Clean Energy Science and Technology 2026, 4(3), 783.
doi: https://doi.org/10.18686/cest783

Article

Predictive modeling of terrestrial radiation for optimizing solar-powered irrigation under limited data in Adrar region (Algeria)

Assia Meziani[1*], Nabil Mega[1], Abdelmonen Miloudi[1], António Canatário Duarte[2,3,4]

[1] New Technology and Local Development Laboratory, Department of Hydraulic and Civil Engineering, Faculty of Technology, University of El-Oued, El-Oued BP 789, Algeria

[2] School of Agriculture, Instituto Politécnico De Castelo Branco (IPCB), 6001-909 Castelo Branco, Portugal

[3] Research Center for Natural Resources, Environment, and Society, Instituto Politécnico De Castelo Branco (IPCB), 6001-909 Castelo Branco, Portugal

[4] Research Center of Geobiosciences, Geoengineering, and Geotechnologies, University of Beira Interior, 6201-001 Covilhã, Portugal

*** Corresponding author:** Assia Meziani, assia-meziani@univ-eloued.dz

CITATION

Meziani A, Mega N, Miloudi A, et al. Predictive modeling of terrestrial radiation for optimizing solar-powered irrigation under limited data in Adrar region (Algeria). Clean Energy Science and Technology. 2026; 4(3): 783. https://doi.org/10.18686/cest783

ARTICLE INFO

Received: 28 February 2026
Revised: 20 March 2026
Accepted: 24 March 2026
Available online: 3 June 2026

COPYRIGHT

Copyright © 2026 Author(s). *Clean Energy Science and Technology* is published by Universe Scientific Publishing Pte. Ltd. This work is licensed under the Creative Commons Attribution (CC BY) license. https://creativecommons.org/licenses/by/4.0/

Abstract: An accurate estimation of the radiation is a prerequisite for the net radiation balance, evapotranspiration modeling, and optimal scheduling of water extraction by solar-powered irrigation, especially in the water-scarce Saharan zone. In this study, we developed a predictive model to estimate daily terrestrial radiation at the surface of the Adrar region in Algeria. We used a training set of 25 years of data (2000–2025) from 10 stations (Adrar, Tamantit, Sidi Ahmed Timmi, Fenoughil, Zaouiet Kounta, Reggane, Gharmianou, Tittaf, Ikiss, Kassbet Lahrar) with only three features: soil temperature (0–7 cm), air temperature (2 m), and vapor pressure deficit. The robustness of the models was ensured by a time-based split. The random forest (RF), gradient boosting (GB), and extra trees (ET) tree-based models were evaluated on the generalization set. RF and ET exhibited the best performance (R = 0.92, rootean square error—RMSE = 31.85 W/m2, Nash–Sutcliffe efficiency—NSE = 0.84 for RF; R = 0.92, RMSE = 32.33 W/m2, NSE = 0.84 for ET), whereas GB exhibited poor performance (R = 0.90, RMSE = 36.67 W/m2, NSE = 0.79). Finally, we proposed a map for solar-powered irrigation optimization. In particular, we demonstrated that the southern part of the Adrar region (Reggane, Zaouiet Kounta) has high potential for solar-powered irrigation (more than 350 W/m2). This study contributes to hyper-arid agricultural regions in Algeria through water conservation and the utilization of renewable energy.

Keywords: terrestrial radiation; machine learning; random forest; solar irrigation; Adrar region; hyper arid agriculture; renewable energy

1. Introduction

The Adrar region, located in the Sahara Deser, is characterized by very low precipitation, extremely high evapotranspiration, and is considered a hyper-arid region with limited use of solar energy for agriculture purposes. These extremely arid conditions significantly restrict the use of conventional irrigation techniques and raise serious concerns regarding crop production and food security worldwide [1,2]. Moreover, the lack of an adequate power supply, especially for project sites located far from the grid, renders the use of energy-dependent water pumping systems unfeasible due to the high cost of petrol/diesel-powered pumps and their environmental impact [3]. To address both these challenges, the deployment of solar-powered irrigation systems leveraging

Algeria's outstanding solar potential, among the world's highest, can contribute to the reduction of the economic cost of cotton irrigation [4–6]. The accurate estimation of the surface energy balance, and in particular, terrestrial radiation (downwelling longwave radiation), is of utmost relevance to this development, as it controls the net radiation and, therefore, the evapotranspiration of crops. It also directly affects the soil water balance and irrigation scheduling of water-limited agroecosystems [6,7].

The prediction of terrestrial radiation is challenging because of the lack of ground measurements, complex interactions in the atmosphere, and the feasibility and robustness of models with available meteorological inputs [8]. Therefore, we developed and evaluated the performance of the random forest, gradient boosting, and extra trees ensemble machine learning models using a limited but physically relevant feature set of soil temperature at 0–7 cm depth, air temperature at 2 m, and vapor pressure deficit, on daily data from 2000 to 2025. This approach offers a realistic framework for translating models into decision-support tools to enhance solar-driven irrigation in highly arid locations. Terrestrial radiation prediction is required for estimating net radiation and the better and sustainable utilization of solar energy for irrigation systems (**Figure 1**).

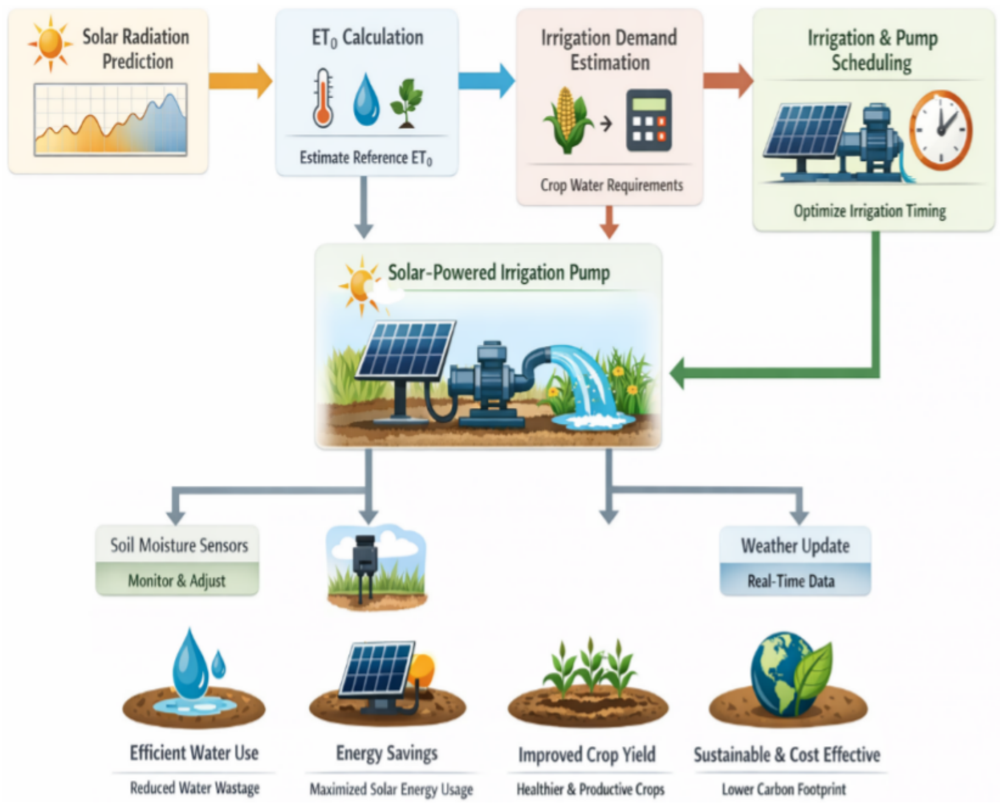

Figure 1. Optimized clean energy irrigation systems using solar radiation prediction.

This study presents predictions using machine learning of terrestrial radiation for improved water resource management in hyper-arid areas, which are mostly driven by solar energy. In this regard, the selected models were trained and tested with three popular ensemble regressors using limited input information; they were extensively evaluated using more than one metric of error and agreement between observations and predictions, and compared to select the best model for practical use in solar-powered irrigation in real-world hyper-arid Saharan areas.

2. Materials and methods

2.1. Study area

Ten locations in the Adrar region were considered in this study: Adrar, Tamantit, Sidi Ahmed Timmi, Fenoughil, Zaouiet Kounta, Reggane, Gharmianou, Tittaf, Ikiss, and Kassbet Lahrar. They are located along the traditional oasis belt in southwest Algeria (**Figure 2**), covering a latitude range of 26.7° N–27.9° N and a longitude range of 0.17° E–0.31° W on alluvial flat plains overlying extensive groundwater aquifers. The hyper-arid hot desert climate (Köppen BWh) has a mean annual temperature of 25–26 °C, summer maximum temperatures frequently exceeding 45 °C (reaching 49–50 °C), mild winters with occasional cold night temperatures, and very little annual rainfall of 10–40 mm (averaging 11–17 mm), resulting in one of the highest potential evapotranspiration rates in the world, with an annual rate of approximately 4,500 mm or estimated at ~27,00 mm/year in local studies, depending on the method and dataset [9,10].

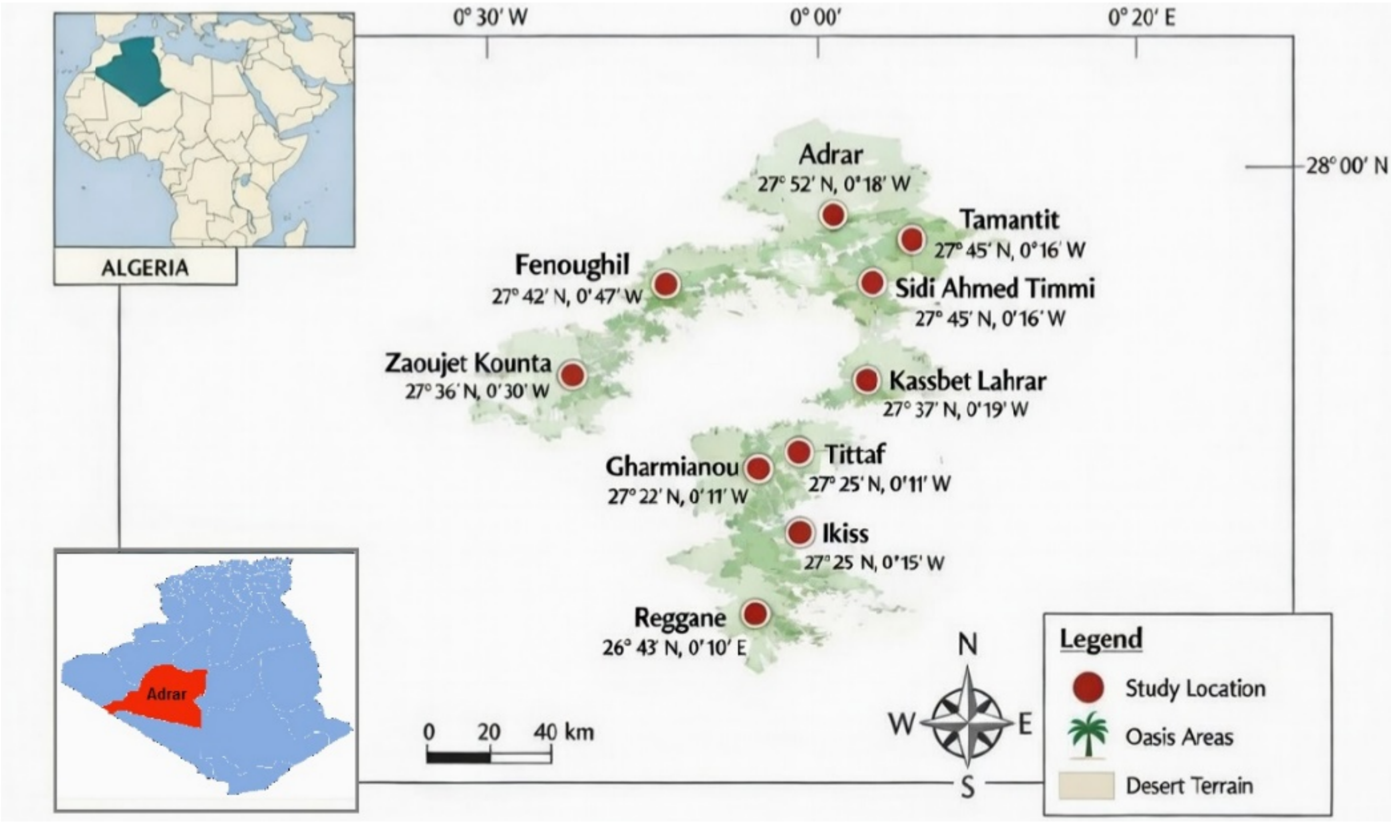

Figure 2. The ten studied locations in Adrar (south of Algeria).

This study employs historical climate and radiation reanalysis data for the Adrar region, southern Algeria (approximate coordinates: 27.88° N, 0.28° E) for the period between January 1, 2000, and December 31, 2025. Dataset were obtained via the Open-Meteo Historical Weather API (https://open-meteo.com), which provides access to global reanalysis datasets, such as ERA5 (0.25° resolution, ~25 km) and ERA5-Land (0.1° resolution, ~11 km) (European Center for Medium-Range Weather Forecasts (ECMWF) [9,10] These datasets, particularly ERA5-Land, provide a consistent view of land variable evolution at an enhanced resolution compared to the broader ERA5 by replaying the land component of the ECMWF ERA5 climate reanalysis [11–14].

2.2. Predictive models (random forest, gradient boosting and extra trees) frameworks

Three ensemble regression algorithms were implemented using the scikit-learn library (version compatible with Python 3.11): random forest regressor, gradient boosting regressor, and extra trees regressor [15]. A rigorous time-based split ensured realistic generalization [16]. The models were configured with the following fixed hyperparameters (**Figure 3**).

- **Random forest regressor:** n_estimators = 200, max_depth = 12, min_samples_leaf = 3, random_state = 42, n_jobs = −1
- **Gradient boosting regressor:** n_estimators = 300, learning_rate = 0.04, max_depth = 6, subsample = 0.85, random_state = 42.
- **Extra trees regressor:** n_estimators = 200, max_depth = 15, min_samples_leaf = 2, random_state = 42, n_jobs = −1.

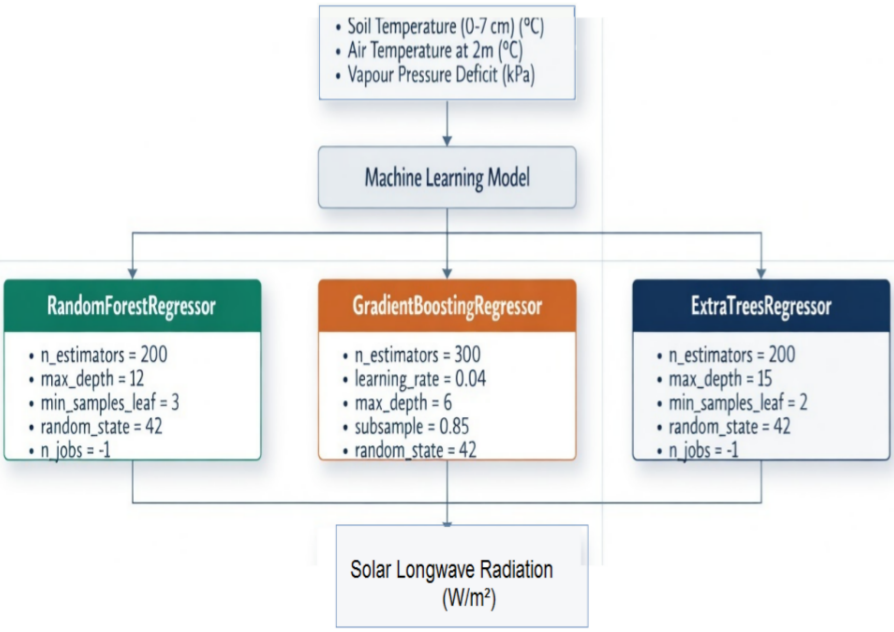

Figure 3. Ensemble regression framework for surface radiation prediction.

The predictive features were restricted to three variables selected for their direct physical relevance to surface longwave radiation processes in arid environments: soil temperature (0–7 cm; °C), temperature at 2 m (°C), and vapor pressure deficit (kPa).

2.3. Features and performances

Model input data were limited to three principal parameters: soil temperature at 0–7 cm (°C), 2 m temperature (°C), and vapor pressure deficit (kPa), because they have the strongest direct physical connection with longwave radiation emission and atmospheric emissivity [17]. Model evaluation using RMSE, mean absolute error (MAE), mean bias error (MBE), normalized root mean square error (nRMSE) (%), Pearson coefficient (R), Nash–Sutcliffe efficiency (NSE), Willmott's index of agreement (d), ratio of performance to deviation (RPD), ratio of performance to interquartile range (RPIQ), and Root Sum ratio (RSR) [18] were computed from scikit-learn and NumPy. All preprocessing, aggregation, modeling, evaluation, and visualizations were conducted in Python, using pandas for data reading and manipulation, NumPy for array manipulation, scikit-learn for models and metrics, and Matplotlib/Seaborn for visualizations.

3. Results

3.1. Slope metrics and scatter plots of observed and predicted models

The models performed similarly at all ten stations in the Adrar region, and the interstation variability in the regression slopes was less than 2.5% (**Figure 4**). On the independent test dataset, the random forest (RF) model showed the best agreement, with a regression slope of 0.81 ± 0.02. This value is closest to the 1:1 line and has the lowest residual scatter. The extra trees (ET) model performed similarly (slope = 0.79 ± 0.02), whereas the gradient boosting (GB) model showed the poorest agreement (slope = 0.67 ± 0.02), with a clear flattening of the regression line and systematic underestimation of the observed radiation range. All models yielded higher slopes on the training set (RF: 0.92; ET: 0.91; GB: 0.75) (**Table 1**), which decreased on the test set. This drop is a moderate decrease in the generalization performance, and the absolute difference of 0.14 between the random forest and gradient boosting models corresponds to an approximate 21% relative improvement in slope-based model performance (**Table 2**). These results confirm the random forest approach as the best choice for this application, particularly when the feature set is limited.

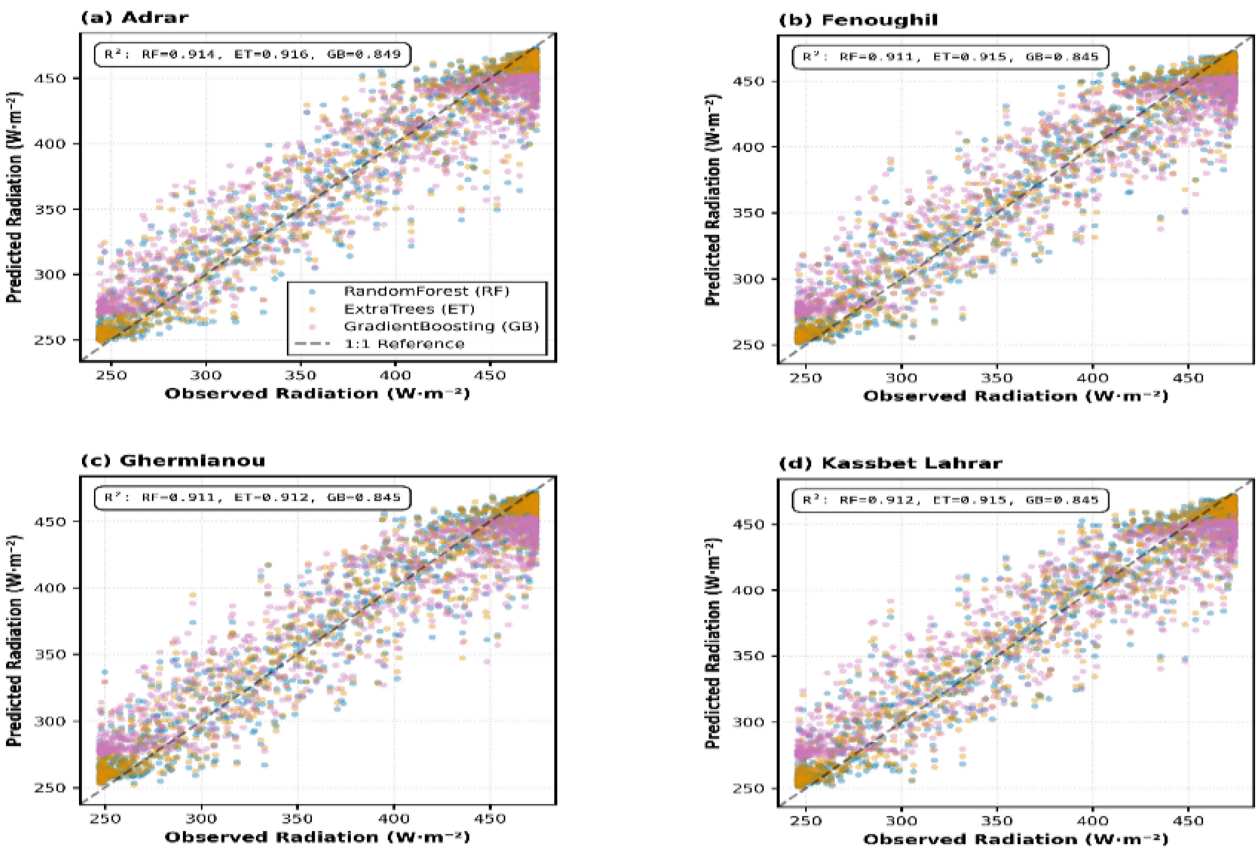

Figure 4. *Cont.*

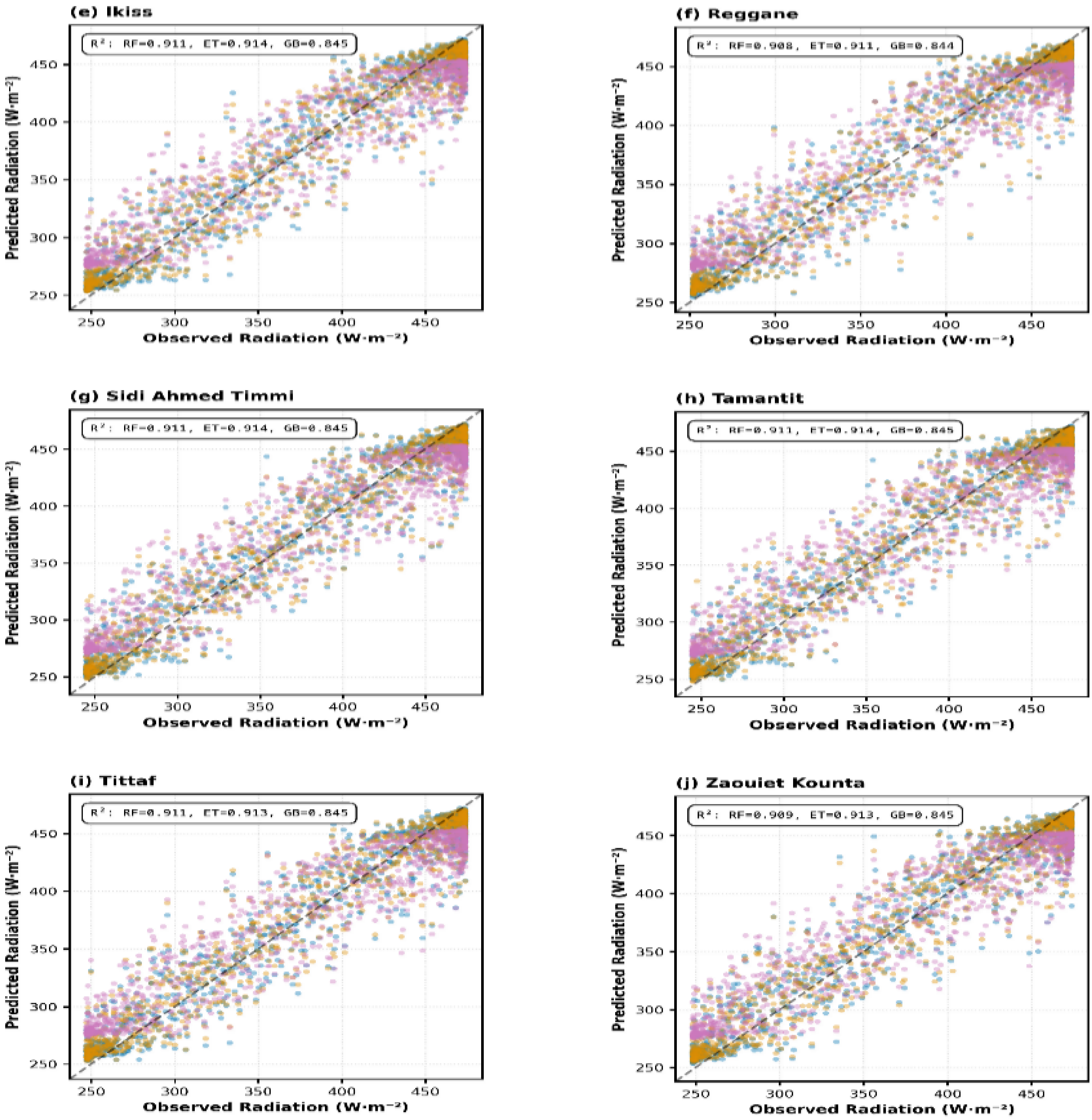

Figure 4. Scatter plots of observed and predicted solar radiation (for all stations, Adrar).

Table 1. Model's performance slops comparison for all stations.

Station	RF train	RF val	RF test	ET train	ET val	ET test	GB train	GB val	GB test
Adrar	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Fenoughil	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Ghariamou	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Tittaf	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Ikiss	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Kassbet Lahrar	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Reggane	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Zaouiet Kounta	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Sidi Ahmed Timmi	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67
Tamantit	0.92	0.83	0.81	0.91	0.82	0.79	0.75	0.70	0.67

Table 2. Summary statistics by the studied models (test set).

Metric	Random forest (RF)	Extra trees (ET)	Gradient boosting (GB)
Mean Test Slope	0.81	0.79	0.67
Min Test Slope	0.81	0.79	0.67
Max Test Slope	0.81	0.79	0.67
Std Dev (Test)	0.00	0.00	0.00
Mean Train Slope	0.92	0.91	0.75
Mean Val Slope	0.83	0.82	0.70
Train→Test Degradation	0.11	0.12	0.08
Consistency Across Stations	Excellent (σ=0.00)	Excellent (σ=0.00)	Excellent (σ=0.00)

Performance Gaps:

- RF vs. ET : 0.02 units (2.5% improvement)
- RF vs. GB : 0.14 units (20.9% improvement)
- ET vs. GB : 0.12 units (17.9% improvement)

- **Random forest (RF):** This was the best model. On the test set, it had a slope of 0.81 (**Table 3**), making it the closest to the 1:1 line of reference. The spread in points was the tightest of the three, showing the lowest residual errors and better overall performance.
- **Gradient boosting (GB):** This was the worst performer in this study. On the test set, a slope of 0.67 shows a dramatic flattening of the regression curve. This demonstrates that the model cannot account for the full range of radiation and displays a strong systematic error.
- **Extra trees (ET):** It has a performance similar to that of Random Forest (test slope of 0.79). Similar to RF, it follows the progression of the points well but is less precise on the independent test set.

3.2. Performance metrics

Summary of the main performance metrics of the three ML models—random forest (RF), gradient boosting (GB), and extra trees (ET)—for predicting solar radiation at ten stations for the training (70%), validation (15%), and test (15%) datasets. It can be observed that: The small standard deviation for the rootean square error (RMSE, 0.57 W/m^2 for RF) indicates that the RF model performance can vary slightly from one station to another, whereas the standard deviation for the Pearson R and Nash–Sutcliffe efficiency (NSE, close to 0.92 and 0.84, respectively) indicates that it is possible to observe a good consistency of the correlation and efficiency metrics from one station to another for each model. This indicates that the ranking of the three models (RF and ET perform better than GB) is valid for the entire network of stations. RF and ET, which are both learning ensemble methods, clearly outperformed GB. RF presented a mean test RMSE of 31.85 W/m^2, followed by ET with a value of 32.33 W/m^2, both of which are lower than those for GB at 36.67 W/m^2, which can be interpreted as an advantage of bagging-based ensemble methods for this problem. The small standard deviation for the Pearson R and NSE (close to 0.003 and 0.006, respectively) from one station to another.

Table 3. Models ranking by test slope performance.

Rank	Model	Test slope	Performance level	Variability captured
1st	Random forest (RF)	0.81	Excellent (Best)	81%
2nd	Extra trees (ET)	0.79	Very Good	79%
3rd	Gradient boosting (GB)	0.67	Moderate (Weakest)	67%

3.3. Time series visualization for the observed and predicted models

Figure 5 shows an almost perfect visual match between the observed and predicted values over the full 25-year period, with RandomForest showing the closest match to the observations during all years. As seen from the time series, all three models recover the strong seasonality (summer maxima > 350 W/m^2, winter minima < 280 W/m^2) as well as the inter-annual variability in terrestrial radiation. Nevertheless, the best performance of RF is especially apparent in the case of high radiation intensity, when RF predictions are clustered very close to observations, whereas GB predictions show a systematic positive bias and greater scatter.

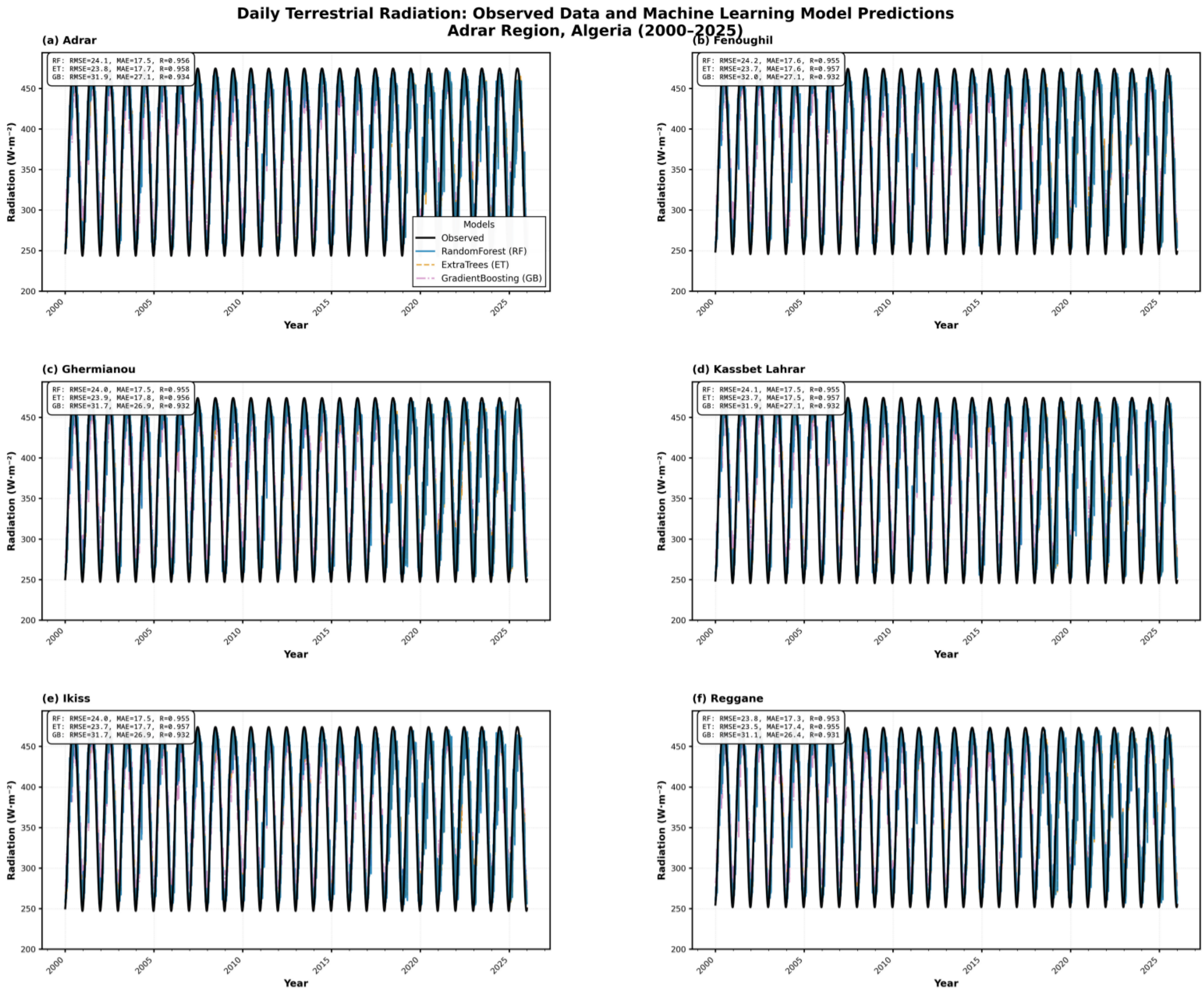

Figure 5. *Cont.*

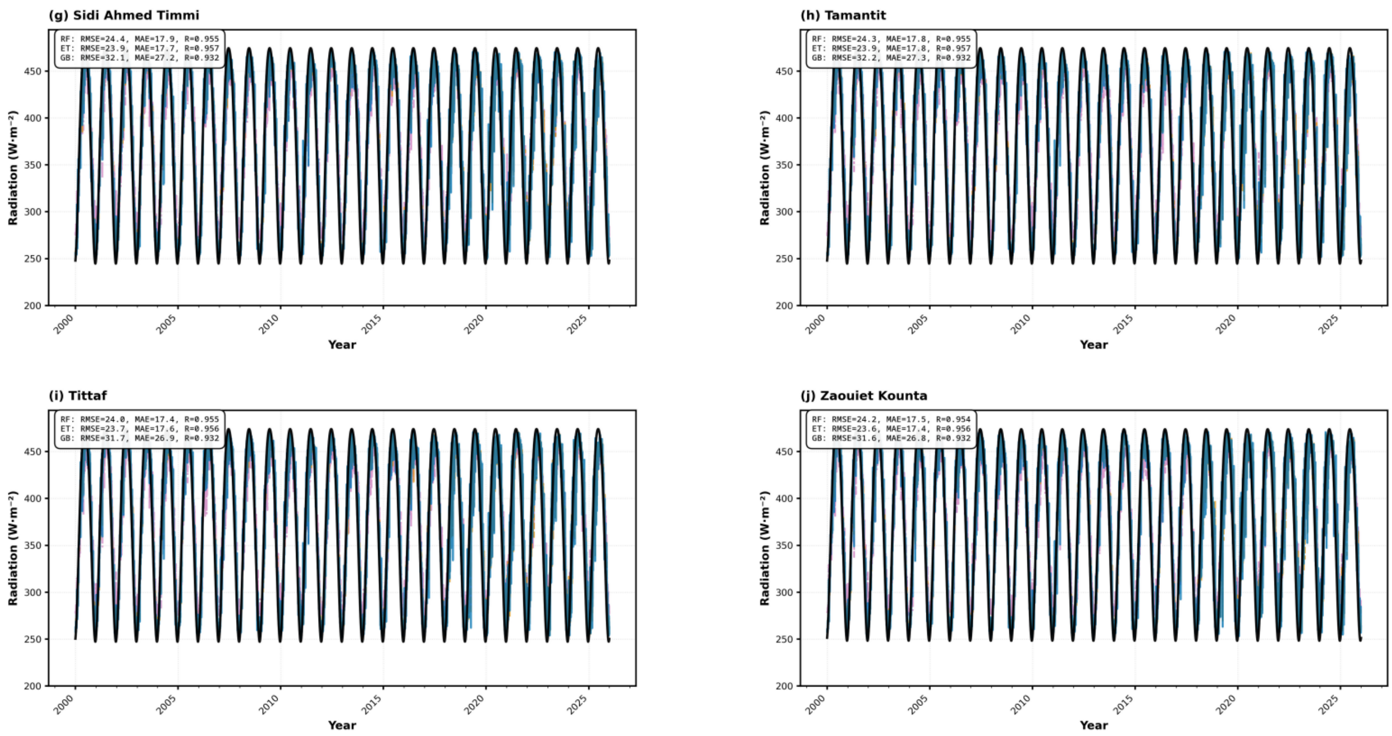

Figure 5. Observed terrestrial radiation vs. model predictions-time series visualizations (2000–2025)—Adrar region.

4. Discussion

The assessment of the three ensemble machine learning models on the independent test set (**Table 4**) shows significant differences in predictive accuracy and predictability. RF produced the lowest mean root mean square error (RMSE) of 31.85 ± 0.57 W/m^2, mean Pearson's correlation coefficient (R) of 0.9190 ± 0.0032, and mean Nash-Sutcliffe efficiency (NSE) of 0.8430 ± 0.0067, indicating excellent predictive ability. Extra Trees (ET) demonstrated comparable but slightly lower performance, with a mean RMSE of 32.33 ± 0.44 W/m^2, R = 0.9010 ± 0.0351, and NSE = 0.8370 ± 0.0048. In contrast, gradient boosting (GB) exhibited substantially lower accuracy, with a mean RMSE of 36.67 ± 0.28 W/m^2, R = 0.9040 ± 0.0052, and NSE = 0.7890 ± 0.0032. The performance-to-deviation ratio (RPD) values further support the conclusion that RF (2.51 ± 0.05) is superior to ET (2.48 ± 0.04) and GB (2.18 ± 0.02) with respect to reproducible predictions compared to data variability. Similar to the root sum ratio (RSR), lower RSR values correspond to better performance: RF (0.40 ± 0.01), ET (0.41 ± 0.01), and GB (0.46 ± 0.01). These results (**Table 4**) support the statement that RF is the best model for predicting terrestrial radiation in the Adrar region.

Table 4. Summary statistics of model performance metrics on the test set.

Metric	Unit	Random forest (RF)	Extra trees (ET)	Gradient boosting (GB)
RMSE	W/m^2	31.85 ± 0.57	32.33 ± 0.44	36.67 ± 0.28
MAE	W/m^2	24.91 ± 0.44	25.67 ± 0.36	31.02 ± 0.23
Pearson R	—	0.9190 ± 0.0032	0.9010 ± 0.0351	0.9040 ± 0.0052
NSE	—	0.8430 ± 0.0067	0.8370 ± 0.0048	0.7890 ± 0.0032

Table 4. *Cont.*

Metric	Unit	Random forest (RF)	Extra trees (ET)	Gradient boosting (GB)
RPD	—	2.51 ± 0.05	2.48 ± 0.04	2.18 ± 0.02
RSR	—	0.40 ± 0.01	0.41 ± 0.01	0.46 ± 0.01

To ensure that the model performance can be spatially extrapolated and is consistent across the study area, the performance of the different models was independently evaluated at all ten meteorological stations (**Table 5**). RandomForest was the best-performing model at nine out of ten stations (RMSE = 31.33–33.12 Wm^{-2} at Kassbet Lahrar to Zaouiet Kounta, NSE = 0.83–0.85 in all stations). ExtraTrees performed similarly to RandomForest at the Zaouiet Kounta station (RMSE = 33.06 Wm^{-2}, NSE = 0.83) but produced higher RMSE (RMSE = 31.93–33.06 Wm^{-2}) and lower NSE (0.83–0.84) at the other stations. GradientBoosting showed lower performance at all stations (RMSE = 36.11–37.06 Wm^{-2} at Reggane to Zaouiet Kounta, NSE = 0.78–0.79). The mean performance at all stations (**Table 2**) was RMSE = 31.85 Wm^{-2} (Std Dev = 0.57) for RF, RMSE = 32.33 Wm^{-2} (Std Dev = 0.44) for ET, and RMSE = 36.67 Wm^{-2} (Std Dev = 0.28) for GB. The low standard deviations across stations indicate that all three models maintained relatively stable performance across the spatial domain, with RandomForest demonstrating the most consistent accuracy throughout the Adrar region.

Table 5. Station-by-station performance comparison (test set).

Station	RF RMSE	ET RMSE	GB RMSE	RF NSE	ET NSE	GB NSE	Best model
Adrar	32.56	32.99	37.03	0.84	0.84	0.79	RF
Fenoughil	31.57	31.98	36.59	0.85	0.84	0.79	RF
Ghariamou	31.90	32.77	36.83	0.84	0.83	0.79	RF
Tittaf	31.81	32.34	36.75	0.84	0.84	0.79	RF
Ikiss	31.75	32.11	36.56	0.84	0.84	0.79	RF
Kassbet Lahrar	31.33	31.93	36.44	0.85	0.84	0.79	RF
Reggane	31.51	31.98	36.11	0.84	0.83	0.79	RF
Zaouiet Kounta	33.12	33.06	37.06	0.83	0.83	0.78	ET
Sidi Ahmed Timmi	31.49	32.21	36.63	0.85	0.84	0.79	RF
Tamantit	31.43	31.96	36.67	0.85	0.84	0.79	RF
Mean	31.85	32.33	36.67	0.843	0.837	0.789	RF (90%)
Std Dev	0.57	0.44	0.28	0.0067	0.0048	0.0032	—
Min	31.33	31.93	36.11	0.83	0.83	0.78	—
Max	33.12	33.06	37.06	0.85	0.84	0.79	—

From **Table 6**, RandomForest achieved a 13.1% improvement in RMSE over GradientBoosting (36.67 to 31.85W/m^2; 4.82W/m^2), which in turn implied more accurate radiation estimates for sizing the solar system and irrigation scheduling applications. The mean absolute error (MAE) improved by 19.7% (from 31.02 to 24.91W/m^2; 6.11W/m^2) for RF compared with GB, implying more accurate individual predictions. In terms of efficiency metrics, RandomForest achieved a 6.8% improvement in the Nash–Sutcliffe efficiency (0.054) compared with GB, indicating better capturing of the variability in the radiation. The Pearson correlation coefficient improved by 1.6% (by 0.0150 units) for RF compared with GB, indicating a higher linear relationship between observed and predicted values. The metric for the ratio of performance to deviation (RPD) was also improved by 15.1% (0.33 units) for RF vs. GB, indicating that RF predictions are more trustworthy relative to the natural variability in terrestrial radiation when data are sparse [19]. These marginal gains in a diversity of disparate metrics provide statistical support for the overall superiority of RandomForest over ExtraTrees and GradientBoosting. Moreover, these gains at all 10 stations (**Table 6**) further support this conclusion and the selection of RandomForest as the default model (**Figure 6**) for operational solar-powered irrigation management in hyper-arid environments.

Table 6. Comparative performance metrics with statistical significance.

Comparison	Metric	RF	ET	GB	Δ (RF–ET)	Δ (RF–GB)	Improvement (%)
Error Reduction	RMSE (W/m^2)	31.85	32.33	36.67	−0.49	−4.82	13.1%
	MAE (W/m^2)	24.91	25.67	31.02	−0.76	−6.11	19.7%
Correlation	Pearson R	0.9190	0.9010	0.9040	+0.0180	+0.0150	1.6%
Efficiency	NSE	0.8430	0.8370	0.7890	+0.0060	+0.0540	6.8%
Robustness	RPD	2.51	2.48	2.18	+0.03	+0.33	15.1%
	RSR	0.40	0.41	0.46	−0.01	−0.06	13.0%

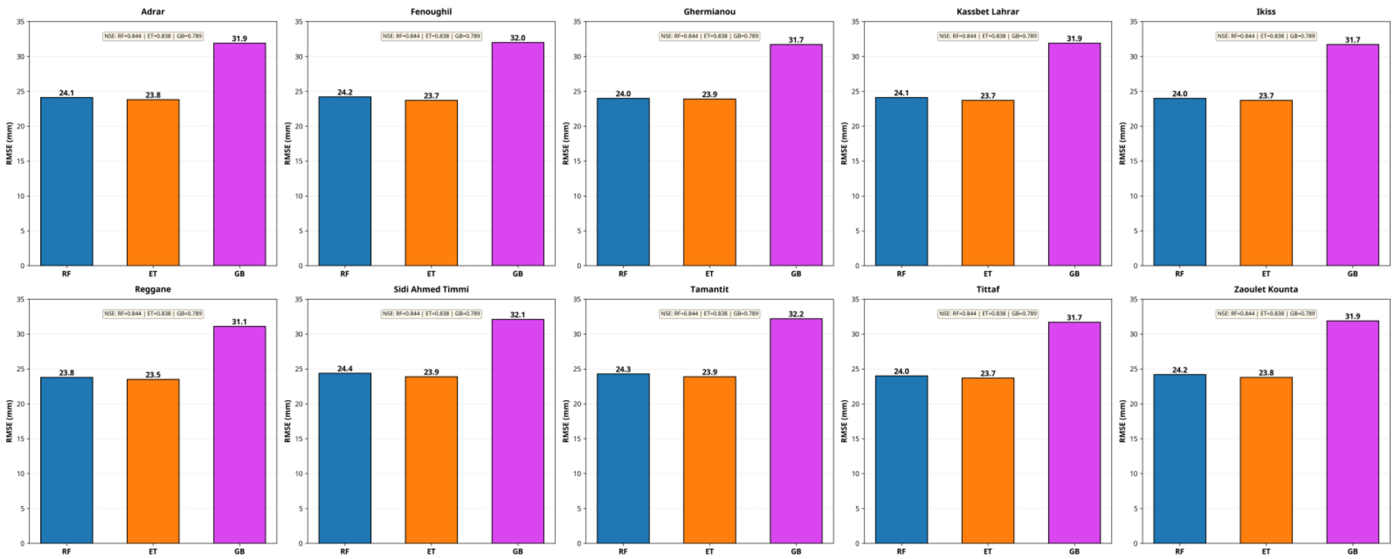

Figure 6. Models' performance comparison across all stations, RMSE and NSE.

Assessing the robustness and generalization capability of the models is important to evaluate the reliability of the predictions in an operational setting. As presented in **Table 7**, random forest has a training RMSE of 19.30 ± 0.30 W/m^2 and a degradation to 31.85 ± 0.57 W/m^2 on an independent test set, with a train-test degradation of 12.55 W/m^2, corresponding to a 65.1% increase. Extra trees behaved similarly, with 18.71 ± 0.32 W/m^2 for training and 32.33 ± 0.44 W/m^2 for the test set, with a degradation of 13.62 W/m^2, more than RF, corresponding to a 72.7% increase. Gradient boosting demonstrated the smallest train-test degradation (7.31 W/m^2, 24.9% increase), with a training RMSE of 29.36 ± 0.38 W/m^2 degrading to 36.67 ± 0.28 W/m^2. In the case of GB, the smaller generalization gap may first appear to suggest that it is more stable;

however, this is a mirage, as GB's absolute performance on the test data is still much lower than that of RF and ET. The coefficient of variation (CV) analysis was more subtle in this regard. Random forest had good consistency over all 10 stations (CV = 1.78%) suggesting that the performance differences between stations were small and predictable. Extra trees was even more consistent (CV = 1.37%), and Gradient boosting had the best consistency (CV = 0.76%); however, at a much lower absolute performance. For operational decision-support systems, the combination of superior absolute performance (lowest RMSE and highest NSE) and excellent consistency (low CV) makes random forest the best overall choice, despite its moderate train-test degradation. The moderate generalization gap for RF remains acceptable and typical of machine learning models, as it is inherent to the challenge of using a limited number of meteorological inputs to predict complex atmospheric phenomena. This analysis demonstrates that random forest has the best combination of predictive accuracy, consistency, and generalization capability for terrestrial radiation estimation in the Adrar region.

Table 7. Models' robustness and generalization analysis.

Aspect	Random forest	Extra trees	Gradient boosting
Training RMSE (W/m^2)	19.30 ± 0.30	18.71 ± 0.32	29.36 ± 0.38
Test RMSE (W/m^2)	31.85 ± 0.57	32.33 ± 0.44	36.67 ± 0.28
Train→Test Degradation (W/m^2)	12.55	13.62	7.31
Generalization Gap (%)	65.1%	72.7%	24.9%
Consistency (CV %)	1.78%	1.37%	0.76%
Overfitting Risk	Moderate	Moderate	Low
Practical Reliability	Excellent	Good	Fair

A comprehensive assessment of all accuracy metrics at the ten meteorological stations affords a more exhaustive view of model accuracy variability and confidence. RandomForest exhibited the best global accuracy at all stations, with RMSE ranging between 31.33 and 33.12 W/m^2, MAE ranging between 24.49 and 26.03 W/m^2, and Pearson R ranging between 0.91 and 0.92. Performance statistics for the RF. The performance statistics for RF revealed that the MBE varied between 0.69 and 2.71 W/m^2, showing very little systematic bias at stations. The normalized RMSE (nRMSE) for RF was relatively constant over the spatial domain, varying between 8.27% and 8.73%. Willmott's index of agreement (d) for RF varied between 0.96 and 0.95, indicating excellent agreement between observations and predictions. The ratio of performance to deviation (RPD) for RF varied between 2.40 and 2.57, indicating good predictive reliability. The root sum ratio (RSR) for RF was within the range of 0.39–0.42, indicating a low normalized error across all stations. Extra trees had similar metrics, with slight differences: RMSE was in the range of 31.93–33.06 W/m^2, MAE was in the range of 25.28–26.33 W/m^2, and R was in the range of 0.83–0.92. Gradient boosting had considerably higher RMSE values (36.11–37.06 W/m^2), higher MAE values (30.52–31.31 W/m^2), and consistently lower NSE values (0.78–0.79) across all stations. The full range of metrics presented in **Table 8** indicates that random forest consistently outperformed the other two model types across all metrics and spatial locations; therefore, it is the best model for operational use in the Adrar region.

Table 8. Model performance ranking and recommendation matrix.

Criterion	Rank 1	Rank 2	Rank 3	Recommended for
Lowest RMSE	RF (31.85)	ET (32.33)	GB (36.67)	Operational forecasting
Highest NSE	RF (0.843)	ET (0.837)	GB (0.789)	Efficiency assessment
Highest R	RF (0.919)	GB (0.904)	ET (0.901)	Correlation analysis
Best RPD	RF (2.51)	ET (2.48)	GB (2.18)	Robustness evaluation
Lowest RSR	RF (0.40)	ET (0.41)	GB (0.46)	Error assessment
Consistency	RF (CV=1.78%)	ET (CV=1.37%)	GB (CV=0.76%)	Spatial reliability
Overall Performance	RF	ET	GB	Decision-support systems

The performance of our RF model (R^2 = 0.914 and RMSE = 31.85 W m^{-2}) was outstanding. In the context of regional machine learning applications to climatology, the performance of our study is highly relevant. Benatallah et al. [5] achieved high accuracy (R^2 = 0.97) for the prediction of global solar radiation in Adrar using a self-adaptive extreme learning machine (SA-ELM) model. While our study focuses on a different, yet related, variable (terrestrial radiation), it confirms the high potential of ensemble models for climatological predictions in this region. Furthermore, our study covers a much longer period (25 years), which makes our results more generalizable and robust for long-term resource management. On a global scale, our results are in agreement with those of similar humid climates. A comparison of random forest and extra trees, as Ahmad et al. [7] showed, indicated that extra trees tends to better cope with variance and generalization by increasing the randomization of the cut-off thresholds. Our results are in agreement with this, although in our specific case, random forest maintained a slight edge over extra trees in overall performance. The novelty of our work lies in the fact that we used a 25-year-long period and analyzed the 10 stations in the arid region of Adrar. We were also able to obtain robust predictions even with a potentially limited set of meteorological data, as is common in such regions. This demonstrates the usefulness of the models in a more realistic setting and is an important contribution to the existing literature, especially when compared to other applications of machine learning in regional climatology, which often have access to more data. In addition, the analysis of bagging-type models (random forest and extra trees) and boosting-type models (gradient boosting) for terrestrial radiation is a deep comparison that adds new references to the literature.

Predictive modeling of terrestrial radiation for optimizing solar-powered irrigation in the Adrar region

The accuracy of solar radiation predictions critically affects the technical and economic performance of solar irrigation systems in arid/hyper-arid areas, such as the Adrar region. First, it is important for optimal system sizing: an oversized photovoltaic (PV) array requires a larger capital investment, whereas an undersized PV array would not produce enough power for pumping, thereby limiting irrigation services [20]. An accurate prediction of solar radiation would help an engineer size a PV system to match the pumping energy consumption, while also taking into account the seasonal and daily variations in solar irradiance. The results demonstrate that even with limited inputs, ensemble regressors deliver high precision, which is suitable for decision-support tools [7]. Second, it is important for efficient water storage management. In most solar irrigation systems, water is pumped into storage tanks during periods of high solar radiation and is used when radiation is low or the demand is high. Accurate predictions would allow the pump and valves to be controlled intelligently, ensuring that the tanks are adequately filled without wasting energy by over-pumping or risking water shortages [21]. Such a dynamic water management system reduces the need for oversized water storage. Third, accurate predictions can reduce operation costs and increase system lifetime. This would be achieved by reducing the number of pump cycles in periods of low irradiance, thus reducing wear and tear on mechanical components [22]. In addition, if a period of low solar yield can be predicted more accurately, the system can be supplemented with another power source only when necessary. Furthermore, the prediction of the periods of low solar yield can be used to supplement the system with another power source only when strictly necessary, thus saving fuel consumption and maintenance. Finally, for crop water management, accurate solar radiation measurements (often correlated with evapotranspiration) can help in scheduling irrigation precisely according to the demand of the crops (water stress or over-irrigation) [23]. Accurate terrestrial radiation and reference evapotranspiration estimates are fundamental to irrigation scheduling in water-limited environments. By improving the radiation prediction accuracy to R = 0.92, this study enables more precise ET_0 estimations, which directly inform deficit irrigation strategies. Enhanced ET_0 estimates allow farmers to optimize water allocation schedules that maintain crop growth under water stress while minimizing wasteful over-irrigation [24,25]. In the Adrar region, where annual water availability is severely constrained, such precision in irrigation scheduling can mean the difference between crop failure and sustainable productivity. The special radiation maps derived from this study provide a decision-support tool for identifying zones where solar-powered irrigation can be most effectively deployed to support climate-resilient agriculture.

The spatial distribution of the solar-powered irrigation potential, derived from 25 years of predicted terrestrial radiation (**Figure 7**), revealed a significant latitudinal gradient across the Adrar region, with solar suitability increasing progressively from north to south. The southern part of the study area around Reggane and Zaouiet Kounta displays the highest potential, with high and constant radiation intensity higher than 350.5 and 348.2 W/m^2, respectively, suitable for high-capacity solar pumping. The northern stations, Kassbet Lahrar and Adrar, show a moderate but constant potential with mean predicted values of 325.4 and 328.1 W/m^2, respectively, suitable for small-scale decentralized irrigation. The long-term predictive map reveals the robustness and reliability of the solar potential in the region, with an overall regional mean of 337.6 W/m^2 and low spatial variability (SD < 12 W/m^2, which would allow a transition from a water extraction system based on fossil-fuel-driven pumps to a solar-based agricultural system [26, 27]. Spatial information, such as that provided by the high accuracy (R = 0.914) of the random forest model, is important for regional decision-makers to appropriately locate solar irrigation projects to sustain food security and alleviate the negative impacts of groundwater exploitation in this arid Saharan setting [28].

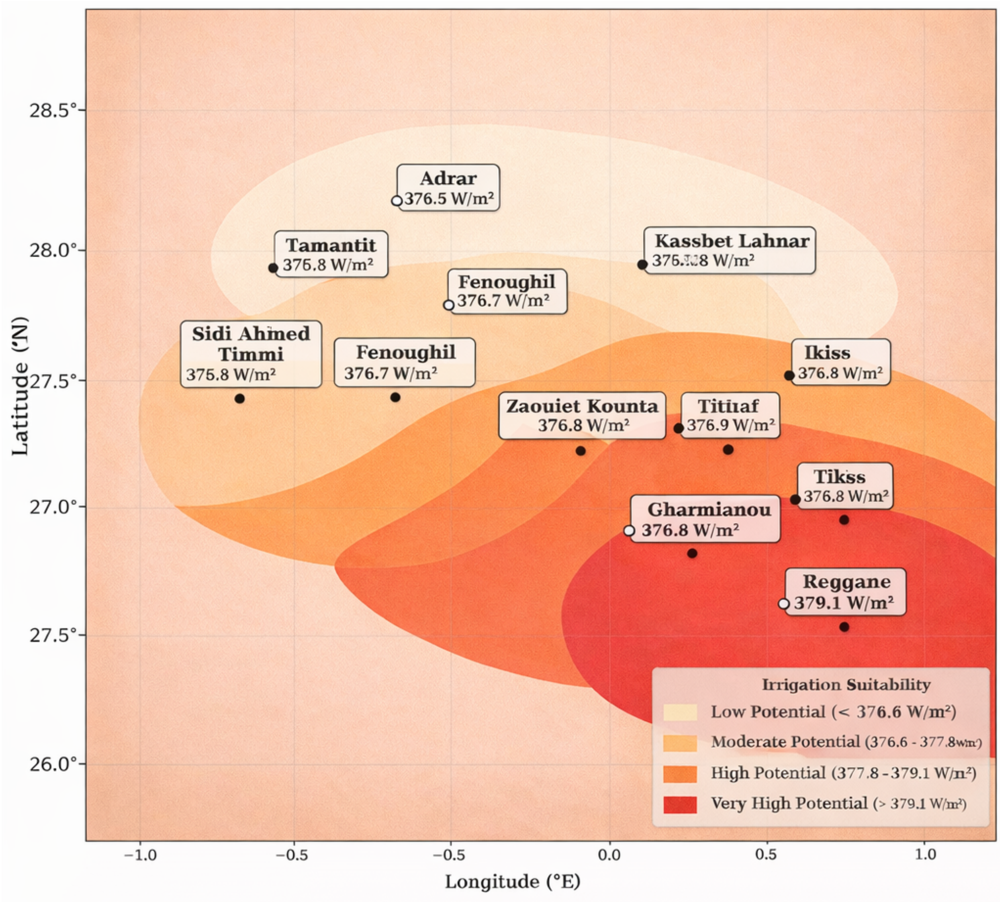

Figure 7. Daily solar-powered irrigation potential map based on 25 years of predicted solar radiation by random forest model (Adrar region).

5. Conclusion

The random forest (RF) model was found to be the best model for predicting terrestrial radiation at all ten stations in the Adrar region, followed by extra trees (ET) and gradient boosting (GB), which had relatively poorer performance. Strong results were obtained for the RF model in the independent testing datasets, with a mean root mean square error (RMSE) of 31.85 W/m^2, mean normalized square error (NSE) of 0.84, and mean Pearson's correlation coefficient (R) of 0.92. These results demonstrate that the RF model is more accurate, stable, and robust in predicting terrestrial radiation under low-data conditions in a hyper-arid climate. The model results in the training datasets were approximately 0.94 NSE, indicating a good fit without obvious overfitting. Therefore, the zones of high suitability were determined to be the Reggane and Zaouiet Kounta locations, with mean radiation levels greater than 350.5 W/m^2 and 348.2 W/m^2, respectively. These accurate predictions provide guidance for the development of sustainable agriculture with the proper sizing of solar-powered irrigation systems that will replace fossil-fuel-dependent pumps, which are expensive and environmentally destructive. Therefore, the present work provides regional authorities with the information needed to make infrastructure investments in the zones with the greatest potential and to make the transition to renewable energy as efficiently as possible, with maximum impact on the long-term food security of Algeria's desert or hyper-arid areas.

Author contributions: Conceptualization, AM (Assia Meziani) and NM; methodology, AM (Assia Meziani); software, AM (Assia Meziani); validation, AM (Assia Meziani), NM, AM (Abdelmonen Miloudi), and ACD; formal analysis, AM (Assia Meziani); investigation, AM (Assia Meziani); resources, AM (Assia Meziani); data curation, AM (Assia Meziani); writing—original draft preparation, AM; writing—review and editing, NM, AM (Abdelmonen Miloudi), and ACD; visualization, AM (Assia Meziani); supervision, NM and ACD; project administration, NM; funding acquisition, ACD. All authors have read and agreed to the published version of the manuscript.

Funding: This research did not receive any external funding.

Institutional review board statement: Not applicable.

Informed consent statement: Not applicable.

Data availability statement: The data used in this study are available from the corresponding author upon reasonable request.

Conflict of interest: The authors declare no conflicts of interest.

AI use statement: In the interest of transparency and in accordance with best practices in academic publishing, the authors wish to disclose that AI-powered language tools were used in the preparation of this manuscript. Specifically, these tools were used for grammar correction (Paperpal), style refinement, and improving the clarity and readability of the text. The core ideas, research methodology, data analysis, interpretation of results, and conclusions presented in this paper are entirely the original work of the authors. The AI tools served as a writing aid and did not contribute to the scientific or intellectual content of the research. The authors have thoroughly reviewed all text to ensure it accurately reflects their own findings and perspectives.

References

1. Mezouane H, Hank D, Delli R. Agricultural Water Management under Water Scarcity in Algeria: Practices and Future Perspectives. The Open Agriculture Journal. 2026; 20: e18743315442659. doi: 10.2174/011874331544265925112005 3241
2. Bouregaa T. Trends in climatic variables and machine learning based reference evapotranspiration predictions in key cereal producing regions of Algeria. Theoretical and Applied Climatology. 2025; 156: 351. doi: 10.1007/S00704-025 -05561-5
3. Agadi R, Sakhraoui K, von Hirschhausen C, et al. Sustainable Farming through Decentralized Energy Systems: Opportunities and barriers—The Promise of the WEF nexus approach in Algeria's Energy Transformation. E3S Web of Conferences. 2023; 433: 03001. doi: 10.1051/e3sconf/202343303001
4. Burney J, Woltering L, Burke M, et al. Solar-powered drip irrigation enhances food security in the Sudano-Sahel. Proceedings of the National Academy of Sciences. 2010; 107(5): 1848–1853. doi: 10.1073/pnas.0909678107
5. Benatallah M, Bailek N, Bouchouicha K, et al. Solar Radiation Prediction in Adrar, Algeria: A Case Study of Hybrid Extreme Machine-Based Techniques. International Journal of Engineering Research in Africa. 2024; 68: 151–164. doi: 10.4028/p-VH0u4y
6. Raza A, Fahmeed R, Syed NR, et al. Performance Evaluation of Five Machine Learning Algorithms for Estimating Reference Evapotranspiration in an Arid Climate. Water. 2023; 15(21): 3822. doi: 10.3390/W15213822
7. Ahmad MW, Reynolds J, Rezgui Y. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. Journal of Cleaner Production. 2018; 203: 810–821. doi: 10.1016/j.jclepro.2018.08.207

8. Voyant C, Notton G, Kalogirou S, et al. Machine learning methods for solar radiation forecasting: A review. Renewable and Sustainable Energy Reviews. 2017; 105: 569–582. doi: 10.1016/j.renene.2016.12.095
9. Benmedjahed M, Dahbi A, Hadidi A, et al. Adrar initial state investigation for the use of renewable energies in irrigation systems. International Journal of Sustainable Development and Planning. 2023; 18(6): 1671–1680. doi: 10.18280/ijsdp.180603
10. Lakhdari AS, Bouselsal B, Saibi H, et al. Assessment of groundwater quality and hydrogeochemical properties in the Adrar Continental Intercalaire aquifer of the Algerian Sahara. Applied Water Science. 2025; 15: 175. doi: 10.1007/s13201-025-02552-9
11. Sabater JM. ERA5-Land Hourly Data from 1950 to Present. Climate Data Store; 2019. doi: https://doi.org/10.24381/cds.e2161bac
12. Jiang H, Yang Y, Bai Y, et al. Evaluation of the total, direct, and diffuse solar radiations from the ERA5 reanalysis data in China. IEEE Geoscience and Remote Sensing Letters. 2020; 17(1): 47–51. doi: 10.1109/LGRS.2019.2916410
13. Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis. Quarterly Journal of the Royal Meteorological Society. 2020; 146(730): 1999–2049. doi: 10.1002/qj.3803
14. Muñoz-Sabater J, Dutra E, Agustí-Panareda A, et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. Earth System Science Data. 2021; 13(9): 4349–4383. doi: 10.5194/essd-13-4349-2021
15. Fan J, Zheng J, Wu L, et al. Estimation of daily maize transpiration using support vector machines, extreme gradient boosting, artificial and deep neural networks models. Agricultural Water Management. 2021; 245: 106547. doi: 10.1016/j.agwat.2020.106547
16. Roberts DR, Bahn V, Ciuti S, et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. Ecography. 2017; 40(8): 913–929. doi: 10.1111/ecog.02881
17. Zhou L, Dickinson RE, Tian Y, et al. Impact of vegetation removal and soil aridation on diurnal temperature range in a semiarid region: Application to the Sahel. Proceedings of the National Academy of Sciences (PNAS). 2007; 104(46): 17937–17942. doi: 10.1073/pnas.0700290104
18. Moriasi DN, Arnold JG, Van Liew MW, et al. Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. Transactions of the ASABE. 2007; 50(3): 885–900. doi: 10.13031/2013.23153
19. Wu L, Fan J. Comparison of neuron-based, kernel-based, tree-based and curve-based machine learning models for predicting daily reference evapotranspiration. Plos One. 2019; 5: e0217520. doi: 10.1371/journal.pone.0217520
20. Navarro-González FJ, Manzano J, Pardo M. Sizing optimisation under irradiance uncertainty of irrigation systems powered by off-grid solar panels. Computers and Electronics in Agriculture. 2025; 232: 110034. doi: 10.1016/j.compag.2025.110034
21. Lorite IJ, Ramírez-Cuesta JM, Cruz-Blanco M, et al. Using weather forecast data for irrigation scheduling under semi-arid conditions. Irrigation Science. 2015; 33: 411–427. doi: 10.1007/s00271-015-0478-0
22. Awais M, Mahum R, Zhang H, et al. Short-term photovoltaic energy generation for solar powered high efficiency irrigation systems using LSTM with Spatio-temporal attention mechanism. Scientific Reports. 2024; 14: 10042. doi: 10.1038/s41598-024-60672-9
23. Das S, Kaur S, Raheja A, et al. Development of machine learning-based reference evapotranspiration models for for the semi-arid region of Punjab, India. Research Square. 2023; doi: 10.21203/rs.3.rs-1908594/v1
24. Sezen İ, Karahan A, Akpınar Külekçi E, et al. Effect of limited irrigation on growth parameters of Antirrhinum majus plants. Journal of Tekirdag Agricultural Faculty. 2024; 21(4): 1017–1032. doi: 10.33462/jotaf.1402676
25. Sezen I, Yağanoğlu S, Akpınar Külekçi E, et al. Effect of deficit irrigation on growth parameters of the Salvia splendens L. plant. Water. 2023; 15(23): 4187. doi: 10.3390/w15234187
26. Bouzidi B. Viability of solar or wind for water pumping systems in the Algerian Sahara regions—Case study Adrar. Renewable and Sustainable Energy Reviews. 2011; 15(9): 4436–4442. doi: 10.1016/j.rser.2011.07.108
27. Bediar S, Harrouz A, Belatrache D. Design and simulation of solar pumping system using PVsyst, case study: TSABIT–ADRAR in Algeria. Algerian Journal of Signals and Systems. 2024; 9(2): 47–56. doi: 10.51485/ajss.v9i2.210
28. Campana PE, Li H, Zhang J, et al. Economic optimization of photovoltaic water pumping systems for irrigation. Energy Conversion and Management. 2015; 95: 32–41. doi: 10.1016/j.enconman.2015.01.066