

Comparative Machine Learning Bias Correction of Satellite-Derived Solar Irradiance at Tropical Stations in Central Java, Indonesia

Eko Riyanto¹, Iwan Setiawan², Uky Yudatama³

¹School of Postgraduate Studies, Diponegoro University, Semarang, Indonesia

²Department of Electrical Engineering, Diponegoro University, Semarang, Indonesia

³Department of Informatics, Universitas Muhammadiyah Magelang, Magelang, Indonesia

Received:

June 24, 2026

Revised:

July 19, 2026

Accepted:

September 2, 2026

Published:

October 24, 2026

Corresponding Author:

Author Name*:

Eko Riyanto

Email*:

Ekoriyanto89@gmail.com

DOI:

10.63158/journalisi.v8i5.1778

© 2026 Journal of Information Systems and Informatics. This open access article is distributed under a (CC-BY License)



Abstract. Satellite-derived Global Horizontal Irradiance products carry systematic bias in tropical regions, so site-level correction is a prerequisite for their use where ground networks are sparse. This study compares bias-correction methods for NASA POWER daily irradiance in Central Java, Indonesia, and asks how much reported accuracy reflects algebraic dependence between predictors and the ground reference. Ten PUSDATARU stations were analysed over 2020–2024 (17,072 station-days; 8,459 from five exactly co-located stations). Random Forest, XGBoost, LightGBM, Elastic Net, and Quantile Mapping were trained chronologically (2020–2022/2023/2024) against persistence and climatology baselines, with preprocessing fitted on training data only, paired significance tests, leave-one-station-out validation, and SHAP attribution. Random Forest reduced RMSE from 3.398 to 2.674 MJ m⁻² day⁻¹ (21.3%), lowered mean bias from 0.988 to 0.221, and raised R² from 0.266 to 0.546; across ten stations XGBoost reduced error from 5.921 to 2.831. Differences among tree ensembles were not statistically robust, and the weighted ensemble gave no gain at the highest inference cost. Adding ground-observed sunshine as a predictor raised the coefficient of determination to 0.9994, identified as recovery of the Ångström–Prescott expression defining the reference (70.9% attribution). Near-perfect accuracies against sunshine-derived references therefore measure leakage, not skill; correction also failed to transfer spatially, motivating sub-degree geostationary products as the next step.

Keywords: Bias Correction, Global Horizontal Irradiance, NASA POWER, Site Adaptation, Tropical Climate

1. INTRODUCTION

Accurate estimation of Global Horizontal Irradiance (GHI) underpins photovoltaic yield forecasting, dispatch planning, and renewable energy integration [1], [2]. Where ground networks are sparse, satellite-derived products such as NASA POWER are often the only continuous source of irradiance information, so their accuracy directly conditions the quality of feasibility studies and grid planning decisions [3], [4].

The problem is acute in the tropics. Rapid convective cloud development, high atmospheric moisture, and elevated aerosol loading degrade satellite retrievals and produce systematic departures from surface conditions [2], [5], [6] [2]. Validation studies for Southeast Asia report that NASA POWER daily GHI carries a positive mean bias of the order of one megajoule per square metre per day and a relative RMSE approaching twenty per cent, with the largest departures during the wet season and under broken cloud [3], [7], [8]. Correction at the site level is therefore a prerequisite for operational use.

Bias correction is formulated here as a supervised regression problem in which a model f maps a vector of satellite-derived and engineered predictors $\mathbf{x}_{s,t}$ for station s on day t onto the ground-based reference irradiance:

$$\widehat{GHI}_{s,t}^{corr} = f(\mathbf{x}_{s,t}) \quad (1)$$

As shown in Equation (1), the models predict corrected GHI directly rather than the residual $GHI_{s,t}^{ref} - GHI_{s,t}^{POWER}$, so that the tree-based learners, the regularized linear model, and the quantile-mapping benchmark are compared on an identical target. This is a site-adaptation study. The objective is to correct satellite GHI at instrumented tropical locations and, through leave-one-station-out testing, to establish whether such a correction transfers to sites that contributed no training data. The design does not support claims of continuous regional correction, and none are made.

Machine learning has been applied widely to this problem. Random Forest reduces variance through bootstrap aggregation and handles correlated meteorological inputs robustly [9], [10], [11]. Gradient boosting fits trees sequentially against residual error and

has outperformed both bagging and linear methods in several irradiance studies [1], [6], [12], [13], with LightGBM offering leaf-wise growth at lower computational cost [14]. Elastic Net addresses multicollinearity through combined L1 and L2 penalties [15], [16], although its linear structure limits its capacity to represent cloud-driven nonlinearity. Quantile Mapping remains the standard statistical benchmark, matching the cumulative distribution of the satellite product to that of the reference [17], [18], [19], but applies a static transformation irrespective of atmospheric state.

Three gaps motivate this study. First, most published comparisons evaluate one or two correction methods, so the relative behaviour of bagging, boosting, regularized regression, and distribution matching under a single protocol is not well established [17], [20]. Second, most site-adaptation work has been conducted in temperate or arid climates [15], [16], [19]; evidence from equatorial Indonesia remains limited [5], [21], [22], and studies that do report tropical results rarely test spatial transfer. Third, reported accuracies in this literature are seldom accompanied by leakage diagnostics, even though the reference irradiance series used in many tropical studies are themselves derived from sunshine duration.

Therefore, this research compares five bias-correction approaches for NASA POWER daily GHI at PUSDATARU stations in Central Java under a common chronological protocol. The objectives are: (i) to quantify the accuracy of Random Forest, XGBoost, LightGBM, Elastic Net, and Quantile Mapping against the uncorrected product and against persistence and climatology baselines; (ii) to test whether a weighted-average ensemble improves upon the best individual model; (iii) to determine whether a fitted correction transfers to stations excluded from training; and (iv) to quantify how much of the accuracy reported in this class of study is attributable to algebraic dependence between predictors and the reference series.

2. METHODS

2.1. Study Sites and Data

Figure 1 summarizes the eight-stage workflow followed in this study, from raw data acquisition through the diagnostic checks that motivate its main conclusions. Each stage is described in detail in the corresponding subsection.

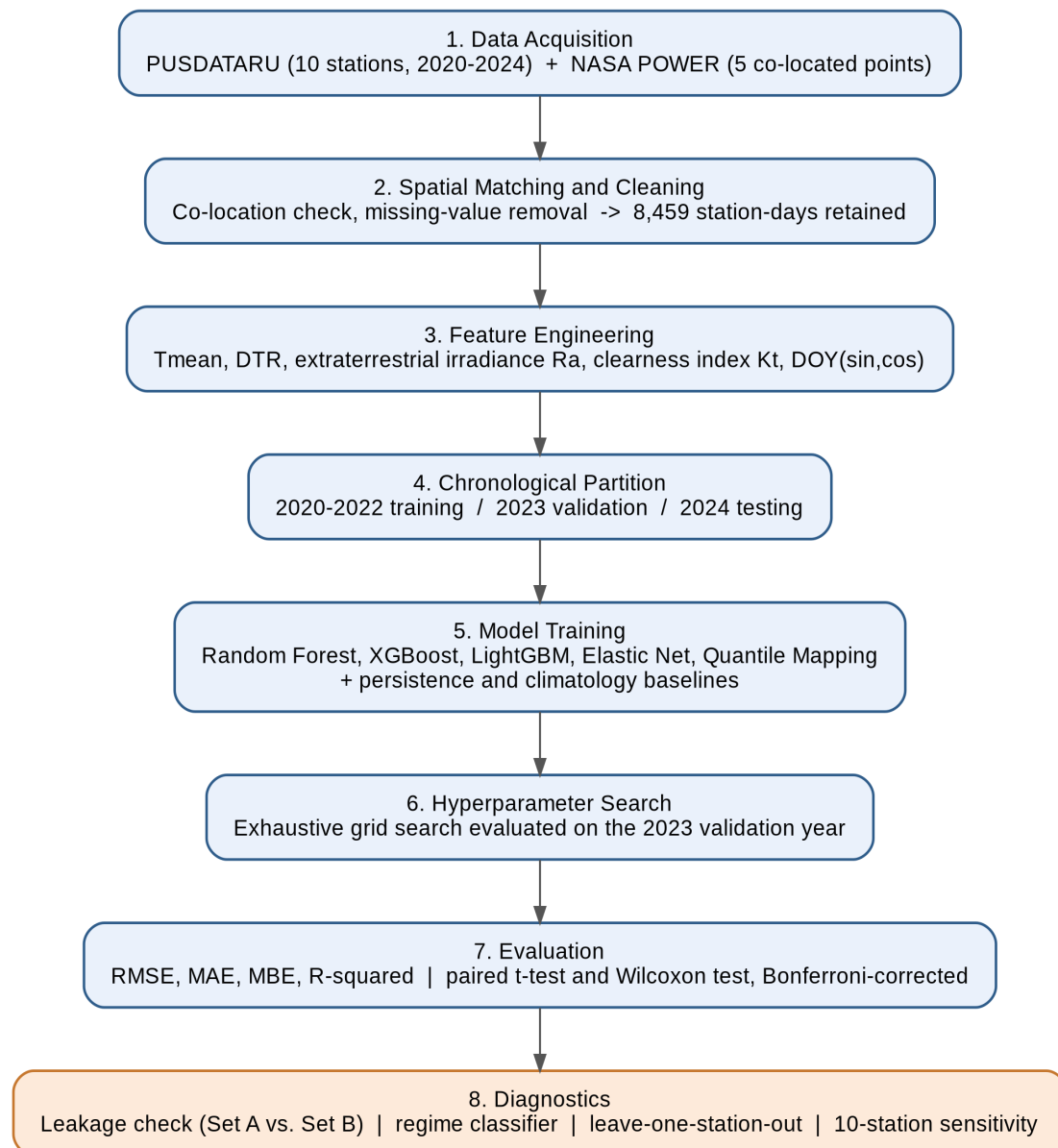


Figure 1. Overview of the research workflow, from data acquisition through diagnostic analysis

Ground observations were obtained from ten climatological stations operated by PUSDATARU (Dinas Pekerjaan Umum Sumber Daya Air dan Penataan Ruang), Province of Central Java. NASA POWER daily data were retrieved for five point locations whose coordinates coincide exactly with five of these stations. The remaining five stations lie 0.16° to 0.52° from the nearest retrieved point.

The analysis is structured accordingly. The primary analysis (below) uses the five exactly co-located stations, so that the reported correction accuracy is not confounded by spatial mismatch between the station and the satellite retrieval point. A sensitivity analysis ("Sensitivity to Station–Grid Offset," below) repeats the evaluation across all ten stations using nearest-point matching, which tests how the conclusions behave when the station and the retrieval point are separated by up to 58 km, as listed in Table 1.

Table 1. PUSDATARU stations and their matched NASA POWER retrieval points

Station	Latitude	Longitude	Grid elevation (m)	Offset (°)	Days
DANGI	7.0263° S	110.7190° E	137.6	0.000	1737
JRAGUNG	7.1548° S	110.5612° E	137.6	0.000	1737
KEDUNGOMBO	7.2642° S	110.8480° E	356.3	0.000	1582
SEMPU	7.4110° S	110.2370° E	625.1	0.000	1768
TEMPURAN	7.9099° S	110.4647° E	159.1	0.000	1635
KETRO	7.3806° S	110.9603° E	356.3	0.162	1779
CENGKLIK	7.5178° S	110.7317° E	356.3	0.279	1728
KEDUNGULING	7.9417° S	110.8392° E	159.1	0.376	1682
KRADENAN	7.7254° S	109.9935° E	625.1	0.398	1746
NGANCAR	7.9892° S	110.9789° E	159.1	0.520	1678

The record covers 1 January 2020 to 31 December 2024. NASA POWER supplies 1827 days per point. After joining and removing days with missing values, 8459 station-days remain in the primary set, distributed as 4975 training days (2020–2022), 1755 validation days (2023), and 1729 test days (2024). The ten-station set contains 17,072 station-days (10,089 / 3487 / 3496). Satellite irradiance variables were converted from $\text{kW h m}^{-2} \text{ day}^{-1}$ to $\text{MJ m}^{-2} \text{ day}^{-1}$ by multiplying by 3.6.

One property of the retrieved data must be stated. The CERES SYN1deg radiation fields supplied by NASA POWER are identical at all five retrieval points on all 1827 days. The points span 0.98° of latitude and 0.61° of longitude but fall within a single radiation grid cell, so the satellite irradiance predictor carries no spatial information across the study area. Only the MERRA-2 meteorological fields, grid elevation, and latitude vary between sites. This property bears directly on the cross-station results in "Seasonal, Regime, and Cross-Station Behaviour,".

2.2. The Reference Series and the Leakage It Creates

The ground-based reference irradiance is not a pyranometer measurement. It is computed from observed sunshine duration through the Ångström–Prescott relation given in Equation (2):

$$R_{S_{ref}} = R_a \left(a + b \frac{n}{N} \right), \quad a = 0.25, b = 0.50 \quad (2)$$

where R_a is extraterrestrial irradiance, n observed sunshine duration, and N astronomical day length. Reproducing this expression from the supplied R_a and n/N values recovers the reference series to within $8 \times 10^{-9} \text{ MJ m}^{-2} \text{ day}^{-1}$, that is, exactly. Two consequences follow, and both are reported here because neither is routinely disclosed in comparable studies.

First, the reference is a sunshine-based estimate of irradiance, not a direct measurement, and inherits the assumption of fixed coefficients a and b that were not calibrated locally. This is standard practice where pyranometers are unavailable and is defensible, but the reference and the satellite product are not measurements of the same quantity obtained in the same way. Second, any predictor set containing n/N or a monotone function of it makes the target algebraically recoverable. A model given the sunshine ratio and R_a need not learn atmospheric physics; it need only fit a two-term linear expression, which it can do to arbitrary precision. “Quantifying the Leakage in the Sunshine-Inclusive Configuration,” below, quantifies the consequence. The primary analysis therefore uses a predictor set containing no ground-observed sunshine information. This is also the operationally realistic configuration: a system correcting satellite irradiance in near-real time has no access to same-day ground sunshine observations, because if it did, it would have no need of the satellite product.

2.3. Predictor Sets

1) Set A (primary)

Sixteen NASA POWER variables — all-sky, clear-sky, direct normal and diffuse surface irradiance, all-sky clearness index, top-of-atmosphere irradiance, cloud amount, aerosol optical depth at 550 nm, temperature at 2 m with its daily maximum and minimum, relative and specific humidity, surface pressure, wind speed, and corrected precipitation — together with the engineered features described under “Feature Engineering” and site

descriptors (latitude, grid elevation). Twenty-three predictors in total, none ground-observed.

2) Set B (replication of the original configuration)

Set A plus the ground-observed sunshine ratio n/N , used only in "Quantifying the Leakage in the Sunshine-Inclusive Configuration" to quantify leakage. A third configuration adding ground-observed relative humidity, maximum and minimum temperature, and wind run, but still excluding sunshine, is reported as an ablation in "Predictor-Set Ablation."

2.4. Feature Engineering

Mean daily temperature and diurnal temperature range were computed from the reanalysis fields as in Equation (3):

$$T_{mean} = \frac{T_{max} + T_{min}}{2}, \quad DTR = T_{max} - T_{min} \quad (3)$$

Extraterrestrial irradiance on a horizontal surface was computed from latitude φ , solar declination δ , and sunset hour angle ω_s for day of year J :

$$R_a = \frac{86400 G_{sc}}{\pi} \left(1 + 0.033 \cos \frac{2\pi J}{365} \right) (\cos \varphi \cos \delta \sin \omega_s + \omega_s \sin \varphi \sin \delta) \quad (4)$$

with $G_{sc} = 1367 \text{ W m}^{-2}$. Annual periodicity was encoded as continuous cyclical components in Equation (5):

$$DOY_{\sin} = \sin \left(\frac{2\pi J}{365} \right), \quad DOY_{\cos} = \cos \left(\frac{2\pi J}{365} \right) \quad (5)$$

2.5. Weather Regimes

Regimes were defined on the observed clearness index $K_t^{obs} = R_{s_{ref}}/R_a$, which ranges from 0.254 to 0.750 in this dataset with the resulting class boundaries given in Table 2.

Table 2. Weather Regimes Index

Regime	Criterion	Days (full record)	Days (2024 test)
Clear	$K_t^{obs} \geq 0.55$	2713	570
Transitional	$0.40 \leq K_t^{obs} < 0.55$	3595	738
Overcast	$K_t^{obs} < 0.40$	2151	421

Defining regimes on the observed rather than the satellite clearness index is deliberate. When regimes are defined on the satellite index, which is itself a Set A predictor, the classification task described under “Weather-Regime Classification” becomes trivially solvable and a classifier attains perfect accuracy — an artefact rather than a result. Regime labels are used for error stratification (Table 7) and as the target of a diagnostic classifier (Figure 4); they are not supplied as inputs to the correction models.

As a diagnostic, a LightGBM multiclass classifier (400 estimators, default learning rate, regime label as target) was trained on the same twenty-three Set A predictors used for the regression models, fitted on the 2020–2022 training partition and evaluated on the 2024 test partition. Its purpose is to quantify how well atmospheric state can be recovered from satellite predictors alone — a ceiling on achievable correction skill — not to feed regime information into the correction models themselves. Results (accuracy terch, macro-F1, and the confusion matrix) are reported under “Weather-Regime Classification,” below.

2.6. Models, Baselines, and Hyperparameters

Two naïve baselines establish the minimum skill any correction must exceed. The persistence baseline predicts each day's reference irradiance from the previous day at the same station; the climatology baseline uses the training-period mean for the corresponding day of year, following standard practice for solar irradiance benchmarking [1], [18]. Random Forest [11], XGBoost [23], LightGBM [14], and Elastic Net[16] were implemented with their standard formulations; hyperparameters for each were selected on the 2023 validation year by exhaustive grid search (Table 3).

Table 3. Search space and selected hyperparameters (validation RMSE, MJ m⁻² day⁻¹)

Model	Search space	Selected configuration	Val. RMSE
Random Forest	n_estimators 300; max_depth {None, 12}; min_samples_leaf {1, 5}; max_features {sqrt, 0.5}	max_depth 12, min_samples_leaf 5, max_features 0.5	2.640
XGBoost	n_estimators {400, 800}; learning_rate {0.03, 0.08}; max_depth {4, 6}; subsample	n_estimators 400, learning_rate 0.03, max_depth 4	2.590

Model	Search space	Selected configuration	Val. RMSE
	0.8; colsample_bytree 0.8; reg_lambda 1.0		
LightGBM	n_estimators {400, 800}; learning_rate {0.03, 0.08}; num_leaves {31, 63}; min_child_samples {10, 30}	n_estimators 400, learning_rate 0.03, num_leaves 31, min_child_samples 30	2.673
Elastic Net	α {0.001, 0.01, 0.1, 1.0}; l1_ratio {0.1, 0.5, 0.9}	α 0.1, l1_ratio 0.1	2.832
Quantile Mapping	—	empirical CDF, 200 quantile bins, linear interpolation	—

The ensemble combines the five corrected series by weighted averaging, as defined in Equation (6):

$$\widehat{GHI}^{ens} = \sum_{m=1}^5 w_m \widehat{GHI}^{(m)}, \quad w_m = \frac{1/RMSE_m^{val}}{\sum_k 1/RMSE_k^{val}} \quad (6)$$

Weights were estimated on the validation year alone: Random Forest 0.2142, XGBoost 0.2183, LightGBM 0.2115, Elastic Net 0.1997, Quantile Mapping 0.1563.

2.7. Leakage Control and Evaluation Protocol

All preprocessing parameters — normalization means and standard deviations, and the Quantile Mapping transfer function — were estimated on the 2020–2022 training partition only and applied unchanged to validation and test data. The (7) on is chronological: 2020–2022 training, 2023 validation, 2024 testing. Results under a random 15 per cent hold-out are reported alongside in “Quantifying the Leakage in the Sunshine-Inclusive Configuration” for comparison. Performance was assessed with the four metrics defined in Equations (7) and (8):

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (7)$$

$$MBE = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i), \quad R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (8)$$

All reported in $\text{MJ m}^{-2} \text{ day}^{-1}$, with relative RMSE normalized by the test-period mean reference irradiance ($17.323 \text{ MJ m}^{-2} \text{ day}^{-1}$ for the primary set). Pairwise differences between models were tested with a paired t -test on squared errors and a Wilcoxon signed-rank test on absolute errors, with a Bonferroni-corrected significance threshold applied given the six pairwise comparisons performed [38].

2.8. Computational Environment

Experiments were run on an Intel Xeon processor at 2.10 GHz with a single allocated core and 3 GB of RAM, under 64-bit Linux and Python 3.12, using scikit-learn, XGBoost 3.4.0, LightGBM 4.7.0, SHAP 0.52.0, NumPy, pandas, and SciPy. Timing figures in Table 12 refer to this configuration.

3. RESULTS AND DISCUSSION

3.1. Accuracy Under Chronological Evaluation

This subsection reports the headline accuracy result: how each correction method, together with the two naive baselines, performs on the 2024 test year under the chronological protocol described above. Table 4 gives RMSE, relative RMSE, MAE, MBE, and R^2 in physical units for the uncorrected satellite product, both baselines, and all five correction methods.

Table 4. Test-year (2024) performance, Set A predictors, $\text{MJ m}^{-2} \text{ day}^{-1}$ ($n = 1729$)

Model	RMSE	rRMSE (%)	MAE	MBE	R^2
Uncorrected NASA POWER	3.398	19.62	2.723	+0.988	0.266
Persistence baseline	3.659	21.13	2.698	+0.021	0.150
Climatology baseline	3.873	22.35	3.173	-0.403	0.047
Quantile Mapping	3.579	20.66	2.806	+0.230	0.186
Elastic Net	2.868	16.56	2.275	+0.362	0.477
Random Forest	2.674	15.43	2.061	+0.221	0.546
Ensemble	2.695	15.56	2.093	+0.324	0.539
XGBoost	2.702	15.60	2.063	+0.394	0.536
LightGBM	2.723	15.72	2.085	+0.391	0.529

Correction is clearly worthwhile. Random Forest reduces RMSE by 21.3 per cent relative to the uncorrected product, cuts relative RMSE from 19.6 to 15.4 per cent, removes 78 per cent of the mean bias, and doubles the explained variance from 0.27 to 0.55. Both naïve baselines are beaten comfortably, and both are in fact worse than the raw satellite product — persistence and climatology carry no information about the specific day's cloud state, which is exactly what the satellite provides.

Two results run against the framing of earlier work in this area. Quantile Mapping *increases* RMSE relative to the uncorrected series, from 3.398 to 3.579 MJ m⁻² day⁻¹, while reducing mean bias from +0.988 to +0.230. This is the expected behaviour of a distribution-matching method: it corrects the marginal distribution, which is what it is designed to do, at the cost of day-to-day accuracy, which it makes no attempt to preserve. Reporting only RMSE would mischaracterize the method; both metrics are needed. More consequentially, LightGBM is the weakest of the three tree ensembles here, not the strongest.

3.2. Differences Between the Tree Ensembles Are Not Significant

Table 4 shows Random Forest, XGBoost, LightGBM, and the ensemble clustered within a narrow band of RMSE. Whether that gap is statistically meaningful is tested formally here. Table 5 reports paired significance tests for every model against the best performer, using both a parametric and a non-parametric test so that the two can be cross-checked against one another.

Table 5. Paired tests against the best model (Random Forest), 2024 test year

Model	RMSE	Δ RMSE vs best	Paired t (squared errors), p	Wilcoxon (absolute errors), p
Random Forest	2.674	—	—	—
Ensemble	2.695	+0.021	0.129	0.019
XGBoost	2.702	+0.028	0.081	0.836
LightGBM	2.723	+0.050	0.009	0.394
Elastic Net	2.868	+0.195	< 0.001	< 0.001
Uncorrected	3.398	+0.724	< 0.001	< 0.001
Quantile Mapping	3.579	+0.905	< 0.001	< 0.001

The three tree ensembles and their weighted average are separated by $0.05 \text{ MJ m}^{-2} \text{ day}^{-1}$, which is 1.8 per cent of the best model's RMSE and roughly one fiftieth of the residual standard deviation. The two tests disagree about whether even that gap is meaningful — the paired t -test rejects equality for LightGBM while the Wilcoxon test does not, and the reverse holds for the ensemble — which is itself the clearest evidence that the ordering among them is not robust. The defensible conclusion is that Random Forest, XGBoost, LightGBM, and their ensemble perform equivalently on this dataset, and that all four are significantly better than Elastic Net, the uncorrected product, and Quantile Mapping. The practical implication is that algorithm selection within the tree-ensemble family is not the decisive design choice; predictor provenance, addressed in the leakage-quantification and predictor-ablation analyses below, is.

This comparison involves six pairwise tests against the best model, so a Bonferroni-corrected threshold of $\alpha = 0.05/6 \approx 0.008$ [38] is the more defensible criterion for declaring a difference significant. Under that threshold, only Elastic Net, the uncorrected product, and Quantile Mapping remain distinguishable from Random Forest; none of XGBoost, LightGBM, or the ensemble survives correction. This reinforces, rather than weakens, the conclusion that the tree ensembles are practically interchangeable here.

3.3. Residual Behaviour

Aggregate error metrics can conceal how errors are distributed across days. This subsection examines the full residual distribution for each method rather than its summary statistics alone. Table 6 reports the mean, standard deviation, median, interquartile range, skewness, and 95th-percentile absolute error of the 2024 residuals for every model.

Table 6. Distribution of 2024 residuals ($\hat{y} - y$), $\text{MJ m}^{-2} \text{ day}^{-1}$

Model	Mean	SD	Median	IQR	Skewness	p95 abs. error
Uncorrected	0.988	3.252	0.877	4.657	0.070	6.511
Random Forest	0.221	2.665	0.003	3.393	0.396	5.460
XGBoost	0.394	2.674	0.145	3.340	0.412	5.563
LightGBM	0.391	2.696	0.155	3.390	0.339	5.663
Elastic Net	0.362	2.846	0.222	3.868	0.180	5.697
Quantile Mapping	0.230	3.572	0.074	4.584	-0.017	6.722
Ensemble	0.324	2.676	0.230	3.427	0.238	5.421

Correction narrows the interquartile range from 4.66 to 3.39 MJ m⁻² day⁻¹ and lowers the 95th percentile absolute error from 6.51 to 5.46. The corrected residuals acquire a mild positive skew absent from the raw product, indicating that the models retain a tendency to overestimate on a minority of days, predominantly heavily overcast days, as “Seasonal, Regime, and Cross-Station Behaviour” shows. The histogram in Figure 2 compresses the overall shape of the residual distribution into a single panel; Figure 3 breaks that same distribution down by model as a set of box plots, which makes the differences in spread across the six correction methods easier to compare directly.

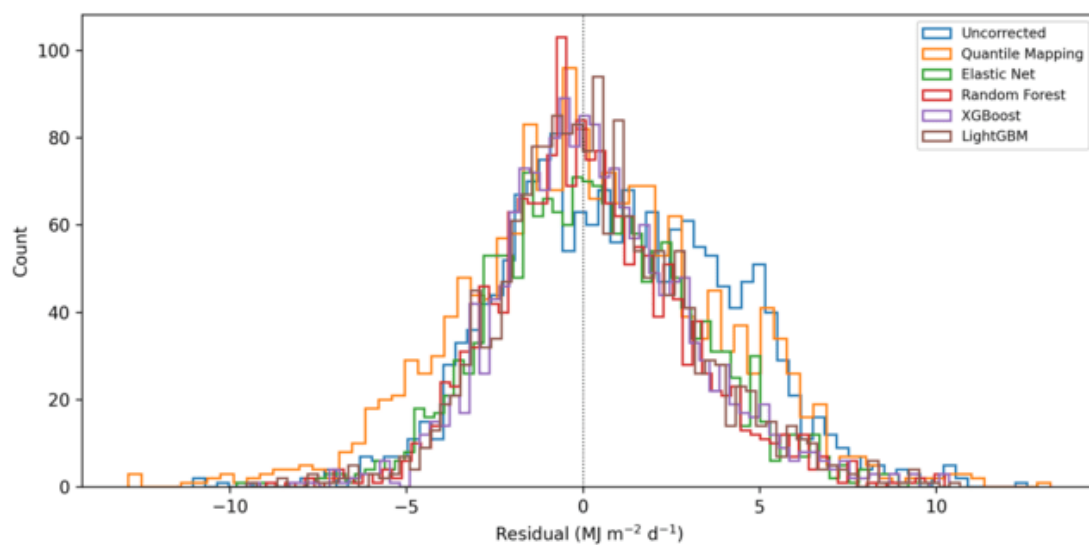


Figure 2. Residual error distribution, 2024 test year.

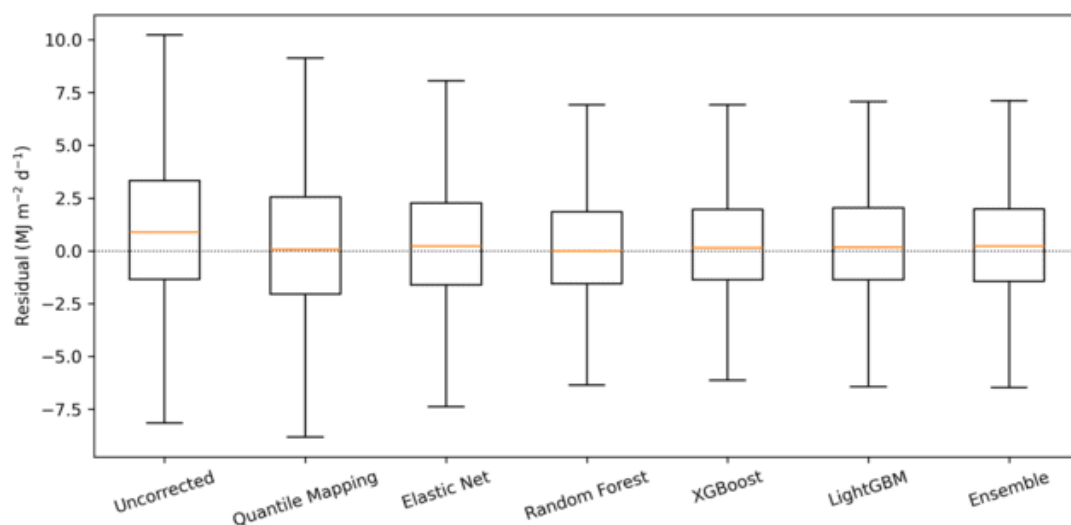


Figure 3. Error spread by model, 2024 test year.

Both of the preceding figures summarize the residuals as one-dimensional distributions. Figure 4 instead plots corrected against observed GHI directly for each method, which reveals systematic over- or under-prediction patterns that a residual histogram or box plot alone cannot show.

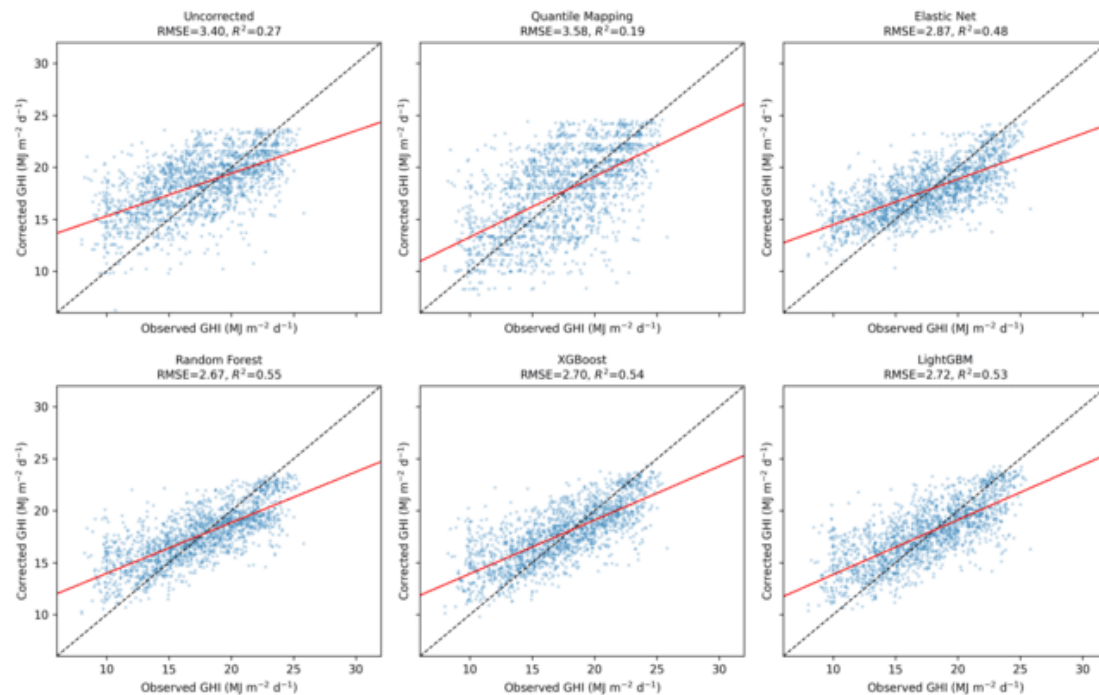


Figure 4. Observed against corrected GHI for the 2024 test year, with the 1:1 line (dashed) and the fitted regression line (solid).

The scatter plots make the limits of the correction visible in a way the summary statistics do not. All methods compress the predicted range relative to the observed range, a regression-to-the-mean effect characteristic of models fitted under squared-error loss, so that the highest observed irradiance days are systematically underpredicted and the lowest overpredicted. This is a real limitation for applications concerned with extremes rather than with means.

3.4. Weather-Regime Classification

As described above, a diagnostic classifier was trained to recover the observed weather regime from the same predictors used for irradiance correction. Figure 5 presents its confusion matrix on the 2024 test partition, giving the count of correct and misclassified days for each of the three regimes.

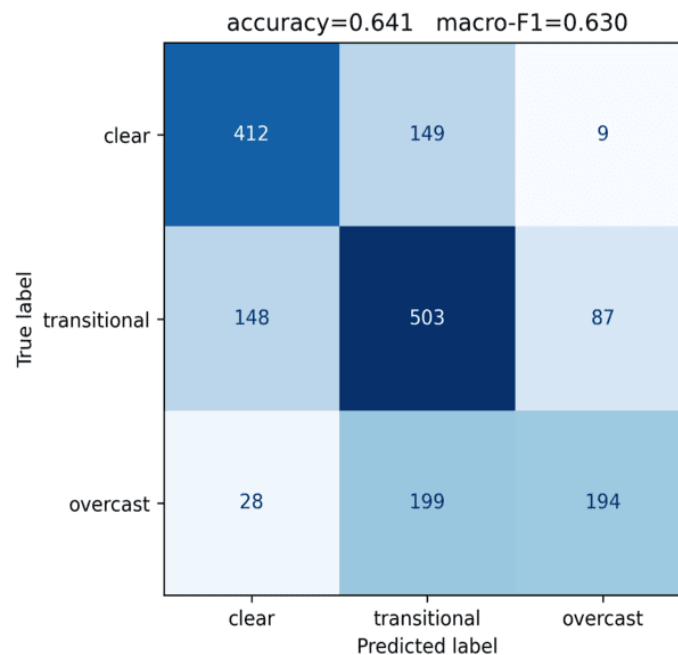


Figure 5. Confusion matrix for weather-regime classification from Set A predictors, 2024 test year. Overall accuracy 0.641, macro-F1 0.630.

A LightGBM classifier trained on the same predictors recovers the observed regime with 64.1 per cent accuracy and a macro-F1 of 0.630 (clear: precision 0.701, recall 0.723; transitional: 0.591, 0.682; overcast: 0.669, 0.461). Performance is weakest for the overcast class, which the classifier under-detects, and errors occur almost entirely between adjacent classes. This is a moderate result, and it is informative precisely because it is moderate. Satellite predictors recover the observed sky state at a Central Java station only imperfectly, which is a direct measure of the atmospheric information the satellite product fails to capture, and therefore of the ceiling on what any correction method fitted to those predictors can achieve.

3.5. Seasonal, Regime, and Cross-Station Behaviour

The accuracy figures reported so far are pooled averages over the full test year. This subsection disaggregates them to examine where correction skill is concentrated. Table 7 reports RMSE separately by season and by the observed weather regime defined earlier, for every method.

Table 7. RMSE by season and by observed weather regime, 2024 test year ($\text{MJ m}^{-2} \text{ day}^{-1}$)

Model	DJF	MAM	JJA	SON	Clear	Transitional	Overcast
Uncorrected	3.176	3.758	3.414	3.231	2.611	3.052	4.675
Random Forest	2.902	3.001	2.062	2.677	2.505	1.838	3.865
XGBoost	2.981	3.057	2.025	2.675	2.331	1.930	4.011
LightGBM	3.013	3.057	2.041	2.713	2.357	2.034	3.960
Elastic Net	2.877	3.093	2.727	2.783	2.714	2.161	3.954
Quantile Mapping	3.524	3.834	3.466	3.497	3.214	3.572	4.030
Ensemble	2.862	3.013	2.194	2.680	2.473	2.028	3.786

Correction gains are strongly regime-dependent. RMSE falls by 40 per cent under transitional conditions (3.052 to 1.838), by 17 per cent under overcast conditions, and by only 4 per cent under clear skies. This is coherent: the satellite product is already accurate on clear days, and the largest recoverable error lies in the partly cloudy conditions that dominate the tropical record. Seasonally, correction is most effective in the dry season (JJA, 40 per cent reduction) and least in the wet-season transition (MAM, 20 per cent), consistent with the greater unpredictability of convective cloud during the monsoon.

Table 8. Leave-one-station-out validation across the five co-located stations, Random Forest trained on the other four

Held-out station	n	Uncorrected RMSE	Corrected RMSE	rRMSE (%)	MBE	R^2
DANGI	1737	4.002	4.031	22.54	+0.309	0.193
JRAGUNG	1737	2.827	3.744	20.53	-0.308	0.074
KEDUNGOMBO	1582	3.093	4.265	23.00	-3.040	-0.252
SEMPU	1768	3.725	3.344	22.18	+2.191	-0.204
TEMPURAN	1635	3.626	3.578	21.46	+1.406	0.058

The correction does not transfer. At four of five held-out stations the corrected RMSE is no better than, or substantially worse than, the uncorrected satellite product, and at three stations the corrected series has negative or near-zero R^2 . The bias introduced is severe at some sites: $-3.04 \text{ MJ m}^{-2} \text{ day}^{-1}$ at KEDUNGOMBO and $+2.19$ at SEMPU.

The “Study Sites and Data” section, above, explains why. NASA POWER supplies an identical radiation series to all five stations, so the model can distinguish sites only through MERRA-2 meteorology, elevation, and latitude. What it learns is therefore a station-specific offset between one shared satellite series and each local reference, and applied to a station it has not seen, that offset is simply the wrong one. Any claim of regional generalization from single-cell satellite radiation is unsupportable, and this study makes none.

3.6. Quantifying the Leakage in the Sunshine-Inclusive Configuration

The preceding subsections used only satellite and reanalysis predictors. This subsection restores the ground-observed sunshine ratio to the predictor set, replicating the configuration used in the original submission, so that the leakage mechanism identified in the Methods can be measured directly. Table 9 reports the resulting accuracy under that configuration, for comparison against Table 4.

Table 9. Test-year performance with the ground-observed sunshine ratio included as a predictor (Set B)

Model	RMSE	rRMSE (%)	MAE	R ²
Random Forest	0.191	1.10	0.127	0.9977
XGBoost	0.113	0.65	0.083	0.9992
LightGBM	0.094	0.54	0.049	0.9994
Ensemble	0.100	0.58	0.070	0.9994
Elastic Net	0.290	1.67	0.226	0.9947
Quantile Mapping	3.579	20.66	2.806	0.1862

Adding a single ground-observed predictor lowers LightGBM's RMSE by a factor of 29, from 2.723 to 0.094 MJ m⁻² day⁻¹, and raises R² from 0.529 to 0.9994. Quantile Mapping is unaffected, because it uses only the satellite series and cannot exploit the additional column. That selective response is itself diagnostic: a genuine improvement in atmospheric modelling would not leave the statistical benchmark untouched while transforming every learner able to read the extra variable.

The mechanism is confirmed by attribution analysis. Under Set B, the sunshine ratio alone accounts for 70.9 per cent of total mean absolute SHAP attribution, with R_a — the other

term of the Ångström–Prescott expression – accounting for most of the remainder. Under Set A no predictor exceeds 22.3 per cent, and attribution is distributed across all-sky irradiance (0.901), direct normal irradiance (0.546), surface pressure (0.274), latitude (0.270), maximum temperature (0.248), and R_a (0.227).

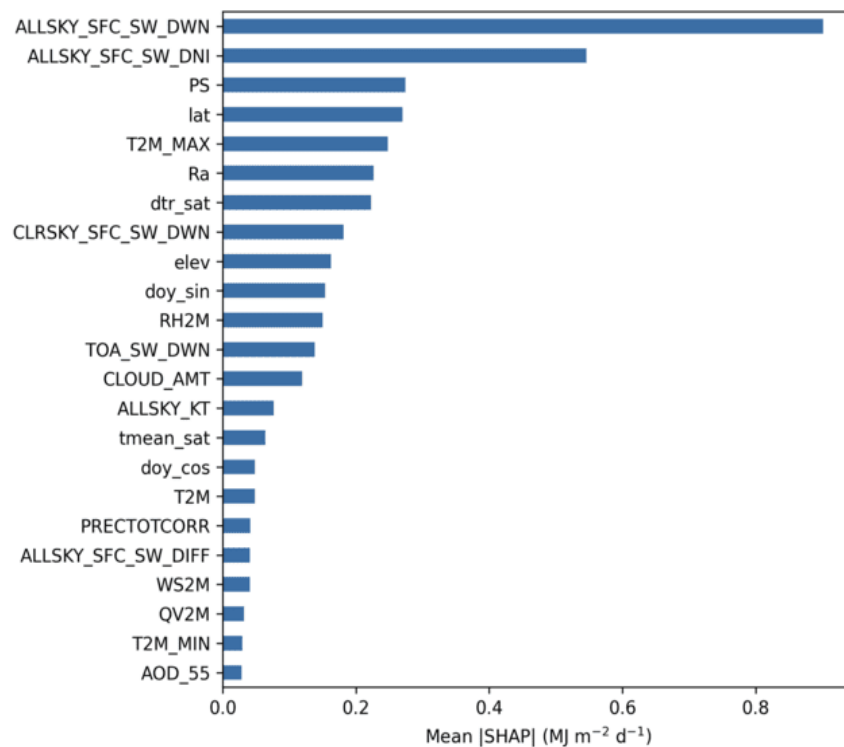


Figure 6. Mean absolute SHAP attribution, Random Forest, Set A predictors, 2024 test year

Table 10. Sensitivity to the partitioning strategy (RMSE, MJ m⁻² day⁻¹)

Model	Random 15% hold-out	Chronological (2024)	Ratio
Random Forest (Set A)	2.756	2.674	0.97
XGBoost (Set A)	2.663	2.702	1.01
LightGBM (Set A)	2.631	2.723	1.03
Elastic Net (Set A)	2.981	2.868	0.96
Random Forest (Set B)	0.812	0.191	0.24
XGBoost (Set B)	0.124	0.113	0.91
LightGBM (Set B)	0.063	0.094	1.48
Elastic Net (Set B)	0.297	0.290	0.98

Table 10 carries a nuance worth stating plainly. Under Set A, random and chronological splitting give results within three per cent of each other; on this dataset, random splitting on its own was not an important source of optimism. Near-perfect accuracy in this class of study is attributable almost entirely to algebraic dependence between a sunshine predictor and an Ångström–Prescott reference, not to the partitioning strategy. The chronological protocol is retained because it is the correct design for autocorrelated series and because it supports the temporal generalization claim the study needs to make, not because it changed the numbers.

3.7. Predictor-Set Ablation

Table 9 showed that adding one ground-observed variable changes accuracy dramatically. This subsection isolates that effect through a controlled ablation, adding predictor groups one at a time. Table 11 reports Random Forest accuracy under four progressively richer predictor sets, from the uncorrected satellite product to the full sunshine-inclusive configuration.

Table 11. Effect of predictor provenance, Random Forest, 2024 test year ($n = 1556$ after requiring complete ground meteorology)

Predictor set	RMSE	rRMSE (%)	MAE	MBE	R ²
Uncorrected satellite	3.385	19.45	2.718	+0.921	0.279
Satellite and reanalysis only	2.710	15.57	2.104	−0.026	0.538
Plus ground temperature, humidity, wind	2.669	15.34	2.070	+0.007	0.552
Plus ground sunshine ratio	0.249	1.43	0.166	+0.013	0.996

Adding ground-observed temperature, humidity, and wind improves RMSE by 1.5 per cent, a marginal gain that does not justify the operational cost of requiring same-day station data. Adding ground sunshine improves it by 91 per cent. The discontinuity between those two rows is the signature of leakage, not of information content: sunshine duration is not ten times more informative about irradiance than humidity and cloud amount combined; it is the variable from which the reference was constructed.

3.8. Computational Cost

Accuracy alone does not determine whether a method is suitable for routine operational use; training and inference cost also matter, particularly for the ensemble, whose fairness

in prior reporting was raised as a concern. Table 12 reports training time, total inference time, and per-sample inference time for every method on the 1729-day test set.

Table 12. Training and inference time (1729 test samples)

Model	Training (s)	Inference (s)	Inference per sample (ms)
Random Forest	5.977	0.0594	0.0343
LightGBM	0.733	0.0489	0.0283
XGBoost	0.472	0.0074	0.0043
Elastic Net	0.0054	0.00015	0.00009
Quantile Mapping	0.0015	0.00025	0.00014
Ensemble (aggregation only)	—	0.00009	0.00005
Ensemble (total, incl. base-model inference)	7.189	0.1162	0.0672

The ensemble is reported in both forms. The aggregation step alone takes 0.00009 s, but the operationally meaningful total is 0.116 s, because inference from all five base models is required. This makes the ensemble the slowest configuration tested, not the fastest, and since it offers no accuracy advantage (Table 5), it has no operational justification here. All methods are nonetheless fast enough in absolute terms that computational cost is not a discriminating criterion at daily resolution: every model corrects a full year of five-station data in under 0.12 s.

3.9. Interpretation

Under Set A the models attribute their corrections principally to the satellite irradiance fields themselves — all-sky and direct normal irradiance together account for roughly a third of total attribution — with secondary contributions from surface pressure, latitude, temperature, and extraterrestrial irradiance. The corrections are therefore anchored in the radiative variables one would expect on physical grounds, and the regime stratification of Table 7 shows the expected pattern of largest gains under partly cloudy conditions. To that extent the models learn a physically interpretable attenuation relationship rather than an arbitrary mapping. The limits of that interpretation are equally clear. Explained variance reaches only 0.55, the correction compresses the predicted range at both extremes, and it does not transfer to unseen stations. The models capture

a systematic, site-specific component of satellite error; they do not resolve the day-to-day cloud variability that satellite retrieval misses in the first place.

The finding that gradient boosting and bagging perform comparably is consistent with Damiani et al. [19] and Shi et al. [10] in showing tree ensembles to be effective for satellite radiation correction, but it does not support the stronger claims of algorithm superiority found in parts of that literature. Where this study departs from both is in cross-station transfer: [24] and [25] report spatially generalizable correction using satellite products with genuine sub-degree spatial resolution, whereas the single CERES grid cell covering the present sites provides no such variation. The contrast suggests that spatial transferability in this class of work is governed more by the spatial resolution of the satellite product than by the choice of learning algorithm. Finally, bias correction of this kind cannot replace ground observation networks. It depends on them entirely for calibration, its skill vanishes at unseen sites (Table 8), and the reference series used here is itself a ground-instrument product.

3.10. Sensitivity to Station–Grid Offset

The primary analysis used only stations whose coordinates coincide exactly with a NASA POWER retrieval point. Table 13 repeats the evaluation across all ten PUSDATARU stations with nearest-point matching, where the offset reaches 0.52° (approximately 58 km).

Table 13. Ten-station sensitivity analysis, 2024 test year ($n = 3496$), $\text{MJ m}^{-2} \text{ day}^{-1}$

Model	RMSE	rRMSE (%)	MAE	MBE	R ²
Uncorrected NASA POWER	5.921	40.39	4.749	+3.628	−0.529
Persistence baseline	3.016	20.59	1.855	+0.018	0.603
Climatology baseline	4.735	32.30	4.186	−0.025	0.022
Quantile Mapping	5.588	38.11	4.376	+0.769	−0.362
Elastic Net	4.118	28.09	3.607	+0.540	0.260
XGBoost	2.831	19.31	1.907	+0.362	0.650
Random Forest	2.906	19.82	1.897	+0.399	0.632
Ensemble	2.997	20.44	2.256	+0.458	0.608
LightGBM	3.010	20.53	1.946	+0.420	0.605

Three observations follow. First, spatial mismatch is costly: uncorrected RMSE rises from 3.398 to 5.921 MJ m⁻² day⁻¹ and mean bias from +0.988 to +3.628 when stations up to 58 km from the retrieval point are included, and the raw product's R² becomes negative. Second, correction is correspondingly more valuable in that setting, reducing RMSE by 52 per cent, but the *corrected* accuracy is not better than in the primary analysis (2.831 versus 2.674 MJ m⁻² day⁻¹); most of the larger apparent gain reflects a worse starting point rather than a better endpoint. Third, the qualitative conclusions are unchanged: tree ensembles cluster tightly (2.831 to 3.010), the weighted ensemble again fails to improve on the best single model, Quantile Mapping again fails to reduce RMSE, and the identity of the best tree ensemble changes — XGBoost here, Random Forest in the primary analysis — which reinforces the earlier conclusion, in “Differences Between the Tree Ensembles Are Not Significant,” that the ordering among them is not stable.

Leave-one-station-out validation across all ten stations gives corrected R² below zero at seven sites, with RMSE beating the uncorrected product at six of ten only because the raw product is so poor at the offset stations. The failure of spatial transfer reported in “Seasonal, Regime, and Cross-Station Behaviour” therefore holds for the larger network as well.

3.11. Discussion

The results show that machine-learning correction improves daily global horizontal irradiance (GHI) estimates when the models are trained and evaluated within the same station network. Under the chronological 2024 evaluation, Random Forest produced the lowest RMSE (2.674 MJ m⁻² day⁻¹), reducing error by 21.3% relative to the uncorrected NASA POWER product and increasing explained variance from 0.266 to 0.546. Although Random Forest ranked first numerically, the differences among Random Forest, XGBoost, LightGBM, and the weighted ensemble were small and were not robust after multiple-comparison correction. Therefore, the principal finding is not that one tree-ensemble algorithm is universally superior, but that tree-based correction models consistently outperform the raw satellite product, naïve baselines, Elastic Net, and Quantile Mapping when trained with satellite and reanalysis predictors.

The predictor-attribution results support a physically plausible interpretation of the Set A models. All-sky irradiance and direct normal irradiance together account for

approximately one-third of the total SHAP attribution, while surface pressure, temperature, extraterrestrial irradiance, and location-related variables provide secondary contributions. This pattern indicates that the models rely primarily on radiative information already contained in the satellite product and use meteorological variables to adjust systematic discrepancies between satellite estimates and local observations. However, this interpretation should not be overstated as a causal representation of atmospheric processes. In particular, latitude and other site-related predictors may partly encode station-specific bias patterns rather than purely physical radiative mechanisms. Thus, the correction models should be interpreted as learning a physically informed but empirically calibrated relationship between satellite-derived irradiance and local ground observations.

The regime-specific analysis further clarifies where this correction is most useful. The largest reduction in RMSE occurred during transitional conditions, where Random Forest reduced RMSE from 3.052 to 1.838 MJ m⁻² day⁻¹, followed by overcast conditions. In contrast, the improvement under clear skies was limited because the uncorrected satellite product was already relatively accurate. This pattern is consistent with the physical difficulty of estimating irradiance during partly cloudy conditions, when rapidly changing cloud cover can cause substantial deviations between satellite retrievals and surface observations. Nevertheless, the moderate weather-regime classification accuracy (64.1%) shows that the available satellite and reanalysis predictors do not fully recover local cloud-state variability. This limitation helps explain why the corrected models achieve only moderate explained variance and why substantial errors remain on heavily overcast days.

Residual analysis confirms that correction improves typical and high-error days but does not eliminate systematic limitations. Compared with the uncorrected product, Random Forest reduced the interquartile range of residuals from 4.657 to 3.393 MJ m⁻² day⁻¹ and lowered the 95th-percentile absolute error from 6.511 to 5.460 MJ m⁻² day⁻¹. However, all correction models compressed the prediction range, tending to overpredict low-irradiance days and underpredict high-irradiance days. This regression-to-the-mean behaviour is expected for models optimized under squared-error loss, but it limits their suitability for applications that depend strongly on extreme irradiance conditions, such as peak photovoltaic generation analysis or extreme-weather risk assessment.

The comparison with Quantile Mapping is also important. Quantile Mapping reduced mean bias but increased RMSE relative to the uncorrected product. This result is not necessarily evidence that the method failed; rather, it reflects its methodological objective. Quantile Mapping adjusts the marginal distribution of the satellite series, whereas the machine-learning models attempt to preserve day-specific variation through additional predictors. Consequently, Quantile Mapping may be useful where long-term distributional consistency or bias reduction is the primary objective, but it is less appropriate when accurate day-to-day irradiance prediction is required.

A critical finding concerns predictor provenance. Adding ground-observed temperature, humidity, and wind yielded only a marginal improvement over satellite and reanalysis predictors. By contrast, adding the ground-observed sunshine ratio caused RMSE to collapse from 2.710 to 0.249 MJ m⁻² day⁻¹ in the ablation analysis, while the full Set B configuration produced near-perfect performance. This apparent improvement is not evidence of a substantially better atmospheric correction model. Because the reference irradiance series is derived using the Ångström–Prescott relationship, the sunshine ratio is algebraically related to the target. The dominance of sunshine ratio in the SHAP analysis therefore identifies target leakage rather than genuine predictive skill. Models using this predictor should not be interpreted as independent satellite-correction models or used to support claims of operational performance where same-day ground sunshine observations are unavailable.

The cross-station validation provides the strongest boundary on the study's conclusions. Correction did not transfer reliably to unseen stations, and several held-out stations showed negative corrected $\langle R^2 \rangle$ values or substantial bias. This outcome is consistent with the structure of the input data: NASA POWER provides the same CERES radiation grid-cell value to the co-located stations, while the model learns the local discrepancy between that shared satellite signal and each station's ground reference. When applied to a new station, the learned station-specific correction is no longer valid. The ten-station sensitivity analysis reinforces this conclusion. Larger station-to-grid offsets increase the error of the raw satellite product and make correction appear more beneficial in relative terms, but they do not improve the final corrected accuracy or establish spatial generalizability.

These findings are broadly consistent with prior studies showing that tree ensembles can improve satellite-based solar-radiation estimates, including Damiani et al. [24] and Shi et al [25]. However, the present results do not support strong claims that a particular tree-based algorithm is inherently superior. The identity of the best-performing model changed between the primary and station-offset analyses, whereas the overall performance of the tree ensembles remained closely clustered. More importantly, the results suggest that spatial transferability depends less on whether bagging or boosting is selected and more on whether the satellite product contains sufficient spatially resolved information to distinguish local atmospheric and geographic conditions.

Overall, the models capture a systematic and station-specific component of satellite irradiance error, particularly under transitional cloud conditions. They do not fully resolve the local, day-to-day cloud variability absent from the satellite and reanalysis predictors, nor can they replace ground-observation networks. Their most defensible role is therefore as locally calibrated bias-correction tools for stations with historical ground observations, rather than as universally transferable models for estimating irradiance at unmonitored locations.

4. CONCLUSION

This study evaluated six bias-correction approaches for NASA POWER daily global horizontal irradiance (GHI) against ground-based reference irradiance at PUSDATARU stations in Central Java during 2020–2024 using a chronological evaluation protocol and physically interpretable error metrics. Correction was beneficial at the five stations exactly co-located with a NASA POWER retrieval point: Random Forest achieved the lowest point-estimate error, reducing RMSE from 3.398 to 2.674 MJ m⁻² day⁻¹ (21.3%), relative RMSE from 19.6% to 15.4%, and mean bias from +0.988 to +0.221 MJ m⁻² day⁻¹, while increasing R^2 from 0.266 to 0.546. Across the full ten-station sensitivity analysis, XGBoost achieved the lowest RMSE, reducing error from 5.921 to 2.831 MJ m⁻² day⁻¹; however, the ranking among Random Forest, XGBoost, LightGBM, and the weighted ensemble was not statistically robust and changed between station sets. Accordingly, the substantive conclusion is that a single tuned tree ensemble is an effective local correction tool, whereas the selection among tree-ensemble algorithms is secondary. The weighted ensemble provided no reliable accuracy advantage while incurring the greatest

total inference cost, and Quantile Mapping, although effective in reducing mean bias, increased RMSE and is therefore unsuitable where day-specific irradiance accuracy is the principal objective. Two negative findings are equally important: the learned correction did not transfer reliably to stations excluded from training because all co-located sites shared the same CERES radiation grid-cell value, and the near-perfect performance obtained when ground-observed sunshine ratio was included as a predictor ($R^2=0.9994$) reflects recovery of the Ångström–Prescott relationship used to construct the reference series rather than independent predictive skill. The conclusions are bounded by the use of an Ångström–Prescott reference with uncalibrated coefficients ($a=0.25$), ($b=0.50$) rather than pyranometer observations, a five-year daily hindcast record, threshold-dependent weather-regime labels, the absence of photovoltaic-output validation, and retrieval-point offsets of up to 58 km for five stations. Future work should therefore prioritize spatially resolved satellite products, such as Himawari-9 or GK2A, direct pyranometer-based reference measurements and local coefficient calibration, multi-station model development, and validation against measured photovoltaic generation, which together would provide a more rigorous assessment of whether satellite-radiation correction can support transferable and operational solar-energy applications.

REFERENCES

- [1] W. Tercha, S. A. Tadjer, F. Chekired, and L. Canale, "Machine Learning-Based Forecasting of Temperature and Solar Irradiance for Photovoltaic Systems," *Energies (Basel)*, vol. 17, no. 5, p. 1124, 2024, doi: 10.3390/en17051124.
- [2] S. Al-Dahidi, B. Rinchi, R. Dababseh, O. Ayadi, and M. Alrbai, "A geographic multi-scale machine learning framework for predicting solar irradiation on tilted surfaces," *Energy*, vol. 313, Dec. 2024, doi: 10.1016/j.energy.2024.133767.
- [3] M. U. Khan and M. A. Jama, "Evaluation and correction of solar irradiance in Somaliland using ground measurements and global reanalysis products," *Heliyon*, vol. 10, no. 16, p. e35256, 2024, doi: 10.1016/j.heliyon.2024.e35256.
- [4] *Artificial Intelligence for Renewable Energy Systems*. Elsevier, 2022. doi: 10.1016/C2020-0-03645-5.

- [5] I. A. Aditya, P. M. Putra, R. A. Rajagukguk, and Y. Asri, "Estimating Irradiance through Satellite-Driven Deep Learning Models," in *Proceedings - International Seminar on Intelligent Technology and its Applications, ISITIA*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 48–52. doi: 10.1109/ISITIA63062.2024.10667941.
- [6] T. Nguyen-Duy, T. Ngo-Duc, M. Choisy, I. P. Fernandez, and S. Sparrow, "Performance and added value of a high-resolution (2 km) rainfall product based on WRF-downscaled ERA5 for Ho Chi Minh City, Vietnam," *Theor. Appl. Climatol.*, vol. 156, no. 12, Dec. 2025, doi: 10.1007/s00704-025-05894-1.
- [7] D. Puspita *et al.*, "Hybrid ANN WRF solar radiation forecasting in data limited tropical region," *Renewable and Sustainable Energy Transition*, vol. 8, Dec. 2025, doi: 10.1016/j.rset.2025.100124.
- [8] N. S. Sitohang, I. Budi, K. G. H. Mangunkusumo, F. Albar, and G. R. Afdholy, "Assessing Satellite-Based GHI Accuracy via Statistical and LSTM Approaches: Nusa Penida PVPP Case Study," in *Proc. ICT-PEP*, IEEE, 2025, pp. 43–48. doi: 10.1109/ICT-PEP67281.2025.11232049.
- [9] J. Zhu, X. Shi, and S. Zhang, "Machine Learning-Based Grammar Error Detection Method in English Composition," *Sci. Program.*, vol. 2021, 2021, doi: 10.1155/2021/4213791.
- [10] D. Selvamuthu, V. Kumar, and A. Mishra, "Indian stock market prediction using artificial neural networks on tick data," *Financial Innovation*, vol. 5, no. 1, 2019, doi: 10.1186/s40854-019-0131-7.
- [11] L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [12] M. S. Alam, F. S. Al-Ismael, M. S. Hossain, and S. M. Rahman, "Ensemble Machine-Learning Models for Accurate Prediction of Solar Irradiation in Bangladesh," *Processes*, vol. 11, no. 3, p. 908, 2023, doi: 10.3390/pr11030908.
- [13] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [14] G. Ke *et al.*, "LightGBM: A Highly Efficient Gradient Boosting Decision Tree," in *Advances in Neural Information Processing Systems 30*, 2017, pp. 3146–3154.
- [15] R. Ledesma and G. Salazar, "Machine Learning for Site Adaptation of Satellite-Derived Solar Irradiance in Northwestern Argentina," 2025. [Online]. Available: <https://datasasaf.isasvcs.ipma.pt>

- [16] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005, doi: 10.1111/j.1467-9868.2005.00503.x.
- [17] L. Yezeguelian, C. Vernay, T. Carrière, and P. Blanc, "Characterizing the convergence and robustness of the kernel density mapping method for site-adaptation of global horizontal irradiation in Western Europe," HAL, 2021. [Online]. Available: <https://hal.science/hal-03337523>
- [18] M. D. Bailey, D. Nychka, M. Sengupta, and S. Bandyopadhyay, "Adapting Quantile Mapping to Bias Correct Solar Radiation Data," arXiv, 2024. doi: 10.48550/arXiv.2405.20352.
- [19] L. Gudmundsson, J. B. Bremnes, J. E. Haugen, and T. Engen Skaugen, "Technical Note: Downscaling RCM precipitation to the station scale using quantile mapping – a comparison of methods," May 15, 2012. doi: 10.5194/hessd-9-6185-2012.
- [20] S. Zainali, D. Yang, T. Landelius, and P. E. Campana, "Site adaptation with machine learning for a Northern Europe gridded global solar irradiance product," *Energy and AI*, vol. 15, Jan. 2024, doi: 10.1016/j.egyai.2023.100331.
- [21] F. A. Suwadana *et al.*, "Solar Irradiance Estimation in Tropical Regions Using Recurrent Neural Networks and WRF Models," *Energies (Basel)*, vol. 18, no. 4, p. 925, 2025, doi: 10.3390/en18040925.
- [22] P. M. P. Garniwa *et al.*, "Comparison of Semi-Empirical Models in Estimating Global Horizontal Irradiance for South Korea and Indonesia," *Resources*, vol. 14, no. 11, p. 170, 2025, doi: 10.3390/resources14110170.
- [23] S. Chen, C. Li, Y. Xie, and M. Li, "Global and direct solar irradiance estimation using deep learning and selected spectral satellite images," *Appl. Energy*, vol. 352, Dec. 2023, doi: 10.1016/j.apenergy.2023.121979.
- [24] A. Damiani *et al.*, "Spatially generalizable bias correction of satellite solar radiation for regional climate assessment – a case study in Japan," *International Journal of Applied Earth Observation and Geoinformation*, vol. 145, p. 104947, 2025, doi: 10.1016/j.jag.2025.104947.
- [25] H. Shi *et al.*, "First estimation of high-resolution solar photovoltaic resource maps over China with Fengyun-4A satellite and machine learning," *Renewable and Sustainable Energy Reviews*, vol. 184, p. 113549, 2023, doi: 10.1016/j.rser.2023.113549.