Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Hourly solar-radiation variability, forecastability and ENSO modulation in Pernambuco, northeastern Brazil

Fábio Sandro dos Santos

fabio.santos@ufpi.edu.br

Fábio Sandro dos Santos

Fernando Henrique Antunes de Araújo

Instituto Federal de Educação Ciência e Tecnologia da Paraíba

Kerolly Kedma Felix Nascimento

Universidade Regional do Cariri

Research Article

Keywords: hourly solar radiation, Pernambuco, permutation entropy, Fisher information, ENSO, northeastern Brazil

Posted Date: August 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10047768/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Hourly solar-radiation variability, forecastability and ENSO modulation in Pernambuco, northeastern Brazil

Fábio Sandro dos Santos (ORCID: https://orcid.org/0000-0002-0135-4981)

Fernando Henrique Antunes de Araújo (ORCID: https://orcid.org/0000-0001-5417-0513)

Kerolly Kedma Felix do Nascimento (ORCID: https://orcid.org/0000-0001-7360-0961)

Abstract

Hourly solar radiation controls surface energy exchange, cloud-climate diagnostics and photovoltaic predictability, yet station-scale assessments often emphasize mean irradiance rather than temporal organization. This study examines 630,298 quality-controlled hourly global solar-radiation records from twelve automatic stations in Pernambuco, northeastern Brazil, spanning coastal, Zona da Mata, Agreste and semi-arid Sertao environments. We combine descriptive climatology, monthly, seasonal and annual diagnostics, Bandt-Pompe permutation entropy, Fisher information, Fisher-Shannon representation, inverse-distance spatial mapping, ENSO stratification, common-period sensitivity and predictive validation. Mean radiation is highest in the semi-arid interior, but magnitude and temporal organization are not equivalent. Cabrobo records the highest mean hourly irradiation (456.2 Wh m^-2), whereas Palmares has the lowest mean (366.6 Wh m^-2) and the highest coefficient of variation. Recife, Palmares and Surubim are closest to the stochastic reference, while Petrolina, Ibimirim, Cabrobo and Serra Talhada show stronger ordinal organization. The Sertao has the largest relative forecastability index, and this ranking remains stable in a common 2018 window. ENSO stratification indicates spatially heterogeneous anomalies, with the clearest signal along the coast during El Nino months. The results show that solar-radiation forecastability is spatially decoupled from mean resource magnitude, providing a meteorological basis for climate-aware radiation assessment in northeastern Brazil.

Keywords: hourly solar radiation; Pernambuco; permutation entropy; Fisher information; ENSO; northeastern Brazil

1. Introduction

Solar generation is moving from an additional renewable option to a strategic component of power-system planning. The global energy transition now depends on rapid electrification, renewable expansion and the ability of grids to integrate large shares of weather-dependent generation. The International Energy Agency projects renewable capacity to grow strongly by 2030, with solar photovoltaic and wind technologies accounting for most new renewable additions, but the main-case trajectory still remains below the global tripling pledge agreed at COP28

(International Energy Agency 2024). In Brazil, this transition occurs inside an electricity system that is already predominantly renewable but increasingly exposed to weather-dependent variability. The Brazilian Energy Balance 2025 reports that wind and solar photovoltaic generation together accounted for 23.7% of national electricity generation in 2024, and national planning keeps renewable participation above 85% while expanding solar, wind and distributed generation (Empresa de Pesquisa Energética 2025a, 2025b). ANEEL's generation information system further shows that solar assets are now part of both centralized and distributed expansion decisions (Agência Nacional de Energia Elétrica 2026). Solar-resource assessment has therefore become a renewable-energy integration problem linked to forecasting, storage, grid flexibility and climate-risk screening, not only a climatological description.

The Brazilian Northeast is central to this question. The region combines high solar availability, large semi-arid areas, rapid renewable development, vulnerable distribution grids and transmission corridors that must absorb larger shares of variable generation. The Brazilian Solar Energy Atlas identifies the Northeast as one of the most favourable solar-resource regions in the country (Pereira et al. 2017). Pernambuco is especially useful as a state-scale test case because it compresses strong environmental contrasts into a relatively small territory: humid Atlantic coast, Zona da Mata, Agreste transition belt and semi-arid Sertao. These sectors may show similar mean radiation during some periods but differ in cloud modulation, humidity, seasonality, convective variability and clear-sky recurrence. For renewable-energy planning in Pernambuco, the relevant question is not only where annual irradiation is high, but where hourly radiation is regular enough to support photovoltaic forecasting, storage dispatch, voltage control and future hydrogen or irrigation-linked electricity demand.

The scientific gap is that many solar-resource assessments still privilege magnitude: annual or monthly mean irradiance, seasonal totals and spatial interpolation. These metrics are indispensable, but they do not tell whether the hourly radiation sequence is dynamically organized or dominated by a broad diversity of short-term patterns. For photovoltaic operation, this matters. Forecasting difficulty, ramp frequency, inverter stress, storage cycling, reserve needs and electrolyser scheduling are affected by temporal structure as much as by mean resource (Pedro and Coimbra 2012; Inman et al. 2013; Voyant et al. 2017; Kleissl 2013). Cloud-induced variability and aerosol effects can alter radiation at time scales that are operationally relevant for renewable-energy systems with high photovoltaic penetration (Gueymard 2012; Lave and Kleissl 2013).

Information-theory-based quantifiers provide a rigorous way to add this missing dimension. Bandt–Pompe permutation entropy converts a time series into ordinal patterns and measures how evenly those patterns are visited (Bandt and Pompe 2002; Riedl et al. 2013). Fisher information, estimated on the same ordinal distribution, measures local concentration and structural sharpness. The Fisher–Shannon representation can distinguish stochastic signals from organized dynamics even when conventional moments are similar (Martin et al. 2006; Rosso et al. 2007; Zunino et al. 2010). This makes the approach well suited for solar radiation, where deterministic astronomical forcing is modulated by cloud, humidity and regional climate variability.

This study contributes a Pernambuco-focused framework for renewable-energy planning that goes beyond the earlier magnitude-only view. The novelty is fivefold: (i) it builds a robust hourly station database with explicit unit conversion and station-year filtering; (ii) it evaluates mean radiation together with ordinal organization and forecastability; (iii) it separates monthly, seasonal and annual behaviour; (iv) it stratifies anomalies by El Niño, La Niña and neutral months using the NOAA Oceanic Niño Index; and (v) it adds common-period sensitivity, predictive validation and conservative exploratory scenarios to 2030. The central hypothesis is that solar-resource magnitude and short-term temporal forecastability are spatially decoupled in Pernambuco. Testing this hypothesis matters because sites with similar irradiation can impose different requirements for forecasting, storage and grid operation. The objective is not to replace satellite products, physical irradiance models or operational PV forecasts, but to provide a reproducible screening layer that identifies where solar radiation is abundant, organized, stochastic or climate-phase sensitive before deeper grid-integration, storage and techno-economic studies are performed.

2. Study area and data

Pernambuco is located in northeastern Brazil and extends from the Atlantic coast to the interior semi-arid zone. The studied network contains twelve automatic meteorological stations: Arcoverde, Cabrobo, Floresta, Garanhuns, Ibimirim, Ouricuri, Palmares, Petrolina, Recife, Salgueiro, Serra Talhada and Surubim. Hourly observations were read directly from the CSV files in the Pernambuco data folder and are consistent with the structure of Brazilian automatic-station radiation records (Instituto Nacional de Meteorologia 2026). The analysis used the global solar-radiation field, timestamp fields and station metadata. The radiation field is supplied as hourly accumulated energy in kJ m^{-2}; for energy-sector interpretation it was also converted to Wh m^{-2} using $R_{\mathrm{Wh}} = R_{\mathrm{kJ}}/3.6$.

The stations were grouped into four climatic sectors. Recife represents the coast; Palmares represents the Zona da Mata; Arcoverde, Garanhuns and Surubim represent the Agreste; and the remaining seven stations represent the

Sertao. These labels are practical regional classes rather than rigid meteorological boundaries. They allow interpretation of the transition from humid maritime influence to semi-arid interior regimes.

Quality control removed missing and non-numeric radiation values and retained observations in the interval $0 \leq R_{\mathrm{kJ}} \leq 5000$. This broad envelope excludes implausible values while preserving the observed daytime distribution. The first-pass database contains 692,560 valid hourly records. To reduce bias from unequal coverage, station-year combinations with fewer than 3000 valid hourly records or annual mean radiation below 500 kJ m^{-2} were excluded from the statistical and information-theory-based analyses. The resulting analysis-ready database contains 630,298 hourly records. Table 1 summarizes station statistics, and Fig. 1 shows the study area, elevation, station locations and regional sectors used in the analysis.

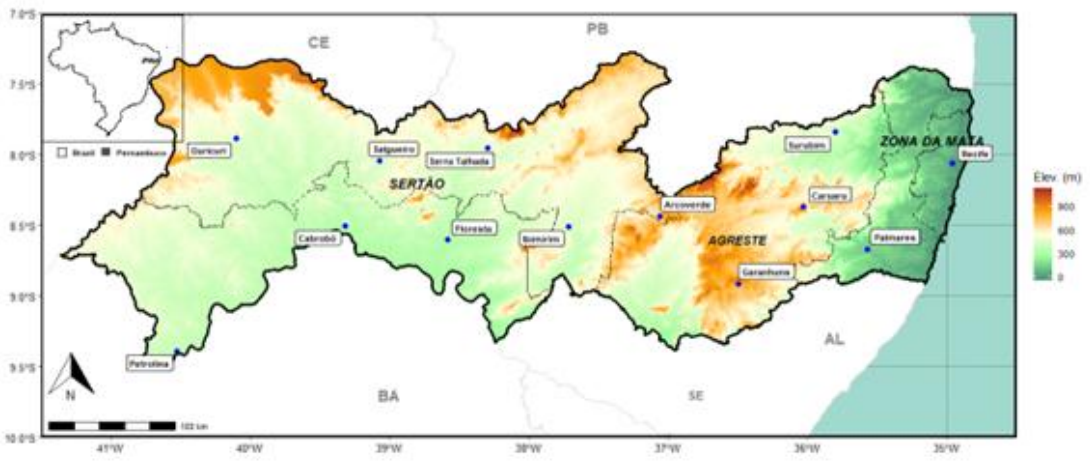


Fig. 1 *Study area in Pernambuco, northeastern Brazil, showing elevation, automatic meteorological stations and the regional sectors used in the analysis*

Table 1 *Analysis-ready hourly solar-radiation statistics for the Pernambuco station network. Mean radiation is reported in both kJ m^{-2} and Wh m^{-2}*

Station	**Region**	**N**	Mean kJ m^{-2}	Mean Wh m^{-2}	SD kJ m^{-2}	**CV**	**Skew.**
Arcoverde	Agreste	73235	1553.7	431.6	1125.2	0.724	0.185
Cabrobo	Sertao	60157	1642.4	456.2	1162.0	0.708	0.161
Floresta	Sertao	53386	1490.5	414.0	1172.7	0.787	0.232

Station	Region	N	Mean kJ m^{-2}	Mean Wh m^{-2}	SD kJ m^{-2}	CV	Skew.
Garanhuns	Agreste	60014	1454.6	404.1	1096.7	0.754	0.312
Ibimirim	Sertao	56387	1629.8	452.7	1170.9	0.718	0.129
Ouricuri	Sertao	44443	1623.8	451.1	1134.5	0.699	0.140
Palmares	Zona da Mata	44769	1319.9	366.6	1002.8	0.760	0.310
Petrolina	Sertao	57647	1477.7	410.5	1057.3	0.716	0.350
Recife	Coast	56323	1613.4	448.2	1170.4	0.725	0.225
Salgueiro	Sertao	8779	1639.3	455.4	1170.9	0.714	0.144
Serra Talhada	Sertao	56041	1605.3	445.9	1182.0	0.736	0.155
Surubim	Agreste	59117	1581.0	439.2	1095.9	0.693	0.146

3. Methods

The analysis was designed as a screening framework with four linked blocks: descriptive resource assessment, ordinal diagnostics, climate-period stratification and exploratory future scenarios. The methodological sequence keeps the emphasis on reproducible processing steps, statistical definitions and interpretation limits.

3.1. Descriptive climatology and temporal scales

For each station, the mean $\bar{R}$, standard deviation σ_R, coefficient of variation $CV = \sigma_R/\bar{R}$, skewness and kurtosis were computed after quality control. Hourly profiles were summarized by median and interquartile range. Monthly means were computed by station, year and month, and seasonal means followed austral definitions: summer (DJF), autumn (MAM), winter (JJA) and spring (SON). Annual means were retained when at least nine valid monthly means were available. Relationships among seasonal and annual means were evaluated using Pearson correlation. Unless explicitly stated otherwise, temporal-scale comparisons are expressed in Wh m^{-2}.

3.2. Ordinal probabilities and permutation entropy

Let $X = \{x_1, x_2, \dots, x_N\}$ be a station radiation series. For embedding dimension D and delay τ, vectors are formed as

$$\boldsymbol{x}_i^{(D,\tau)} = (x_i, x_{i+\tau}, \dots, x_{i+(D-1)\tau}).$$

The Bandt–Pompe symbolic approach is widely used because it captures temporal ordering without assuming a parametric distribution and is comparatively robust to monotonic transformations and observational noise (Bandt and Pompe 2002; Riedl et al. 2013; Rosso et al. 2007; Zunino et al. 2010). Each vector is replaced by the permutation π that orders its values. The relative frequencies of all observed permutations define the ordinal probability distribution $P = \{p_j\}$. The normalized permutation entropy is

$$H(P) = -\frac{1}{\ln(D!)}\sum_{j} p_j \ln p_j.$$

Values close to one indicate a broad and nearly uniform use of ordinal patterns. Lower values indicate recurrent temporal order and stronger short-term structure. This study used $D = 5$ and $\tau = 1$, capturing hourly ordering while maintaining enough ordinal samples for station-scale estimation.

3.3. Fisher information and forecastability index

Fisher information was estimated on the ordinal distribution as

$$F(P) = \frac{1}{2}\sum_{j=1}^{M-1}\left(\sqrt{p_{j+1}} - \sqrt{p_j}\right)^2,$$

where M is the number of observed ordinal states. Ordinal states were arranged in lexicographic order before computing F, following Fisher–Shannon applications to Bandt–Pompe distributions (Martin et al. 2006; Rosso et al. 2007). Fisher information increases when probability mass is locally concentrated, complementing entropy's global disorder measure.

The pair (H, F) was compared with the stochastic reference $(1,0)$. The Euclidean distance

$$d_i = \sqrt{(1 - H_i)^2 + F_i^2}$$

measures the departure of station i from a high-entropy, low-structure reference. A relative forecastability index was defined as

$$\Phi_i = \frac{d_i - \min(d)}{\max(d) - \min(d)}.$$

The index varies from zero to one within the station network. It is not an out-of-sample forecast skill score. It is an ordinal organization indicator: high values mean that the observed hourly sequence contains more recurrent ordering and may therefore be more favourable for short-term forecasting and scheduling.

Two additional checks were used to evaluate whether the ordinal ranking was robust and operationally meaningful. First, Φ was recomputed for the common 2018 window, the year with simultaneous coverage across all stations, and compared with the full robust-period ranking using Spearman correlation. Second, a station-wise ExtraTrees benchmark (Geurts et al. 2006) was trained with cyclic hour and month terms plus one-hour, three-hour and 24-hour lagged radiation features. ExtraTrees was selected as a compact nonlinear benchmark because it captures interactions among diurnal, seasonal and lagged predictors with limited parameter tuning. The last 20% of each station record was held out chronologically, and the benchmark was compared with one-hour persistence and month–hour climatology baselines. This exercise is not proposed as an operational forecast system or as an exhaustive model-selection study; it tests whether stations with larger Φ tend to have lower out-of-sample errors under a transparent predictive protocol.

3.4. Spatial diagnostics, ENSO stratification and scenarios

Inverse-distance weighting was used to produce exploratory spatial fields for mean radiation, variability and information-theory-based quantities (Shepard 1968). The maps are station-constrained diagnostics rather than final gridded climatologies.

ENSO influence was evaluated with monthly Oceanic Niño Index values from NOAA (NOAA Climate Prediction Center 2026). Months with ONI ≥ 0.5 were classified as El Niño, months with ONI ≤ -0.5 as La Niña and the remaining months as neutral. This is a monthly phase-stratification diagnostic rather than an official event-by-event ENSO classification. For each station and month, anomalies were computed relative to that station's monthly climatology. These anomalies were then aggregated by station, region and ENSO phase, with bootstrap 95% intervals estimated by resampling monthly anomalies (Efron and Tibshirani 1993). This design avoids confusing a dry-season month with an ENSO effect because the monthly seasonal cycle is removed before phase comparison.

Finally, exploratory scenarios to 2030 were produced from annual regional means. Linear trend extension with bootstrap residual uncertainty was used only when the historical record supported a stable extrapolation. When station coverage or trend behaviour made extrapolation unstable, a conservative climatological-persistence scenario was used. These scenarios are screening products, not deterministic climate projections, bankable yield estimates or climate-change simulations.

4. Results

4.1. Radiation magnitude and variability

The network displays a strong daily cycle at all stations, with near-zero nighttime radiation and daytime peaks from late morning to afternoon (Fig. 2). The hourly boxplots show both the central daytime distribution and the spread around the hourly mean. The semi-arid stations preserve a broader and more regular daytime envelope, while the coast and Zona da Mata show more evident modulation. Cabrobo has the highest mean hourly irradiation (456.2 Wh m^{-2}), while Palmares has the lowest mean (366.6 Wh m^{-2}) and highest coefficient of variation (0.760), consistent with stronger cloud and humidity effects. Surubim has the lowest coefficient of variation (0.693), indicating a comparatively regular daily envelope within the Agreste.

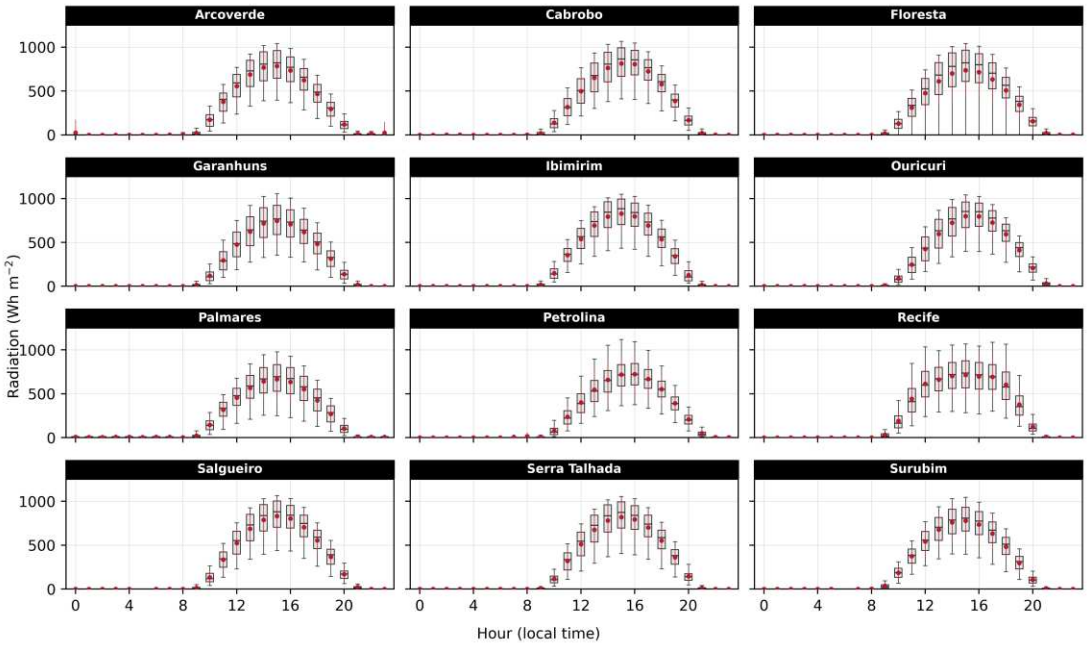


Fig. 2 *Diurnal solar-radiation profiles by station. Boxes show the hourly interquartile range, whiskers show the 5th-95th percentile interval, and red markers with vertical bars show the hourly mean ± one standard deviation*

The month–hour heatmaps reinforce the sectoral contrast (Fig. 3). The Sertao shows a persistent daytime radiation structure, especially in spring, while the Zona da Mata has lower and more variable values. The Agreste occupies a transition position. These results are important because PV operation depends on the shape and seasonal timing of the daytime window, not only on the annual mean.

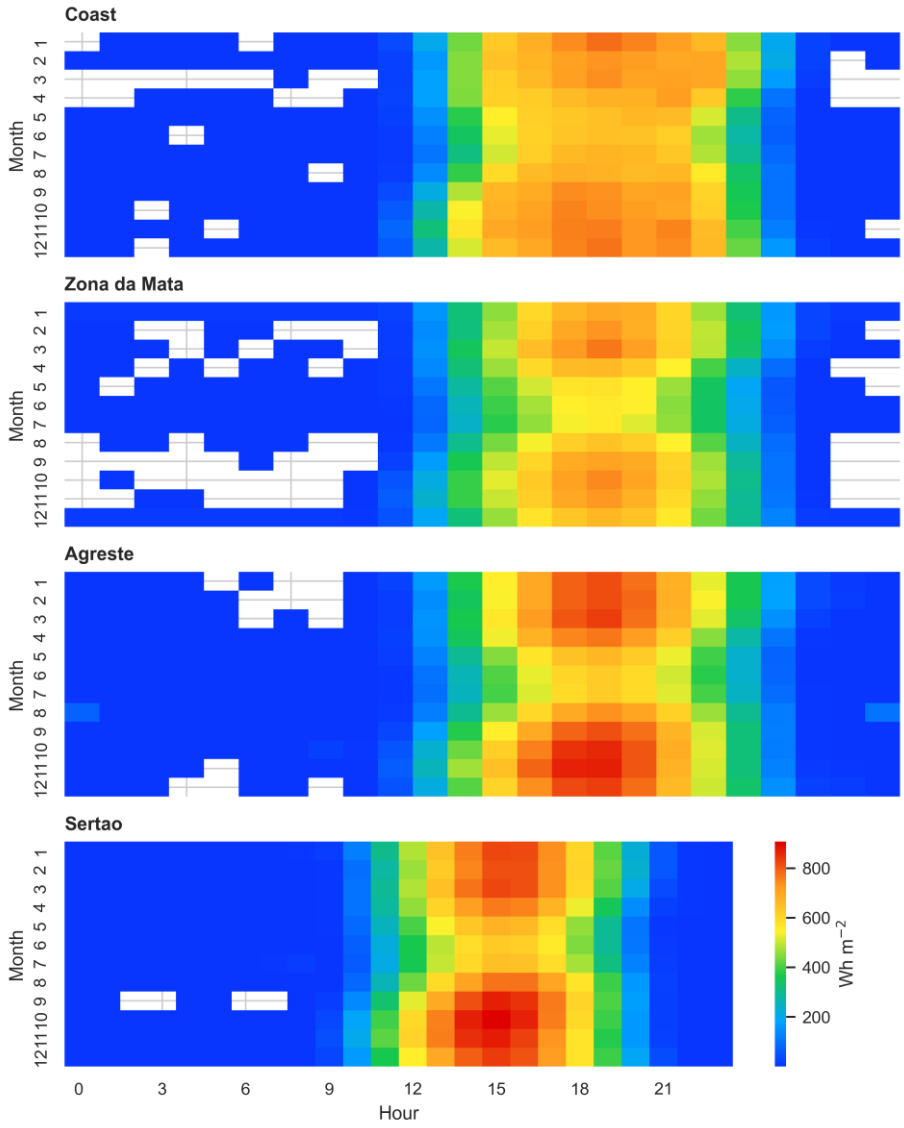


Fig. 3 *Month-hour mean radiation by climatic sector. The colour scale is expressed in Wh m^{-2}*

Spatial diagnostics of descriptive moments show that mean radiation and variability do not follow exactly the same geography (Fig. 4). Palmares forms a low-mean and high-CV point in the humid sector, whereas high mean values appear in the western and central interior. This mismatch is operationally relevant: storage sizing and ramp management respond strongly to variability, while annual production responds strongly to mean resource.

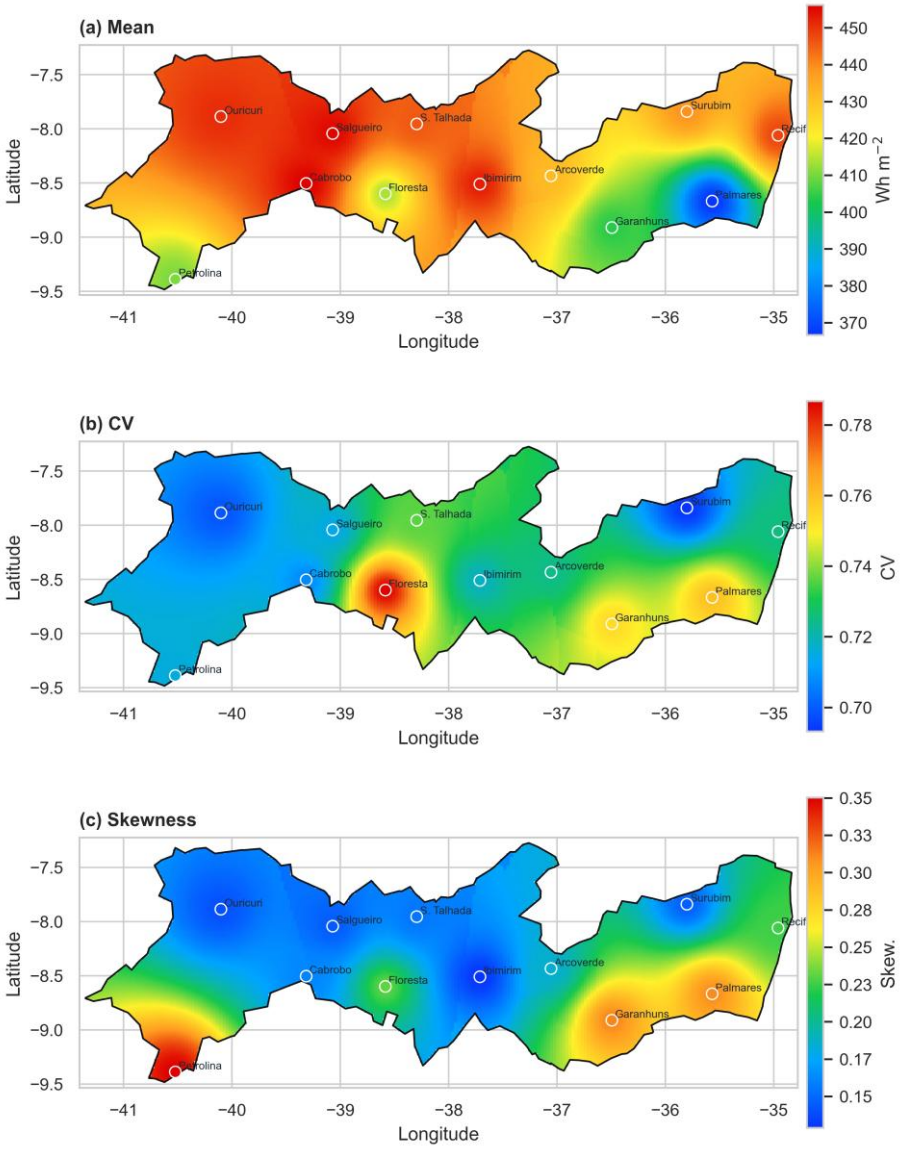


Fig. 4 Spatial diagnostics of mean hourly radiation, coefficient of variation and skewness. The blue-green-yellow-red palette emphasizes low-to-high gradients across Pernambuco

4.2. Entropy-Fisher ranking and forecastability

Table 2 gives the information-theory-based ranking. Recife is closest to the stochastic reference, followed by Palmares and Surubim. These stations have comparatively high entropy and low distance to (1,0), meaning that their ordinal sequences are more diverse within the network. At the opposite end, Petrolina, Ibimirim, Cabrobo and Serra Talhada have larger distances and higher Φ, indicating stronger temporal organization.

Table 2 *Information-theory-based ranking from normalized permutation entropy H, Fisher information F, distance d to the stochastic reference and relative forecastability Φ*

Rank	Station	H	F	d	Φ
1	Recife	0.757	0.164	0.293	0.000
2	Palmares	0.732	0.164	0.314	0.174
3	Surubim	0.715	0.176	0.335	0.348
4	Garanhuns	0.699	0.187	0.354	0.507
5	Floresta	0.687	0.194	0.368	0.622
6	Arcoverde	0.686	0.197	0.371	0.645
7	Salgueiro	0.673	0.207	0.387	0.782
8	Ouricuri	0.673	0.208	0.388	0.790
9	Serra Talhada	0.667	0.217	0.398	0.870
10	Cabrobo	0.667	0.217	0.398	0.871
11	Ibimirim	0.662	0.223	0.405	0.929
12	Petrolina	0.654	0.227	0.413	1.000

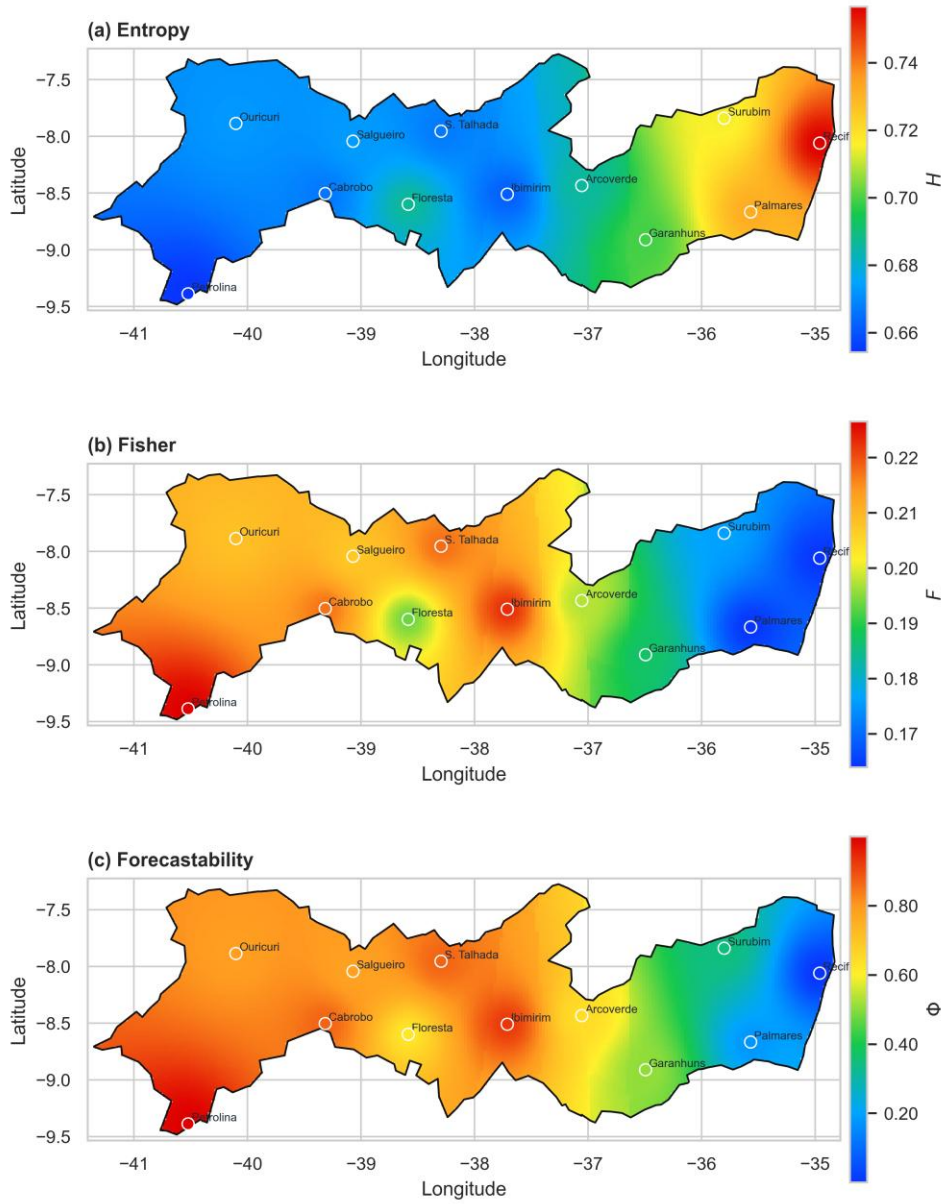
Because station records differ in length, the ordinal ranking was also tested on the common 2018 window (Table 3). The full robust-period and common-window rankings are strongly consistent ($\rho = 0.923$), indicating that the main forecastability gradient is not an artefact of unequal temporal coverage.

Table 3 *Common-period sensitivity of the relative forecastability index. The 2018 window is the year with simultaneous station coverage across the network*

Station	Φ robust	Φ 2018	Difference
Recife	0.000	0.000	+0.000
Palmares	0.174	0.588	+0.414

Station	Φ robust	Φ 2018	Difference
Surubim	0.348	0.551	+0.203
Garanhuns	0.507	0.571	+0.064
Floresta	0.622	0.860	+0.238
Arcoverde	0.645	0.868	+0.222
Salgueiro	0.782	0.790	+0.008
Ouricuri	0.790	0.875	+0.085
Serra Talhada	0.870	1.000	+0.129
Cabrobo	0.871	0.936	+0.066
Ibimirim	0.929	1.000	+0.071
Petrolina	1.000	0.968	-0.032

The spatial fields of entropy, Fisher information and forecastability show an interior tendency toward stronger ordinal organization (Fig. 5). This does not imply that coastal solar generation is unattractive. It means that the coastal and forest-zone sectors are more likely to require nowcasting, local flexibility or more conservative uncertainty margins.



203

204 ***Fig. 5*** *Information-theory-based organization of hourly radiation. Larger forecastability values indicate stronger*
205 *recurrent ordinal structure within the observed network*

206 The Fisher–Shannon plane clarifies why the magnitude and organization rankings differ (Fig. 6). Stations close to
207 the stochastic reference have broader ordinal diversity, while interior stations with larger distance from (1,0)
208 preserve stronger repeated ordering. This separation is the main added value of the method: it detects temporal
209 organization that is not visible in mean radiation alone.

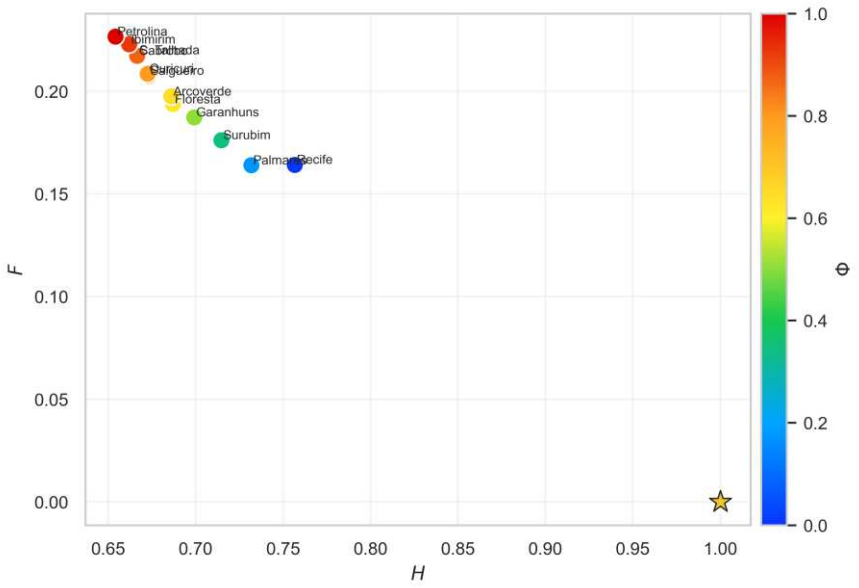


Fig. 6 *Fisher-Shannon representation of the stations. Colour indicates the relative ordinal forecastability index, and the star marks the stochastic reference (H,F)=(1,0)*

The predictive benchmark provides an independent operational check (Table 4; Fig. 7). Higher Φ is associated with lower ExtraTrees errors across stations, with Spearman correlations of -0.734 for RMSE ($p = 0.007$) and -0.783 for normalized MAE ($p = 0.003$). The benchmark also improves on simple baselines: mean RMSE is 93.5 Wh m^{-2} for ExtraTrees, compared with 173.6 Wh m^{-2} for one-hour persistence and 136.3 Wh m^{-2} for month–hour climatology. This corresponds to average RMSE reductions of 46.2% and 31.4%, respectively. These results support the interpretation of Φ as a useful screening indicator. Site-specific PV forecasting would still require dedicated predictors, sky imagery, satellite products or power-output data.

Table 4 *Out-of-sample predictive validation using a station-wise ExtraTrees benchmark. Errors are reported in Wh m^{-2}, and nMAE is normalized by each test-period mean. Persistence and month-hour climatology were used as simple baseline checks*

Station	Φ	RMSE	MAE	nMAE	R^2
Salgueiro	0.782	82.5	53.1	0.103	0.942
Petrolina	1.000	64.6	40.9	0.115	0.928
Ibimirim	0.929	86.5	56.0	0.121	0.930

Station	Φ	RMSE	MAE	nMAE	R^2
Floresta	0.622	87.6	56.7	0.125	0.926
Cabrobo	0.871	91.8	57.9	0.127	0.918
Serra Talhada	0.870	92.2	58.4	0.129	0.924
Arcoverde	0.645	94.9	61.7	0.145	0.903
Ouricuri	0.790	100.3	67.0	0.145	0.895
Surubim	0.348	102.0	70.9	0.156	0.886
Garanhuns	0.507	98.1	66.2	0.160	0.899
Palmares	0.174	96.4	65.0	0.184	0.878
Recife	0.000	124.8	82.8	0.186	0.847

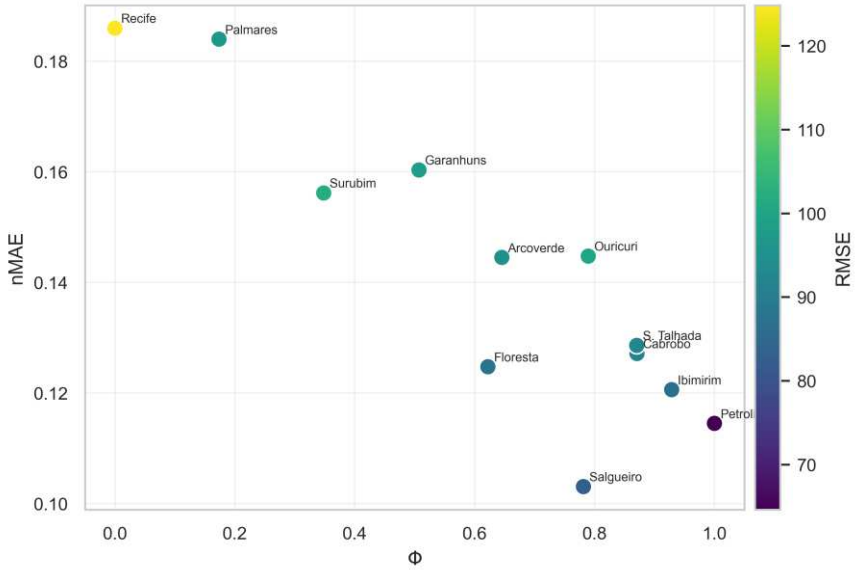


Fig. *7 Predictive validation of the ordinal forecastability index. Colour indicates station RMSE, and lower normalized MAE corresponds to easier short-term prediction in the benchmark model*

4.3. Monthly, seasonal and annual relationships

Table 5 and Fig. 8 show that seasonal timing is a major planning variable. Spring is the strongest season in the Agreste and Sertao, reaching mean regional values of 474.1 and 490.4 Wh m^{-2}, respectively. Winter is the weakest

229 season in all sectors, especially in the Zona da Mata (315.6 Wh m^{-2}). The coast has a strong summer signal (482.9

230 Wh m^{-2}), while the Sertao combines a high spring maximum with a comparatively resilient winter.

231 ***Table 5*** *Regional monthly-derived seasonal and annual radiation means in Wh* m^{-2}

Region	Summer	Autumn	Winter	Spring	Annual
Agreste	464.4	413.1	356.2	474.1	426.6
Coast	482.9	444.9	398.6	469.2	448.9
Sertao	467.9	416.4	376.2	490.4	437.4
Zona da Mata	383.6	369.4	315.6	406.8	369.3

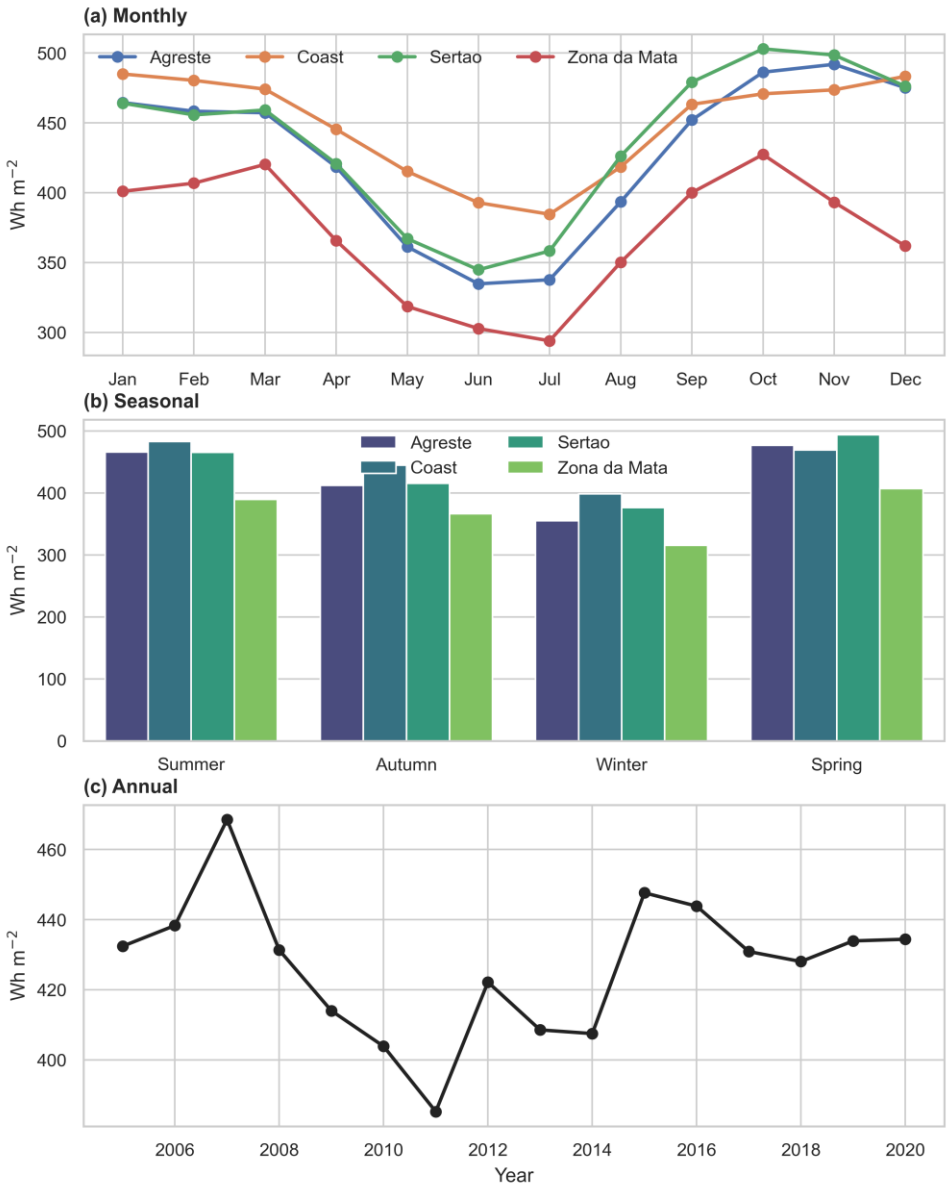


Fig. 8 *Monthly climatology, seasonal means and annual network mean*

The correlation matrix in Fig. 9 shows how periods are related. Annual mean radiation is most strongly connected with the high-radiation seasons, indicating that spring and summer conditions have disproportionate influence on annual resource assessment. This is important for planning because errors in characterizing peak-season regularity can propagate into annual yield and storage estimates.

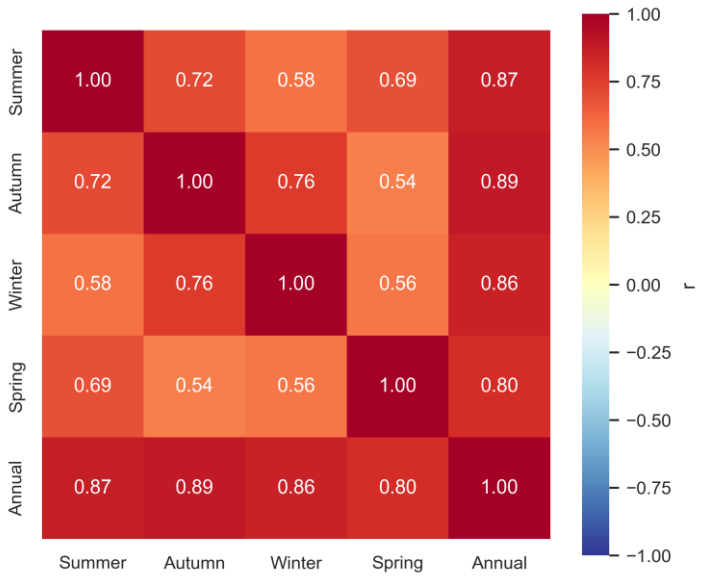


Fig. 9 *Relationships among seasonal and annual means based on station-year records*

4.4. ENSO influence and 2030 exploratory scenarios

ENSO-phase stratification reveals that climate teleconnections do not affect Pernambuco uniformly (Table 6; Fig. 10). El Niño months show the clearest positive anomaly on the coast (+6.73%, 95% interval +3.09 to +10.32%). Sertao El Niño anomalies are also positive on average (+1.23%), but the interval nearly includes zero. Most other regional phase intervals cross zero, including La Niña anomalies in the Agreste, coast, Sertao and Zona da Mata. The robust conclusion is therefore spatial heterogeneity with strong uncertainty outside the coastal El Niño signal. This response is consistent with the broader literature showing that Northeast Brazil climate variability is strongly affected by tropical Pacific and Atlantic forcing, but the sign and magnitude of local anomalies depend on season, circulation and regional moisture pathways (Hastenrath and Heller 1977; Andreoli and Kayano 2005; Marengo et al. 2017).

Table 6 *Mean monthly radiation anomaly by ENSO phase after removing each station's monthly climatology. Values are percentages, with bootstrap 95% intervals in brackets*

Region	La Niña	Neutral	El Niño
Agreste	+1.35 [-0.14,+2.87]	-0.52 [-1.53,+0.53]	-0.66 [-1.87,+0.51]
Coast	-2.77 [-6.46,+0.93]	-2.31 [-6.08,+1.70]	+6.73 [+3.09,+10.32]
Sertao	-1.90 [-4.91,+0.85]	+0.35 [-1.06,+1.65]	+1.23 [-0.02,+2.54]

Region	La Niña	Neutral	El Niño
Zona da Mata	-0.70 [-6.70,+4.46]	+1.85 [-0.46,+4.19]	-2.02 [-10.25,+4.54]

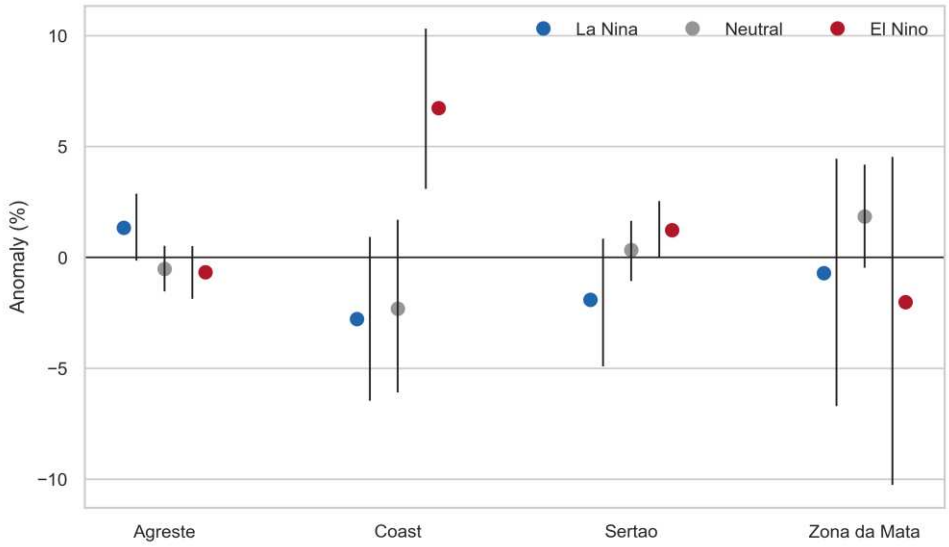


Fig. 10 *ENSO-linked regional radiation anomalies based on NOAA ONI. Points show mean anomalies and vertical bars show bootstrap 95% intervals*

The 2030 scenarios in Table 7 and Fig. 11 should be read as planning-oriented stress tests rather than forecasts. They do not impose climate-model forcing, electricity-demand growth, technology learning or land-use constraints. The Sertao scenario remains close to the historical mean, suggesting persistence of a high-resource interior regime under the simple extrapolation. The Agreste shows a slight downward tendency, while the coast has a wider uncertainty interval because it is represented by a single station and exhibits strong interannual variability. The Zona da Mata scenario was treated conservatively because late-record station artefacts make unconstrained extrapolation unstable.

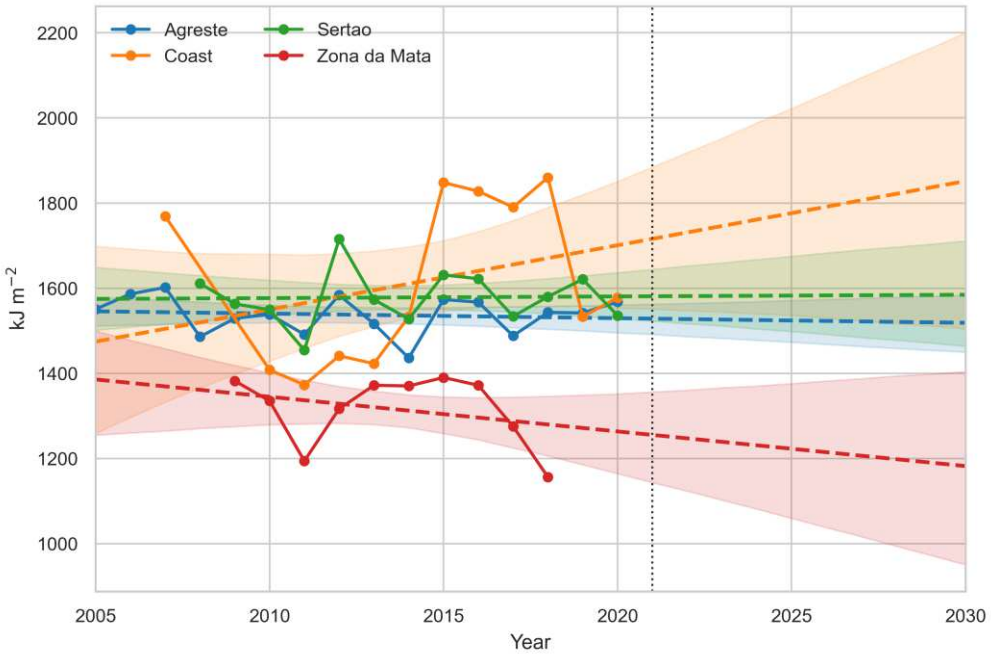


Fig. 11 *Exploratory regional radiation scenarios to 2030. Dashed lines are scenario extensions, shaded areas are uncertainty intervals, and the vertical dotted line marks the end of observations*

Table 7 *Exploratory regional radiation scenarios to 2030 in kJ m^{-2}. P05 and P95 denote bootstrap uncertainty bounds or historical variability bounds for conservative persistence cases*

Region	Year	Scenario mean	P05	P95
Agreste	2025	1524.1	1473.1	1571.2
Agreste	2030	1518.7	1449.8	1583.6
Coast	2025	1776.2	1536.2	2021.8
Coast	2030	1851.7	1504.4	2200.2
Sertao	2025	1582.6	1498.4	1674.8
Sertao	2030	1584.6	1463.6	1710.8
Zona da Mata	2025	1222.8	1059.4	1375.9
Zona da Mata	2030	1182.1	950.6	1404.1

5. Discussion

5.1. Atmospheric and renewable-resource relevance

The main advance of this study is to separate three dimensions that are often blended in conventional renewable-resource assessment: magnitude, temporal organization and climate-phase sensitivity. Solar atlases and GIS-based studies are fundamental for identifying high-resource areas (Pereira et al. 2017). Forecasting reviews and renewable-energy integration studies show that operational value depends on short-term variability and predictability (Inman et al. 2013; Pedro and Coimbra 2012; Voyant et al. 2017). This study connects these two traditions by showing that a station can be attractive in mean radiation but still close to a stochastic ordinal regime. Recife is a clear example: it has high mean radiation but the lowest Φ in the network. The predictive benchmark strengthens this interpretation because stations with larger Φ tend to produce lower normalized forecast errors under the same learning architecture, while also outperforming persistence and climatology baselines.

These results matter for Brazil because photovoltaic generation is no longer a marginal technology. BEN 2025 shows that solar and wind already form a substantial share of national electricity generation, while PDE 2034 expects the system to preserve a highly renewable structure and continue adding solar, wind and distributed generation (Empresa de Pesquisa Energética 2025a, 2025b). This national challenge is part of a broader global transition: current policy and market trajectories expand renewable capacity rapidly, but they still leave an implementation gap relative to the 2030 tripling benchmark (Fig. 12). This creates a new planning problem: high renewable participation is desirable, but it increases the value of forecasting, storage, grid flexibility and climate-risk screening. Entropy–Fisher diagnostics offer a low-cost station-based layer for renewable-energy planning, especially in regions where SCADA power data, dense pyranometer networks or high-resolution sky-imaging systems are not yet broadly available.

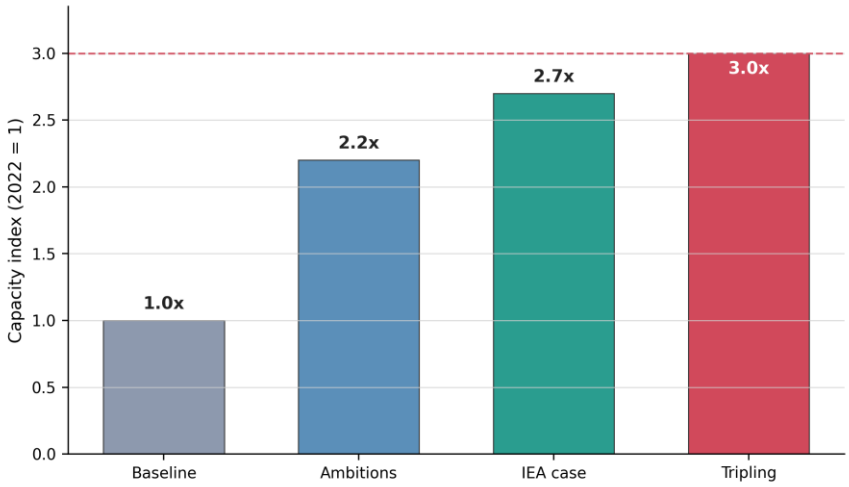


Fig. 12 *Global renewable-power capacity pathways to 2030, indexed to 2022. Bars summarize government ambitions, the IEA main case and the tripling pledge; the dashed line marks the tripling benchmark*

5.2. Pernambuco as a meteorological transition region

Pernambuco is valuable because it contains several Northeast climate regimes within one state. The coast and Zona da Mata are more exposed to maritime humidity and cloud modulation; the Agreste is a transition zone; and the Sertao is more strongly controlled by semi-arid clear-sky persistence. The observed pattern is physically coherent: humid sectors show stronger intermittency and lower ordinal organization, while interior stations show more recurrent daytime ordering. This supports the interpretation that the Sertao is not only a high-resource solar corridor but also a region where the hourly radiation sequence may be more operationally manageable.

For Pernambuco, the practical implication is direct. The state is positioned between coastal load centres, interior renewable corridors and the broader Sao Francisco Valley development axis. Petrolina and neighbouring semi-arid municipalities are relevant for solar expansion, irrigation economies and possible hydrogen-linked demand, while the Recife metropolitan area concentrates distribution-grid and rooftop-PV challenges. The present analysis adds that semi-arid stations such as Petrolina, Ibimirim, Cabrobo and Serra Talhada have strong ordinal organization. These locations deserve deeper PV-production, grid-access and storage studies because they combine resource availability with temporal regularity. Conversely, the Zona da Mata and coastal sectors may still support distributed PV, but system design should emphasize uncertainty management, nowcasting and voltage-control flexibility.

5.3. Seasonality, ENSO and radiation predictability

The seasonal analysis shows that annual means can hide when the resource is operationally strongest. Spring dominates the Sertao and Agreste, while winter weakens all sectors. This matters because storage sizing and grid operation are seasonal problems: a high annual mean is less useful if the weak season coincides with demand peaks, hydrological stress or grid constraints. The correlation analysis further indicates that high-radiation seasons exert strong influence on annual values, so monitoring spring and summer regularity is important for annual-yield confidence.

The ENSO analysis strengthens the climate interpretation, but also shows why uncertainty needs to be explicit. Northeast Brazil droughts and rainfall anomalies are known to respond to tropical Pacific and Atlantic variability (Hastenrath and Heller 1977; Andreoli and Kayano 2005; Marengo et al. 2017). Solar radiation responds indirectly through cloudiness, humidity and rainfall controls. The station results suggest that El Niño does not produce a single state-wide radiation response; the coastal positive signal is the most robust, whereas several interior and forest-zone intervals overlap zero. This spatial heterogeneity is important for portfolio planning. A geographically diversified PV portfolio inside Pernambuco may not respond uniformly to ENSO, reducing or redistributing climate risk, but operational use of this signal should be combined with Atlantic-gradient, cloud and rainfall diagnostics.

5.4. Use and limits of the 2030 scenarios

The 2030 scenarios are intentionally conservative. They are not climate-model projections and should not be interpreted as guarantees of future radiation, bankable yield estimates or policy-capacity forecasts. Their value is screening: they identify whether historical station behaviour suggests persistence, weak trend or large uncertainty under simple assumptions. For planning-oriented renewable-energy studies, a carefully framed future component helps clarify the range of plausible operating conditions without overstating predictability. Here, the safest interpretation is that the Sertao remains robust under a simple scenario extension, while single-station humid sectors require better monitoring before strong future claims are made.

5.5. Limitations

The analysis uses station radiation records rather than satellite-derived irradiance fields, homogenized pyranometer networks or PV power measurements. The spatial interpolation is exploratory and constrained by twelve stations. The forecastability index is an ordinal screening measure, not a replacement for operational irradiance forecasting, even though the ExtraTrees validation shows that it is related to out-of-sample errors. ENSO classification uses

monthly ONI and does not separate Atlantic sea-surface temperature gradients, Madden–Julian variability or local cloud microphysics. Future work should compare these diagnostics with satellite products, clear-sky indices, probabilistic ramp forecasts, PV power output, storage-dispatch simulations, hybrid renewable configurations and transmission constraints.

6. Conclusions

This study advances solar-resource assessment in Pernambuco by combining hourly radiation climatology, ordinal information theory, climate-phase stratification, common-period sensitivity, predictive validation and exploratory 2030 scenarios in a single planning-oriented framework. The novelty is the joint interpretation of resource magnitude, temporal organization, seasonality and climate sensitivity, which allows photovoltaic planning to distinguish high-radiation sites from sites with stronger short-term forecastability. The main conclusions are:

1. The first-pass database contains 692,560 valid hourly records, and the robust analysis-ready database contains 630,298 records from twelve Pernambuco stations.
2. Mean radiation and ordinal organization produce different rankings, showing that magnitude alone is insufficient for solar-resource screening.
3. Cabrobo has the highest mean hourly irradiation, while Palmares has the lowest mean and largest coefficient of variation.
4. Petrolina, Ibimirim, Cabrobo and Serra Talhada show the largest relative ordinal forecastability, indicating strong recurrent hourly organization in the semi-arid interior.
5. Seasonal analysis shows a strong spring maximum in the Sertao and Agreste and a weak winter signal across all sectors.
6. Common-period sensitivity and ExtraTrees validation support the interpretation of Φ as a useful forecastability-screening index.
7. ENSO stratification reveals spatially heterogeneous effects, with the clearest bootstrap-supported signal being positive El Niño anomalies on the coast.
8. Conservative 2030 scenarios suggest persistence of a robust semi-arid resource but large uncertainty in single-station humid sectors.

The study's central message is that Brazil, the Northeast and Pernambuco need renewable-energy planning that integrates solar-resource magnitude, temporal organization, seasonality and climate variability. Such information is

essential for photovoltaic forecasting, storage design, smart-grid operation and green-hydrogen planning in a renewable-rich electricity system.

Statements and Declarations

Author contributions

Fábio Sandro dos Santos: Conceptualization, Methodology, Software, Formal analysis, Visualization, Writing – original draft, Writing – review and editing. Fernando Henrique Antunes de Araújo: Methodology, Validation, Writing – review and editing. Kerolly Kedma Felix do Nascimento: Data curation, Investigation, Validation, Writing – review and editing.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Competing interests

The authors have no relevant financial or non-financial interests to disclose.

Data availability

The processed tables, figures and quality-controlled hourly radiation file are provided with the submission package. The original automatic-station data are available from INMET/BDMEP and were supplied in the project folder used for this study. The ONI values used for ENSO stratification are based on the NOAA Climate Prediction Center Oceanic Niño Index.

Acknowledgements

The authors acknowledge the public meteorological data infrastructure that enables regional renewable-energy assessment in Brazil.

References

Agência Nacional de Energia Elétrica. 2026. *Sistema de Informações de Geração Da ANEEL (SIGA)*. https://www.gov.br/aneel/pt-br/centrais-de-conteudos/relatorios-e-indicadores/geracao.

Andreoli, Rita V., and Mary Toshie Kayano. 2005. “ENSO-Related Rainfall Anomalies in South America and Associated Circulation Features During Warm and Cold Pacific Decadal Oscillation Regimes.” *International Journal of Climatology* 25 (15): 2017–30. https://doi.org/10.1002/joc.1222.

Bandt, Christoph, and Bernd Pompe. 2002. “Permutation Entropy: A Natural Complexity Measure for Time Series.” *Physical Review Letters* 88 (17): 174102. https://doi.org/10.1103/PhysRevLett.88.174102.

Efron, Bradley, and Robert J. Tibshirani. 1993. *An Introduction to the Bootstrap*. Chapman; Hall/CRC.

Empresa de Pesquisa Energética. 2025a. *Balanço Energético Nacional 2025: Ano Base 2024*. https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/balanco-energetico-nacional-2025.

Empresa de Pesquisa Energética. 2025b. *Plano Decenal de Expansão de Energia 2034*. https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/plano-decenal-de-expansao-de-energia-2034.

Geurts, Pierre, Damien Ernst, and Louis Wehenkel. 2006. “Extremely Randomized Trees.” *Machine Learning* 63 (1): 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Gueymard, Christian A. 2012. “Temporal Variability in Direct and Global Irradiance at Various Time Scales as Affected by Aerosols.” *Solar Energy* 86 (12): 3544–53. https://doi.org/10.1016/j.solener.2012.01.013.

Hastenrath, Stefan, and Linda Heller. 1977. “Dynamics of Climatic Hazards in Northeast Brazil.” *Quarterly Journal of the Royal Meteorological Society* 103 (435): 77–92. https://doi.org/10.1002/qj.49710343505.

Inman, Richard H., Hugo T. C. Pedro, and Carlos F. M. Coimbra. 2013. “Solar Forecasting Methods for Renewable Energy Integration.” *Progress in Energy and Combustion Science* 39 (6): 535–76. https://doi.org/10.1016/j.pecs.2013.06.002.

Instituto Nacional de Meteorologia. 2026. *Banco de Dados Meteorológicos Para Ensino e Pesquisa*. https://bdmep.inmet.gov.br/.

International Energy Agency. 2024. *Renewables 2024: Analysis and Forecast to 2030*. https://www.iea.org/reports/renewables-2024.

Kleissl, Jan, ed. 2013. *Solar Energy Forecasting and Resource Assessment*. Academic Press. https://doi.org/10.1016/C2011-0-07022-9.

Lave, Matthew, and Jan Kleissl. 2013. “Cloud Speed Impact on Solar Variability Scaling–Application to the Wavelet Variability Model.” *Solar Energy* 91: 11–21. https://doi.org/10.1016/j.solener.2013.01.023.

Marengo, Jose A., Roger Rodrigues Torres, and Lincoln Muniz Alves. 2017. “Drought in Northeast Brazil–Past, Present, and Future.” *Theoretical and Applied Climatology* 129: 1189–200. https://doi.org/10.1007/s00704-016-1840-8.

Martin, Maria T., Angel Plastino, and Osvaldo A. Rosso. 2006. “Generalized Statistical Complexity Measures: Geometrical and Analytical Properties.” *Physica A: Statistical Mechanics and Its Applications* 369 (2): 439–62. https://doi.org/10.1016/j.physa.2005.11.053.

NOAA Climate Prediction Center. 2026. *Oceanic Niño Index (ONI)*. https://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ONI_v5.php.

Pedro, Hugo T. C., and Carlos F. M. Coimbra. 2012. “Assessment of Forecasting Techniques for Solar Power Production with No Exogenous Inputs.” *Solar Energy* 86 (7): 2017–28. https://doi.org/10.1016/j.solener.2012.04.004.

Pereira, Enio Bueno, Fernando Ramos Martins, André Rodrigues Gonçalves, et al. 2017. *Atlas Brasileiro de Energia Solar*. 2nd ed. Instituto Nacional de Pesquisas Espaciais.

Riedl, Maik, Andreas Müller, and Niels Wessel. 2013. “Practical Considerations of Permutation Entropy.” *The European Physical Journal Special Topics* 222: 249–62. https://doi.org/10.1140/epjst/e2013-01862-7.

Rosso, Osvaldo A., Hilda A. Larrondo, Maria T. Martin, Angel Plastino, and Miguel A. Fuentes. 2007. “Distinguishing Noise from Chaos.” *Physical Review Letters* 99 (15): 154102. https://doi.org/10.1103/PhysRevLett.99.154102.

Shepard, Donald. 1968. “A Two-Dimensional Interpolation Function for Irregularly-Spaced Data.” *Proceedings of the 1968 ACM National Conference*, 517–24. https://doi.org/10.1145/800186.810616.

Voyant, Cyril, Gilles Notton, Soteris Kalogirou, et al. 2017. “Machine Learning Methods for Solar Radiation Forecasting: A Review.” *Renewable Energy* 105: 569–82. https://doi.org/10.1016/j.renene.2016.12.095.

Zunino, Luciano, Miguel C. Soriano, Ingo Fischer, Osvaldo A. Rosso, and Claudio R. Mirasso. 2010. “Permutation-Information-Theory Approach to Unveil Delay Dynamics from Time-Series Analysis.” *Physical Review E* 82 (4): 046212. https://doi.org/10.1103/PhysRevE.82.046212.