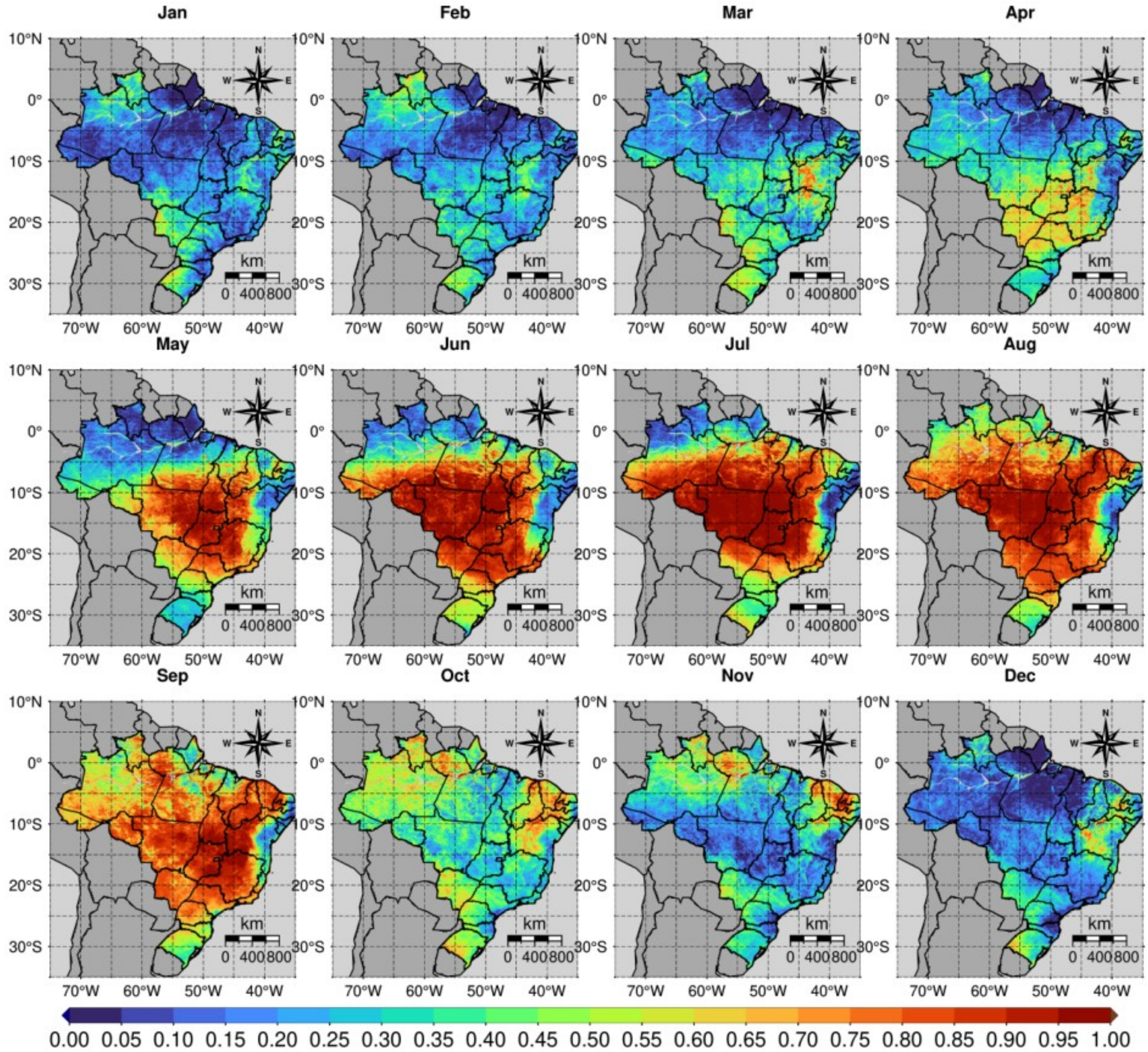

Fig. 4: Brazil maps with monthly *CPD* values for the 2020-2024 period.

To understand the interannual behavior of *CPD* values, Figure 5 displays annual maps of Brazil from 2020 to 2024. These maps suggest a consistent pattern of *CPD* highs and lows across years, likely influenced by recurring synoptic conditions that affect cloud cover. While the spatial distribution tends to remain stable, the intensity of *CPD* values may vary more locally.

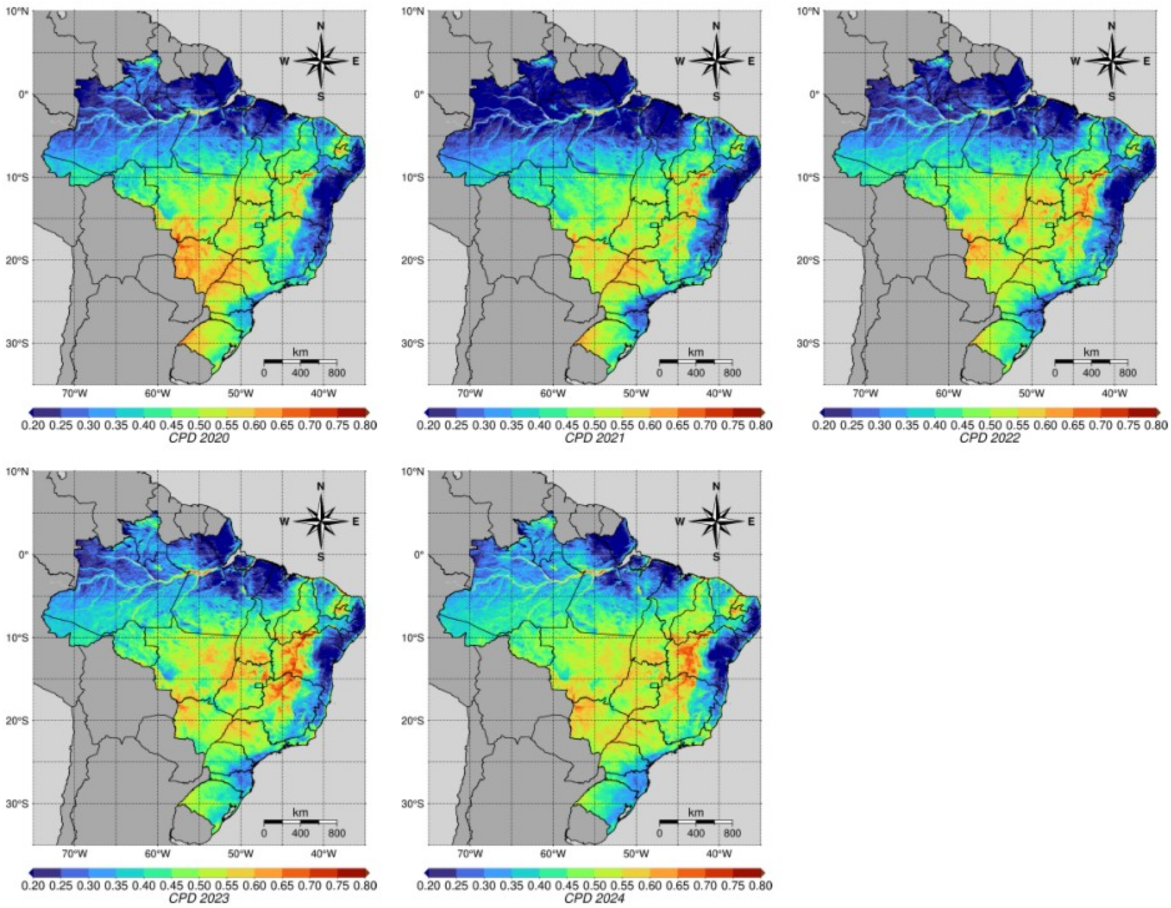

Fig. 5: Brazil maps with yearly *CPD* values for the 2020-2024 period.

Figure 6 highlights the differences in *CPD* for each year compared to the average *CPD* over the full period (2020–2024), emphasizing this variability. The figure reveals a dipole pattern in *CPD* variation between the northern and central regions of Brazil versus the southern region. During 2023–2024, *CPD* values in the North and Center were above average, while the South experienced below-average values. In contrast, from 2020 to 2022, this pattern reversed: the South showed above-average *CPD* values, and the North and Center fell below average.

These fluctuations appear to be linked to the *ENSO* Oscillation, which drives the *El Niño* (2023–2024) and *La Niña* (2020–2022) phenomena. *El Niño* periods are typically associated with drier conditions and reduced cloud cover in northern and central Brazil, while the South experiences wetter conditions and increased cloud cover. This shift in atmospheric circulation affects regional cloud patterns and may explain the observed *CPD* variations. During *La Niña*, the pattern reverses—bringing more rain to the North and Center and drier conditions to the South, thereby flipping the dipole.

Another noteworthy insight from Figure 5 is that, depending on the year, the number of days suitable for solar energy generation can differ by up to two months—either more or fewer. This level of interannual variation plays a significant role in assessing the economic viability of solar energy projects.

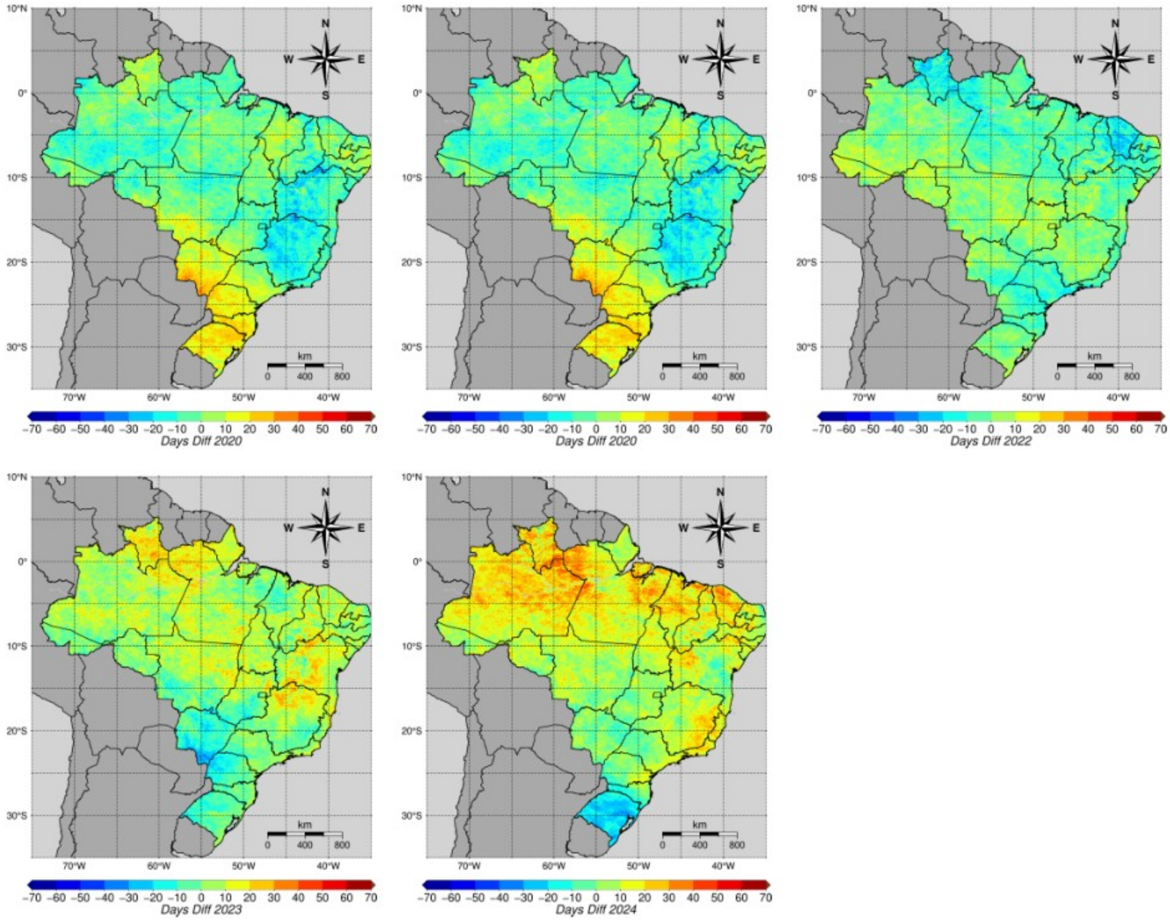

Fig. 6: Maps of Brazil displaying annual deviations in *CPD* relative to the average *CPD* (Figure 3), with values converted to days (*CPD* × 365), covering the 2020-2024 period.

5. Conclusions

This section of the paper presents the results of a study evaluating the most suitable regions in Brazil for solar thermal power generation during the 2020–2024 period. The analysis considered observed variability in cloud cover using the *DARR* index and cloud cover fraction, applying the C_{eff} calculation. Cluster analysis was employed to objectively identify the optimal sample space for C_{eff} and *DARR* values in relation to power generation. Additionally, the *CPD* index was used as a metric to assess the frequency of favorable conditions throughout the year.

The findings indicate that the northwestern regions of Minas Gerais and western Bahia (*CPD* > 0.65) offered the most favorable conditions for year-round solar power generation. These areas exhibited consistently low cloud cover fractions ($C_{eff} < 0.2$) and minimal variability (*DARR* < 5). Although the central region of Brazil also showed high *CPD* values, these were primarily concentrated during the dry season due to favorable synoptic conditions—specifically, the weakening of the Intertropical Convergence Zone (*ITCZ*) and the South Atlantic Convergence Zone (*SACZ*).

Conversely, the North region, as well as the eastern and coastal portions of the Northeast and Southeast regions, and the South region, experienced the least favorable conditions for solar power generation throughout the year (*CPD* < 0.25). This was attributed to the greater activity of convective systems and oceanic influences. The interannual variation in *CPD* revealed a dipole pattern between the North and Central regions and the South, likely linked to the *El Niño* and *La Niña* phenomena.

These results demonstrate that the methodology is sensitive to climatic variations on timescales longer than one year. For future research, it is recommended to extend this approach to periods of 20 to 30 years to better capture and understand interannual effects.

6. Acknowledgments

The National Council for Scientific and Technological Development (CNPq), Brazil supported the authors [grant numbers 300816/20235, 301337/2023-3, 311620/2023-0, 4097122/2021-7]. The authors thank DIIAV/INPE and Unifesp for their institutional support of the research infrastructure. The authors also thanks to National Institute of Science and Technology for Climate Change – INCT-MC Project Phase 2 (Grants FAPESP 2014/50848-9, CNPq, Brazil 465501/2014-1, and CAPES/FAPS, Brazil No 16/2014) and FINEP ‘‘SOLTER - União por difusão de metais leves em acumuladores de energia solar compactos para a geração de eletricidade por fonte termossolar no Brasil’’ (FINEP 0018/21).

7. References

Da Rocha, Vinicius Roggério; Costa, Rodrigo Santos; Martins, Fernando Ramos; Gonçalves, André Rodrigues; Pereira, Enio Bueno. Variability Index of Solar Resource Based on Data from Surface and Satellite. Renewable Energy, v. 201, p. 354–378, dez. 2022. ISSN 09601481. DOI: 10.1016/j.renene.2022.10.093.

De Siqueira, Ricardo Almeida; Gonçalves, André Rodrigues; Costa, Rodrigo Santos; Martins, Fernando Ramos. Sensitivity Analysis of Solar Irradiance Estimates over a Tropical Region by Cloud Index Models. Solar Energy, v. 287, p. 113176, 1 fev. 2025. ISSN 0038-092X. DOI: 10.1016/j.solener.2024.113176.

Stein, Joshua; HANSEN, Clifford; RENO, Matthew J. The Variability Index: A New And Novel Metric For Quantifying Irradiance And Pv Output Variability. [S. l.], 1 mar. 2012. https://www.osti.gov/biblio/1068417.

van Haaren, Rob; Morjaria, Mahesh; Fthenakis, Vasilis. Empirical Assessment of Short Term Variability from Utility-Scale Solar PV Plants. Progress in Photovoltaics: Research and Applications, v. 22, n. 5, p. 548–559, 2014. ISSN 1099-159X. DOI: 10.1002/pip.2302.