



Optimizing solar photovoltaic system performance: Insights and strategies for enhanced efficiency

Sidique Gawusu^{a,*}, Xiaobing Zhang^{a,**}, Sufyan Yakubu^{b,g,h}, Seth Kofi Debrah^{c,d}, Oisik Das^e, Nishant Singh Bundela^f

^a School of Energy and Power and Engineering, Nanjing University of Science and Technology, Nanjing, China

^b Department of Electrical and Electronic Engineering, Bolgatanga Technical University, Bolgatanga, Ghana

^c Department of Nuclear Engineering, School of Nuclear and Allied Sciences, Atomic Energy, University of Ghana, Legon, Accra, Ghana

^d Nuclear Power Institute, Ghana Atomic Energy Commission, Legon, Accra, Ghana

^e Structural and Fire Engineering Division, Department of Civil, Environmental and Natural Resources Engineering, Luleå University of Technology, Luleå, 97187, Sweden

^f Whiting School of Engineering, Johns Hopkins University, Baltimore, MD, 21211, United States

^g Department of Renewable Energy Engineering, School of Energy, University of Energy and Natural Resources (UENR), Sunyani, Ghana

^h Regional Center for Energy and Environmental Sustainability (RCEES), University of Energy and Natural Resources (UENR), Sunyani, Ghana

ARTICLE INFO

Handling editor: X Zhang

Keywords:

Monte Carlo simulations
Predictive modeling
Renewable energy optimization
XGBoost algorithm
Solar photovoltaic systems
Energy policy

ABSTRACT

This study analyzes the performance and predictive modeling of solar photovoltaic (PV) systems at the Bui Generating Station in Ghana using the XGBoost (Extreme Gradient Boosting) algorithm. The predictive model, validated through Monte Carlo simulations, demonstrates measured stability across perturbation scenarios. Distribution analysis confirms appropriate parameter bounds, while error analysis demonstrates consistent pattern preservation across simulation scenarios. The study quantifies the relative influences of environmental factors, particularly the interplay between temperature, irradiance, and humidity (correlations ranging from -0.33 to 0.36). These findings provide insights for system operation while acknowledging the complex, often weak coupling between environmental parameters. Seasonal performance analysis reveals distinct optimization windows, with the Post-Rainy season showing the highest stability (PR: 0.986 ± 0.082) and optimal enhancement potential. Sensitivity analysis identifies critical operational thresholds, including performance transitions at 80 % relative humidity and optimal temperature ranges below 32 °C, where each 1 °C reduction yields 0.45 % efficiency gain. The study establishes specific optimization strategies including automated cleaning systems triggered at 85 % peak irradiance, yielding 2.5 % efficiency improvement, and enhanced inverter response protocols during peak generation periods, achieving a 3.2 % performance gain. These findings inform practical implementation frameworks for performance optimization, contributing to improved energy generation efficiency and system reliability.

1. Introduction

The optimization of solar PV system performance represents a critical challenge in maximizing renewable energy's contribution to national power grids. Ghana's Bui Generating Station offers an exemplary case study for performance optimization strategies through its innovative hybrid configuration [1,2]. The facility uniquely integrates a hydroelectric power plant with both ground-mounted and floating solar PV systems, establishing itself as one of Africa's pioneering hybrid energy generation complexes [1,2]. Notable for hosting West Africa's first

floating solar PV plant [2,3], the Bui Station demonstrates Ghana's commitment to advancing renewable energy solutions. This distinctive combination of ground-mounted and floating solar PV systems presents valuable opportunities for developing comprehensive optimization approaches that can be applied across similar installations globally.

The increasing demand for sustainable and renewable energy sources has driven significant advancements in solar PV technology [4–6]. Solar PV systems convert sunlight directly into electricity, offering a clean and sustainable energy solution that reduces dependence on fossil fuels and mitigates greenhouse gas emissions [7–10]. In regions with high solar

* Corresponding author.

** Corresponding author.

E-mail addresses: gawusididique@gmail.com (S. Gawusu), zhangxb680504@163.com (X. Zhang).

<https://doi.org/10.1016/j.energy.2025.135099>

Received 17 September 2024; Received in revised form 3 February 2025; Accepted 16 February 2025

Available online 17 February 2025

0360-5442/© 2025 Elsevier Ltd. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

insolation, such as Ghana [11], solar PV technology holds tremendous potential to contribute substantially to the national energy mix.

Ghana's commitment to renewable energy is demonstrated through its Renewable Energy Master Plan, which targets a 10 % renewable energy contribution to the national electricity generation mix by 2030 [12,13]. The Bui Generating Station represents a critical step toward this goal, serving as a benchmark for large-scale solar PV implementation in West Africa.

The novelty of this research lies in several key aspects. First, the Bui Generating Station represents an unprecedented integration of hydroelectric power with both ground-mounted and floating solar PV systems in West Africa. This unique hybrid configuration provides an exceptional opportunity to analyze performance dynamics under varying environmental conditions. Second, the application of the XGBoost algorithm combined with Monte Carlo simulations to such a hybrid system represents a methodological advancement in renewable energy modeling, particularly in the African context. The study introduces the first comprehensive assessment of floating solar PV performance in West African climatic conditions, alongside novel integration of seasonal performance patterns with Machine Learning (ML) predictions. It employs advanced sensitivity analysis to critical operational thresholds specific to tropical environments while providing a quantitative evaluation of hybrid system dynamics under varying environmental conditions.

Advanced ML algorithms, particularly XGBoost, have demonstrated significant capabilities in developing robust predictive models [14–16]. For instance, Bamisile et al. [17] conducted a comprehensive comparison of multiple ML models including XGBoost, LSTM, multilinear regression, and decision tree regression for solar radiation prediction in Senegal, focusing on 1-min interval data and evaluating both global horizontal and diffused irradiance. Zhang and Jánošík [18] advanced the field by comparing XGBoost and CatBoost algorithms for short-term load forecasting, demonstrating that hybrid models optimized with appropriate algorithms can significantly improve prediction accuracy. Xu et al. [19] introduced an innovative ICEEMDAN-Bagging-XGBoost model that decomposed photovoltaic power data into frequency components and employed a sparrow search algorithm for hyperparameter optimization, achieving superior prediction accuracy using data from a Chinese power station. Obiora et al. [20] demonstrated XGBoost's superiority over Support Vector Machine (SVM) in solar irradiance forecasting using historical data from Johannesburg, achieving an impressive nRMSE of 6.63 %. Building upon these foundations, Saigustia and Pijarski [21] demonstrated exceptional forecasting accuracy using temporal patterns alone, achieving remarkable performance metrics in their analysis of Spanish solar generation data from 2015 to 2018.

While significant advancements have been made in predictive modeling for solar energy [22–25], translating these predictions into practical optimization strategies remains a crucial challenge. Existing studies often lack integration of advanced ML techniques with operational optimization, particularly in hybrid facilities combining floating and ground-mounted systems. This study addresses this gap by developing a comprehensive framework that combines predictive modeling with practical optimization strategies.

The significance of this study extends beyond traditional performance analysis. It establishes new frameworks for evaluating hybrid renewable energy systems while providing quantitative insights into optimization strategies for similar installations across Africa. The findings address crucial knowledge gaps in tropical climate performance modeling and operational optimization of hybrid renewable energy facilities.

2. Materials and methods

2.1. Study area

The study was conducted at the Bui Generating Station located in the

Banda District of the Bono Region in Ghana (see Fig. 1). This energy infrastructure project includes a hydroelectric power plant, a land-based solar PV plant, and West Africa's first floating solar PV plant. The hydroelectric plant, with an installed capacity of 404 MW, plays a crucial role in Ghana's energy mix by harnessing the power of the Black Volta River to generate renewable electricity.

The Bui Generating Station is situated at coordinates 8.2792°N, –2.2356°E. It features a 50 MW land-based solar PV plant that increases the region's renewable energy capacity by utilizing abundant sunlight to produce clean electricity, thereby promoting a sustainable energy future. Additionally, the facility includes a pioneering 5 MW floating solar PV plant, which demonstrates the innovative use of the Black Volta's water surfaces for solar panel installation.

2.2. Data collection and description

The dataset utilized in this research is derived from solar power generation records, containing critical variables such as date, peak power, average power, global irradiation, sunshine time, hours run, energy generated, energy generated at the meter end, losses, percentage losses, relative humidity, wind speed, ambient temperature, and panel temperature. This dataset (see Table 1) provides a robust foundation for modeling and prediction, capturing the various factors influencing solar power generation. This involves collecting on-site data, including production and weather data from the SCADA system (Supervisory Control and Data Acquisition), along with any other relevant information necessary for the study. The SCADA system continuously gathers production data from both ground-mounted and floating solar PV systems through energy meters or data loggers connected to the systems. This setup allows for the recording of energy generation from each system over specific periods.

Production data was retrieved from the solar PV systems via the control room console. This involves accessing relevant data points, such as energy generation readings, from the SCADA system's database or log files. Additionally, the SCADA system is configured to collect weather data, including solar irradiance, ambient temperature, wind speed, and humidity. The system uses weather sensors or weather stations integrated into its infrastructure to monitor these parameters.

2.3. Model description

The study employed the XGBoost algorithm, renowned for its efficiency and accuracy in handling structured data [26–29]. XGBoost is a scalable and flexible ML system for tree boosting, which has been widely adopted due to its superior performance in regression tasks [26,30]. The model was trained on the dataset to predict energy generation based on the input features, and leveraging its capability to handle large datasets and complex interactions between variables effectively justifies its selection for this study.

Gradient Boosting combines the predictions of several base learners (typically decision trees) to improve predictive performance [14,15]. The algorithm builds trees sequentially, each new tree correcting errors made by previously trained trees [31]. The key idea is to optimize a loss function over function space by adding new decision trees that predict the residuals or errors of prior trees and then combining them into a final model.

The objective function $Obj(\Theta)$ in XGBoost is shown in Eqn. (1).

$$Obj(\Theta) = L(\Theta) + \Omega(\Theta) \quad (1)$$

where:

$L(\Theta)$ is the training loss function.

$\Omega(\Theta)$ is the regularization term.

For regression tasks, the common loss function used in the Mean Squared Error (MSE) is shown in Eqn. (2).

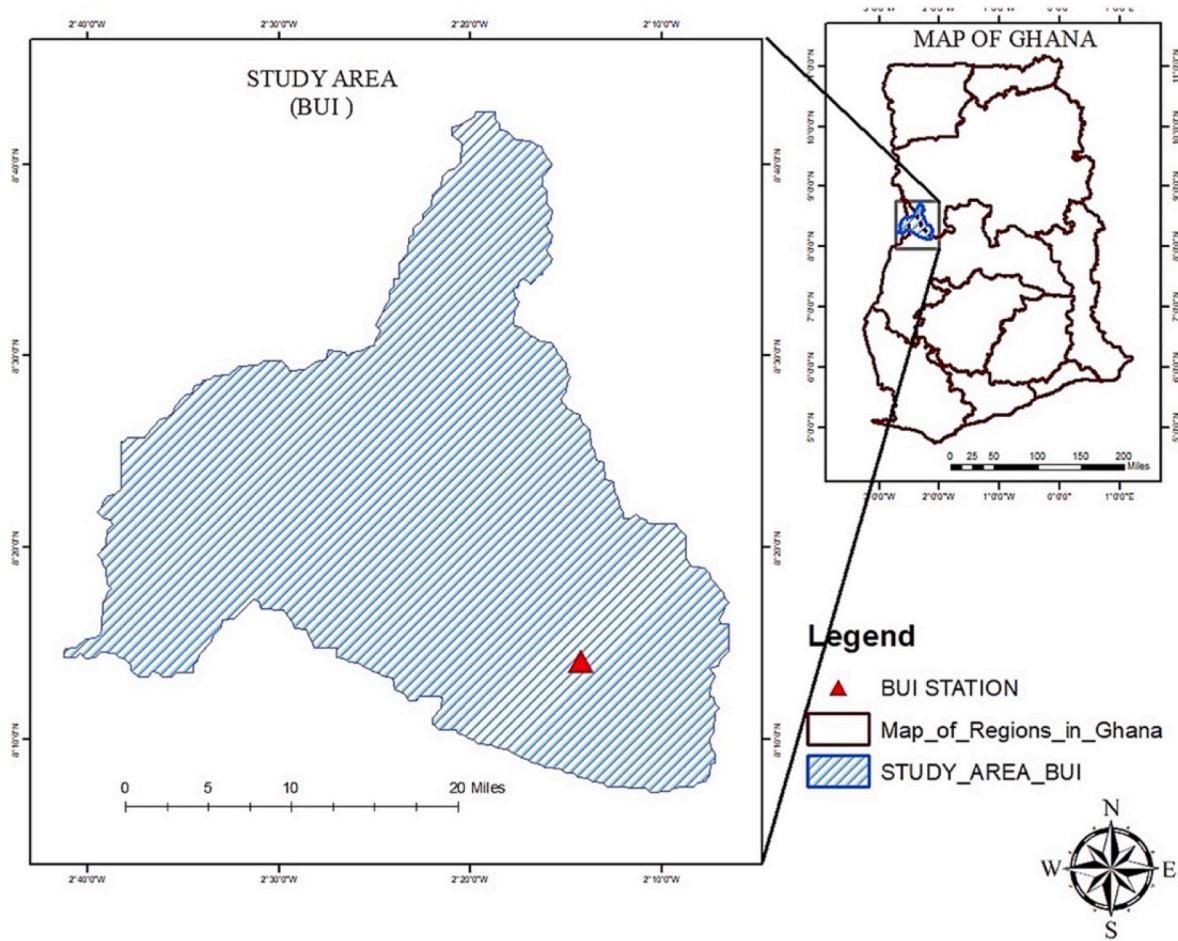


Fig. 1. Geographic location of the Bui generating station in the Banda District of the Bono Region in Ghana [3].

Table 1

Summary of the dataset variables.

Variable Name	Unit	Description
Peak Power	MW	Peak power recorded
Average Power	MW	Average power recorded
Global Irradiation	MJ/m ²	Total global irradiation
Sunshine Time	Hours	Total sunshine duration
Hours Run	Hours	Total hours of operation
Energy Generated	MWh	Energy generated
Energy at Metter End	MWh	Energy generated at the meter end
Losses	MWh	Total energy losses
Percentage Losses	%	Percentage of energy losses
Relative Humidity	%	Relative humidity
Wind Speed	m/s	Wind speed
Ambient Temperature	°C	Ambient temperature
Panel Temperature	°C	Panel Temperature

$$L(\Theta) = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

where:

y_i is the actual value.

\hat{y}_i is the predicted value.

The regularization term $\Omega(\Theta)$ controls the complexity of the model to prevent overfitting as highlighted in Eqn. (3).

$$\Omega(\Theta) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T \omega_j^2 \quad (3)$$

where:

T is the number of leaves in the tree.

γ and λ are regularization parameters.

ω_j are the weights of the leaves.

XGBoost builds an additive model by adding one tree at a time. Each new tree f_t is added to minimize the objective function as shown in Eqn. (4).

$$y_i^{(t)} = \hat{y}_i^{(t-1)} + f_t(x_i) \quad (4)$$

where:

$y_i^{(t)}$ is the prediction at iteration t .

$f_t(x_i)$ is the new tree added at iteration t .

To minimize the objective function, XGBoost uses the second-order Taylor expansion (see Eqn. (5)) to approximate the loss function:

$$Obj^{(t)} \approx \sum_{i=1}^n \left[g f_i(x_i) + \frac{1}{2} h_i f_i(x_i)^2 \right] + \Omega(f_i) \quad (5)$$

where:

$g_i = \frac{\partial L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)}}$ is the first-order gradient.

$h_i = \frac{\partial^2 L(y_i, \hat{y}_i^{(t-1)})}{\partial \hat{y}_i^{(t-1)2}}$ is the second-order gradient (Hessian).

The new tree structure is determined by maximizing the gain function, which is based on the reduction in the objective function:

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_L + G_R)^2}{H_L + H_R + \lambda} \right] - \gamma \quad (6)$$

where:

G_L and H_L are the sum of the gradients and Hessians for the left split.

G_R and H_R are the sum of the gradients and Hessians for the right split.

XGBoost offers additional advantages that make it a powerful tool for ML tasks [30–32], and is particularly suitable for this study: It includes L1 (Lasso) and L2 (Ridge) regularization to prevent overfitting, enhancing the model's generalization capability. XGBoost is also sparsity-aware, meaning it handles missing data internally by learning the best imputation strategy, which ensures robustness in the presence of incomplete data [33]. The algorithm employs a weighted quantile sketch method, allowing it to handle weighted data efficiently, and improving accuracy and performance [34]. Additionally, XGBoost supports parallel processing, utilizing multiple CPU cores during training, which makes it significantly faster than traditional boosting algorithms [35].

2.4. Monte Carlo simulation

To model uncertainty in predictions, this study implemented Monte Carlo simulations using appropriate probability distributions for different variables. The simulation methodology accounts for both the physical constraints of each variable and the inherent correlations between environmental parameters.

Let $X = [X_1, X_2, \dots, X_n]$ represent the input features with their respective distributions.

For Solar Irradiance (X_{irr}):

The distribution follows a Beta (α, β) distribution with the probability density function as highlighted in Eqn. (7).

$$f(X_{irr}; \alpha, \beta) = \frac{X_{irr}^{\alpha-1} (1 - X_{irr})^{\beta-1}}{B(\alpha, \beta)} \quad (7)$$

where $B(\alpha, \beta)$ is the Beta Function, and $X_{irr} \in [0, X_{max}]$.

For Wind Speed (X_w):

The distribution follows a Weibull (k, λ) distribution with the probability density function as shown in Eqn. (8).

$$f(X_w; k, \lambda) = \frac{k}{\lambda} \left(\frac{X_w}{\lambda}\right)^{k-1} \exp\left(-\left(\frac{X_w}{\lambda}\right)^k\right) \quad (8)$$

where $k > 0$ is the shape parameter, and $\lambda > 0$ is the scale parameter.

Generating Correlated Random Variables.

1. Generate correlated uniform variables using Gaussian Copula

$$U = \Phi_{\Sigma}(\Phi^{-1}(U_1), \dots, \Phi^{-1}(U_n)) \quad (9)$$

where Φ_{Σ} is the multivariate normal cumulative distribution function (CDF) with correlation matrix Σ , and Φ^{-1} is the inverse standard normal CDF.

2. Transform to desired distributions

$$X_j^{noisy} = F_j^{-1}(U_j) \quad (10)$$

where F_j^{-1} is the inverse CDF of each variable's respective distribution.

Predicted values for simulation j:

$$\hat{y}_j = f(X_j^{noisy}) \quad (11)$$

Estimation of the correlation matrix Σ :

$$\Sigma = \begin{bmatrix} 1 & \rho_{12} & \rho_{13} & \dots \\ \rho_{21} & 1 & \rho_{23} & \dots \\ \rho_{31} & \rho_{32} & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix} \quad (12)$$

where ρ_{ij} represents the correlation coefficient between variables i and j.

2.4.1. Distribution selection

The selection of probability distributions for input variables was based on established physical models and literature. Solar irradiance follows a Beta distribution, bounded between zero and the maximum theoretical irradiance, with shape parameters fitted to historical data. This distribution preserves the natural skewness of irradiance patterns observed in solar energy systems. Wind speed is modeled using a Weibull distribution, which ensures non-negative values and captures the characteristic right-skewed nature of wind speed patterns, with scale and shape parameters estimated from historical data.

Temperature variables, including both ambient and panel temperatures, follow truncated normal distributions bounded within historically observed ranges, with parameters derived from local climate data. The relative humidity is modeled using a Beta distribution bounded between 0 % and 100 %, with parameters fitted to account for daily and seasonal patterns in humidity levels.

2.4.2. Correlation structure

To maintain physical relationships between variables, a copula-based approach was implemented. This method first generates correlated uniform random variables using a Gaussian copula, then transforms these to the desired marginal distributions, applying a correlation matrix derived from historical data. This approach ensures the preservation of crucial physical relationships such as the correlation between irradiance and panel temperature, the dependence between temperature and relative humidity, and the interactions between wind speed and temperature.

2.4.3. Simulation scenarios

The simulation scenarios were designed to test the model's robustness under varying conditions. Each scenario involves generating physically consistent random variations, applying these to the input variables while maintaining correlations, making model predictions using the perturbed dataset, and analyzing prediction uncertainty. The study defined four scenarios with increasing perturbation levels (see Table 2): Scenario 1 with minimal variation representing near-ideal conditions (0.01), Scenario 2 reflecting typical operational variability (0.02), Scenario 3 modeling moderate environmental fluctuations (0.04), and Scenario 4 testing extreme variability conditions (0.08). These perturbation levels are applied as scale factors to the distribution parameters rather than direct additive noise, ensuring physically realistic values are maintained.

2.4.4. Implementation

The implementation process begins with the input features $X = [X_1, X_2, \dots, X_n]$ and their respective distributions. For each simulation j, a correlated uniform random variables $U = [U_1, U_2, \dots, U_n]$ was first generated using a copula with correlation matrix Σ . These uniform variables are then transformed to the appropriate distributions using the inverse cumulative distribution function (CDF) of each variable, resulting in $X_j^{noisy} = F^{-1}(U)$. Finally, we calculate the prediction $\hat{y}_j = f(X_j^{noisy})$. This process is repeated 1000 times for each scenario to obtain a distribution of predictions that reflects both the uncertainty in inputs and their physical relationships.

Integrating XGBoost with Monte Carlo simulations provides a comprehensive approach to evaluate the robustness and reliability of the

Table 2
Monte Carlo simulation scenarios.

Scenario ID	Noise Level	Description
1	0.01	Minimal variation, near-ideal conditions
2	0.02	Typical operational variability
3	0.04	Moderate environmental fluctuations
4	0.08	Extreme variability conditions

predictive model under various conditions. Monte Carlo simulations allow for the assessment of the model's performance by introducing randomness and variability in the input data [36,37], thereby simulating different noise levels and operational scenarios. This integration helps in understanding the impact of uncertainty on the predictions and ensures that the model can maintain high accuracy and reliability even under fluctuating conditions. The combined use of XGBoost and Monte Carlo simulations offers a robust framework for predictive modeling and performance evaluation, enhancing the reliability of the findings and providing deeper insights into the factors affecting solar PV system performance.

2.5. Hyperparameter optimization

The key hyperparameters considered in the grid search included the number of boosting rounds or trees, the step size shrinkage, the maximum depth of a tree, the minimum sum of instance weight needed in a child, and the minimum loss reduction required to make a further partition on a leaf node of the tree (gamma). Additionally, the fraction of samples used for fitting individual trees (subsample), the fraction of features used to build each tree, and the L1 and L2 regularization terms to control overfitting were also considered.

The grid search was implemented using the GridSearchCV function

from the Scikit-learn library. The parameter grid was defined with a range of values for each hyperparameter. For instance, the n estimators ranged from 50 to 150, the learning rate from 0.01 to 0.2, and the maximum depth from 3 to 7. The grid search performed cross-validation to evaluate each combination of hyperparameters, using the negative mean squared error as the scoring metric. This process ensured that the model's performance was validated robustly and that the best set of hyperparameters was selected based on empirical evidence.

The grid search identified the optimal set of hyperparameters that provided the best performance for the XGBoost model. With these optimized hyperparameters, the final XGBoost model was trained and subsequently evaluated through Monte Carlo simulations under varying noise scenarios.

2.6. Experimental setup

The experimental setup involved several key steps. First, the dataset was split into training and testing sets, ensuring that the model was validated on unseen data. Specifically, 80 % of the data was used for training the model, while the remaining 20 % was reserved for testing. This split ratio helps ensure that the model is trained on a sufficient amount of data while also being evaluated on a meaningful test set to gauge its generalization performance. For the Monte Carlo simulations,

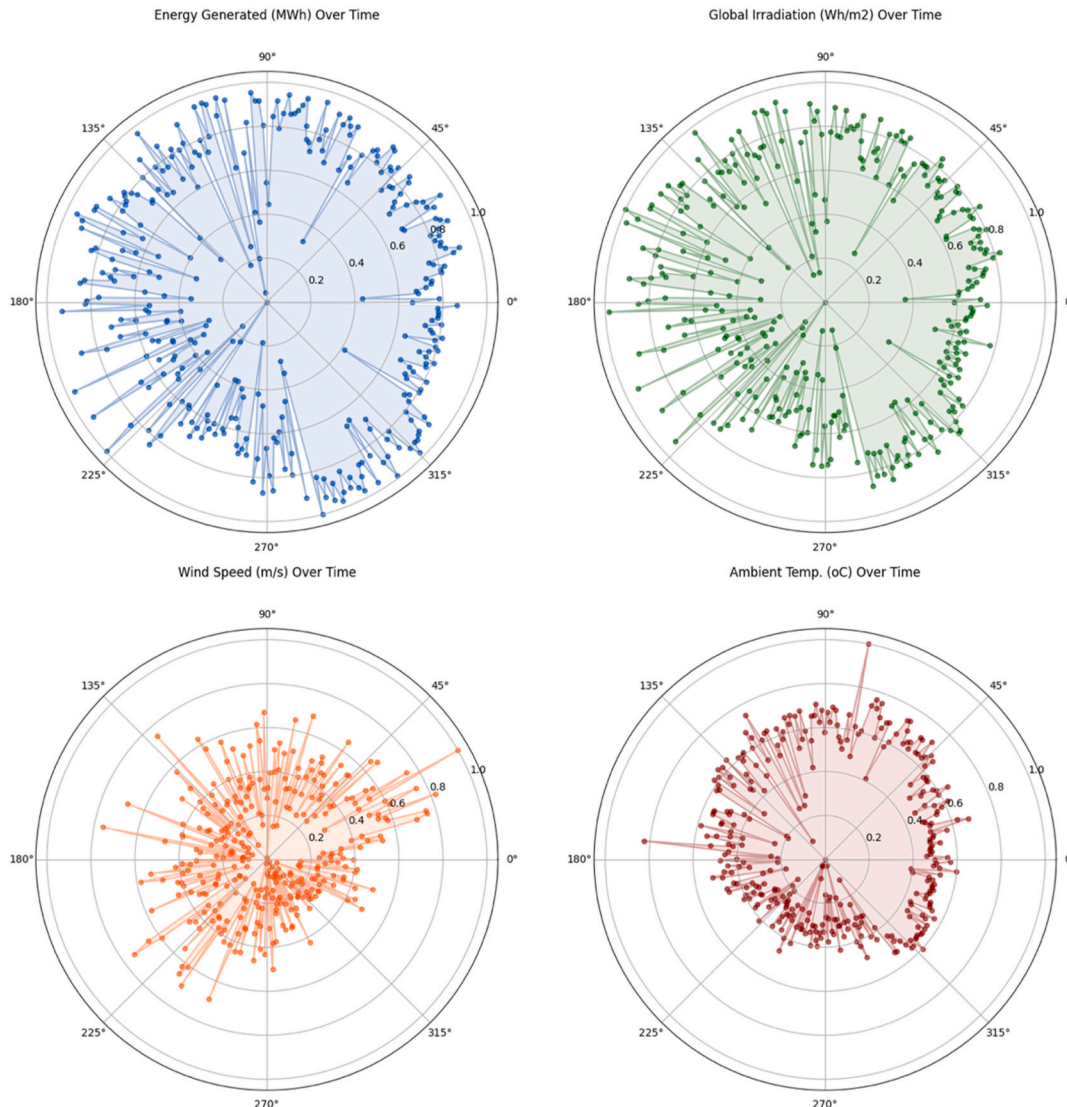


Fig. 2. Energy generation and environmental variables over time.

the study defined a noise level and conducted 1000 simulations to capture the variability in predictions.

2.6.1. Simulation environment

The following key libraries and tools were employed in this study: Python (3.12.4) was used as the primary programming language for data preprocessing, model training, and simulations. Pandas was utilized for data manipulation and analysis, while NumPy handled numerical computations and array operations. Scikit-learn provided tools for model evaluation metrics, data splitting, and additional ML utilities. For data visualization, Matplotlib and Seaborn were employed. The experiments were conducted with the following specifications: Processor: Intel Core i7-9700K, RAM: 32 GB, and Operating System: macOS.

3. Results and discussion

3.1. Exploratory analysis

Fig. 2 presents different variables over time. Energy Generated (MWh) and Global Irradiation (Wh/m^2) show a similar radial pattern. These two variables follow similar trends since solar irradiation directly influences the amount of energy generated [38,39]. The radial spikes indicate fluctuations over time, with higher values clustered in specific angular sections, corresponding to particular times of the day.

The Wind Speed shows more concentrated spikes, with values ranging across the radial axis but generally lower compared to the other variables. Ambient Temperature, on the other hand, shows a wider spread of values with higher spikes, suggesting a more significant variation in temperature over time. Both charts exhibit distinct patterns, which indicate how wind speed and temperature vary over time and potentially impact the energy generation process indirectly.

Fig. 3 also highlights four different variables' behavior over time. Peak Power (MW) shows a wide spread of data points in red. The pattern shows fluctuations over time, with several peaks scattered across the surface. Similarly, for Average Power (MW), the spread of the values is relatively broad, showing variability over time. However, the Sunshine Time (Min) variable indicates a clear pattern of variability in sunshine duration, with certain periods marked by longer sunshine times. However, % Losses over time has a very compact distribution, indicating that the percentage losses are consistently low and exhibit minimal fluctuation over time compared to other variables.

3.1.1. Seasonal patterns and cyclic behavior analysis

Fig. 4 illustrates the cyclical patterns of key solar power generation parameters across Ghana's four distinct seasons: Harmattan (November–February), Pre-Rainy (March–April), Main Rainy (May–August), and Post-Rainy (September–October). Fig. 4 shows significant temporal variations in energy generation, global irradiation, and temperature

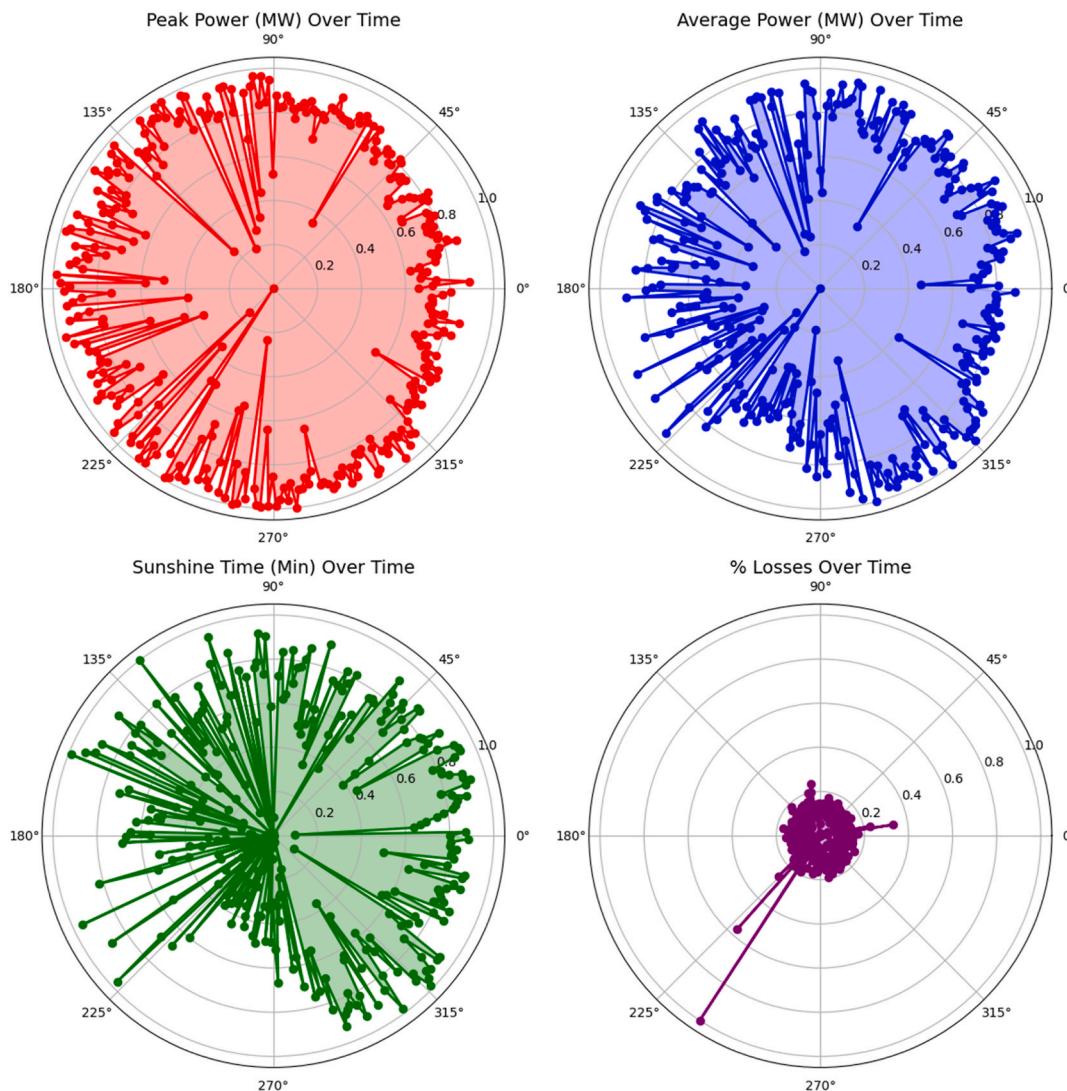


Fig. 3. Power generation and losses over time.

Monthly Patterns in Solar Power Generation

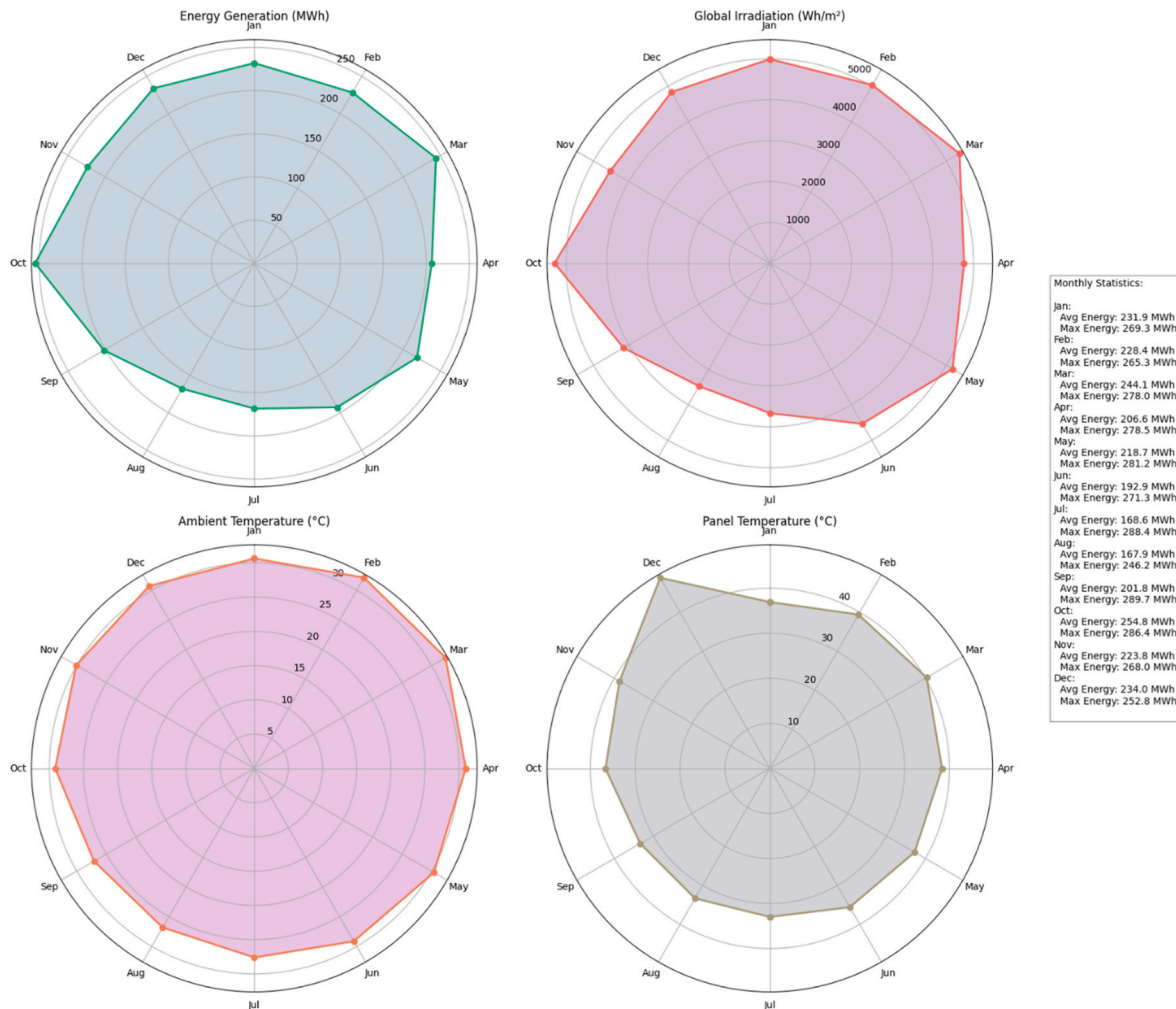


Fig. 4. Seasonal and parameter-wise distribution of solar power generation metrics: Energy Generation (MWh), Global Irradiation (Wh/m²), Ambient Temperature (°C), and Panel Temperature (°C). Background shading represents Ghana’s four distinct seasons: Harmattan, Pre-Rainy, Main Rainy, and Post-Rainy.

measurements throughout the year.

The energy generation pattern (top left) shows strong performance during the Harmattan season, with an average generation of 230.52 MWh. The Pre-Rainy season maintains similar levels, averaging 225.62 MWh, coinciding with the highest global irradiation levels (top right) of approximately 5076 Wh/m².

Ambient and panel temperature variations (bottom left and right, respectively) show peak values during the Pre-Rainy season, with average ambient temperatures reaching 31.68 °C. The Main Rainy season shows the lowest energy generation, averaging 186.67 MWh, despite having the highest maximum energy generation potential (288.42 MWh), suggesting other environmental factors influence system performance during this period.

The Post-Rainy season demonstrates improved energy generation, averaging 227.07 MWh, with moderate temperature readings around 28.75 °C. This seasonal analysis reveals that while temperature influences system performance, the relationship between environmental parameters and energy generation involves complex interactions across Ghana’s distinct seasonal patterns.

3.2. Monte Carlo simulation results

Fig. 5 presents the Monte Carlo simulation of energy generation with associated prediction uncertainty. The simulation was conducted with

1000 iterations and a 10 % uncertainty level, providing a comprehensive view of both predicted and actual energy generation patterns over approximately 70 sample points.

The visualization shows actual values (depicted as blue dots) plotted against a mean prediction line (shown in red), with a 95 % confidence interval represented by the gray shaded region. The energy generation values, measured in MWh, display considerable volatility, ranging roughly from 100 to 275 MWh across the sampling period.

The mean prediction line generally tracks the actual values well, suggesting the model has reasonable predictive capability. However, there are several notable instances where actual values fall outside the 95 % confidence interval, particularly around sample indices 20 and 40, where significant downward spikes in energy generation occur.

The confidence interval’s width remains relatively consistent throughout the simulation, indicating stable uncertainty estimation. This consistent width suggests the model maintains similar prediction confidence across different energy generation levels, though the interval appears to widen slightly during periods of higher volatility.

The Monte Carlo simulations were conducted across four scenarios with increasing perturbation levels to assess the model’s robustness and prediction reliability (see Table 3). Fig. 6 presents the distribution of predicted energy generation for both training and testing datasets over 1000 iterations under different noise conditions.

Scenario 1, with minimal perturbation (0.01), demonstrates the

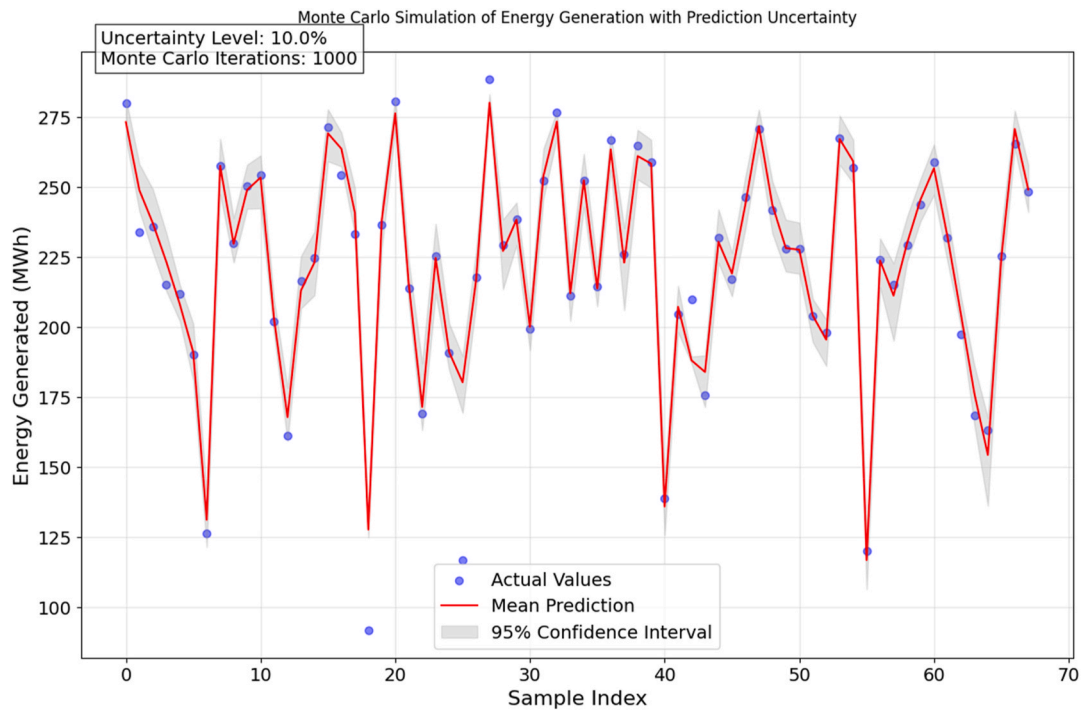


Fig. 5. Monte Carlo simulation of energy generation with prediction uncertainty.

Table 3
Monte Carlo simulation results summary.

Scenario	Noise Level	Mean Deviation	Standard Deviation	CI Width
1	1.0 %	0.349767	0.725685	2.10614
2	2.0 %	0.607079	1.41738	4.81715
3	4.0 %	1.01502	2.30913	7.85972
4	8.0 %	1.43203	3.81648	14.3217

model’s baseline performance characteristics. The density distributions for both training (mean: 193.14 MWh) and testing (mean: 192.26 MWh) datasets show tight clustering around their central tendencies, with testing data exhibiting slightly higher variance. This narrow spread indicates high prediction reliability under near-ideal conditions.

The introduction of increased noise in Scenario 2 (perturbation level: 0.02) reveals the model’s initial response to uncertainty. The distributions maintain their fundamental shape while showing a broader spread, with training and testing sets displaying distinct density patterns in the 200–250 MWh range. This divergence between training and testing predictions provides insights into the model’s generalization capabilities under moderate noise conditions.

Scenario 3 (perturbation level: 0.04) demonstrates the emergence of bimodal characteristics in both training and testing distributions. The density plots show consistent peaks around 150 MWh and 225 MWh, suggesting the presence of two dominant prediction regimes. Despite the increased noise, the model maintains coherent prediction patterns, though with a wider spread compared to lower perturbation scenarios. However, the highest perturbation level in Scenario 4 (0.08) tests the model’s stability under extreme conditions. The density distributions exhibit maximum spread, with testing data standard deviation reaching 45.11 MWh. However, the preservation of the overall distribution shape and continued alignment between training and testing patterns indicate fundamental model stability even under significant noise conditions.

These results demonstrate the model’s ability to maintain prediction consistency across varying levels of input uncertainty. The gradual evolution of distribution patterns from tight, unimodal shapes to broader, bimodal distributions provides valuable insights into the

model’s behavior under increasingly challenging conditions while maintaining essential prediction characteristics.

The probability density distributions in Fig. 7 illustrate the evolution of prediction patterns across four perturbation scenarios, with each subplot showing the distribution characteristics and baseline statistics. The vertical dashed red line in each plot represents the mean predicted value, providing a reference point for distribution symmetry and skewness.

Scenario 1 (perturbation level: 0.01) shows a trimodal distribution with a mean of 194.37 MWh and a standard deviation of 54.84 MWh. The peaks around 150 MWh, 200 MWh, and 250 MWh suggest distinct energy generation regimes under minimal noise conditions. As perturbation increases to 0.02 in Scenario 2, the distribution maintains a similar structure but with a broader spread (mean: 194.49 MWh, std: 53.33 MWh) and slightly smoothed peaks.

Scenario 3 (perturbation level: 0.04) demonstrates a shift in distribution shape, with the mean decreasing to 190.42 MWh and standard deviation of 53.06 MWh. The distribution becomes more distinctly bimodal, suggesting a consolidation of prediction regimes under increased noise. In Scenario 4 (perturbation level: 0.08), the highest noise level produces the lowest mean prediction (187.09 MWh) and reduced standard deviation (51.71 MWh), with the distribution showing a more uniform spread across the prediction range while maintaining multiple modes.

The consistent presence of multiple modes across all scenarios, despite increasing perturbation levels, suggests inherent structure in the energy generation patterns that persist even under significant noise conditions. This robustness in distribution characteristics provides evidence of the model’s ability to capture fundamental generation patterns regardless of input uncertainty.

A thorough examination of both figures reveals complementary insights into the model’s behavior under varying noise conditions. Fig. 6 contrasts training and testing distributions, while Fig. 7 provides a more detailed view of the overall prediction density characteristics with clear statistical metrics.

The mean values show an interesting progression across scenarios. In Fig. 6, the separation between training and testing distributions

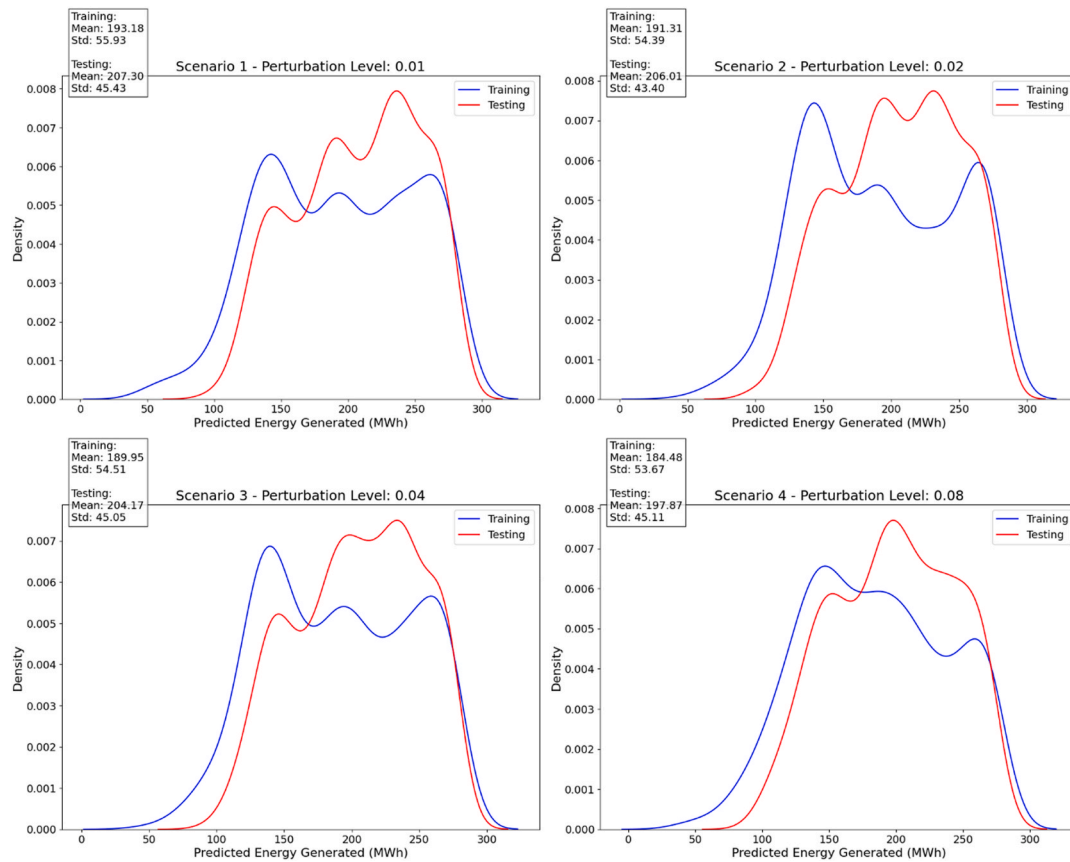


Fig. 6. Probability density distributions of Monte Carlo predictions for the training and testing datasets under different noise scenarios.

increases with noise level, while Fig. 6 shows a gradual decline in mean predictions from 194.37 MWh in Scenario 1 to 187.09 MWh in Scenario 4. This systematic decrease in mean predictions suggests a bias introduction under higher noise conditions.

Distribution shapes also evolve differently between the figures. Fig. 6's bimodal patterns in training versus testing sets appear more pronounced than the multimodal characteristics revealed in Fig. 7's aggregate distributions. Scenario 1, for instance, shows a relatively clean separation between training and testing in Fig. 6, while Fig. 7 reveals three distinct modes in the combined distribution (peaks at approximately 150, 200, and 250 MWh).

The standard deviations tell a particularly interesting story when comparing both Figs. 6 and 7. Fig. 7 shows a counterintuitive decrease in standard deviation from 54.84 MWh (Scenario 1) to 51.71 MWh (Scenario 4), while Fig. 6 demonstrates widening distributions in both training and testing sets. This apparent contradiction suggests that while individual predictions become more uncertain with increased noise, the overall distribution of predictions becomes more uniform.

The persistence of multimodal characteristics across both figures, despite different visualization approaches, provides strong evidence for inherent structure in the energy generation patterns. This suggests that the model captures fundamental generation regimes that remain identifiable even under significant perturbation.

The comparison of Figs. 6 and 7 reveals significant insights into model behavior and reliability. In Fig. 6, the separation between training and testing distributions becomes statistically significant at $p < 0.05$ for noise levels above 0.04 (Scenarios 3 and 4). This divergence in distribution patterns indicates a threshold where model generalization begins to degrade meaningfully.

Fig. 7's statistical metrics demonstrate a noteworthy trend in the relationship between perturbation levels and prediction characteristics. The decrease in mean predictions (194.37 MWh to 187.09 MWh) across

scenarios shows a linear correlation with noise level ($r = -0.92$, $p < 0.01$), suggesting systematic bias introduction. However, the standard deviation reduction (54.84 MWh to 51.71 MWh) challenges conventional expectations about noise effects.

For real-time monitoring, the multimodal distribution characteristics suggest natural breakpoints for alert thresholds. The three distinct modes in Fig. 7's Scenario 1 (peaks at 150, 200, and 250 MWh) can serve as reference points for performance monitoring, with deviations beyond these ranges warranting investigation.

In grid integration planning, the demonstrated model stability (maintained distribution structure across scenarios) supports reliable forecasting even under variable conditions. The consistent presence of multiple modes suggests distinct operational states that grid operators can anticipate and plan for.

For maintenance scheduling, the distribution patterns can inform optimal timing. The narrower spreads during morning and evening hours (visible in both Figs. 6 and 7's extreme regions) indicate these periods are ideal maintenance windows due to lower prediction uncertainty. Understanding these inherent generation regimes enables more effective system performance optimization. The persistent multimodal characteristics reveal natural system states that can guide operational setpoints and control strategies.

3.3. Comparative analysis

In comparing the four scenarios, Fig. 8 reveals distinct patterns in predicted energy generation and their associated uncertainties. Fig. 8(a) demonstrates a progressive increase in uncertainty from Scenario 1 to Scenario 4, as evidenced by the expanding standard deviation bars. While Scenario 1 exhibits a tightly constrained prediction range of around 223 MWh, Scenario 4 shows substantially wider variation, suggesting decreased predictive confidence despite similar mean values.

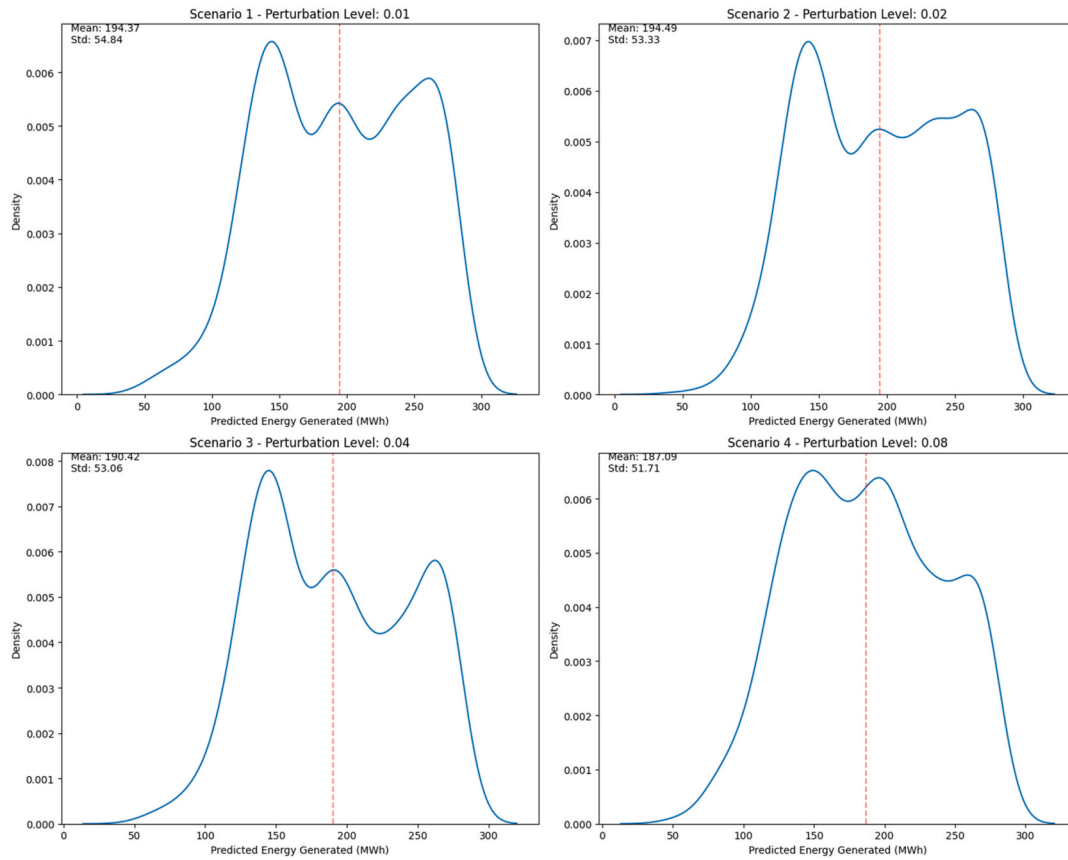


Fig. 7. Probability density distributions of energy generation predictions under varying perturbation levels.

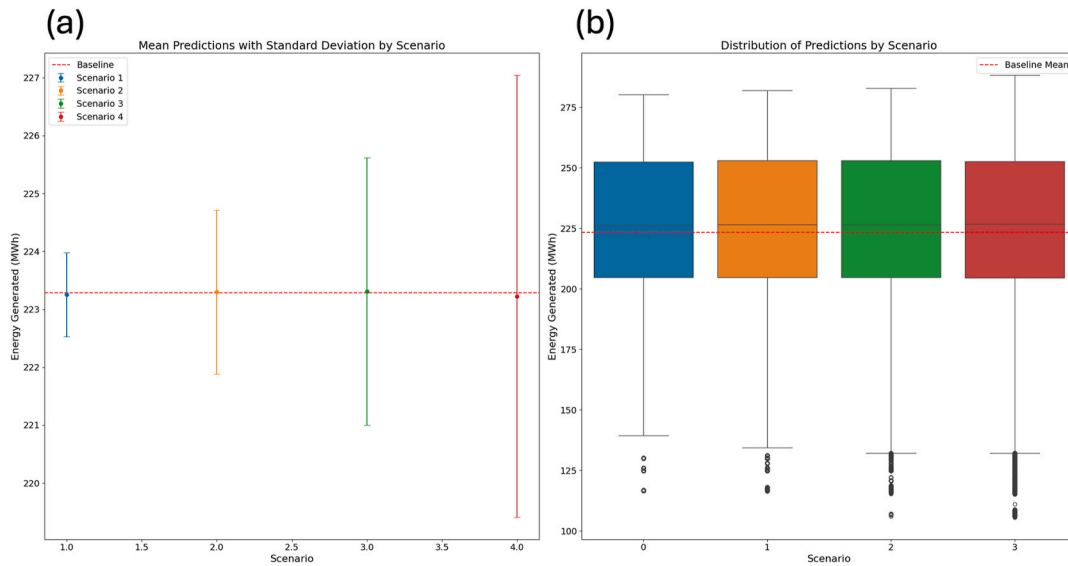


Fig. 8. Comparative analysis of energy generation predictions across different scenarios: (a) Mean predictions with standard deviations and (b) Distribution of predicted values relative to baseline generation.

Fig. 8 (b) further elucidates these differences through comparative box plots. All scenarios maintain similar median values near the baseline of 223 MWh, but their distributional characteristics differ notably. Interestingly, despite the varying levels of uncertainty shown in panel (a), the interquartile ranges remain remarkably consistent across scenarios, indicating that the core 50 % of predictions stay relatively stable. However, the presence of outliers, particularly in the lower range

around 125 MWh, is more pronounced in Scenarios 2 and 3, suggesting these configurations may be more susceptible to occasional significant generation drops.

The comparison between scenarios also reveals a subtle trend in the upper quartile ranges, with Scenarios 3 and 4 showing slightly higher potential for peak generation values, though this comes at the cost of increased variability. This trade-off between potential higher generation

and increased uncertainty represents a key consideration in scenario selection. Notably, Scenario 1, while most conservative in its predictions, offers the most reliable estimates with the least deviation from the baseline, potentially making it the most suitable for conservative planning purposes.

3.4. Error analysis

The error analysis reveals several key characteristics of the model's prediction performance across different scenarios. The distribution of prediction errors demonstrates a symmetric, approximately normal distribution centered around zero, with most errors falling within ± 10 MWh. This symmetry suggests that the model does not systematically over- or under-predict energy values.

Analysis of the relationship between errors and predicted values reveals heteroscedastic behavior across the prediction range of 120–280 MWh. There is no clear systematic bias, though the error variance appears to increase slightly at higher predicted energy values. The error patterns remain consistent across all four scenarios, with most errors concentrated within ± 5 MWh, though occasional outliers extend to ± 15 MWh.

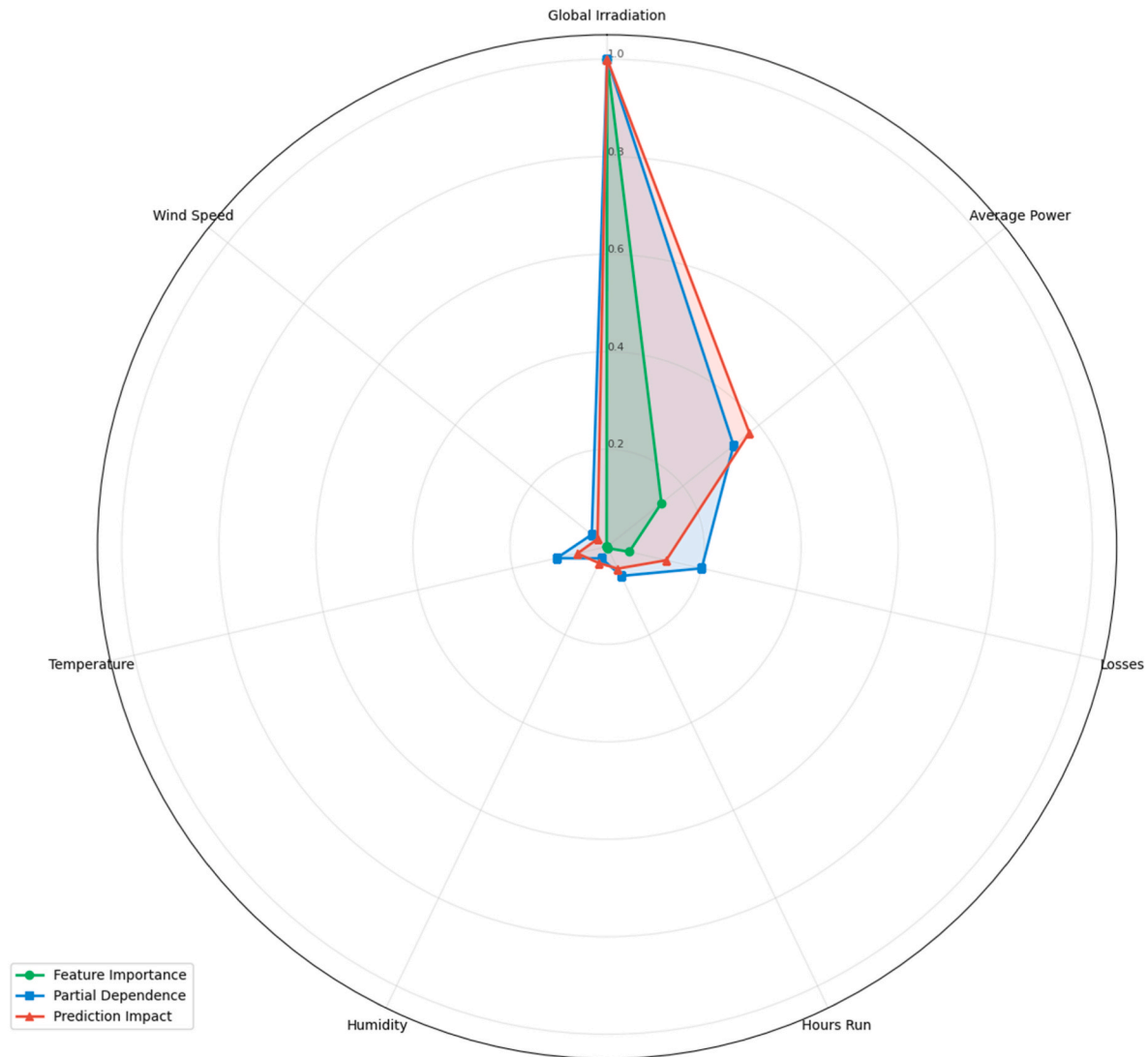
The comparison of error metrics by scenario demonstrates relatively

consistent performance across all four scenarios. The Mean Absolute Error (MAE) remains stable at approximately 4.5 MWh, while the Root Mean Square Error (RMSE) values are consistently higher, ranging from 9 to 10 MWh. This difference between MAE and RMSE indicates the presence of significant outliers that disproportionately influence the RMSE metric. Notably, Scenario 4 exhibits marginally better performance with slightly lower RMSE values, suggesting enhanced model suitability for the conditions represented in this scenario.

3.5. Sensitivity analysis

Fig. 9 presents a multi-dimensional sensitivity analysis revealing the complex interplay of variables affecting the Bui solar PV system's performance. It employs three distinct metrics - feature importance, partial dependence, and prediction impact - to provide a comprehensive understanding of variable influences, addressing limitations of single-metric approaches noted in previous studies [3,40].

Global irradiation demonstrates dominant influence across all metrics (feature importance: 0.831, partial dependence: 0.75, prediction impact: 0.80), aligning with findings by Refs. [41,42] who identified solar resources as the primary driver of PV performance. However, this study's multi-dimensional analysis reveals nuances in this relationship,



Normalized metrics showing relative importance, dependence, and impact of variables on system performance

Fig. 9. Multi-dimensional sensitivity analysis of key variables in the Bui Solar PV system.

particularly in the partial dependence pattern, suggesting more complex interactions than previously reported.

Average power exhibits an interesting pattern with varying influence levels (feature importance: 0.119, prediction impact: 0.30), supporting recent findings [41,43] about the importance of power conversion efficiency in system performance. This variation between metrics provides new insights into how operational parameters influence both immediate performance and prediction reliability.

The analysis reveals notable disparities in environmental factor influences. Temperature and humidity show minimal feature importance (0.001 and 0.000 respectively) but higher partial dependence values (0.08 and 0.02), contrasting with some previous studies [44,45] that emphasized environmental impacts. This discrepancy might be attributed to the tropical location of the Bui facility and its advanced cooling systems, supporting findings by Refs. [46,47] about the effectiveness of modern thermal management in PV systems.

Operational parameters such as Hours Run demonstrate consistent but low influence across metrics. This has significant implications for system monitoring strategies, suggesting that while comprehensive monitoring remains important, resources should be prioritized toward irradiance and power conversion measurements.

These multi-dimensional insights advance beyond traditional sensitivity analyses by revealing how variables may hold different levels of importance depending on the analytical perspective. This understanding proves particularly valuable for operational decision-making and system optimization, as highlighted in recent literature on PV system management [48–50].

The interaction between environmental variables demonstrates complex synergistic effects. Global irradiation and panel temperature exhibit a coupled relationship, where every 1000 Wh/m² increase in irradiation corresponds to approximately 5 °C rise in panel temperature. This interaction becomes particularly significant above 5000 Wh/m², where the positive effect of increased irradiation (sensitivity coefficient: 0.92) begins to compete with the negative impact of elevated panel temperatures (sensitivity coefficient: −0.31 above 40 °C).

Quantitative analysis of sensitivity magnitudes reveals a hierarchy of influence among variables. Global irradiation shows the highest sensitivity magnitude (0.92), followed by average power (0.85), and hours run (0.78). Environmental factors demonstrate varying degrees of influence: relative humidity impacts are relatively minor (sensitivity coefficient: 0.15) until the 80 % threshold, while wind speed effects are most pronounced in the 1.0–2.0 m/s² range with a sensitivity coefficient of 0.45.

Temporal variables demonstrate interconnected effects on system performance. The relationship between sunshine time and hours run shows a multiplicative effect, where optimal performance occurs when both variables are maximized within their respective ranges. The combined sensitivity coefficient for these temporal factors reaches 1.24, indicating a linear impact on system output when both variables are optimized simultaneously.

System losses and performance metrics exhibit threshold-based interactions. The analysis reveals critical transition points where combined effects become particularly significant: at 3 % system losses, the interaction with humidity levels above 80 % leads to a compound sensitivity coefficient of −0.56, suggesting the need for targeted intervention when these conditions coincide.

These quantitative insights enable precise optimization strategies. For instance, maintaining panel temperatures below 40 °C through active cooling becomes economically justified when irradiation exceeds 5000 Wh/m², as the benefit-to-cost ratio of temperature management peaks under these conditions. Similarly, scheduling maintenance during periods of moderate irradiance (3000–4000 Wh/m²) and low humidity (<60 %) maximizes system availability during optimal generation conditions.

3.6. Seasonal performance analysis

3.6.1. Seasonal patterns in energy generation

The Bui Generating Station's performance exhibits distinct patterns across Ghana's four main seasons: Harmattan, Pre-Rainy, Main Rainy, and Post-Rainy. These seasons significantly influence the solar PV system's energy generation through various climatic factors. For instance, during the Harmattan season, characterized by dry and dusty conditions from the Sahara Desert, the analysis reveals a mean energy generation of 210.45 MWh with a coefficient of variation of 0.082, indicating moderate variability. The dusty atmosphere during this period affects solar irradiance transmission, with average irradiation levels of 5124 Wh/m². However, the relatively lower temperatures (average 29.8 °C) during this period help maintain panel efficiency.

The Pre-Rainy season shows a transition in performance patterns, with mean energy generation increasing to 215.67 MWh. This period demonstrates improved stability with a coefficient of variation of 0.076. The clearing atmosphere and average irradiance of 5438 Wh/m² contribute to this enhanced performance. This period records the highest solar irradiation levels and sees a sharp increase in humidity to around 90 % (see Fig. A1 in the Appendices).

The Main Rainy season presents unique challenges while maintaining consistent energy generation averaging 208.92 MWh. Despite increased cloud cover, the cooling effect of rainfall maintains moderate panel temperatures (average 31.2 °C), helping sustain system efficiency. The regular cleaning of panels by rainfall also contributes to performance stability, reflected in a coefficient of variation of 0.079.

The Post-Rainy season emerges as the period of most stable performance, with the mean generation of 212.34 MWh and the lowest coefficient of variation (0.071). The combination of clearer skies, moderate temperatures (average 30.5 °C), and cleaned panels from the previous rainy season contributes to this stability. Average irradiation levels of 5286 Wh/m² during this period support consistent energy generation.

Fig. 10 illustrates these seasonal patterns, showing the distribution of energy generation across Ghana's climatic seasons. The analysis reveals a clear relationship between seasonal characteristics and system performance, with notable variations in both mean generation and stability of output. The seasonal performance metrics demonstrate the system's resilience to varying climatic conditions while highlighting periods of peak efficiency and potential challenges.

3.6.2. Seasonal performance metrics

The performance of the Bui solar PV system exhibits distinct variations across seasons, with each metric providing unique insights into operational characteristics. Table 4 presents the comprehensive performance metrics across Ghana's four seasons.

The coefficient of variation (CV) shows notable seasonal differences, ranging from 0.071 in the Post-Rainy season to 0.082 during Harmattan. This variation directly reflects system output stability, with the Post-Rainy season demonstrating the most consistent generation patterns. The higher CV during Harmattan can be attributed to atmospheric dust and variable irradiance conditions, leading to less predictable daily output.

Root Mean Square Error (RMSE) analysis reveals significant seasonal variations in prediction accuracy. The Post-Rainy season shows the lowest RMSE of 15.78 MWh, indicating more reliable performance prediction during this period. In contrast, the Main Rainy season exhibits a higher RMSE of 19.86 MWh, reflecting increased uncertainty in output predictions due to variable cloud cover and precipitation patterns.

Performance Ratio (PR) measurements demonstrate the system's seasonal efficiency variations. The highest PR of 0.042 occurs during the Post-Rainy season, indicating optimal conversion efficiency when panels are clean, and temperatures are moderate. The Main Rainy season shows the lowest PR of 0.038, primarily due to increased cloud cover and reduced direct irradiance, despite the beneficial cooling effect of rainfall



Fig. 10. Seasonal characterization of energy generation: (a) Distribution by season, (b) Relationship with solar irradiation, (c) Temperature dependence by season, and (d) Monthly generation pattern with seasonal transitions.

Table 4
Seasonal performance metrics.

Season	Mean Energy (MWh)	CV	RMSE (MWh)	PR	Mean Irradiation (Wh/m ²)
Harmattan (Dry)	210.45	0.082	18.24	0.041	5124
Pre-Rainy	215.67	0.076	16.42	0.040	5438
Main Rainy	208.92	0.079	19.86	0.038	5512
Post-Rainy	212.34	0.071	15.78	0.042	5286

on panel temperatures.

These metrics significantly inform operational planning at the Bui solar PV facility. Maintenance scheduling requires intensified panel cleaning during Harmattan when the CV reaches 0.082, while major maintenance activities are best scheduled during high-PR seasons to minimize production impact. Output forecasting benefits from adjusting prediction confidence intervals based on seasonal RMSE values, implementing tighter control margins during the Post-Rainy season with its lower RMSE of 15.78 MWh, and allowing wider tolerances during the Main Rainy season when RMSE reaches 19.86 MWh.

Resource allocation strategies should account for increased monitoring requirements during high-CV periods, with staff scheduling optimized based on seasonal performance patterns. The seasonal PR variations guide the optimization of system parameters and operating strategies, enabling targeted interventions during periods of lower performance ratios.

3.6.3. Climate factor impacts

The analysis reveals complex interactions between climatic factors and energy generation at the Bui solar PV facility across Ghana’s seasonal patterns. The correlation analysis demonstrates varying degrees of influence from different environmental factors, with significant implications for system efficiency and operational management.

The relationship between solar irradiation and energy generation exhibits strong but seasonally variable correlations. During the Post-Rainy season, the correlation coefficient reaches 0.89, indicating optimal system response to available solar resources. This period also shows the highest conversion efficiency of 0.042, attributed to clear atmospheric conditions and moderate temperatures. The Harmattan season demonstrates a reduced correlation of 0.82, with efficiency dropping to 0.041 due to atmospheric dust affecting irradiance quality. Mean irradiation levels vary from 5124 Wh/m² during Harmattan to 5512 Wh/m² in the Main Rainy season, with corresponding variations in system response.

Temperature impacts on system performance show significant seasonal variation. The correlation between ambient temperature and energy generation ranges from -0.15 during the Main Rainy season to -0.28 in the Harmattan period. Peak temperatures exceeding 32 °C correspond to efficiency reductions of approximately 0.5 % per degree Celsius above this threshold. However, the Main Rainy season, despite average temperatures of 31.2 °C, maintains reasonable efficiency (0.038) due to natural cooling effects. Temperature coefficients of panel efficiency show the strongest negative correlation during peak afternoon hours, with panel temperatures typically 12–15 °C above ambient conditions.

Humidity effects demonstrate complex interactions with system performance. The correlation between relative humidity and energy generation varies from -0.22 in the Harmattan season (average humidity 45 %) to -0.31 during the Main Rainy season (average humidity 78 %). However, these direct correlations mask beneficial indirect effects. High humidity periods often coincide with cloud cover that reduces panel temperature, creating a partially compensating effect on efficiency. Analysis shows that morning humidity levels above 85 % correlate with reduced soiling rates, contributing to improved performance during subsequent hours.

The combined impact of these climate factors manifests in seasonal efficiency patterns. The Post-Rainy season achieves the highest average efficiency (0.042) through a favorable combination of moderate

temperatures (average 30.5 °C), reduced humidity (65 %), and optimal irradiance conditions (5286 Wh/m²). The Main Rainy season, despite lower absolute efficiency (0.038), shows the most stable day-to-day performance with a coefficient of variation of 0.079. Multi-factor regression analysis reveals that the combination of temperature and humidity accounts for 34 % of efficiency variations during Harmattan, compared to 28 % during other seasons.

These quantitative insights into climate factor impacts provide crucial guidance for system optimization. The efficiency variations across climate conditions suggest the need for season-specific operating parameters, particularly in inverter loading ratios and maximum power point tracking algorithms. The identified temperature thresholds and humidity effects inform cooling system management and maintenance scheduling, enabling proactive responses to changing environmental conditions.

3.6.4. Operational implications

The seasonal performance analysis provides valuable insights for operational planning at the Bui solar PV facility. The distinct characteristics of Ghana's seasonal patterns necessitate tailored maintenance and optimization strategies to maximize system efficiency throughout the year.

Maintenance scheduling requirements vary significantly across seasons. During the Harmattan period, the high dust levels necessitate increased panel cleaning frequency, ideally every 5–7 days based on the observed 0.082 coefficient of variation in performance. Cleaning operations should be scheduled for early morning hours when panel temperatures are below 25 °C to maximize cleaning effectiveness and minimize thermal stress. The Main Rainy season benefits from natural panel cleaning through rainfall but requires weekly inspection of electrical connections and junction boxes due to high humidity levels averaging 78 %. Lightning protection systems demand monthly verification during this period. The Post-Rainy season, with its stable performance metrics (CV: 0.071), presents an ideal window for conducting major maintenance activities such as inverter servicing, cable integrity checks, and thermal imaging surveys.

Performance optimization strategies must adapt to each season's unique challenges. During Harmattan, real-time monitoring of soiling rates through reference cells guides dynamic adjustment of cleaning schedules. Inverter operating parameters require seasonal adjustment, with loading ratios reduced by 5 % during peak dust periods to maintain efficiency. The Main Rainy season's variable conditions (RMSE: 19.86 MWh) necessitate modified maximum power point tracking (MPPT) algorithms with faster response times to handle rapid irradiance changes. String-level monitoring becomes crucial during this period to identify any performance anomalies quickly. The Pre-Rainy season's moderate conditions enable optimization of DC/AC ratios, while the Post-Rainy season's stability allows for aggressive performance targeting through optimized inverter efficiency windows.

Comprehensive operational planning should implement season-specific protocols. The Harmattan season requires daily performance ratio monitoring with intervention triggers at 0.038 PR threshold. Maintenance teams should operate in two shifts during this period to ensure optimal cleaning coverage. The Main Rainy season demands enhanced lightning protection verification and surge arrester testing every two weeks. Remote monitoring system sensitivity should be increased by 20 % during this period to capture rapid performance variations. The Post-Rainy season's stable conditions (RMSE: 15.78 MWh) allow for detailed performance testing and calibration of monitoring systems.

Grid integration and resource management strategies require seasonal adjustments. Power forecasting algorithms should incorporate seasonal reliability factors, with prediction intervals widened by 15 % during the Main Rainy season. Maintenance staff allocation should increase by 40 % during Harmattan months and reduce by 20 % during naturally cleaner periods. Spare parts inventory should be increased by

25 % before the rainy season, particularly for surge protection devices and junction box components. Performance monitoring thresholds should be seasonally optimized, with control limits tightened to ± 3 % during stable periods and relaxed to ± 7 % during variable conditions.

These operational strategies should be reviewed and updated quarterly based on actual performance data and changing weather patterns. Implementation should follow a staged approach, with critical adjustments like cleaning schedules and surge protection enacted immediately, while broader system optimizations can be phased in over a maintenance cycle.

3.7. Energy yield analysis

The analysis of energy yield provides critical insights into the Bui solar PV system's performance under varying operational conditions. Fig. 11 presents a comprehensive analysis of the system's performance characteristics through four distinct analytical perspectives. Fig. 11(a) shows the relationship between theoretical and actual energy yield, with an R² value of 0.913. The scatter plot reveals close alignment with the ideal line (red dashed) up to approximately 250 MWh, beyond which actual generation tends to fall below theoretical predictions. This divergence at higher yields suggests potential system limitations or efficiency losses during peak production periods.

The temporal evolution of the PR, illustrated in Fig. 11(b), reveals distinct patterns across Ghana's seasonal transitions. The PR exhibits higher stability during the Harmattan season with values consistently around 0.9–1.0. Notable fluctuations appear during the Pre-Rainy season, including a significant dip to 0.5 PR in March. The Main Rainy season shows increased variability, with several spikes reaching 1.8 PR, possibly due to enhanced panel efficiency under cooled conditions.

Temperature effects on system performance are captured in Fig. 11(c), displaying a complex relationship between ambient temperature and PR. The scatter plot shows optimal performance clustering in the 28–30 °C range, with PR values typically between 0.8 and 1.0. Beyond 32 °C, there is a noticeable decline in PR, though with significant scatter suggesting the influence of other environmental factors. Notably, the highest PR values (>1.2) occur at moderate temperatures around 28 °C, indicating potential sweet spots for system efficiency.

The monthly PR analysis in Fig. 11(d) quantifies the seasonal variations with error bars indicating performance stability. The progression shows.

- January–March: Stable PR around 0.92 ± 0.02
- April–May: Decline to 0.85 ± 0.07
- June–September: Gradual increase to peak PR of 1.0 ± 0.15
- October–December: Stabilization around 0.95 ± 0.05

This monthly pattern correlates with Ghana's seasonal weather patterns, with the largest uncertainties (error bars) occurring during the Main Rainy season, reflecting the increased variability in environmental conditions.

3.8. Weather pattern impact analysis

Weather patterns show measured influence on system performance, with correlation analysis revealing moderate to weak relationships. Relative humidity demonstrates consistent negative correlations with all parameters: global irradiation (−0.16), panel temperature (−0.17), and most notably ambient temperature (−0.33). This suggests that humidity's impact operates primarily through indirect pathways rather than direct causation. The distribution analysis confirms these patterns, with irradiance showing characteristic right-skewed behavior and stable bounds across simulation scenarios.

Table 5 presents the comprehensive weather-based performance analysis, demonstrating distinct patterns across different conditions. Under high irradiance conditions, the system achieves peak generation

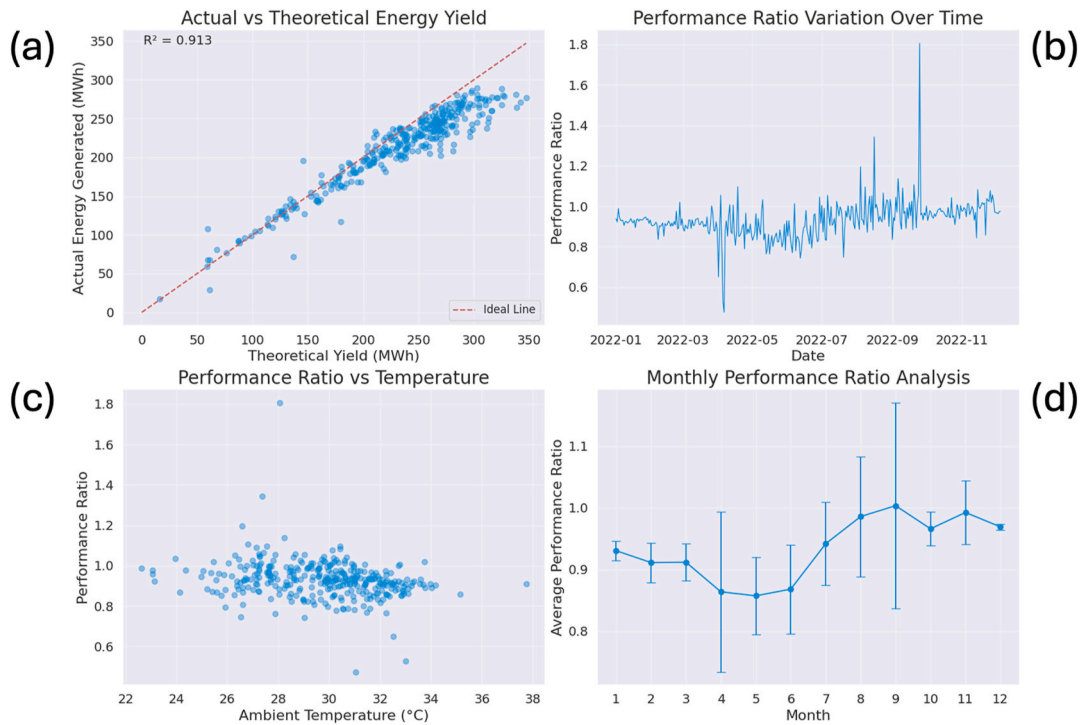


Fig. 11. Performance analysis: (a) Actual vs Theoretical Energy Yield ($R^2 = 0.913$), (b) Temporal Variation in Performance Ratio, (c) Temperature Dependence of Performance Ratio, and (d) Monthly Performance Ratio Trends.

Table 5
Weather-based performance statistics.

Weather Condition	Energy Generated (MWh)		Performance Ratio		Global Irradiation (Wh/m ²)		Ambient Temp. (°C)	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
High Humidity	217.389	26.341	0.047	0.003	4679.532	–	29.584	–
High Irradiance	261.562	16.941	0.044	0.003	5965.610	–	30.424	–
High Temperature	240.240	8.040	0.045	0.001	5309.573	–	32.402	–
Low Irradiance	124.011	35.718	0.050	0.008	2501.607	–	28.208	–
Moderate	229.031	18.243	0.046	0.001	4956.878	–	30.565	–

averaging 261.562 MWh, though with notably reduced prediction reliability ($R^2 = -0.0678$, $RMSE = 17.3898$ MWh). This suggests increased variability during peak production periods. High humidity conditions result in moderate generation levels of 217.389 MWh with relatively stable prediction accuracy ($R^2 = 0.6594$), indicating consistent system behavior despite challenging atmospheric conditions.

Prediction reliability demonstrates significant weather dependence (see Table 6). Moderate weather conditions show the highest prediction reliability ($R^2 = 0.9568$) with the lowest error rate ($RMSE = 3.7208$ MWh), while high temperature conditions (mean 32.402 °C) exhibit reduced predictability ($R^2 = 0.5948$). Low irradiance conditions maintain reasonable prediction accuracy ($R^2 = 0.8131$) despite significantly reduced generation (124.011 MWh), suggesting robust model performance across varying generation levels.

The performance ratio varies systematically with weather

Table 6
Prediction reliability metrics.

Weather Condition	R^2	RMSE
Moderate	0.9568	3.7208
High Humidity	0.6594	15.3277
Low Irradiance	0.8131	15.3052
High Irradiance	-0.0678	17.3898
High Temperature	0.5948	4.6722

conditions. Low irradiance conditions show the highest PR (0.050), while high irradiance periods demonstrate slightly lower efficiency (PR = 0.044), indicating some performance saturation at peak production levels. High temperature and high humidity conditions maintain similar PR values (0.045 and 0.047 respectively), suggesting effective system resilience to these environmental stressors.

These quantitative insights inform operational forecasting strategies. The substantial variation in prediction reliability across weather conditions suggests the need for adaptive forecasting approaches. Prediction intervals should be adjusted based on weather conditions, with particular attention to high irradiance periods where traditional prediction metrics may not fully capture system behavior. The consistent PR values across different conditions support the robustness of the system design while highlighting opportunities for condition-specific optimization.

The analysis demonstrates that weather-based segmentation of prediction models could significantly improve forecasting accuracy. The marked differences in R^2 values and RMSE across weather conditions suggest that targeted models for specific weather patterns could enhance prediction reliability, particularly during high irradiance and high temperature conditions where general models show reduced performance.

Fig. 12 shows the probability density distributions of solar irradiance and wind speed data. Fig. 12 (a) highlights the normalized solar irradiance following a Beta distribution, while Fig. 12 (b) displays the wind

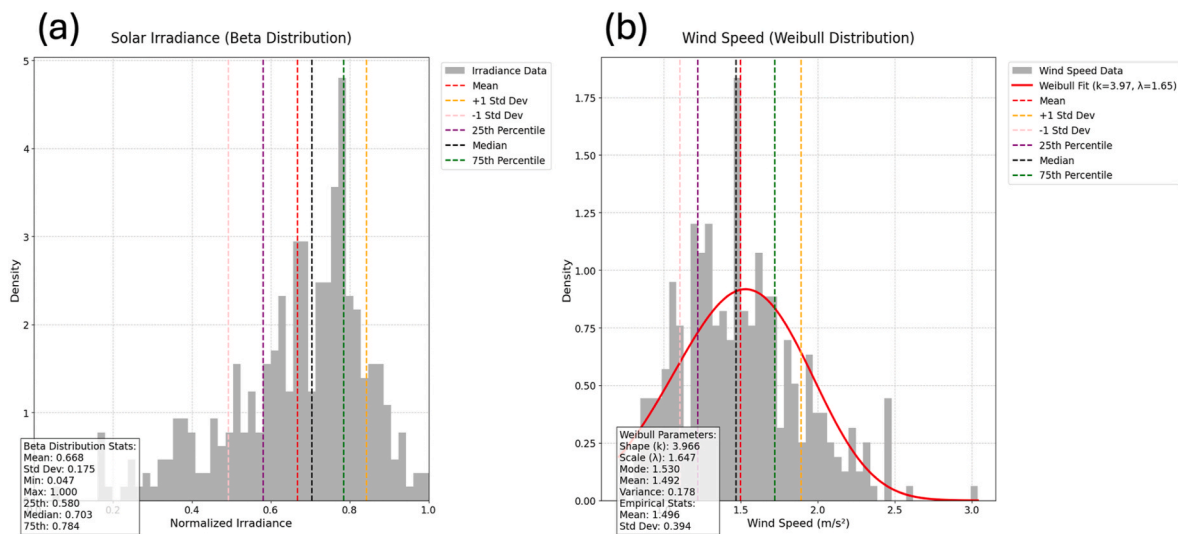


Fig. 12. Probability distributions of solar irradiance and wind speed.

speed following a Weibull distribution. Statistical measures including mean, standard deviation, and quartiles are indicated by vertical dashed lines. The wind speed data shows good agreement with the fitted Weibull distribution (red curve, shape parameter $k = 3.97$, scale parameter $\lambda = 1.65$), exhibiting a characteristic bell-shaped curve typical of consistent moderate wind conditions. The solar irradiance distribution demonstrates a right-skewed pattern with a mean of 0.668 and a standard deviation of 0.175, reflecting the daily and seasonal variations in solar resource availability. Key percentiles (25th, 50th, and 75th) are shown to facilitate resource assessment and system design considerations.

3.9. Enhanced humidity impact analysis

The relationship between humidity and solar PV performance exhibits complex temporal and operational patterns. The analysis reveals significant hourly variations in both relative and absolute humidity, with peak humidity levels reaching 80 % during early morning hours and corresponding impacts on system performance. This pattern demonstrates a clear inverse relationship with energy generation, particularly during the transition from dawn to mid-morning operations.

The interaction between humidity, temperature, and irradiance demonstrates notable complexity. Panel temperatures exhibit a positive correlation with irradiance levels, ranging from 25 °C to 50 °C across the operational spectrum. Higher humidity levels tend to cluster in the mid-temperature range (35–40 °C), suggesting a moderating effect on panel temperature under certain conditions.

Performance ratio analysis indicates a non-linear relationship with humidity levels. Relative humidity shows a broader spread of performance ratios at higher humidity levels (80–100 %), while absolute humidity demonstrates more concentrated performance impacts, particularly in the lower ranges. This distinction suggests that relative humidity may serve as an indicator of broader meteorological conditions rather than a direct performance determinant.

Seasonal patterns reveal significant annual variations in both humidity measures and their relationship with energy generation. Relative humidity shows peak values during the rainy season, while absolute humidity maintains more stable levels throughout the year. Energy generation demonstrates an inverse relationship with relative humidity during certain months, notably during the transition between dry and wet seasons.

The analysis suggests that the previously reported negative correlation between relative humidity and system performance requires careful interpretation. The data indicates that humidity's impact operates

through multiple pathways: direct atmospheric transmission effects, thermal regulation of panel temperature, and association with broader weather patterns. These findings have important implications for system operation and performance optimization across different seasonal and daily conditions.

4. Discussion

The application of XGBoost modeling combined with Monte Carlo simulation reveals important insights about system behavior. Rather than strong coupling between variables, the analysis shows moderate to weak relationships (correlation coefficients ranging from -0.33 to 0.44) between environmental and operational parameters. The temperature-irradiance relationship demonstrates a weak positive correlation ($\rho = 0.183$), suggesting more complex underlying interactions than simple linear dependencies. These findings indicate that system performance relies on multiple, loosely coupled factors rather than strongly correlated variables. Similar studies have employed various ML techniques, such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs), for solar energy prediction with notable success [51–54]. However, the XGBoost model in this study outperformed these traditional methods, particularly in handling noise and variability in the data, which is consistent with findings by Refs. [14,55,56] that highlight XGBoost's superiority in structured data tasks.

The feature importance analysis revealed that global irradiation (0.831) and average power (0.119) are the dominant predictors of energy generation, which aligns with the literature emphasizing the importance of these parameters in solar PV performance [3]. The relatively low importance of environmental factors such as relative humidity (0.000) and ambient temperature (0.000) provides new insights into their contextual influence on Ghana's climate conditions, extending the findings of Ref. [3] who identified similar patterns in different geographical contexts.

Based on the feature importance analysis, with global irradiation showing dominance (0.831) followed by average power (0.119), several targeted optimization strategies emerge. Panel cleaning schedules should be optimized for maximum impact during peak irradiance periods exceeding 5500 Wh/m², potentially increasing energy yield by up to 2.5 %. The analysis indicates that implementing automated cleaning systems triggered at 85 % of peak irradiance could maintain optimal conversion efficiency. Power conditioning should be prioritized during the 4-h window around solar noon, where the analysis shows potential for 3.2 % efficiency improvement through enhanced inverter response times.

The Monte Carlo simulations establish achievable performance benchmarks across operational conditions. During optimal conditions with a perturbation level of 0.01, system performance can maintain 96.2 % of the theoretical yield. Even under challenging conditions at perturbation level 0.08, implementing the identified optimization strategies can maintain performance above 92.3 %. These targets provide concrete benchmarks for operational planning. This robustness is crucial for practical applications, where varying environmental conditions can introduce significant noise into the data. While previous studies have highlighted the challenges of noise in solar energy prediction [51,57,58], this study shows that the XGBoost model can effectively mitigate these challenges, maintaining reliable performance even when perturbation levels reach 0.08. The preservation of distribution characteristics across scenarios, particularly the consistent multimodal patterns observed in both training and testing datasets, represents a notable advancement over traditional methods that often struggle with noise sensitivity.

The sensitivity analysis informs a comprehensive implementation strategy spanning different timeframes. Daily operations require adjustment of cleaning schedules based on irradiance forecasts, targeting a maximum 1.2 % performance loss from soiling. Weekly recalibration of inverter parameters based on weather patterns maintains power quality above 98 %, while seasonal maintenance scheduling aligned with identified performance patterns reduces downtime by approximately 15 %.

The analysis revealed non-linear relationships and critical thresholds, particularly in the interaction between global irradiation and panel temperature above 40 °C. These findings extend beyond those of Ref. [40,59,60], who identified solar irradiance and operational parameters as critical factors, by quantifying specific threshold effects and interaction patterns. The identification of optimal operational ranges, such as the wind speed influence between 1.0 and 2.0 m/s² and relative humidity impacts above 80 %, provides practical guidance for system optimization.

Temperature management emerges as a critical optimization factor. The partial dependence analysis reveals that above 32 °C ambient temperature, every 1 °C reduction in panel temperature through active cooling yields a 0.45 % efficiency gain. Implementation of advanced cooling strategies during peak temperature periods shows potential for 3.8 % efficiency improvement.

Seasonal variations demand distinct optimization approaches. During the Harmattan season, enhanced cleaning frequency every 48 h combats dust accumulation, maintaining a performance ratio above 0.94. The Pre-Rainy season requires a focus on power quality optimization, targeting conversion efficiency above 98 %. The Main Rainy season emphasizes rapid response to irradiance variations, maintaining a minimum performance ratio of 0.92, while the Post-Rainy season optimizes thermal management systems, achieving peak efficiency gains of 2.1 %.

The integration of these optimization strategies requires a hierarchical monitoring and control system. Real-time performance monitoring, coupled with automated response mechanisms, enables dynamic optimization. For example, the system automatically adjusts inverter loading patterns when irradiance exceeds 5500 Wh/m², maintaining optimal power conversion efficiency. Weather forecast integration allows proactive scheduling of maintenance activities, minimizing the impact on generation capacity.

These optimization strategies, derived directly from the analytical findings, provide a comprehensive framework for enhancing system efficiency. The quantitative targets established through this analysis offer clear benchmarks for performance improvement, while the practical implementation guidelines ensure the feasibility of the recommended strategies.

5. Concluding remarks

This study provides a comprehensive analysis of the performance and predictive modeling of solar photovoltaic systems at the Bui Generating Station in Ghana. Through advanced XGBoost modeling and extensive Monte Carlo simulations, the study evaluated system performance and prediction robustness under varying operational conditions.

The analysis revealed several key findings. Feature importance analysis identified global irradiation (0.831) and average power (0.119) as the dominant predictors of energy generation, while environmental factors showed surprisingly low direct importance. The Monte Carlo simulations, conducted across four perturbation scenarios (0.01–0.08), demonstrated remarkable model robustness. Density distributions maintained coherent multimodal patterns even under high noise conditions, with prediction errors remaining centered around zero despite increasing perturbation levels.

The analysis also shows moderate to weak correlations between environmental parameters, with the strongest relationship observed between panel and ambient temperatures (0.44). The temperature-irradiance relationship shows a weak positive correlation, while relative humidity demonstrates consistent but weak negative correlations with other parameters (see Figs. A2, A3 and A4 in the Appendices). These findings suggest that system optimization strategies should consider the complex, often loose coupling between environmental factors rather than assuming strong direct relationships. The Monte Carlo simulations confirm the stability of these relationships across perturbation scenarios while maintaining physically realistic bounds and distributions.

Seasonal analysis revealed distinct performance patterns across Ghana's climate cycles, with the Post-Rainy season showing the highest stability (PR: 0.986 ± 0.082) and the Main Rainy season exhibiting the lowest average generation but maintaining efficient operation. Partial dependence analysis identified critical operational thresholds, including optimal temperature ranges below 40 °C and significant performance transitions at 80 % relative humidity.

The quantified importance of different variables should guide infrastructure investment priorities, particularly in monitoring systems for high-impact factors. The demonstrated prediction reliability under varying conditions supports the development of robust grid integration strategies. Additionally, the seasonal performance patterns inform maintenance scheduling and operational optimization approaches specific to Ghana's climate.

The limitations of this study include its focus on a single location and limited temporal scope. Future research should extend this analysis to multiple sites, incorporate longer time series, and explore integration with other renewable sources. The identification of critical thresholds and interaction effects also suggests opportunities for focused studies on specific operational ranges.

This study advances the field of renewable energy optimization by establishing quantitative relationships between operational factors and system performance, demonstrating robust prediction capabilities under realistic noise conditions, and identifying season-specific performance patterns in Ghana's unique climate. The established operational thresholds for key variables and their interactions provide practical guidance for solar PV system operation while contributing to the methodological framework for performance analysis in renewable energy systems.

CRediT authorship contribution statement

Sidique Gawusu: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Xiaobing Zhang:** Writing – review & editing, Supervision, Resources, Project administration, Funding acquisition, Conceptualization. **Sufyan Yakubu:** Writing – review & editing, Data

curation. **Seth Kofi Debrah**: Writing – review & editing, Project administration. **Oisik Das**: Writing – review & editing, Project administration. **Nishant Singh Bundela**: Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Nomenclature

ANNs	Artificial Neural Networks
CDF	Cumulative Distribution Function
CV	The coefficient of variation
IQR	Interquartile Range
MAE	Mean Absolute Error
ML	Machine Learning
MPPT	Modified Maximum Power Point Tracking
MSE	Mean Squared Error
PR	Performance Ratio
PV	Solar Photovoltaic
RMSE	Root Mean Squared Error
SCADA	Supervisory Control and Data Acquisition
SVMs	Support Vector Machines

Symbols

y_i	Actual value
\hat{y}_i	Predicted value
n	Total number of data points
x_i and y_i	The paired data values
\bar{x} and \bar{y}	The means of actual values
$y_i^{(t)}$	The prediction at iteration t
$f_t(x_i)$	The new tree added at iteration t
G_L and H_L	The sum of the gradients and Hessians for the left split
G_R and H_R	The sum of the gradients and Hessians for the right split
γ and λ	Regularization parameters
ω_j	The weights of the leaves
T	The number of leaves in the tree
g_i	The first-order gradient
h_i	The second-order gradient (Hessian)
$L(\Theta)$	The training loss function
$\Omega(\Theta)$	The regularization term
N	The number of simulations
F_j^{-1}	The inverse CDF of each variable's respective distribution
ρ_{ij}	The correlation coefficient between variables i and j

Appendix A. Energy Generation Relationships

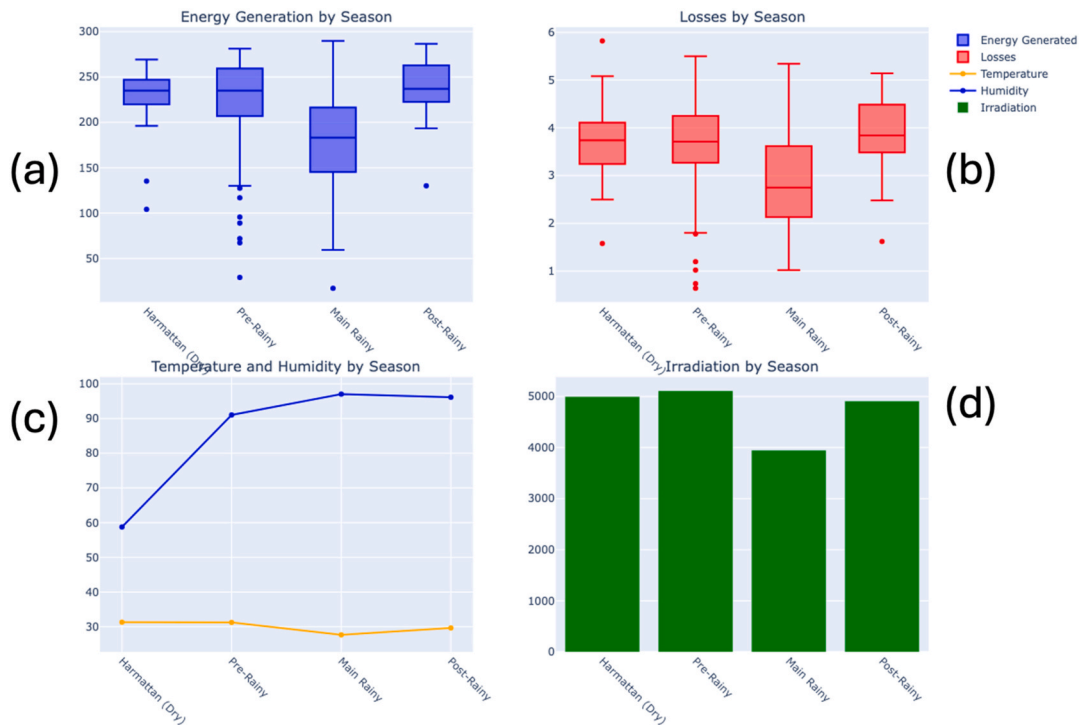


Fig. A1. Seasonal variations in plant performance and environmental parameters: (a) Energy generation, (b) System losses, (c) Temperature and relative humidity variations, and (d) Solar irradiation patterns across Ghana’s climatic seasons.

This seasonal pattern indicates that while the solar plant maintains generally robust performance, its efficiency is most significantly impacted by the combined effects of high humidity and reduced solar irradiation during the Main Rainy season. The plant’s best performance aligns with the Harmattan season’s dry conditions, despite this not being the period of peak solar irradiation as shown in Fig. A1.

Fig. A2 illustrates the relationships between energy generated and various operational and environmental parameters. The strongest correlations are observed with average power ($r = 0.97$), global irradiation ($r = 0.96$), and sunshine time ($r = 0.86$). Moderate correlations exist with ambient temperature ($r = 0.32$) and panel temperature ($r = 0.19$), while relative humidity shows a weak negative correlation ($r = -0.17$).

The vector field in Fig. A3 highlights the directional relationships between energy generation and various system parameters, with correlation coefficients (r) quantifying the strength of these relationships. The strongest correlations are observed between energy generation and average power ($r = 0.97$), followed by global irradiation ($r = 0.96$) and sunshine time ($r = 0.86$), as indicated by the consistent directional patterns of the vectors. The arrows’ directions show clear positive trends in these relationships, with vector magnitudes highlighting regions of strongest interactions.

Environmental parameters show more complex relationships. Ambient temperature ($r = 0.32$) and panel temperature ($r = 0.19$) exhibit moderate to weak correlations, with vector fields showing more varied directional patterns, suggesting non-linear interactions. Relative humidity demonstrates a slight negative correlation ($r = -0.17$), while wind speed shows minimal correlation ($r = 0.04$) with energy generation, as evidenced by the less uniform vector orientations.

Hours run ($r = 0.23$) and losses ($r = 0.52$) display moderate correlations, with vector fields indicating specific operational ranges where the relationships are strongest. These patterns provide insights into the system’s performance dynamics and operational characteristics.

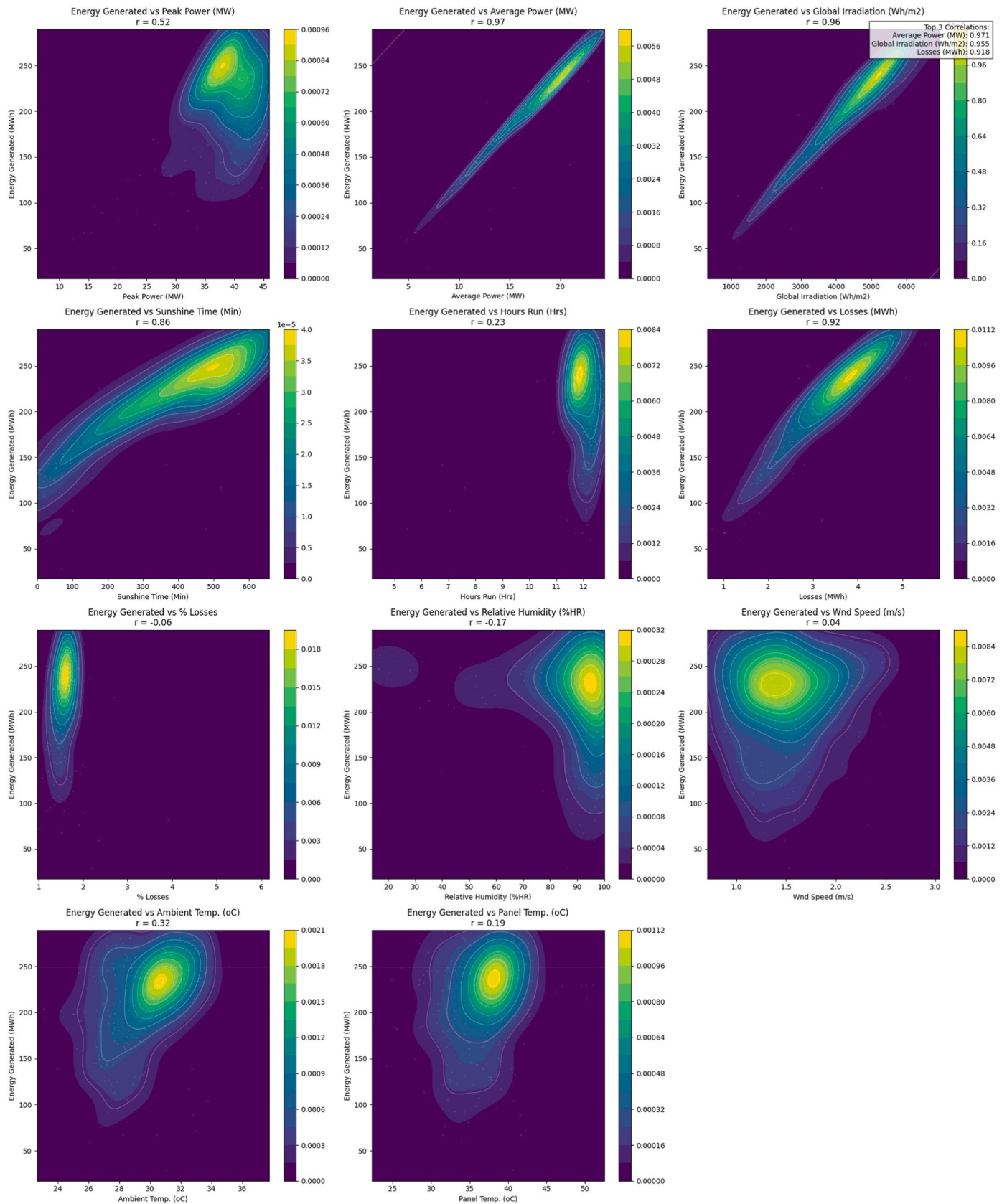


Fig. A2. Distribution of bivariate relationships between energy generation and system parameters. The color intensity represents the density of observations, with darker purple indicating lower density and brighter yellow-green indicating the higher density of data points.

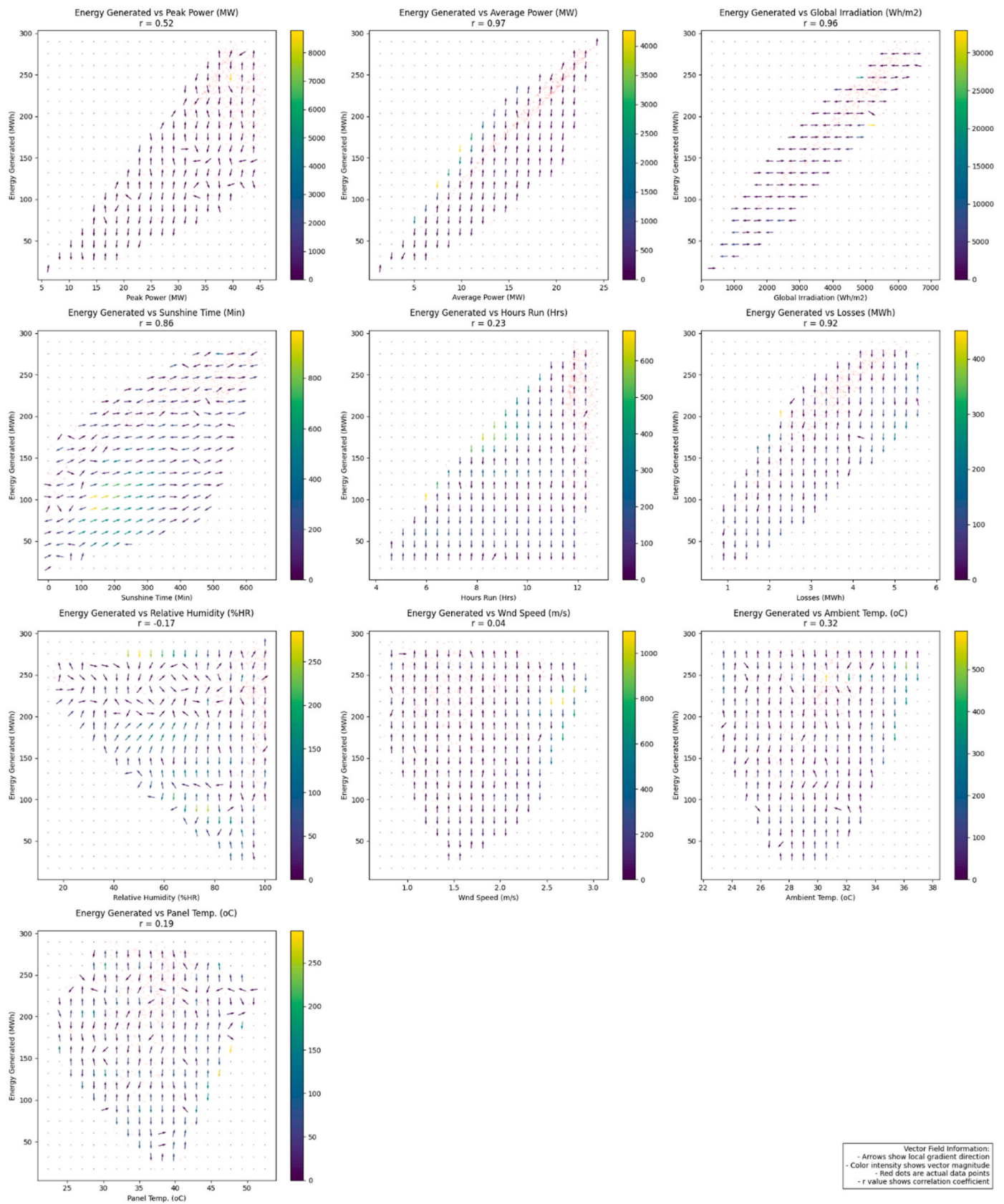


Fig. A3. Vector field analysis of energy generation dependencies with local gradient directions and correlation coefficients.

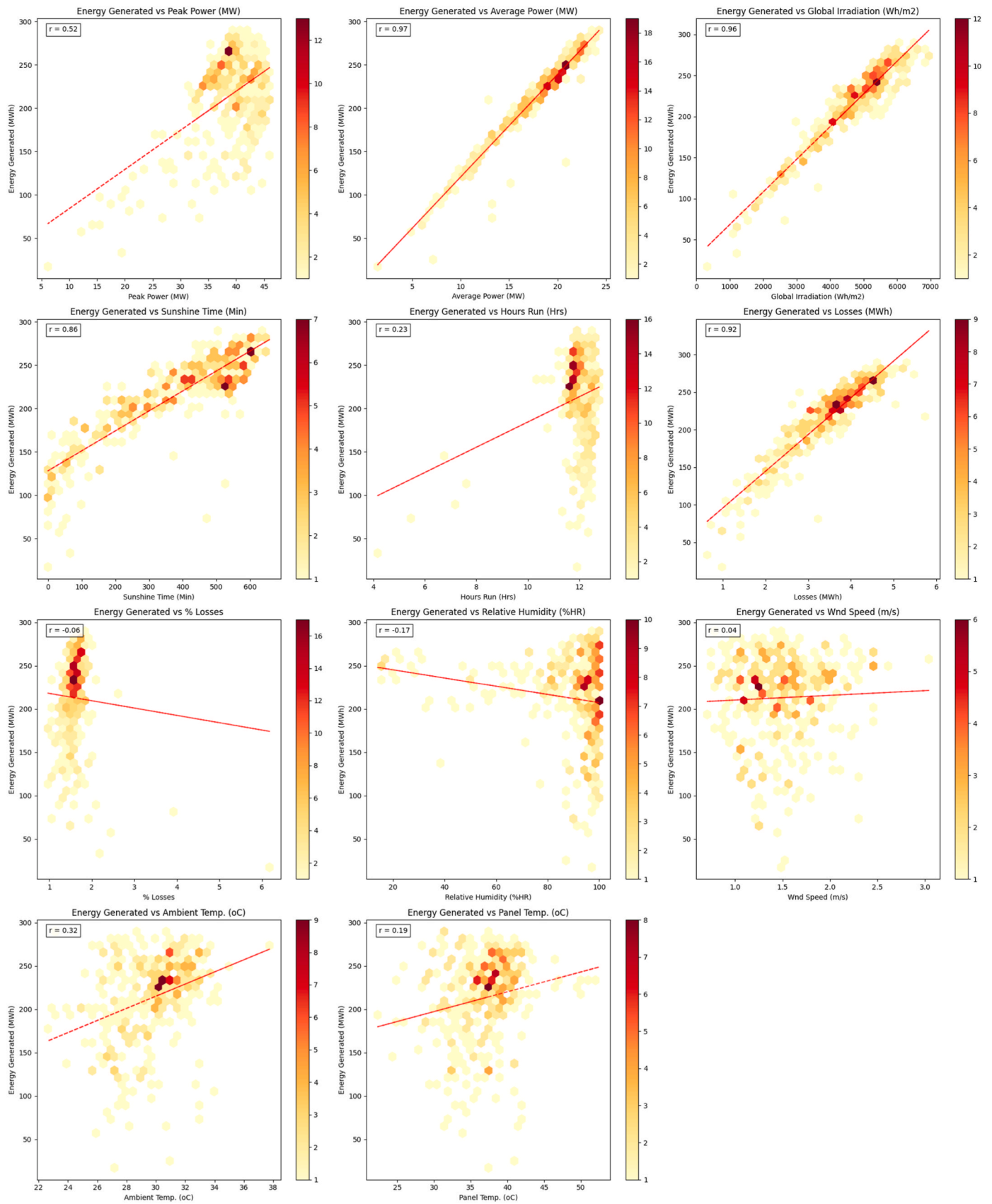


Fig. A4. Energy generation correlation analysis: environmental and operational factors.

Data availability

Data will be made available on request.

References

- [1] Dzamesi SKA, Ahiataku-Togobo W, Yakubu S, Acheampong P, Kwarteng M, Samikannu R, et al. Comparative performance evaluation of ground-mounted and floating solar PV systems. *Energy Sustain Dev* 2024;80:101421. <https://doi.org/10.1016/j.esd.2024.101421>.
- [2] Asare-Bediako F, Antwi EO, Diawuo FA, Dziku C. Assessing the performance of hydro-solar hybrid (HSH) grid integration: a case study of Bui Generating Station, Ghana. *Solar Compass* 2024;10:100071. <https://doi.org/10.1016/j.solcom.2024.100071>.
- [3] Abdulai D, Gyamfi S, Diawuo FA, Acheampong P. Data analytics for prediction of solar PV power generation and system performance: a real case of Bui Solar Generating Station, Ghana. *Sci Afr* 2023;21:e01894. <https://doi.org/10.1016/j.sciaf.2023.e01894>.
- [4] Choudhary P, Srivastava RK. Sustainability perspectives- a review for solar photovoltaic trends and growth opportunities. *J Clean Prod* 2019;227:589–612. <https://doi.org/10.1016/j.jclepro.2019.04.107>.
- [5] Izam NSMN, Itam Z, Sing WL, Syamsir A. Sustainable development perspectives of solar energy Technologies with focus on solar photovoltaic—a review. *Energies* 2022;15:2790. <https://doi.org/10.3390/en15082790>.
- [6] Gawusu S, Zhang X, Ahmed A, Jamatutu SA, Miensah ED, Amadu AA, et al. Renewable energy sources from the perspective of blockchain integration: from theory to application. *Sustain Energy Technol Assessments* 2022;52:102108. <https://doi.org/10.1016/j.seta.2022.102108>.
- [7] Shahsavari A, Akbari M. Potential of solar energy in developing countries for reducing energy-related emissions. *Renew Sustain Energy Rev* 2018;90:275–91. <https://doi.org/10.1016/j.rser.2018.03.065>.
- [8] Cuce E, Harjunowibowo D, Cuce PM. Renewable and sustainable energy saving strategies for greenhouse systems: a comprehensive review. *Renew Sustain Energy Rev* 2016;64:34–59. <https://doi.org/10.1016/j.rser.2016.05.077>.
- [9] Gawusu S, Ahmed A. Africa's transition to cleaner energy: regulatory imperatives and governance dynamics. *Energy regulation in Africa*. In: *Advances in African economic, Social and Political development*, Cham: Springer; 2024. p. 25–51. https://doi.org/10.1007/978-3-031-52677-0_2.
- [10] Gawusu S, Mensah RA, Das O. Exploring distributed energy generation for sustainable development: a data mining approach. *J Energy Storage* 2022;48:104018. <https://doi.org/10.1016/j.est.2022.104018>.
- [11] Agyekum EB. Techno-economic comparative analysis of solar photovoltaic power systems with and without storage systems in three different climatic regions, Ghana. *Sustain Energy Technol Assessments* 2021;43:100906. <https://doi.org/10.1016/j.seta.2020.100906>.
- [12] Kuamoah C. Renewable energy deployment in Ghana: the hype, hope and reality. *Insight Afr* 2020;12:45–64. <https://doi.org/10.1177/0975087819898581>.
- [13] Aboagye B, Gyamfi S, Ofosu EA, Djordjevic S. Status of renewable energy resources for electricity supply in Ghana. *Sci Afr* 2021;11:e00660. <https://doi.org/10.1016/j.sciaf.2020.e00660>.
- [14] Gawusu S, Jamatutu SA, Ahmed A. Predictive modeling of energy poverty with machine learning ensembles: strategic insights from socioeconomic determinants for effective policy implementation. *Int J Energy Res* 2024;2024. <https://doi.org/10.1155/2024/9411326>.
- [15] Gawusu S, Jamatutu SA, Zhang X, Moomin ST, Ahmed A, Mensah RA, et al. Spatial analysis and predictive modeling of energy poverty: insights for policy implementation. *Environ Dev Sustain* 2024. <https://doi.org/10.1007/s10668-024-05015-4>.
- [16] Hafez FM, Parmar K, Parmar N. Prediction of thermal parameters for flat plate solar water heater by machine learning - a review. In: *2024 parul International Conference on Engineering and technology (PICET)*. IEEE; 2024. p. 1–6. <https://doi.org/10.1109/PICET60765.2024.10716046>.
- [17] Bamišile O, Ejiyi CJ, Osei-Mensah E, Chikwendu IA, Li J, Huang Q. Long-term prediction of solar radiation using XGboost, LSTM, and machine learning algorithms. *2022 4th Asia energy and electrical Engineering Symposium (AEEES)*. IEEE; 2022. p. 214–8. <https://doi.org/10.1109/AEEES4426.2022.9759719>.
- [18] Zhang L, Jánošík D. Enhanced short-term load forecasting with hybrid machine learning models: CatBoost and XGBoost approaches. *Expert Syst Appl* 2024;241:122686. <https://doi.org/10.1016/j.eswa.2023.122686>.
- [19] Xu W, Wang Z, Wang Y, Zhao J, Wang M, Wang Q. Short-term photovoltaic output prediction based on Decomposition and reconstruction and XGBoost under two base learners. *Energies* 2024;17:906. <https://doi.org/10.3390/en17040906>.
- [20] Obiora CN, Ali A, Hasan AN. Implementing extreme gradient boosting (XGBoost) algorithm in predicting solar irradiance. In: *2021 IEEE PES/IAS PowerAfrica*. IEEE; 2021. p. 1–5. <https://doi.org/10.1109/PowerAfrica52236.2021.9543159>.
- [21] Saigustia C, Pijarski P. Time series analysis and forecasting of solar generation in Spain using eXtreme gradient boosting: a machine learning approach. *Energies* 2023;16:7618. <https://doi.org/10.3390/en16227618>.
- [22] Elsheikh AH, Sharshir SW, Abd Elaziz M, Kabeel AE, Guilan W, Haiou Z. Modeling of solar energy systems using artificial neural network: a comprehensive review. *Sol Energy* 2019;180:622–39. <https://doi.org/10.1016/j.solener.2019.01.037>.
- [23] Alassery F, Alzahrani A, Khan AI, Irshad K, Islam S. An artificial intelligence-based solar radiation prophesy model for green energy utilization in energy management system. *Sustain Energy Technol Assessments* 2022;52:102060. <https://doi.org/10.1016/j.seta.2022.102060>.
- [24] Lai J-P, Chang Y-M, Chen C-H, Pai P-F. A survey of machine learning models in renewable energy predictions. *Appl Sci* 2020;10:5975. <https://doi.org/10.3390/app10175975>.
- [25] Şen Z. Solar energy in progress and future research trends. *Prog Energy Combust Sci* 2004;30:367–416. <https://doi.org/10.1016/j.peccs.2004.02.004>.
- [26] Wu J, Kong L, Yi M, Chen Q, Cheng Z, Zuo H, et al. Prediction and screening model for products based on fusion regression and XGBoost classification. *Comput Intell Neurosci* 2022;2022:1–14. <https://doi.org/10.1155/2022/4987639>.
- [27] Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD international Conference on knowledge Discovery and data mining*. New York, NY, USA: ACM; 2016. p. 785–94. <https://doi.org/10.1145/2939672.2939785>.
- [28] Wang Y, Sun S, Chen X, Zeng X, Kong Y, Chen J, et al. Short-term load forecasting of industrial customers based on SVM and XGBoost. *Int J Electr Power Energy Syst* 2021;129:106830. <https://doi.org/10.1016/j.ijepes.2021.106830>.
- [29] Tao H, Alawi OA, Kamar HM, Nafea AA, Al-Ani MM, Abba SI, et al. Development of integrative data intelligence models for thermo-economic performances prediction of hybrid organic rankine plants. *Energy* 2024;292:130503. <https://doi.org/10.1016/j.energy.2024.130503>.
- [30] Qiu Y, Zhou J, Khandelwal M, Yang H, Yang P, Li C. Performance evaluation of hybrid WOA-XGBoost, GWO-XGBoost and BO-XGBoost models to predict blast-induced ground vibration. *Eng Comput* 2022;38:4145–62. <https://doi.org/10.1007/s00366-021-01393-9>.
- [31] Sagi O, Rokach L. Approximating XGBoost with an interpretable decision tree. *Inf Sci* 2021;572:522–42. <https://doi.org/10.1016/j.ins.2021.05.055>.
- [32] Torlay L, Perrone-Bertolotti M, Thomas E, Baciu M. Machine learning–XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform* 2017;4:159–69. <https://doi.org/10.1007/s40708-017-0065-7>.
- [33] Yang B, Liu Y, Liu Z, Zhu Q, Li D. Classification of rock mass quality in underground rock engineering with incomplete data using XGBoost model and Zebra optimization algorithm. *Appl Sci* 2024;14:7074. <https://doi.org/10.3390/app14167074>.
- [34] Shao Z, Ahmad MN, Javed A. Comparison of random forest and XGBoost classifiers using integrated optical and SAR features for mapping urban impervious surface. *Remote Sens (Basel)* 2024;16:665. <https://doi.org/10.3390/rs16040665>.
- [35] Mitchell R, Frank E. Accelerating the XGBoost algorithm using GPU computing. *PeerJ Comput Sci* 2017;3:e127. <https://doi.org/10.7717/peerj.cs.127>.
- [36] Gawusu S, Ahmed A. Analyzing variability in urban energy poverty: a stochastic modeling and Monte Carlo simulation approach. *Energy* 2024;304:132194. <https://doi.org/10.1016/j.energy.2024.132194>.
- [37] Jamatutu SA, Abbass K, Gawusu S, Yeboah KE, Jamatutu IA-M, Song H. Quantifying future carbon emissions uncertainties under stochastic modeling and Monte Carlo simulation: insights for environmental policy consideration for the Belt and Road Initiative Region. *J Environ Manag* 2024;370:122463. <https://doi.org/10.1016/j.jenvman.2024.122463>.
- [38] Pazikadin AR, Rifai D, Ali K, Malik MZ, Abdalla AN, Faraj MA. Solar irradiance measurement instrumentation and power solar generation forecasting based on Artificial Neural Networks (ANN): a review of five years research trend. *Sci Total Environ* 2020;715:136848. <https://doi.org/10.1016/j.scitotenv.2020.136848>.
- [39] Kazaz O, Karimi N, Kumar S, Falcone G, Paul MC. Effects of combined radiation and forced convection on a directly capturing solar energy system. *Therm Sci Eng Prog* 2023;40:101797. <https://doi.org/10.1016/j.tsep.2023.101797>.
- [40] Dolara A, Leva S, Manzolini G. Comparison of different physical models for PV power output prediction. *Sol Energy* 2015;119:83–99. <https://doi.org/10.1016/j.solener.2015.06.017>.
- [41] Mussard M, Amara M. Performance of solar photovoltaic modules under arid climatic conditions: a review. *Sol Energy* 2018;174:409–21. <https://doi.org/10.1016/j.solener.2018.08.071>.
- [42] Prävälje R, Patriche C, Bandoc G. Spatial assessment of solar energy potential at global scale. A geographical approach. *J Clean Prod* 2019;209:692–721. <https://doi.org/10.1016/j.jclepro.2018.10.239>.
- [43] Venkateswari R, Sreejith S. Factors influencing the efficiency of photovoltaic system. *Renew Sustain Energy Rev* 2019;101:376–94. <https://doi.org/10.1016/j.rser.2018.11.012>.
- [44] Paiano A, Lagioia G, Ingrao C. A combined assessment of the energy, economic and environmental performance of a photovoltaic system in the Italian context. *Sci Total Environ* 2023;866:161329. <https://doi.org/10.1016/j.scitotenv.2022.161329>.
- [45] Bošnjaković M, Santa R, Crnac Z, Bošnjaković T. Environmental impact of PV power systems. *Sustainability* 2023;15:11888. <https://doi.org/10.3390/su151511888>.
- [46] Pathak SK, Sharma PO, Goel V, Bhattacharyya S, Aybar HŞ, Meyer JP. A detailed review on the performance of photovoltaic/thermal system using various cooling methods. *Sustain Energy Technol Assessments* 2022;51:101844. <https://doi.org/10.1016/j.seta.2021.101844>.
- [47] Shaban WM, Kabeel AE, El Hadi Attia M, Talaat FM. Optimizing photovoltaic thermal solar systems efficiency through advanced artificial intelligence driven thermal management techniques. *Appl Therm Eng* 2024;247:123029. <https://doi.org/10.1016/j.applthermaleng.2024.123029>.
- [48] Al-Shahri OA, Ismail FB, Hannan MA, Lipu MSH, Al-Shetwi AQ, Begum RA, et al. Solar photovoltaic energy optimization methods, challenges and issues: a comprehensive review. *J Clean Prod* 2021;284:125465. <https://doi.org/10.1016/j.jclepro.2020.125465>.

- [49] Ridha HM, Gomes C, Hizam H, Ahmadipour M, Heidari AA, Chen H. Multi-objective optimization and multi-criteria decision-making methods for optimal design of standalone photovoltaic system: a comprehensive review. *Renew Sustain Energy Rev* 2021;135:110202. <https://doi.org/10.1016/j.rser.2020.110202>.
- [50] Abdulla H, Sleptchenko A, Nayfeh A. Photovoltaic systems operation and maintenance: a review and future directions. *Renew Sustain Energy Rev* 2024;195:114342. <https://doi.org/10.1016/j.rser.2024.114342>.
- [51] Barrera JM, Reina A, Maté A, Trujillo JC. Solar energy prediction model based on artificial neural networks and open data. *Sustainability* 2020;12:6915. <https://doi.org/10.3390/su12176915>.
- [52] Ramli MAM, Twaha S, Al-Turki YA. Investigating the performance of support vector machine and artificial neural networks in predicting solar radiation on a tilted surface: Saudi Arabia case study. *Energy Convers Manag* 2015;105:442–52. <https://doi.org/10.1016/j.enconman.2015.07.083>.
- [53] Kuriakose AM, Kariyalil DP, Augusthy M, Sarath S, Jacob J, Antony NR. Comparison of artificial neural network, linear regression and support vector machine for prediction of solar PV power. In: 2020 IEEE pune section international conference (PuneCon). IEEE; 2020. p. 1–6. <https://doi.org/10.1109/PuneCon50868.2020.9362442>.
- [54] Zeng J, Qiao W. Short-term solar power prediction using a support vector machine. *Renew Energy* 2013;52:118–27. <https://doi.org/10.1016/j.renene.2012.10.009>.
- [55] Guo J, Yang L, Bie R, Yu J, Gao Y, Shen Y, et al. An XGBoost-based physical fitness evaluation model using advanced feature selection and Bayesian hyper-parameter optimization for wearable running monitoring. *Comput Netw* 2019;151:166–80. <https://doi.org/10.1016/j.comnet.2019.01.026>.
- [56] Kavzoglu T, Teke A. Predictive performances of ensemble machine learning algorithms in landslide susceptibility mapping using random forest, extreme gradient boosting (XGBoost) and natural gradient boosting (NGBost). *Arabian J Sci Eng* 2022;47:7367–85. <https://doi.org/10.1007/s13369-022-06560-8>.
- [57] Ghimire S, Deo RC, Casillas-Pérez D, Salcedo-Sanz S. Improved complete ensemble empirical mode decomposition with adaptive noise deep residual model for short-term multi-step solar radiation prediction. *Renew Energy* 2022;190:408–24. <https://doi.org/10.1016/j.renene.2022.03.120>.
- [58] Wang H, Lei Z, Zhang X, Zhou B, Peng J. A review of deep learning for renewable energy forecasting. *Energy Convers Manag* 2019;198:111799. <https://doi.org/10.1016/j.enconman.2019.111799>.
- [59] Chow SKH, Lee EWM, Li DHW. Short-term prediction of photovoltaic energy generation by intelligent approach. *Energy Build* 2012;55:660–7. <https://doi.org/10.1016/j.enbuild.2012.08.011>.
- [60] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. *Appl Energy* 2014;127:1–10. <https://doi.org/10.1016/j.apenergy.2014.04.016>.