



(RESEARCH ARTICLE)



MULTIVARIATE ANALYSIS OF METEOROLOGICAL REGIMES AND SOLAR IRRADIANCE VARIABILITY IN DOSSO, SAHELIAN NIGER: IMPLICATIONS FOR PHOTOVOLTAIC APPLICATIONS

Mahamadou Hamidine ^{1,*}, Aminou Baare Ali ^{2,3}, Aboubakar Amadou Yansambou Mohamed ^{2,3} and Makinta Boukar ^{2,3}

¹ Department of Physics, Faculty of Science and Technology, Dan Dicko Dankoulodo University of Maradi, BP 465, Maradi, Niger.

² Department of Physics, Faculty of Science and Technology, Abdou Moumouni University, Niamey, Niger.

³ Energy, Electronics, Electrical Engineering, Automation and Industrial Computing Laboratory (L3EA2I), Abdou Moumouni University, Niamey, Niger.

* Corresponding author; Emails: hamidine2000@gmail.com

ORCID Details

Mahamadou Hamidine: <https://orcid.org/0009-0000-3945-5019>

GSC Advanced Research and Reviews, 2026, 28(03), 182–189

Publication history: Received on 07 August 2026; revised on 16 September 2026; accepted on 18 September 2026

Article DOI: <https://doi.org/10.30574/gscarr.2026.28.3.0238>

Copyright © 2026 Author(s) retain the copyright of this article. This article is published under the terms of the Creative Commons Attribution License 4.0

ABSTRACT

Solar irradiance in the Sahel covaries with several meteorological variables, making pooled bivariate correlations vulnerable to seasonal confounding. This study characterizes the joint meteorological structure associated with daytime global solar irradiance in Dosso, Niger, using field observations from May and August 2021. Sixty-two daily records were derived from hourly measurements of irradiance, ambient temperature, relative humidity and wind speed. After quality control, Pearson correlations were evaluated for the pooled dataset and separately by month; standardized variables were then analyzed using principal component analysis (PCA) and k-means clustering. Mean irradiance decreased from 822.75 W m⁻² in May to 651.52 W m⁻² in August, while mean temperature decreased from 38.23 to 29.85°C and relative humidity increased from 35.96 to 74.15%. Across all 62 days, irradiance correlated positively with temperature ($r = 0.658$) and negatively with relative humidity ($r = -0.611$), but the within-month relationships differed markedly. The first two principal components explained 91.28% of total standardized variance. PC1 (64.24%) represented a hot-dry-high-irradiance versus cool-humid-low-irradiance gradient, whereas PC2 (27.04%) was dominated by wind speed. A two-cluster k-means solution achieved a silhouette coefficient of 0.469 and separated all May observations from all August observations. The separation describes two sampled seasonal regimes rather than a universal climatic classification. The results demonstrate that season-aware multivariate analysis provides a more defensible interpretation of short solar-resource records than pooled single-variable correlations and can guide the selection of representative meteorological scenarios for photovoltaic studies in data-sparse Sahelian regions.

Keywords: Solar Resource, Principal Component Analysis, Seasonal Variability, Relative Humidity, K-Means Clustering, Sahel

1 INTRODUCTION

Solar-energy planning in arid and semi-arid regions requires more than an estimate of mean irradiance. It also requires an understanding of the atmospheric conditions that control temporal variability and photovoltaic operating temperature. In the Sahel, clouds, water vapour, aerosols, temperature, convection and seasonal circulation evolve jointly. Ground observations remain sparse, and satellite or reanalysis products can differ appreciably from local measurements, especially where cloud and aerosol conditions change rapidly [1–7]. These limitations increase the value of extracting physically interpretable information from short field campaigns while retaining caution about measurement uncertainty and temporal representativeness.

A methodological difficulty arises when observations from contrasting seasons are pooled. Between-season differences can dominate covariance and produce strong correlations even when within-season associations are weak. Such reversals or changes in association after aggregation are well documented in statistical analysis [8]. The pooled coefficients remain descriptive of the combined sample, but they should not be interpreted as invariant or causal effects. This issue is particularly important for relative humidity, which is thermodynamically coupled to temperature and often changes simultaneously with clouds, precipitation and atmospheric water content.

Multivariate methods provide a suitable framework for this problem. Principal component analysis (PCA) transforms correlated variables into orthogonal components that retain decreasing proportions of variance and has been used to summarize climatic gradients and reduce redundancy in solar-radiation datasets [9,10]. K-means clustering can then identify groups with similar multivariate characteristics without imposing seasonal labels, while the silhouette coefficient provides a compact measure of within-cluster cohesion and between-cluster separation [11,12]. Analyses of photovoltaic performance in arid climates similarly emphasize that meteorological drivers should be interpreted jointly rather than in isolation [13].

An earlier analysis of the Dosso campaign reported an inverse association between relative humidity and solar radiation [14]. The present study extends that work by adopting a multivariate and season-aware framework. Its objectives are to quantify the May–August contrasts in irradiance, temperature, relative humidity and wind speed; compare pooled and within-month correlations; identify the dominant covariance axes using standardized PCA; and test whether unsupervised clustering recovers distinct groups in the four-variable space. The analysis is exploratory and characterizes only the two observed periods; it is not intended to establish a complete climatology of Dosso.

2 MATERIAL AND METHODS

2.1 Study area and field campaign

The measurement site is located in Dosso, south-western Niger, at 13°02'30" N and 3°11'15" E, at a mean altitude of approximately 217 m. The site is flat and free of nearby obstacles that could cast shade over the measurement area. Dosso lies within the West African Sahel, where a long dry season and a shorter monsoon season produce strong annual contrasts in temperature, moisture, cloudiness and aerosol conditions. Field observations were collected during May and August 2021, hourly between 07:00 and 18:00 local time. May represents a hot and relatively dry pre-rainy-season period, whereas August lies within the core rainy season.

Ambient temperature and relative humidity were measured using a SMART SENSOR AS847 thermo-hygrometer. Wind speed was measured with a UT363 BT mini-anemometer positioned approximately 2 m above ground. Global irradiance was measured with an SM206-SOLAR instrument; the readings analyzed here represent irradiance on a horizontal plane measured near ground level. The two observation periods yielded 62 daily records, comprising 31 days in May and 31 days in August.

2.2 Data preparation and quality control

For each day, valid hourly observations were averaged to obtain daytime daily means of global irradiance (G), ambient temperature (T), relative humidity (RH) and wind speed (W). Thus, G is a mean irradiance in $W\ m^{-2}$ over the 07:00–18:00 observation window; it is not daily solar irradiation in $kWh\ m^{-2}\ day^{-1}$. For G , T and RH , zero values that were physically incompatible with the daytime window or adjacent observations were treated as missing. Zero wind-speed observations were retained because they can represent calm conditions. Clearly identifiable decimal-entry errors were converted to numeric values only when the intended value was unambiguous. No temporal interpolation was applied, thereby avoiding the introduction of artificial covariance. Table 1 summarizes the variables, units and daily treatments used in the analysis.

Table 1: Variables included in the multivariate analysis

Variable	Symbol	Unit	Daily treatment
Global horizontal irradiance	G	W m^{-2}	Mean of valid daytime hourly observations
Ambient temperature	T	$^{\circ}\text{C}$	Mean of valid daytime hourly observations
Relative humidity	RH	%	Mean of valid daytime hourly observations
Wind speed	W	m s^{-1}	Hourly mean; zero values retained

Traceable calibration documentation for the SM206-SOLAR instrument was not available in the campaign files. Absolute irradiance magnitudes are therefore treated as field observations rather than reference-standard radiometry, consistent with the importance assigned to calibration and traceability in meteorological measurement guidance [15]. The principal focus is the internal covariance and seasonal structure, although measurement uncertainty may still affect the results.

2.3 Statistical analysis

Monthly means were calculated for all four variables. Pearson correlation coefficients and two-sided significance probabilities were then computed for the complete 62-day dataset and separately for each month. The monthly decomposition was used as a diagnostic for seasonal confounding: a strong pooled correlation can arise from differences in monthly means even when within-month relationships are weak. Because several pairwise tests were examined, the reported p-values are unadjusted and are interpreted as descriptive evidence rather than confirmatory hypothesis tests, following established cautions concerning multiple simultaneous tests [16].

PCA was applied to the four standardized daily variables. Each observation was centered by the sample mean and divided by the corresponding sample standard deviation so that variables expressed in different units contributed comparably. Components were obtained from the eigendecomposition of the correlation matrix. The first two components were retained because their cumulative explained variance exceeded 90%. Component signs are arbitrary; interpretation was based on the relative magnitude and direction of loadings.

K-means clustering was applied to the same standardized four-variable dataset. Candidate solutions with $k = 2, 3, 4$ and 5 clusters were compared using the average silhouette coefficient. The algorithm did not receive month labels; those labels were used only after clustering to interpret group composition. The $k = 2$ solution provided the highest silhouette coefficient among the tested solutions. Because k-means is sensitive to initialization and cluster geometry, its result was interpreted descriptively rather than as a definitive climatic taxonomy [17,18].

All analyses were rerun using Python 3.12.14. Data processing and numerical calculations used pandas 2.2.3, NumPy 2.3.5 and SciPy 1.17.0. PCA, k-means clustering and silhouette analysis were performed using scikit-learn 1.8.0. Figures were produced using Matplotlib 3.10.8 and Seaborn 0.13.2.

3 RESULTS AND DISCUSSION

3.1 Seasonal contrast in the observed variables

Table 2: Mean meteorological conditions during the two sampled periods

Variable	May	August	August relative to May
Global irradiance, G (W m^{-2})	822.75	651.52	-20.81%
Ambient temperature, T ($^{\circ}\text{C}$)	38.23	29.85	-8.38 $^{\circ}\text{C}$
Relative humidity, RH (%)	35.96	74.15	+38.20 percentage points
Wind speed, W (m s^{-1})	1.107	1.084	-2.05%

The two periods differed strongly in irradiance, temperature and humidity, whereas mean wind speed changed little (Table 2). Mean irradiance declined by 20.81% from May to August, temperature decreased by 8.38 $^{\circ}\text{C}$, and relative humidity increased by 38.20 percentage points. Mean wind speed decreased by only 2.05%. The sampled seasonal contrast was therefore primarily radiative and thermo-hygrometric rather than a shift in average ventilation.

3.2 Pooled and within-month correlation structure

For the pooled dataset, irradiance was positively correlated with temperature ($r = 0.658$, $p < 0.001$), negatively correlated with relative humidity ($r = -0.611$, $p < 0.001$), and moderately correlated with wind speed ($r = 0.394$, $p = 0.002$). Temperature and relative humidity showed an exceptionally strong pooled inverse correlation ($r = -0.985$, $p < 0.001$). Figure 1 displays all six pooled pairwise correlations, whereas Table 3 compares the three irradiance relationships and the temperature-humidity diagnostic across the pooled and month-specific datasets.

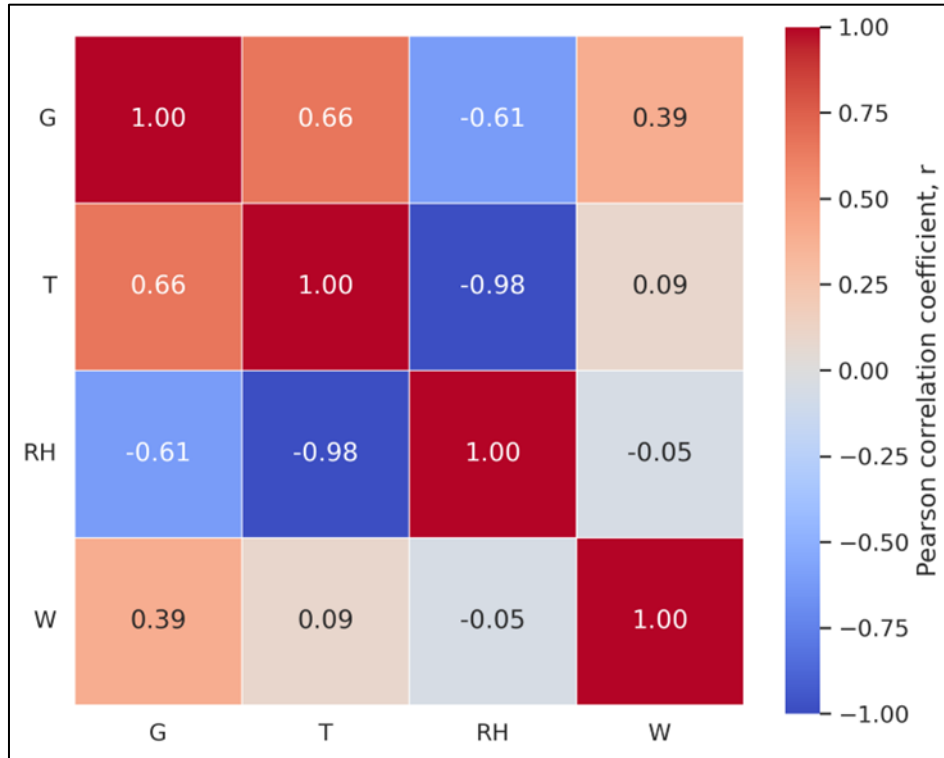


Figure 1: Pearson correlation matrix for the pooled set of 62 daily observations. G: global irradiance; T: ambient temperature; RH: relative humidity; W: wind speed.

The within-month analysis demonstrates that the pooled relationships were not stationary (Table 3). In May, G-T and G-RH correlations were weak and non-significant at the 5% level ($p = 0.273$ and $p = 0.222$, respectively). In August, the corresponding correlations strengthened to 0.809 and -0.602 (both $p < 0.001$). The G-W correlation also increased from 0.284 in May ($p = 0.122$) to 0.532 in August ($p = 0.002$). The pooled coefficients therefore combine within-period covariation with the large difference between the two monthly means.

Table 3: Selected pairwise correlations for pooled and month-specific datasets

Variable pair	All 62 days	May (n = 31)	August (n = 31)
G-T	+0.658***	+0.203	+0.809***
G-RH	-0.611***	-0.226	-0.602***
G-W	+0.394**	+0.284	+0.532**
T-RH	-0.985***	-0.904***	-0.839***

** $p < 0.01$; *** $p < 0.001$ (two-sided tests).

3.3 Principal component structure

PC1 explained 64.24% of total standardized variance and PC2 explained 27.04%; together they accounted for 91.28% (Table 4). PC1 had loadings of approximately +0.522 for G, +0.595 for T, -0.582 for RH and +0.183 for W. Positive PC1 scores therefore represented hotter, drier and more strongly irradiated conditions, whereas negative scores represented cooler, more humid and less irradiated conditions. PC2 was dominated by wind speed, whose loading was approximately +0.889. Wind thus contributed principally to a secondary source of variability rather than the main May-August contrast.

Table 4: Explained variance and interpretation of the retained principal components

Component	Explained variance	Dominant interpretation
PC1	64.24%	Hot-dry-high-irradiance to cool-humid-low-irradiance gradient
PC2	27.04%	Secondary variability dominated by wind speed
PC1 + PC2	91.28%	Two axes capture most standardized variance

3.4 K-means classification of the sampled regimes

Among $k = 2$ to 5, the two-cluster solution produced the highest average silhouette coefficient, equal to 0.469. The clusters corresponded exactly to the two sampled periods: all 31 May observations were assigned to one cluster and all 31 August observations to the other (Figure 2). Since month labels were excluded from the features, the result confirms that the joint structure of G , T , RH and W was sufficient to separate the two periods in this dataset.

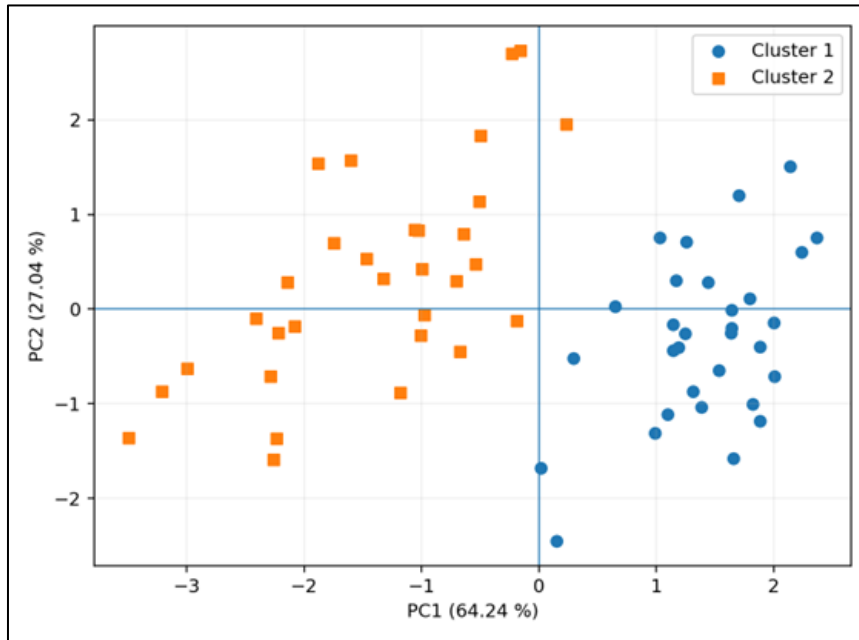


Figure 2: Two-cluster k-means classification projected onto the PC1-PC2 plane. Cluster membership reproduces the May and August groups, respectively; cluster numbers are arbitrary labels.

The silhouette coefficient indicates moderate rather than exceptionally strong separation. Moreover, exact month-cluster correspondence does not establish a universal two-class climate system. The sampling design deliberately includes two contrasting periods, so the result quantifies the multivariate distinction between those periods rather than the full diversity of annual weather states in Dosso.

3.5 Seasonal confounding in pooled correlations

The comparison in Table 3 shows that a pooled coefficient cannot be used as an invariant meteorological response function. Each pooled value reflects both covariation within a month and displacement between the May and August monthly means. The weak May G - T and G - RH coefficients, contrasted with the stronger August values, therefore change the interpretation of the pooled relationships: they describe the sampled seasonal gradient rather than an isolated effect of temperature or humidity on irradiance.

The pattern is physically plausible because the variables are coupled. Relative humidity depends on both moisture content and temperature; cloud formation is linked to humidity and convection; aerosols and dust vary by season; and near-surface temperature responds to the radiative environment [5–7]. The very strong inverse T - RH correlation is therefore best viewed as an indicator of the hot-dry versus cool-humid contrast. Atmospheric variables omitted from the campaign, particularly cloud fraction, precipitation, aerosol optical depth and precipitable water, may account for part of the remaining structure.

3.6 Multivariate description and implications for photovoltaic studies

PCA and clustering provide complementary, rather than duplicate, evidence. PCA identifies the joint gradient that organizes the observations, while k-means tests whether the standardized four-variable space contains coherent groups without receiving month labels. Their agreement supports a marked contrast within the sampled record, but it does not by itself demonstrate that the same two-group structure will recur in other months or years [9,10,17,18].

For photovoltaic modelling, May and August impose distinct but not uniformly favourable boundary conditions. The May regime supplies greater irradiance but also higher ambient temperature, which can increase module temperature and offset part of the irradiance benefit through temperature-related efficiency losses [19,20]. The August regime combines lower irradiance with higher humidity and more variable atmospheric attenuation; humidity, air motion and dust can influence PV operating conditions and optical losses [21,22]. Representative-day selection and model validation should therefore preserve the observed joint meteorological states instead of relying on one monthly irradiance value or one pooled correlation [13]. The campaign did not measure electrical PV output; these implications concern scenario selection and cannot be interpreted as quantified PV-energy losses or gains.

This point is particularly relevant in West Africa, where surface radiation stations are sparse and satellite products can show location- and season-dependent departures from ground observations [1,4,6,7]. Regime-based analysis can support the selection of physically contrasting scenarios for PV yield, thermal behaviour and storage simulations. With longer records, regime-conditioned performance models could also determine whether prediction errors differ systematically between dry, humid, cloudy or dusty states.

3.7 Study limitations and perspectives

The interpretation is limited by the 62-day record, which covers only May and August 2021, and by daily averaging, which masks sub-daily variability. The absence of cloud, precipitation and aerosol measurements restricts physical attribution, while unavailable calibration records and PV-output data preclude reference-grade irradiance assessment or quantified PV-performance conclusions. Future studies should use multi-year, traceably calibrated measurements, include atmospheric indicators and electrical PV output, and evaluate the stability of the multivariate regimes through out-of-sample validation.

4 CONCLUSION

The 62 daily observations reveal a clear multivariate contrast between the sampled May and August 2021 conditions in Dosso. The divergence between pooled and within-month correlations shows that season must be accounted for before meteorological associations are used in solar-resource interpretation. PCA retained 91.28% of standardized variance in two axes, and k-means reproduced the two sampled periods with a moderate silhouette coefficient of 0.469. Together, these results support the use of joint meteorological regimes to select contrasting inputs for PV simulation in data-sparse Sahelian settings. They do not establish causal effects, annual regime frequencies or electrical PV performance. Generalization to annual resource assessment requires longer records, traceable irradiance calibration and direct PV-output measurements.

STATEMENTS ON COMPLIANCE WITH ETHICAL STANDARDS

Acknowledgments

The authors acknowledge the use of Python and its open-source scientific libraries for data processing, statistical analysis and figure preparation.

Disclosure of conflict of interest

The authors declare that they have no conflict of interest.

Author contributions

Mahamadou Hamidine: conceptualization, interpretation of results, and writing - original draft. **Aminou Baare Ali:** investigation, field measurements, and data collection. **Aboubakar Amadou Yansambou Mohamed:** methodology and writing - review and editing. **Makinta Boukar:** supervision and writing - review and editing. All authors have read and approved the final version of the manuscript.

Data availability

The field data supporting this study may be obtained from the corresponding author subject to confirmation of the applicable access conditions.

REFERENCES

- [1] Marchand M, Ghennioui A, Wey E, Wald L. Comparison of several satellite-derived databases of surface solar radiation against ground measurement in Morocco. *Advances in Science and Research*. 2018;15:21-29. doi:10.5194/asr-15-21-2018
- [2] Nwokolo SC. A comprehensive review of empirical models for estimating global solar radiation in Africa. *Renewable and Sustainable Energy Reviews*. 2017;78:955-995. doi:10.1016/j.rser.2017.04.101
- [3] Salazar G, Gueymard CA, Galdino JB, de Castro Vilela O, Fraidenraich N. Solar irradiance time series derived from high-quality measurements, satellite-based models, and reanalyses at a near-equatorial site in Brazil. *Renewable and Sustainable Energy Reviews*. 2020;117:109478. doi:10.1016/j.rser.2019.109478
- [4] Sawadogo W, Bliefernicht J, Fersch B, Salack S, Guug S, Diallo B, Ogunjobi KO, Nakoulma G, Tanu M, Meilinger S, Kunstmann H. Hourly global horizontal irradiance over West Africa: A case study of one-year satellite- and reanalysis-derived estimates versus in situ measurements. *Renewable Energy*. 2023;216:119066. doi:10.1016/j.renene.2023.119066
- [5] Wild M. Global dimming and brightening: A review. *Journal of Geophysical Research: Atmospheres*. 2009;114:D00D16. doi:10.1029/2008JD011470
- [6] Lefèvre M, Oumbe A, Blanc P, Espinar B, Gschwind B, Qu Z, Wald L, Schroedter-Homscheidt M, Hoyer-Klick C, Arola A, Benedetti A, Kaiser JW, Morcrette JJ. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmospheric Measurement Techniques*. 2013;6:2403-2418. doi:10.5194/amt-6-2403-2013
- [7] Sengupta M, Habte A, Gueymard C, Wilbert S, Renné D, Stoffel T. *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications*. 2nd ed. Golden, CO: National Renewable Energy Laboratory; 2017. doi:10.2172/1411856
- [8] Tu YK, Gunnell D, Gilthorpe MS. Simpson's paradox, Lord's paradox, and suppression effects are the same phenomenon - the reversal paradox. *Emerging Themes in Epidemiology*. 2008;5:2. doi:10.1186/1742-7622-5-2
- [9] Jolliffe IT, Cadima J. Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A*. 2016;374(2065):20150202. doi:10.1098/rsta.2015.0202
- [10] Zarzo M, Martí P. Modeling the variability of solar radiation data among weather stations by means of principal components analysis. *Applied Energy*. 2011;88(8):2775-2784. doi:10.1016/j.apenergy.2011.01.070
- [11] MacQueen JB. Some methods for classification and analysis of multivariate observations. In: *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*. Vol. 1. Berkeley: University of California Press; 1967. p. 281-297.
- [12] Rousseeuw PJ. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*. 1987;20:53-65. doi:10.1016/0377-0427(87)90125-7
- [13] Al Shabaan GN, Altarawneh IS. Decoding climatic drivers of solar photovoltaic output in arid climates: Integrating sensitivity, lag dynamics, and principal component analysis for forecasting accuracy. *Sustainable Energy Technologies and Assessments*. 2026;86:104851. doi:10.1016/j.seta.2026.104851
- [14] Baare Ali A, Hamidine M, Sani Dan Noma H, Boukar M, Madougou S. Air humidity influence on global solar radiation on the surface of a photovoltaic module in Niger: Case of the city of Dosso. *International Journal of Innovation and Applied Studies*. 2023;40(1):167-178.
- [15] World Meteorological Organization. *Guide to Instruments and Methods of Observation*. WMO-No. 8. Geneva: World Meteorological Organization; 2018.
- [16] Wilks DS. The stippling shows statistically significant grid points: How research results are routinely overstated and overinterpreted. *Bulletin of the American Meteorological Society*. 2016;97(12):2263-2273. doi:10.1175/BAMS-D-15-00267.1
- [17] Hartigan JA, Wong MA. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society Series C: Applied Statistics*. 1979;28(1):100-108. doi:10.2307/2346830

- [18] Jain AK. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*. 2010;31(8):651-666. doi:10.1016/j.patrec.2009.09.011
- [19] Skoplaki E, Palyvos JA. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Solar Energy*. 2009;83(5):614-624. doi:10.1016/j.solener.2008.10.008
- [20] Dubey S, Sarvaiya JN, Seshadri B. Temperature dependent photovoltaic efficiency and its effect on PV production in the world: A review. *Energy Procedia*. 2013;33:311-321. doi:10.1016/j.egypro.2013.05.072
- [21] Mekhilef S, Saidur R, Kamalisarvestani M. Effect of dust, humidity and air velocity on efficiency of photovoltaic cells. *Renewable and Sustainable Energy Reviews*. 2012;16(5):2920-2925. doi:10.1016/j.rser.2012.02.012
- [22] Mani M, Pillai R. Impact of dust on solar photovoltaic performance: Research status, challenges and recommendations. *Renewable and Sustainable Energy Reviews*. 2010;14(9):3124-3131. doi:10.1016/j.rser.2010.07.065