Analysis of the Spatial Variability of GHI in Rio Grande do Norte Brazil

Aléxia Monteiro Valentim[1 2], Ana Cleide Bezerra Amorim[1 3] Cláudio Moisés Santos e Silva[3], João Vitor de Oliveira[2] and Samira de Azevedo Santos Emiliavaca[1 3]

[1] SENAI Institute for Innovation in Renewable Energy (ISI-ER), Natal (Brazil)

[2] Post-graduate Program in Climate Sciences, Natal (Brazil)

[3] Lecturer in Post-graduate Program in Climate Sciences, Natal (Brazil)

Abstract

This study explores the solar resources of the state of Rio Grande do Norte (RN), located in northeastern Brazil, with the aim of characterizing solar energy availability. Data were obtained from the Brazilian Solar Energy Atlas, and a simulation was performed using the Weather Research and Forecasting (WRF) model for this type of application, WRF-Solar. The Atlas data were divided into 502 observation points, with an irradiation value assigned to each geographic coordinate. Bilinear interpolation was performed using the simulation results. In the cluster analysis, the hierarchical type with Euclidean distance was selected, and the silhouette method indicated that two clusters provided the best fit. The monthly results for irradiation suggest the existence of two groups with similar behavior throughout RN. The BIAS results indicate high reliability for the first quarter of the year for cluster 2 and excellent spatial variability of the WRF-Solar cluster groups in relation to the Atlas data.

Keywords: Solar Energy, Northeast Brazil, Cluster Analysis, WRF, WRF-Solar.

1. Introduction

The Brazilian potential was distributed through studies with heliographic and meteorological stations (Tiba *et al.,* 2002). Brazil has a privileged location regarding available natural resources, notably its solar radiation (Pereira *et al.*, 2020) and photovoltaic potential, which also links to the potential for producing new sources like hydrogen from renewable energy (Solano and Affonso, 2025).

As a result, Brazil's clean energy matrix is more prominent than the global one. In September 2025 alone, the expansion in the sector reached 1.4 GW from renewable sources, totaling 17 solar power plants and adding up to 934.72 MW, according to the National Electric Energy Agency (ANEEL, 2025).

Rio Grande do Norte (RN) is one of the states in Brazil's Northeast Region (NEB), featuring many hours of daily sunshine. Areas with high daily insolation rates are the most suitable for solar energy utilization as a mitigation and adaptation strategy. To characterize and classify the available monthly resource, the study aims to present the spatial distribution of the resource for each group.

Studies show Fovell e Fovell (1993) the clustering of the United States into approximately 14 clusters, which partitions its climatic zones. Mimmack, Mason and Galpin (2001) aimed to define regions with common precipitation regimes, utilizing Southern Africa for clustering regionalization, demonstrating two forms of grouping: using distances like Euclidean (shortest distance between two points) and Mahalanobis (high correlation). As studies reveal similarities and differences to provide relevant information about each group, the application of multivariate methods yields significant results for data classification.

10.18086/swc.2025.04.43 Available at http://proceedings.ises.org

2. Materials and Methods

2.1 Study area

The study area encompasses two different climates: the semi-arid part of Brazil's Northeast Region (NEB), identified as BSh, and the humid tropical climate (Aw), according to the Köppen classification Alvarez *et al.* (2013). The NEB region has high potential year-round, showing a more linear resource compared to other Brazilian regions (Martins *et al*., 2008). Solarimetric resource assessment studies require quality observational data campaigns to enable all types of forecasts, both statistical and numerical techniques, which are better trained with reliable data.

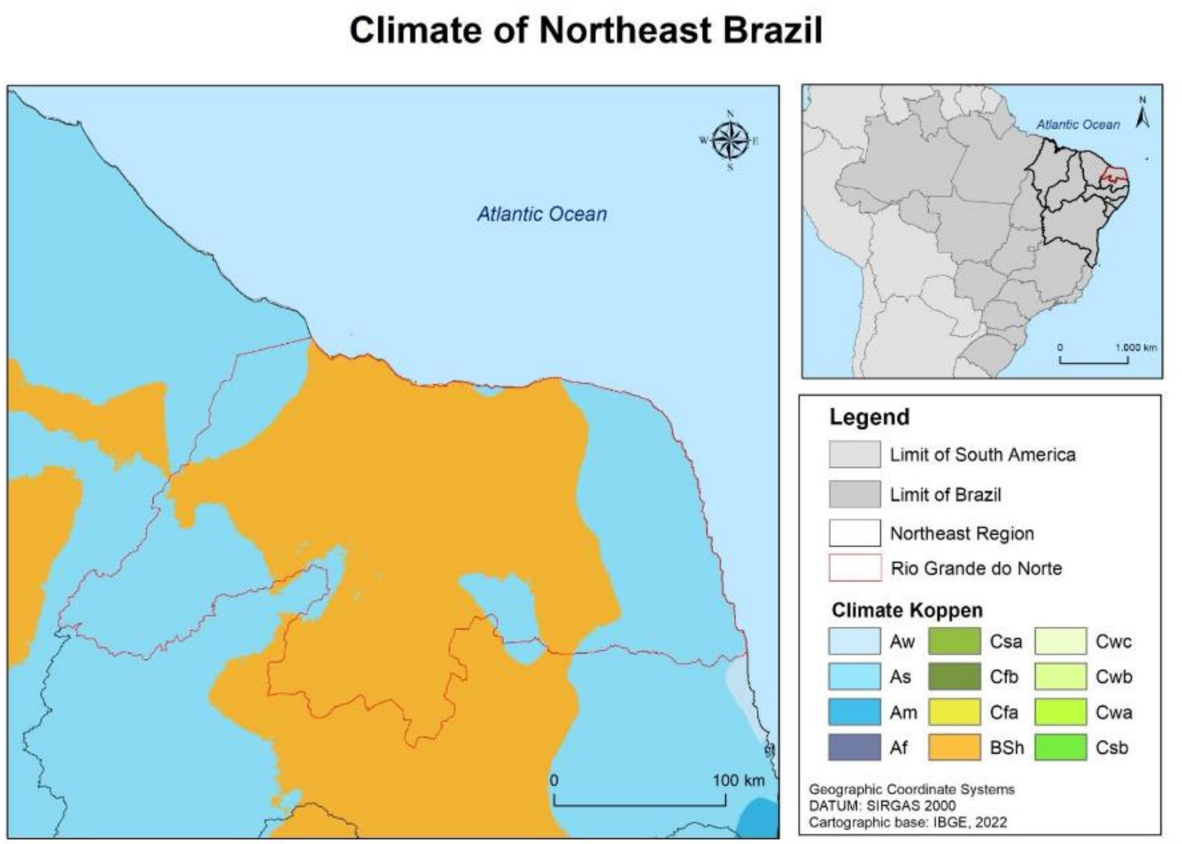


Figure 1: Climate in the state of Rio Grande do Norte.

2.2. Typical Meteorological Year

With the aim of representing the year that most closely approximates the long-term average climatic conditions for a specific variable of interest, The Typical Meteorological Year (TMY) is a multivariate statistical method that combines data from different meteorological variables. To this end, typical years for Global Horizontal Irradiance (G) were estimated using the National Renewable Energy Laboratory (NREL) platform (Sengupta *et al*., 2014, 2018). For March and September, where no observed data were available, an average was calculated across all available years.

2.3.1 Atlas Database

Monthly averages of total daily irradiation in Wh/m² per day with spatial resolution of 0.1° x 0.1° (approximately 10 km x 10 km) provided by the Modeling Laboratory, Renewable Energy Resources Studies (LABREN) and the Earth System Science Center (CCST) and National Institute for Space Research (INPE) – Brazil, through the Brazilian Solar Energy Atlas – Database (Martins *et al.* 2017). The data were qualified and had as input conditions approximately 17 years of satellite images validated by the solarimetric stations of the Environmental Data Organization System Network (SONDA), also from INPE, and included data from the entire available database of the National Institute of Meteorology (INMET) for the application process to the Brazil – SR model. Qualification tests were applied in physical tests, temporal variations, intercomparison between observations and numerical models of clear skies in order to remove data considered suspicious. The Atlas data for the state of RN were collected and added up to approximately 502 points, each with the specific latitude and longitude of each location, complying with the spatial resolution standard.

2.3.2 Observational Database

The data from the National Institute of Meteorology (INMET) database were calculated at hourly intervals for Global Radiation for each of the typical years calculated according to Table 1 and converted into units of kWh/m² per day. The reference observation station was Mossoró, RN (5° 11′ 16″ S, 37° 20′ 38″ W), a municipality with a *BSh* climate in Rio Grande do Norte. Due to the predominance of the climate in the municipality, it was convenient to use it for modeling parameters, such as the typical meteorological year.

Table 1: Typical Meteorological Year

Month	January	February	March	April	May	June
TMY	2008	2017	1999	2021	2021	2013
Month	July	August	September	October	November	December
TMY	2013	2018	2018	2013	1998	2020

2.3.3 Reanalysis ERA-5

The reanalysis data are essential variables for model initialization and were selected from the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-5 hourly data set for atmospheric pressure, relative humidity, u and v components of wind speed at 10 meters, and air temperature at 2 meters (Hersbach *et al.*, 2020).

2.4 WRF Modeling

The model used in this study was the mesoscale Weather Research and Forecasting (WRF) model from the National Oceanic and Atmospheric Administration (NCAR), mainly in its solar version, WRF-Solar (Skamarock *et al.*, 2021; Jimenez *et al.*, 2016). WRF-Solar has features that serve the solar industry, incorporating improvements and the inclusion of other solar components. The simulation was conducted with 100 points to the east and west and 106 points between north and south. Thus, the model has been used in a variety of studies that demonstrate its ability to simulate G (Vieiraa *et al.,* 2020; Da Silva Ramos *et al.,* 2025).

2.4 Parameterizations

Emiliavaca *et al.*, 2024 uses some parameterizations for the study area comprising RN, comparing two parameterizations in the dry and rainy seasons, the best parameters, and based on the results obtained, the parameterizations selected for microphysics were Single-Moment 5-class, for Grell 3D Grell and Devenyi (2002) clouds , in the planetary boundary layer Mynn 2.5 TKE (Janjic, 1994; Mesinger, 1993), for soil the Noah model NIU *et al.* (2011) and shortwave and longwave parameterizations with the Rapid Radiative Transfer Model for Global Climate (RRTMG) Iacono *et al.* (2008).

2.4.2 WRF Post-processing

The study facilitated the obtaining of the best parameterizations for the region, requiring less time to perform sensitivity tests. The results are based on hourly outputs, with the first 6 hours of simulations removed to use only results with the most stable and reliable model. The model results are hourly and were subsequently integrated into monthly values so that the values can be compared. For the INMET reference weather station, the error values are shown in the following table. Metrics such as correlation (r), which helps explain the model in relation to observational data, are essential to show consistency with the simulated data. In addition, the root mean square error (RMSE) metric facilitates understanding of the best results found. Thus, the best correlation values (r) were in February and March, and the same is true for the RMSE metric.

Table 2: Error metrics for Mossoró (RN).

Month	January	February	March	April	May	June
RMSE	97,8	75,2	77,1	107,1	85,4	79,9
r	0,95	0,97	0,97	0,94	0,95	0,96
Month	July	August	September	October	November	December
RMSE	85,0	97,9	120,4	90,2	107	118,2
r	0,95	0,94	0,96	0,96	0,95	0,95

2.5 Cluster analysis for both spatial data

Despotovic *et al*. (2024) clustered weather stations in Spain into clusters for locations lacking long historical time series, with the aim of spatializing irradiance data (G). Based on this premise, spatial datasets such as BRASIL-SR model enable the comparison and clustering of points to provide information in regions lacking sensors. In this study, following the simulations and post-processing of the WRF-Solar model, the G data were interpolated to the grid points of the BRASIL-SR model.

In this study, following the WRF-Solar simulations and post-processing, G data were mapped onto the BRASIL-SR grid points via bilinear interpolation. Subsequently, cluster analysis was applied to both datasets; based on the silhouette plot, two groups were identified as the most representative for the spatial data. Consequently, cluster analysis was then applied to both sets of points from the models (BRASIL-SR and WRF-Solar), and the results were compared spatially and through silhouette plot analysis, which indicated that clustering into two groups most effectively represents the modeled data.

All data were aggregated into a representative monthly average for each point; after characterizing the groups, a monthly average was calculated for the groups of each model and compared.

2.6 Selecting the number of clusters

According to Rousseeuw (1987), the method that best indicates the number of clusters and the quality of clustering is the silhouette (S), which ranges from -1 to 1, where 1 represents the best fit and -1 the worst. Thus, using the value of S, it is possible to determine a representative number of groups, as shown in the following equation.

$$-1 \leq S \leq 1 \qquad \text{(eq. 1)}$$

Based on the analysis of the dendrogram results using Ward's clustering method, it was possible to define the quality of the fit and the number of cluster groups Hervada-Sala e Jarauta-Bragulat (2004). The Figure 2 illustrates the distribution of S values in histograms. Analysis showed that the greater the number of groups, the poorer the fit; for this reason, and because the number of groups with an S value closest to 1 was two, this choice was made.

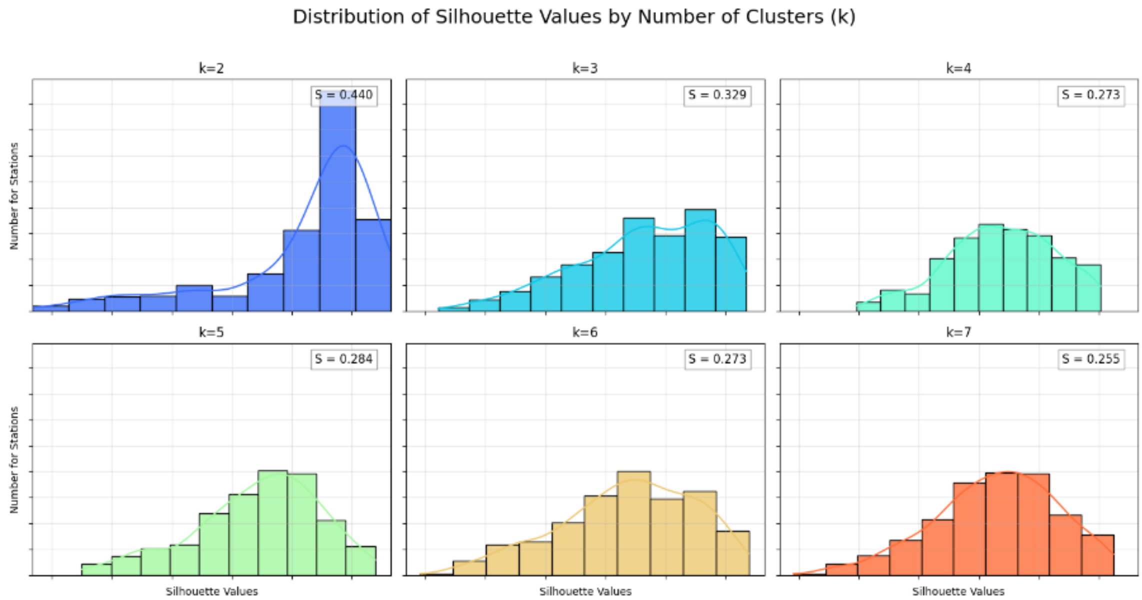
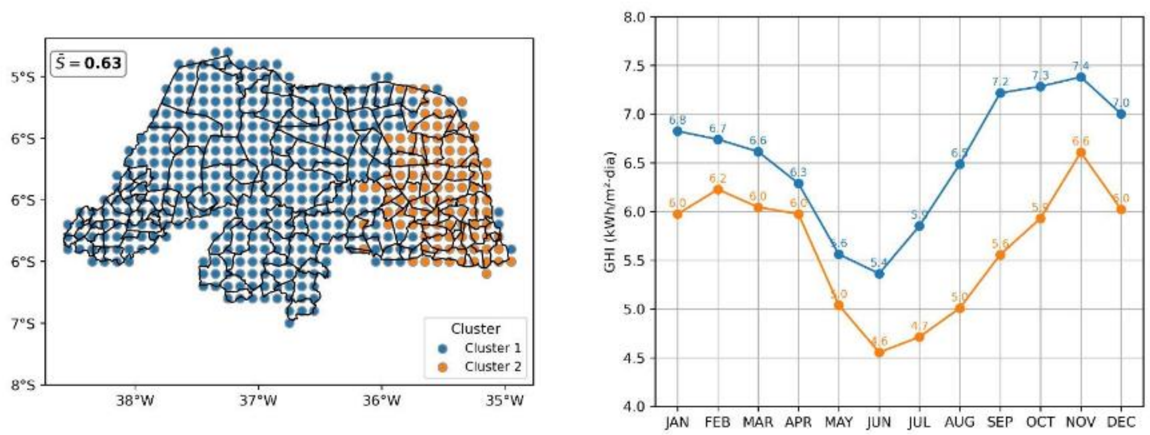


Figure 2: Distribution of silhouette values from Atlas data, according to the number of clusters.

The data were divided into 502 points according to the geographic coordinates of the atlas data points. For each month, a G value was assigned, and in the post-processing modeling data (WRF-Solar), a bilinear interpolation was performed for each coordinate point.

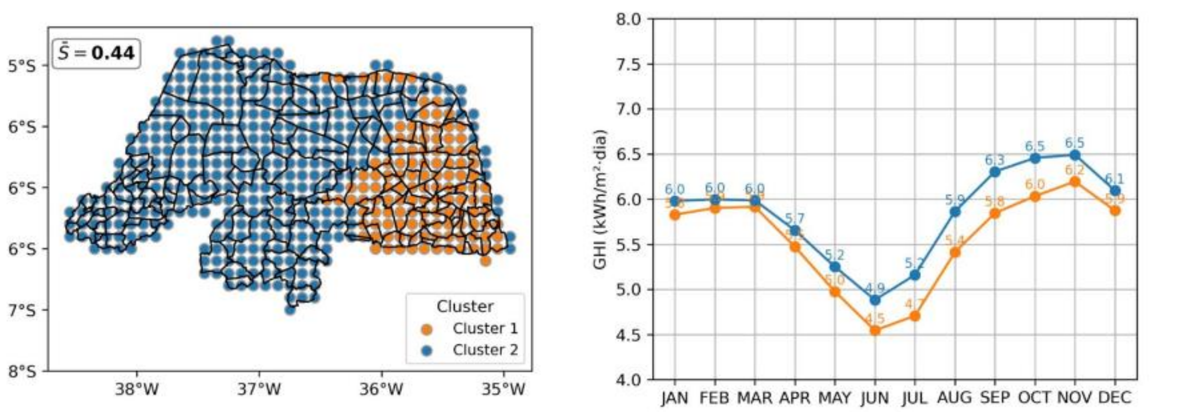
3. Results and Discussion

The study identifies the modeling results as effective in estimating the spatial distribution of the G for the state of RN, even with the accuracy of the simulation performed for an observational station point in RN. The best performance was in cluster 2, where the silhouette value was most consistent with the dataset, with S equal to 0.44 for the Brazil-SR data and 0.63 for the modeling results with WRF-Solar. The silhouette value is consistent with other analyses, that indicate a value of 0.5 as ideal for group classification (Govender *et al*., 2018).



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Cluster 1	6.8	6.7	6.6	6.3	5.6	5.4	5.9	6.5	7.2	7.3	7.4	7.0
Cluster 2	6.0	6.2	6.0	6.0	5.0	4.6	4.7	5.0	5.6	5.9	6.6	6.0

Figure 3: Spatial distribution according to k = 2 and annual cycle of G. Source: Solar Atlas INPE (2017)



	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
Cluster 1	5.8	5.9	5.9	5.5	5.0	4.5	4.7	5.4	5.8	6.0	6.2	5.9
Cluster 2	6.0	6.0	6.0	5.7	5.2	4.9	5.2	5.9	6.3	6.5	6.5	6.1

Figure 4: Spatial distribution according to k = 2 and the annual cycle of the G measured by the WRF-Solar model.

According to Figures 3 and 4, the spatial distribution of the clusters is very similar, indicating that the validated and measured data are consistent with the simulation results. Even though the climate of the central region of RN is semi-arid, the distribution of G throughout the year indicates that only the eastern coastal portion has lower values compared to other portions of the state.

Regarding error analysis, the underestimation of BIAS, on average, each month, there is a difference of 0.2 kWh/m²/day for the second cluster and a monthly average difference of 1 kWh/m²/day for the first cluster in relation to the Atlas data. The largest error in cluster 1 occurred in September and October, with a difference of 1.4 kWh/m²/day, and the smallest between March and May. For the second cluster, the differences were smaller, reaching zero in the first quarter of the year, with underestimation values throughout the months, the largest being in September, reaching 0.9 kWh/m²/day.

In cases of clear skies, WRF-Solar errors decrease significantly compared to intermittent days, thus this factor has a greater influence than domains and resolutions (Prasad and Kay, 2020). As for the monthly results of the simulations, clear and intermittent days are combined, thus the calculated error is acceptable. Some studies perform simulations of global radiation in different seasons to demonstrate the uncertainties associated with the simulations (Lu *et al*., 2023), mainly because the correlation values are close to 0.9.

4. Conclusions

The study demonstrated that WRF-Solar is effective in estimating the spatial distribution of G for the state of RN, even with a simulation performed predominantly from the semi-arid region. The best performance occurred in cluster 2, mainly in the summer (DJFM), while the best performance of cluster 1 was in the period from March to May. The differences highlight the importance of reliability in some months and the quantification of errors in other times of the year, such as summer, which has maximum irradiation mainly in the semi-arid climate present in the state of Rio Grande do Norte and achieves a high-level result through the WRF-Solar model.

Several studies indicate the uncertainties quantified in numerical models Chen (2021), but the objective of the study involved the spatial classification of the resource and the predictive ability of the model. The practical implications of these results for solar energy projects in RN are the availability of the resource throughout the months and similarity in some parts of the state.

5. Acknowledgments

Thanks to the SENAI Institute for Innovation in Renewable Energy (ISI-ER) and the Graduate Program in Climate Sciences at the Federal University of Rio Grande do Norte.

6. References

Alvares, C. A., Stape, J. L., Sentelhas, P. C., Gonçalves, J. D. M., & Sparovek, G. (2013). Köppen's climate classification map for Brazil. Meteorologische zeitschrift, 22(6), 711-728.

Chen, L. (2021). Uncertainties in solar radiation assessment in the United States using climate models. Climate Dynamics, 56(1), 665-678.

Emiliavaca, S. D. A. S., Amorim, A. C. B., Dantas, V. A., de Sousa, A. R., & Freire, J. L. M. (2024, September). SENSITIVITY TESTS USING WRF-SOLAR TO ESTIMATE GLOBAL HORIZONTAL IRRADIATION IN FIVE CLIMATIC REGIONS OF BRAZIL. In Proceedings of the Brazilian Solar Energy Congress (CBENS).

Despotovic, M., Voyant, C., Garcia-Gutierrez, L., Almorox, J., & Notton, G. (2024). Solar irradiance time series forecasting using auto-regressive and extreme learning methods: Influence of transfer learning and clustering. Applied Energy, 365, 123215.

Fovell, R. G., & Fovell, M. Y. C. (1993). Climate zones of the conterminous United States defined using cluster analysis. Journal of climate, 6(11), 2103-2135.

Grell, G. A., & Dévényi, D. (2002). A generalized approach to parameterizing convection combining ensemble and data assimilation techniques. Geophysical Research Letters, 29(14), 38-1.

Govender, P., Brooks, M. J., & Matthews, A. P. (2018). Cluster analysis for classification and forecasting of solar irradiance in Durban, South Africa. Journal of Energy in Southern Africa, 29(2), 51-62.

Janjić, Z. I. (1994). The step-mountain eta coordinate model: Further developments of the convection, viscous sublayer, and turbulence closure schemes. Monthly weather review, 122(5), 927-945.

Jimenez, P. A., Hacker, J. P., Dudhia, J., Haupt, S. E., Ruiz-Arias, J. A., Gueymard, C. A., ... & Deng, A. (2016). WRF-Solar: Description and clear-sky assessment of an augmented NWP model for solar power prediction. Bulletin of the American Meteorological Society, 97(7), 1249-1264.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. Quarterly journal of the royal meteorological society, 146(730), 1999-2049.

Hervada-Sala, C., & Jarauta-Bragulat, E. (2004). A program to perform Ward's clustering method on several regionalized variables. Computers & Geosciences, 30(8), 881-886.

Iacono, M. J., Delamere, J. S., Mlawer, E. J., Shephard, M. W., Clough, S. A., & Collins, W. D. (2008). Radiative forcing by long-lived greenhouse gases: Calculations with the AER radiative transfer models. Journal of Geophysical Research: Atmospheres, 113(D13).

Lu, Y. B., Wang, L. C., Zhou, J. J., Niu, Z. G., Zhang, M., & Qin, W. M. (2023). Assessment of the high-resolution estimations of global and diffuse solar radiation using WRF-Solar. Advances in Climate Change Research, 14(5), 720-731.

Martins, F. R., Pereira, E. B., Silva, S. A. B., Abreu, S. L., & Colle, S. (2008). Solar energy scenarios in Brazil, Part one: Resource assessment. Energy Policy, 36(8), 2853-2864.

Martins, F. R., Pereira, E. B., Gonçalves, A. R., Costa, R. S., Lima, F. J., Rüther, R., ... & Souza, J. G. D. (2017). Atlas brasileiro de energia solar.

Mimmack, G. M., Mason, S. J., & Galpin, J. S. (2001). Choice of distance matrices in cluster analysis: Defining regions. Journal of climate, 14(12), 2790-2797.

Mesinger, F. (1993). Forecasting upper tropospheric turbulence within the framework of the Mellor-Yamada 2.5 closure. Res. Activ. Atmos. Oceanic Mod.

Niu, G. Y., Yang, Z. L., Mitchell, K. E., Chen, F., Ek, M. B., Barlage, M., ... & Xia, Y. (2011). The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. Journal of Geophysical Research: Atmospheres, 116(D12).
NREL. NSRDB. Disponível em: <https://nsrdb.nrel.gov/data-viewer>.

Prasad, A. A., & Kay, M. (2020). Assessment of simulated solar irradiance on days of high intermittency using WRF-solar. Energies, 13(2), 385.

September sees 1.4 GW expansion in the electricity matrix, solely from renewable sources. Available at: <https://www.gov.br/aneel/pt-br/assuntos/noticias/2025/setembro-registra-expansao-de-1-4-gw-na-matriz-eletrica-somente-com-fontes-renovaveis>. Acesso em: 15 out. 2025.

Sengupta, M., Habte, A., Gotseff, P., Weekley, A., Lopez, A., Molling, C., & Heidinger, A. (2014). Physics-Based GOES Satellite Product for Use in NREL's National Solar Radiation Database (No. NREL/CP-5D00-62237). National Renewable Energy Lab.(NREL), Golden, CO (United States).

Sengupta, M., Xie, Y., Lopez, A., Habte, A., Maclaurin, G., & Shelby, J. (2018). The national solar radiation data base (NSRDB). Renewable and sustainable energy reviews, 89, 51-60.

Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Liu, Z., Berner, J., ... & Huang, X. (2021). A Description of the Advanced Research WRF Model Version 4.3, NCAR Tech. Boulder: National Center for Atmospheric Research.

Solano, E. S., & Affonso, C. M. (2025). Wind and solar-based green hydrogen potential in Brazil: Production, costs, and CO2 mitigations. International Journal of Hydrogen Energy, 172, 151212.

da Silva Ramos, D. N., de Lima, F. J. L., Weyll, A. L. C., dos Santos, T. S., Kitagawa, Y. K. L., Jacondino, W. D., ... & Moreira, D. M. (2025). SHORT-TERM FORECAST OF GLOBAL HORIZONTAL IRRADIATION IN THE BRAZILIAN SEMI-ARID REGION UNDER DIFFERENT CLOUDINESS CONDITIONS USING THE WRF-SOLAR MODEL. Brazilian Journal of Solar Energy, 16(1), 9-18.

Rousseeuw, P. J. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics, 20, 53-65.

Vieiraa, C. S., da Silva Palmeiraa, A., Acioli, P. M., Diasa, E. G. S. N., Moreiraa, D. M., & Cimatec, C. U. S. (2020). Solar resource mapping in the vale São Francisco da Bahia using the WRF-Solar model. In VI International Symposium on Innovation and Technology (SIINTEC).

Tiba, C., Fraidenraich, N., Gallegos, H. G., & Lyra, F. J. M. (2002). Solar energy resource assessment—Brazil. Renewable Energy, 27(3), 383-400.