Edelweiss Applied Science and Technology
ISSN: 2576-8484
Vol. 9, No. 1, 278-291
2025
Publisher: Learning Gate
DOI: 10.55214/25768484.v9i1.4112

EEMD-GA-LSTM: An innovative approach for daily solar radiation estimation in southern of Algeria

Ahlam SENOUCI[1*], Ali BENATIALLAH[1], Djelloul BENATIALLAH[2], Kada BOUCHOUICHA[3], Rachid MAOUEDJ[4]
[1]Department of Material Sciences, Faculty of Material Sciences, Mathematics and Computer Science, Laboratory of Energy, Environment, and Information Systems, Ahmed Draia University Adrar, 01000 Adrar, Algeria; Ahlem.snc@gmail.com (A.S.).
[2]Material Sciences Department, Faculty of Material Sciences, Mathematics and Computer Science, Laboratory of Sustainable Development and Computer Science University of Adrar, Adrar, 01000, Algeria.
[3]Center for Renewable Energy Development (CDER), Bouzareah, 16340 Algiers, Algeria.
[4]Research Unit for Renewable Energies in Saharan region (URERMS), Renewable Energy Development Center (CDER), 01000 Adrar, Algeria.

Abstract: In a solar-rich area, like the Saharan climate in southern Algeria, to optimize all diversions of use of this energy to the maximum, it is necessary to accurately evaluate the radiation received from the sun. To improve prediction accuracy of daily solar radiation, we present in this paper a new synergistic model that combines three powerful techniques LSTM (Long Short-Term Memory Network), GA (Genetic Algorithm) and EEMD (Ensemble Empirical Mode Decomposition). EEMD is a technique that breaks down complex and highly non-linear solar radiation data into smaller and more manageable components to identify hidden trends, patterns. The GA is used to optimize hyperparameters of LSTM network so that time relevance can be well captured in the solar radiation data. The EEMD-GA-LSTM model was tested in several southern Algerian regions (Biskra, Tamanrasset, Adrar and Tindouf) with different climates. As compared to other existing models including ANN-GA, ANN-PSO, ANFIS-GA and ANFIS-PSO our method performed markedly good R^2 values and lower RMSE values (RMSE from 3.0125% to 1.554%). The findings underline the model's robustness and reliability in solar radiation prediction, providing valuable information for renewable energy assessment in arid and semi-arid regions. The current study shows the benefit of hybrid models combining metaheuristic optimization and deep learning in a complex environmental data set analysis, paving pathways for future work in both solar energy fields and climate prediction.
Keywords: *Genetic algorithm, Ensemble empirical mode decomposition, Long short-term memory network, Metaheuristic optimization.*

1. Introduction

As renewable energy is attracting increasing interest worldwide, solar energy has been regarded as one of the most prospective green power sources. In countries with high solar radiation like southern Algeria, accurate prediction of solar radiation is essential to optimally design and operate solar energy systems and their applications in agricultural and environmental management. During this, it is seen that solar radiation shows a discontinuous characteristic due to the variability caused by cloud cover as well as the weather itself which makes forecasting of solar radiation very complex matter [1, 2].

Traditional forecasting approaches include the application of both statistical techniques and conventional machine learning models [3] this often falls short in terms of achieving the levels of intricacy for which non-linear patterns are captured in solar radiation data. Though these models could work out the general trends, they do not perform well for dynamic variables and highly variable climates characterized by arid and semi-arid climates [4, 5]. However, there has lately been interest in the more advanced hybrid approaches that handle such complexities for more

History: Received: 11 November 2024; Revised: 23 December 2024; Accepted: 3 January 2025; Published: 7 January 2025
* Correspondence: Ahlem.snc@gmail.com

accurate and reliable predictions [6].

In recent years, hybrid models that combine data decomposition techniques, optimization algorithms, and deep learning have demonstrated significant potential in time-series forecasting [7, 8]. This paper introduces a novel hybrid model—Ensemble Empirical Mode Decomposition (EEMD), Genetic Algorithm (GA), and Long Short-Term Memory (LSTM)—referred to as EEMD-GA-LSTM. Each model component solves one problem in solar radiation forecasting:

- EEMD is a method used for analyzing complex time-series data into simpler IMFs that represent the underlying patterns and trends in the overall variation of solar radiation [9, 10]. GA: Genetic Algorithm is a model inspired by the principles of natural selection that optimizes important LSTM parameters, such as learning rate and sequence length, to improve performance [11, 12]. LSTM: This neural network is specialized in temporal data, and hence the LSTM model effectively grasped the temporal dependencies, which is best for solar radiation forecasting that exhibits complex time-based patterns [13, 14].

The EEMD-GA-LSTM integrates all of these techniques and benefits from their synergistic strengths to achieve more accurate and reliable solar radiation forecasting. In this paper, we applied the model to a number of locations in the south of Algeria, namely, Biskra, Tamanrasset, Adrar, and Tindouf, which have different climatic features. Compared to other models, ANN-GA, ANN-PSO, ANFIS-GA, and ANFIS-PSO [15] it showed that the best model was always the EEMD-GA-LSTM, with good values of R^2 and lower values of RMSE.

This will contribute to the development of the forecast of solar radiation and will show how important hybrid approaches are in capturing environmental patterns. Improved models, like EEMD-GA-LSTM, enhance renewable energy and climate-sensitive planning applications, especially over resource-rich but challenging regions such as the Sahara.

2. Tools and Methodology

2.1. Regions of Study and Data Acquisition Process

Several locations were selected in southern Algeria due to its high solar energy potential and particular climatic condition to assess the performance of the proposed EEMD-GA-LSTM model in forecasting solar radiation. In fact, this region of the country has an arid and semi-arid climate with particular environmental conditions that strongly affect the pattern of solar radiation, thus the correctness of the forecast will be very important for a better exploitation of solar energy.

2.1.1. Study Areas

Four major locations are focused on: Biskra, Tamanrasset, Adrar, and Tindouf. These areas represent diverse environmental characteristics in the Algerian Sahara and thus enable the model to be tested in various climatic conditions [16, 17]:

- Biskra: It is situated in northeastern Algeria. It possesses a hot desert climate with strong solar radiation throughout the year. Its summer temperatures are very high while winters are moderate and rainfall is received rarely, thus making it suitable for solar radiation variability studies [18].
- Tamanrasset: This city is located in southern Algeria in the Hoggar Mountains. Therefore, Tamanrasset has a dry, high-desert climate with large diurnal temperature differences and small cloud cover. The prevailing conditions of solar radiation are quite different in comparison with the others [15, 19].
- Adrar: Extreme arid conditions in the central part of the Algerian Sahara characterize the region, with high solar radiation throughout the year. It is also known to attain some of the highest records as far as insolation rates are concerned; hence it is a useful dataset for assessing solar radiation patterns [15, 20].

ISSN: 2576-8484
Vol. 9, No. 1: 278-291, 2025
DOI: 10.55214/25768484.v9i1.4112

- Tindouf: It is a western part of Algeria. Its arid desert climate makes it highly relevant for studies in solar energy forecasting due to the limited annual rainfall and high levels of radiation received from the sun [15].

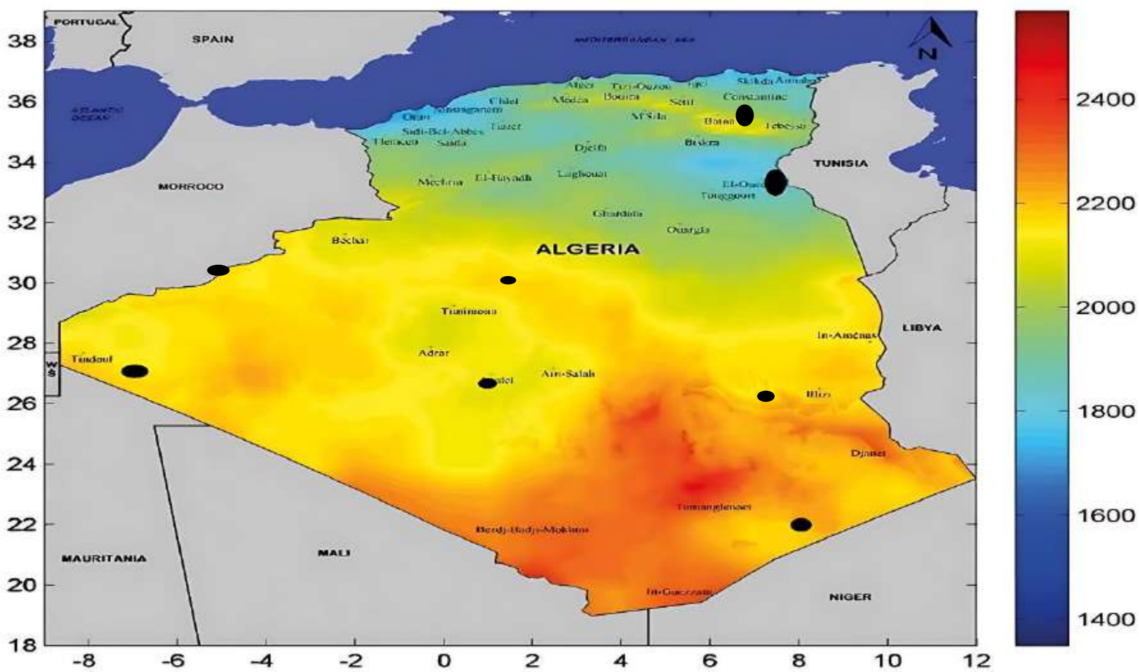


Figure 1.
Algeria's solar potential [21].

2.1.2. Data Collection

For each site shown, daily solar radiation data were collected over a number of years to capture both seasonal and annual fluctuations. The data were primarily obtained from the weather stations run by the Algerian National Meteorological Office.

The dataset includes all the major meteorological variables, which are described as follows:

- Temperature (°C): The data collection considered the average day temperatures because the changes in temperature affect solar radiation since it alters the formation of clouds and the clarity of the atmosphere .
- Rel. Humidity (%): The relative humidity data is essential in understanding the concentrations of moisture in the atmosphere, which have the potential to reduce solar radiation as it passes through the Earth's atmosphere.
- Solar Declination (°): the angle between the sun's rays and the plane of the Earth's equator at any given time of the year that impacts the duration and intensity of sunlight.

The solar hour angle, in degrees, refers to the sun's position with respect to solar noon and is one of the important factors determining radiation variability in intensity at any given location.

Extraterrestrial solar irradiation (Wh/m²) is a reference metric representing the amount of solar energy available outside the Earth's atmosphere that allows for modeling of possible solar gain in the absence of atmospheric disruptions.

The data for the wilaya of Biskra spans from 2010 to 2022, while the data for the wilayas of Adrar, Tamanrasset, and Tindouf covers the period from 2016 to 2021.

2.2. Data Preprocessing

Normalization of the variables between 0 and 1 was one of the most important steps in optimizing the EEMD-GA-LSTM model [13]. The dataset was divided into 80% for training and 20% for validation, which helped in testing the robustness of the model across different climatic conditions of southern Algeria and its generalizability to other arid and semi-arid regions around the world

The normalization usually is a common practice to avoid the variables having large amplitudes from dominating the variables with smaller magnitudes, so the learning algorithm can be misled. This process ensures that all variables contribute equitably to the model's training using the following normalization equation [22]:

$$I_{nor} = \frac{I_{non-norm} - I_{min}}{I_{max} - I_{min}} \quad (1)$$

Table 1.
Key features and training specifications of the EEMD-GA-LSTM model.

Category	Details
Input Data	Daily solar radiation data (Biskra, Tamanrasset, Adrar, Tindouf).
Preprocessing	EEMD to decompose data into IMFs.
Optimization	Genetic Algorithm (GA) for hyperparameter tuning.
Model	LSTM with optimized architecture (layers, neurons).
Metrics	R^2, RMSE and MAE
Hyperparameters	Tuned via GA: layers, neurons, learning rate, batch size.
Training	80% training, 20% testing; 100–200 epochs with early stopping.

3. Predictive Performance Indicators

Evaluation metrics widely used in the literature were applied to determine the performance of the methods under assessment [23]. The indices include the Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2), which are defined below:

3.1. Mean Absolute Error (MAE) [24]:

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

Where y_i represents the experimental values $\hat{y}_i$ the predicted values, and n the total number of observations. MAE quantifies the average magnitude of errors in a set of predictions.

3.2. Root Mean Square Error (RMSE) [25]:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (3)$$

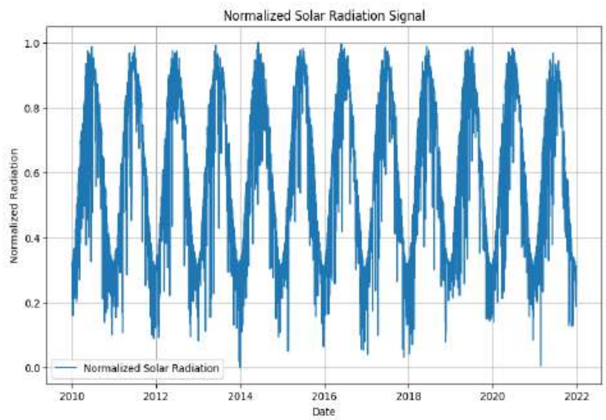
RMSE measures the square root of the average squared differences between predicted and observed values, emphasizing larger errors.

3.3. Coefficient of Determination (R^2) [26]:

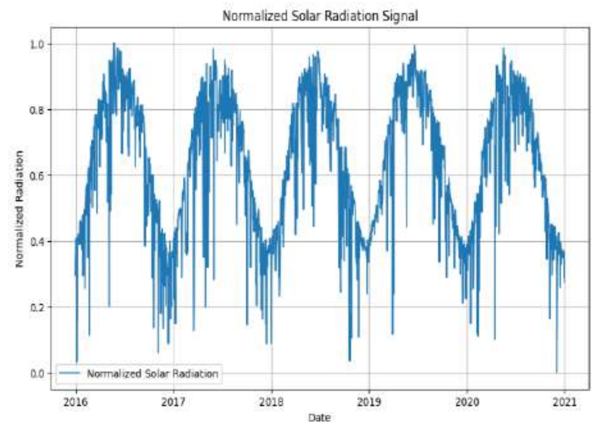
$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \tag{4}$$

Where $\bar{y}_i$ is the mean of observed values. R^2 represents the proportion of variance in the observed data explained by the model; the closer it is to 1, the better the predictive performance of the model is.

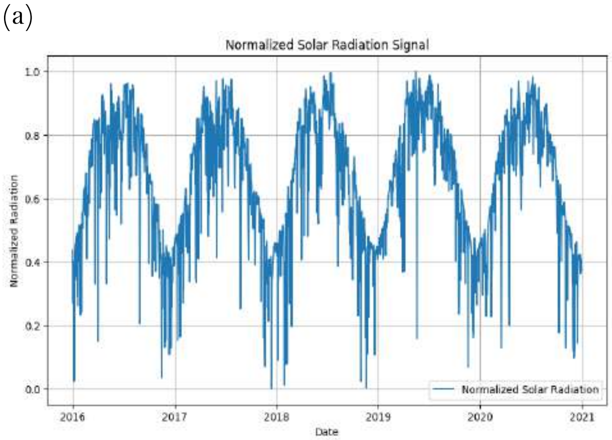
4. Results and Discussion



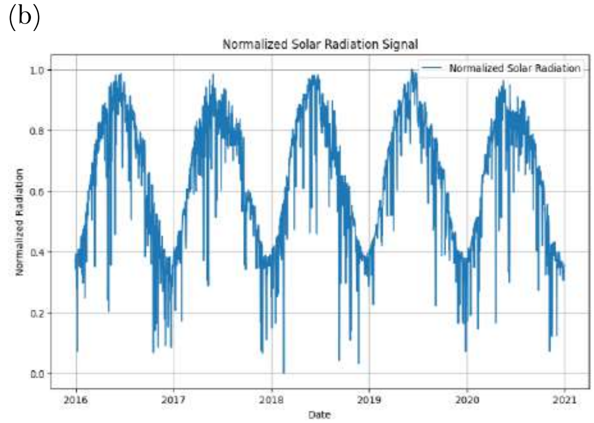
(a)



(b)

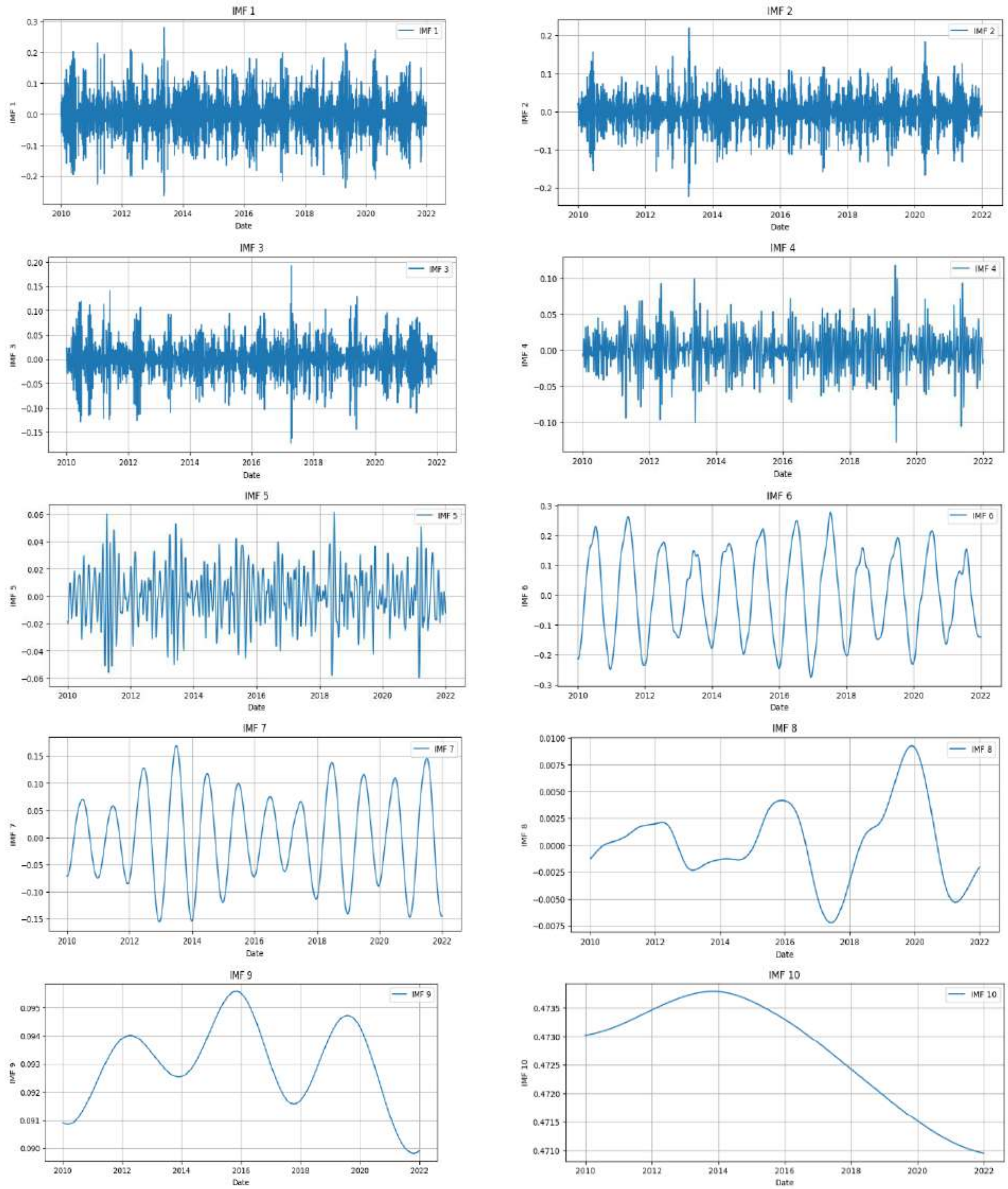


(c)



(d)

Figure 2.
Normalized Solar Radiation Signal in the Regions (a) Biskra, (b) Adrar, (c) Tamanrasset, and (d) Tindouf.



DOI: 10.55214/25768484.v9i1.4112

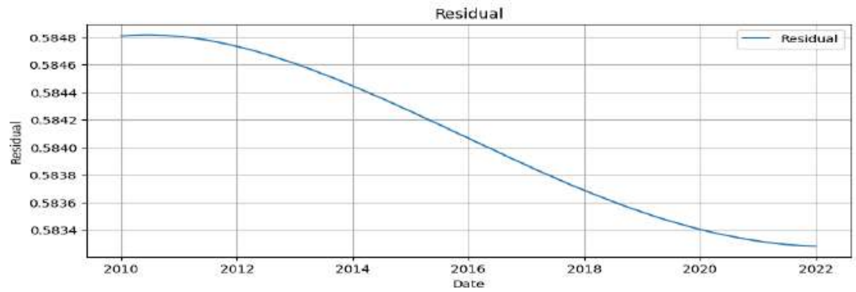
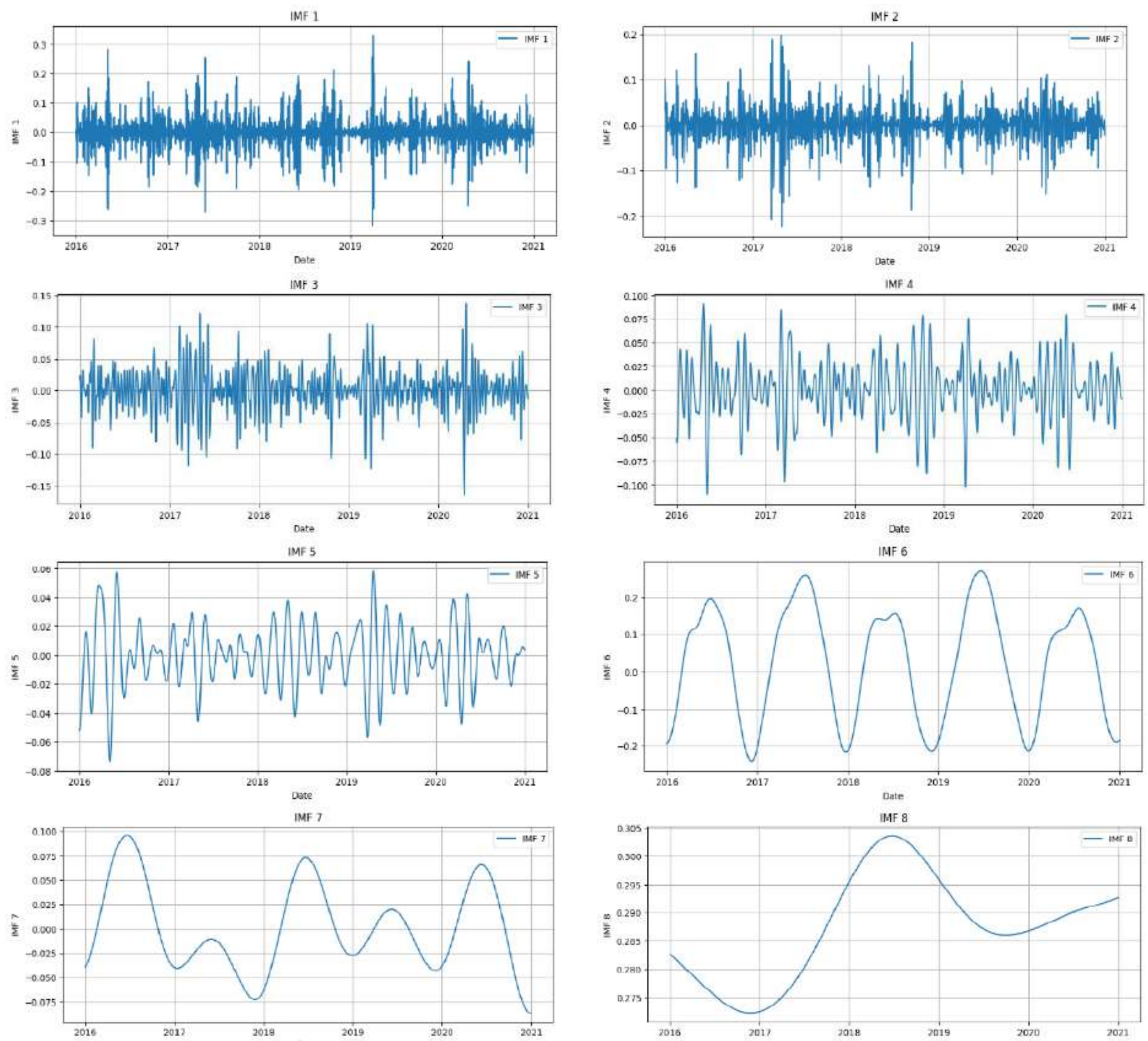


Figure 3.
EEMD Decomposition Results in the Biskra Region.



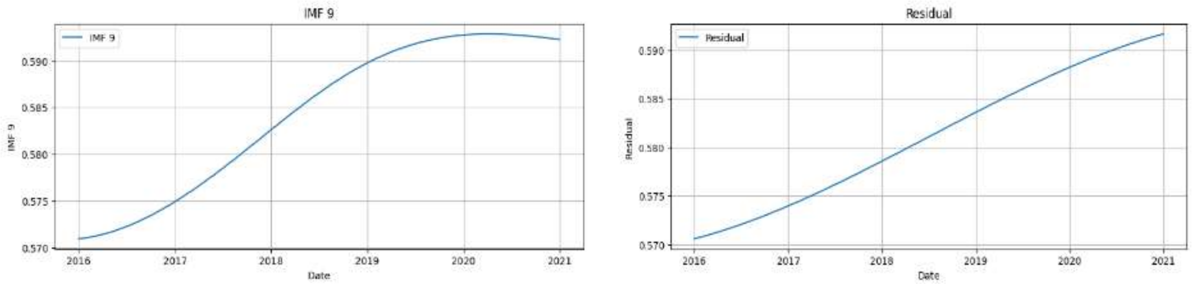
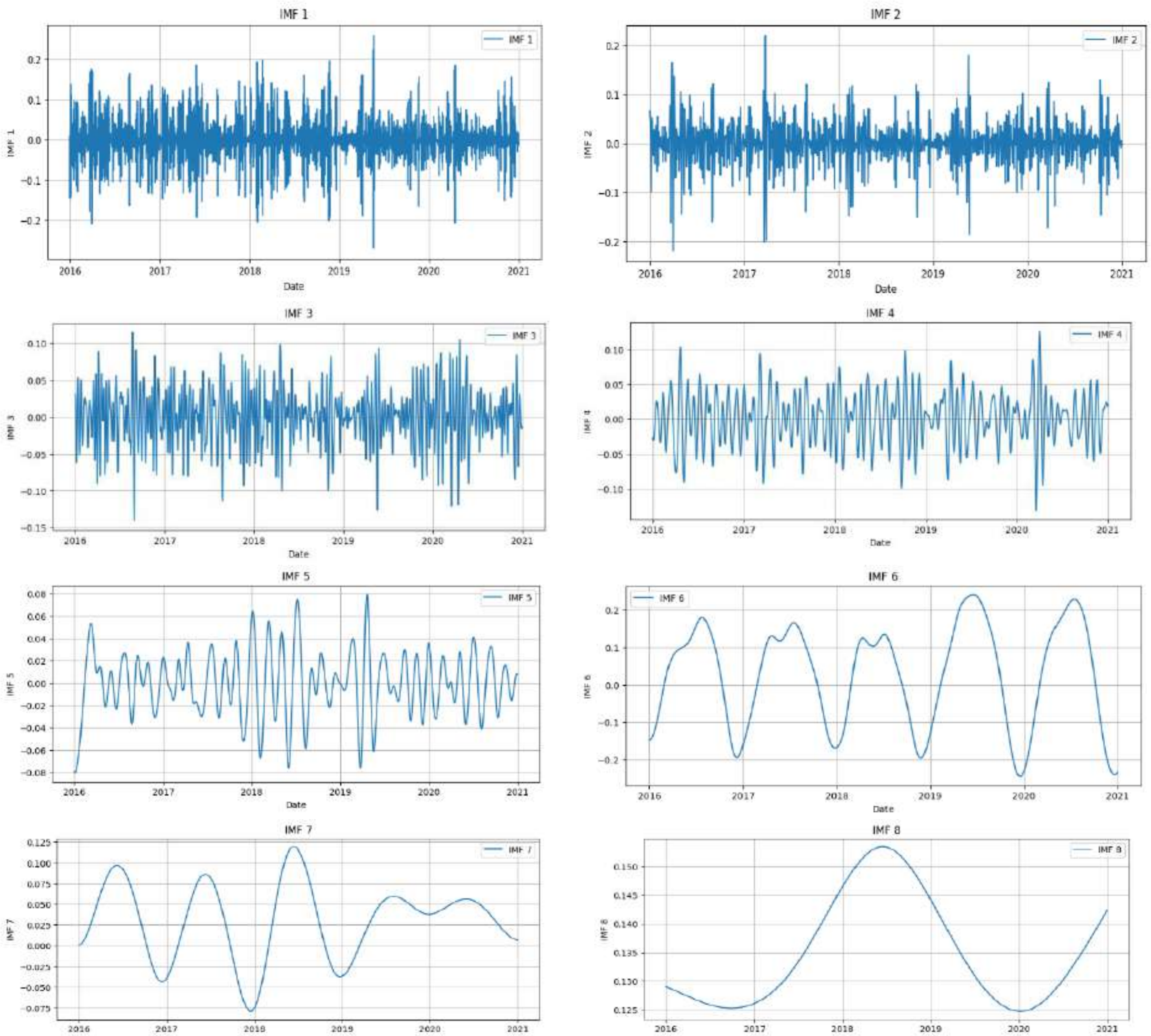


Figure 4.
EEMD Decomposition Results in Adrar Region.



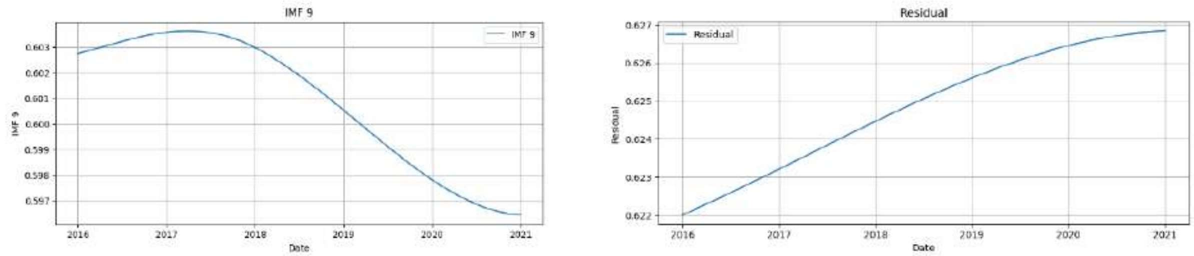


Figure 5.
EEMD Decomposition Results in Tamanrasset Region.

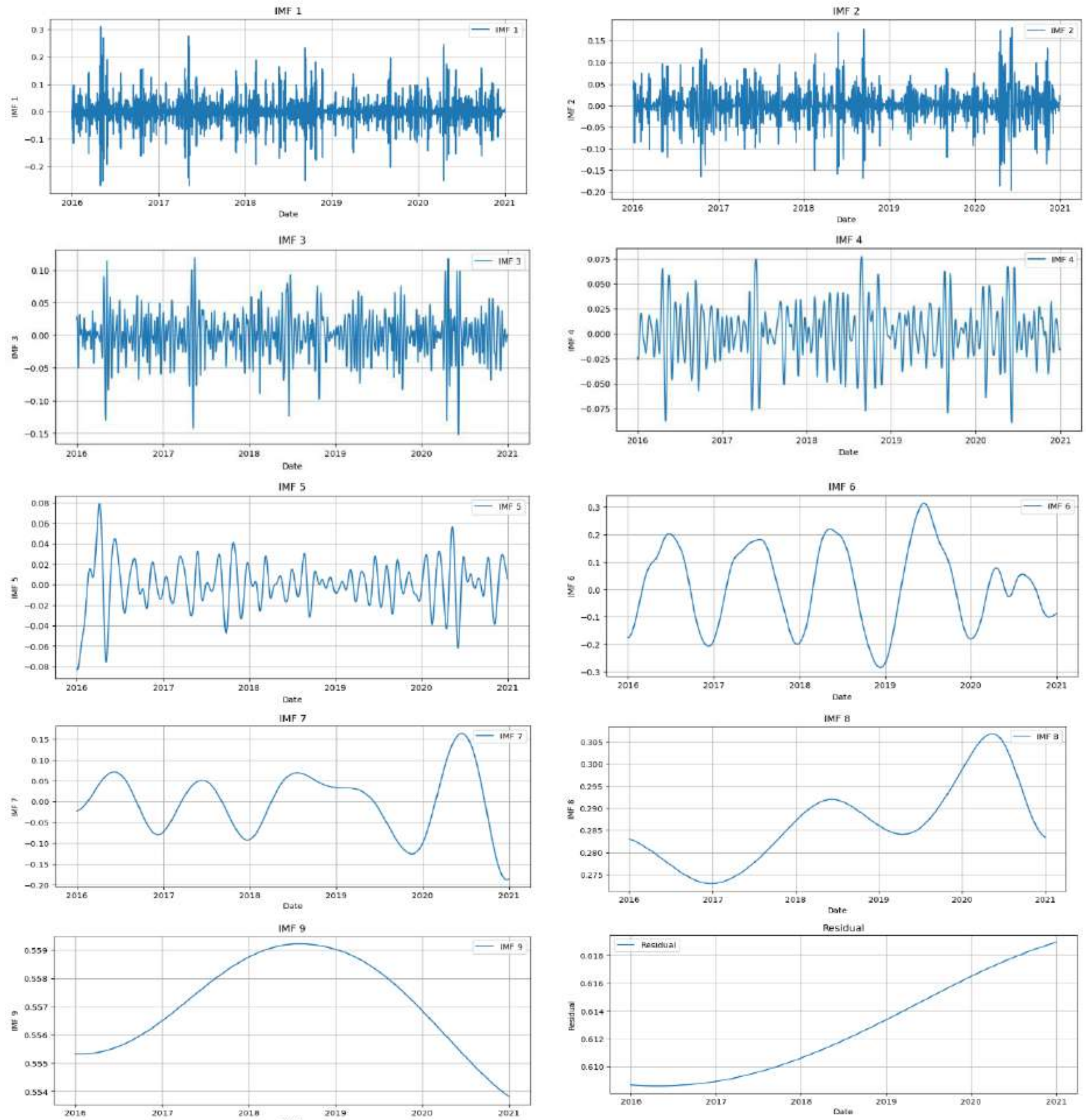


Figure 6.
EEMD Decomposition Results in Tindouf Region.

These results of EEMD decomposition give a deep insight into the dynamics of solar radiation in the studied regions. This would not only enhance the accuracy of the future predictions but also be very important to get valuable insights into the underlying mechanisms controlling solar radiation variations, including the impacts of local climatic conditions, seasonal cycles, and specific geographical features. This approach facilitates the development of more robust models, which are tailored to include the characteristics of each region and thereby increase their applicability in such fields as solar energy and environmental management.

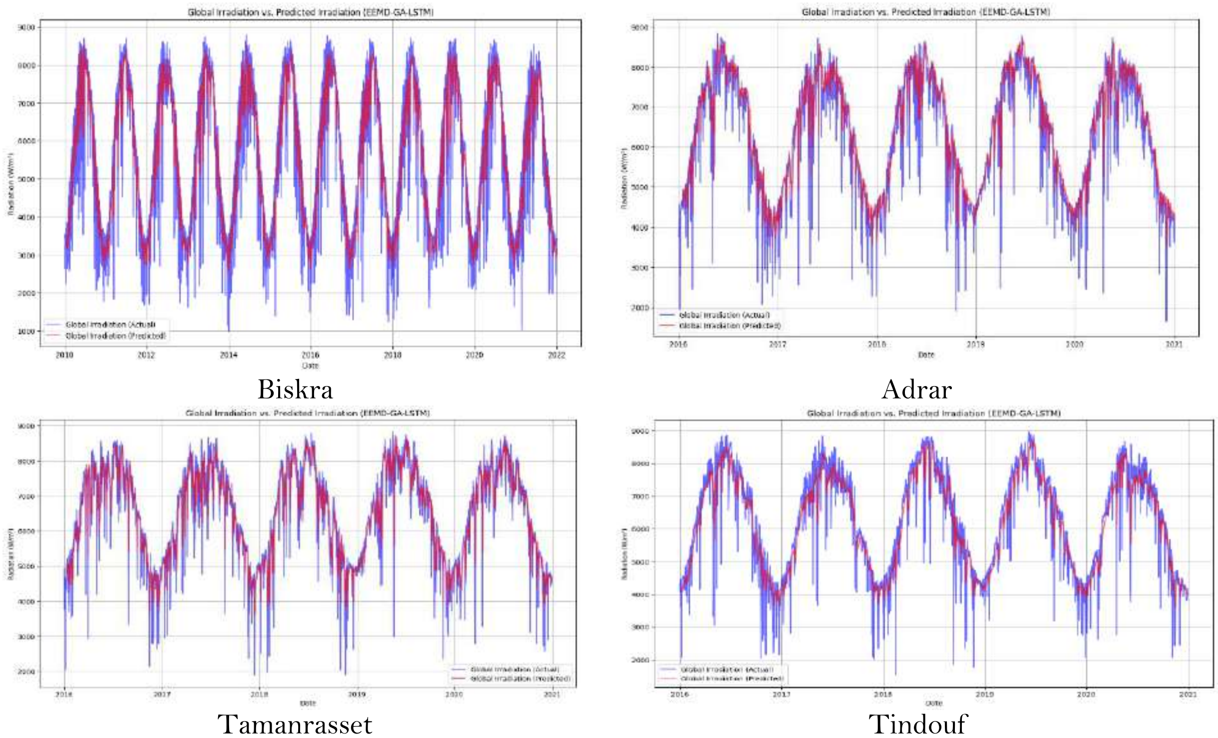


Figure 7.
Observed vs. Predicted Global Irradiation Using the EEMD-GA-LSTM Approach in the Regions of Biskra, Adrar, Tamanrasset, and Tindouf.

Figure 7 presents a comparison between the observed and predicted solar radiation values obtained using the EEMD-GA-LSTM method in the four analyzed regions: Biskra, Adrar, Tamanrasset, and Tindouf. This figure provides a visual assessment of the model's prediction accuracy by contrasting the predicted values with the actual data.

Biskra: The predicted and observed values are very close to each other, indicating high accuracy of the model in this region. The prediction graphs and actual observations almost coincide with each other; therefore, the EEMD-GA-LSTM model is capable of successful solar radiation fluctuation capture. These results confirm the previously determined performance indicators (RMSE, MAE, and R^2), which highlighted the model's robustness in Biskra.

In the regions of Adrar, Tamanrasset, and Tindouf, observed and expected values are in good agreement. Slight differences appear, however, at places that can be explained by more variability in climatic conditions; even so, the accuracy of predictions is quite high, and the model keeps fitting well to data coming from these regions.

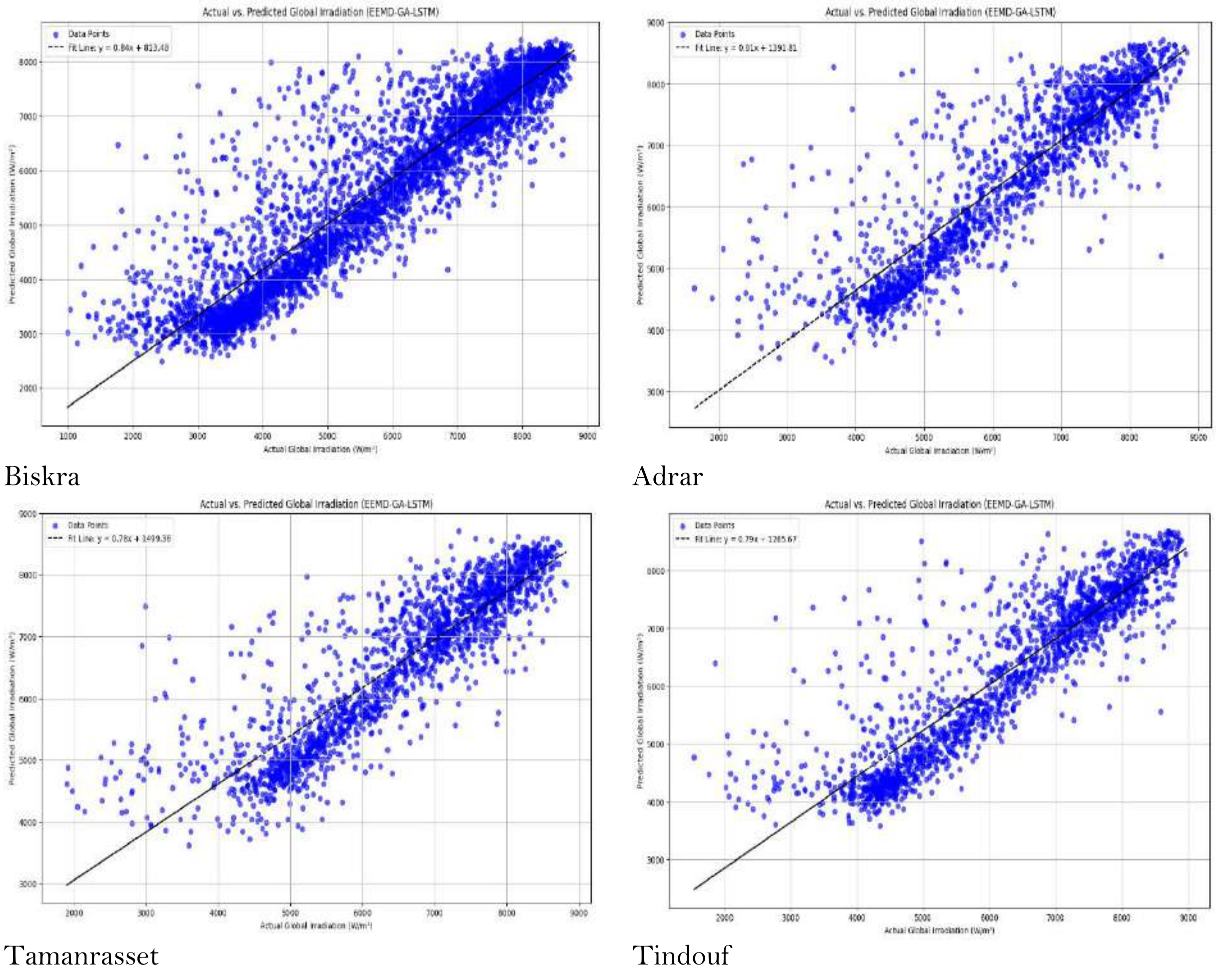


Figure 8.
Predicted Irradiation Using the EEMD-GA-LSTM Approach vs. Actual Irradiation in the Regions of Biskra, Adrar, Tamanrasset, and Tindouf.

Figure 8 represents the correlation between the predicted and actual values of solar radiation based on the EEMD-GA-LSTM approach for the regions of Biskra, Adrar, Tamanrasset, and Tindouf. This kind of graphical representation allows one to evaluate the model's accuracy by checking the extent of correspondence between the predictions and the actual true values. Ideally, the perfect model would have data points shown in the graph falling exactly on the regression line with a slope of 1, meaning a perfect match of the realized and predicted values.

Biskra: In this area, most of the data points are located close to the optimal regression line, which means that the EEMD-GA-LSTM model gives an optimal estimation for solar radiation values with high accuracy. Data points in this area scatter strongly along the diagonal, showing that the differences between observed and predicted values are relatively small. This fact also confirms the model's good performance in this area, confirmed by the above figures.

In the regions of Adrar, Tamanrasset, and Tindouf, most data points are on or close to the regression line, although some deviate further, indicating prediction errors for some cases. However, the relation between observed and predicted values is still strong, showing that the model keeps working well but with a little decrease compared to Biskra.

Table 2.
Statistical indicators result.

	Biskra	Adrar	Tamanrasset	Tindouf
RMSE %	1.554	1.478	1.246	1.423
MAE %	1.074	0.985	0.828	0.786
R^2	0.986	0.964	0.976	0.972

Table 3.
A comparison of the performance of the presented models with results from various studies.

Ref.	model	location	R
[15]	ANN-GA	Adrar	0.9962
	ANN-PSO	Adrar	0.9929
	ANFIS-GA	Adrar	0.9210
	ANFIS-PSO	Adrar	0.9523
	ANN-GA	Tamanrasset	0.9887
	ANN-PSO	Tamanrasset	0.9987
	ANFIS-GA	Tamanrasset	0.8425
	ANFIS-PSO	Tamanrasset	0.9877
	ANN-GA	Tindouf	0.9998
	ANN-PSO	Tindouf	0.9999
	ANFIS-GA	Tindouf	0.9866
	ANFIS-PSO	Tindouf	0.9986
This study	EEMD-GA-LSTM	Biskra-Adrar-Tamanrasset -Tindouf	0.964 - 0.986

The calculated evaluation index (R) is applied to evaluate the predictive performance of the studied methodologies. The results obtained in the present study show a comparative effectiveness with other methods regarding the accuracy and forecasting ability. Moreover, this model does not require a large variation of data and gives good results even with small data. As a result, it is particularly suited to regions where the meteorological databases are not very large because it can function with very minimal data.

5. Conclusion

In summary, the proposed EEMD-GA-LSTM model was an efficient way to conduct solar radiation prediction in these four regions of Biskra, Tamanrasset, Adrar, and Tindouf. The performance of the model, evaluated using the calculated assessment index (R), is of high accuracy and reliability, which is comparable to other methodologies with respect to its precision and forecasting capabilities. Notably, the model performed well even with limited data, which underlines its robustness in handling regions with a lack of meteorological information. Most notably, the model performed almost perfectly for Biskra when validating the observed and predicted values against each other. This results in minor divergences for Adrar, Tamanrasset, and Tindouf, justified by the higher variability of the meteorological condition of these locations. However, it remained highly predictable with a very high degree of accuracy.

This therefore qualifies the EEMD-GA-LSTM model for the most suitable in regions with limited meteorological infrastructure since the performance was with limited data. This characteristic is essential for applications in remote or data-scarce areas, where reliable solar radiation estimates are critical for energy planning and other applications. Overall, the results confirm the model's potential as a powerful tool for solar radiation estimation, with broad applicability in regions with limited access to detailed meteorological data.

Copyright:

References

[1] A. B. Stambouli, Z. Khiat, S. Flazi, and Y. Kitamura, "A review on the renewable energy development in Algeria: Current perspective, energy scenario and sustainability issues," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 7, pp. 4445-4460, 2012. https://doi.org/10.1016/j.rser.2012.04.031

[2] Y. Zahraoui, M. R. B. Khan, I. AlHamrouni, S. Mekhilef, and M. Ahmed, "Current status, scenario, and prospective of renewable energy in algeria: A review," *Energies*, vol. 14, no. 9, p. 2354, 2021. https://doi.org/10.3390/en14092354

[3] S. Makridakis, E. Spiliotis, V. Assimakopoulos, A.-A. Semenoglou, G. Mulder, and K. Nikolopoulos, "Statistical, machine learning and deep learning forecasting methods: Comparisons and ways forward," *Journal of the Operational Research Society*, vol. 74, no. 3, pp. 840-859, 2023. https://doi.org/10.1080/01605682.2022.2118629

[4] E. Spiliotis, S. Makridakis, A.-A. Semenoglou, and V. Assimakopoulos, "Comparison of statistical and machine learning methods for daily SKU demand forecasting," *Operational Research*, vol. 22, no. 3, pp. 3037-3061, 2022. https://doi.org/10.1007/s12351-020-00605-2

[5] T. Ahmad and H. Chen, "A review on machine learning forecasting growth trends and their real-time applications in different energy systems," *Sustainable Cities and Society*, vol. 54, p. 102010, 2020. https://doi.org/10.1016/j.scs.2019.102010

[6] A. Tascikaraoglu and M. Uzunoglu, "A review of combined approaches for prediction of short-term wind speed and power," *Renewable and Sustainable Energy Reviews*, vol. 34, pp. 243-254, 2014. https://doi.org/10.1016/j.rser.2014.03.033

[7] H.-F. Yang and Y.-P. P. Chen, "Hybrid deep learning and empirical mode decomposition model for time series applications," *Expert Systems with Applications*, vol. 120, pp. 128-138, 2019. https://doi.org/10.1016/j.eswa.2018.11.019

[8] S.-X. Lv and L. Wang, "Deep learning combined wind speed forecasting with hybrid time series decomposition and multi-objective parameter optimization," *Applied Energy*, vol. 311, p. 118674, 2022. https://doi.org/10.1016/j.apenergy.2022.118674

[9] P. Hawinkel, E. Swinnen, S. Lhermitte, B. Verbist, J. Van Orshoven, and B. Muys, "A time series processing tool to extract climate-driven interannual vegetation dynamics using ensemble empirical mode decomposition (EEMD)," *Remote Sensing of Environment*, vol. 169, pp. 375-389, 2015. https://doi.org/10.1016/j.rse.2015.08.024

[10] S.-Y. Wang, J. Qiu, and F.-F. Li, "Hybrid decomposition-reconfiguration models for long-term solar radiation prediction only using historical radiation records," *Energies*, vol. 11, no. 6, p. 1376, 2018. https://doi.org/10.3390/en11061376

[11] H. Chung and K.-s. Shin, "Genetic algorithm-optimized long short-term memory network for stock market prediction," *Sustainability*, vol. 10, no. 10, p. 3765, 2018. https://doi.org/10.3390/su10103765

[12] B. Rogers, N. Noman, S. Chalup, and P. Moscato, "A comparative analysis of deep neural network architectures for sentence classification using genetic algorithm," *Evolutionary Intelligence*, vol. 17, no. 3, pp. 1933-1952, 2024. https://doi.org/10.1007/s12065-023-00874-8

[13] A. Gupta and K. Gupta, "Short term solar irradiation prediction framework based on EEMD-GA-LSTM Method," *Strategic Planning for Energy and the Environment*, pp. 255-280, 2022. https://doi.org/10.13052/spee1048-5236.4132

[14] Y. Chen *et al.*, "Short-term wind speed predicting framework based on EEMD-GA-LSTM method under large scaled wind history," *Energy Conversion and Management*, vol. 227, p. 113559, 2021. https://doi.org/10.1016/j.enconman.2020.113559

[15] S. Ahlam, M. Rachid, B. Ali, B. Djelloul, and B. Kada, "Enhanced daily global solar radiation prediction through hybrid artificial neural network and adaptive neuro-fuzzy inference system with meta-heuristic algorithm integration," *Instrumentation, Mesures, Métrologies*, vol. 22, no. 6, p. 241, 2023. https://doi.org/10.18280/i2m.220603

[16] S. Semahi, M. A. Benbouras, W. A. Mahar, N. Zemmouri, and S. Attia, "Development of spatial distribution maps for energy demand and thermal comfort estimation in Algeria," *Sustainability*, vol. 12, no. 15, p. 6066, 2020. https://doi.org/10.3390/su12156066

[17] S. Semahi, "Development of a decision-support model based on fuzzy logic for optimizing of high energy performance (HPE) housing design in Algeria," Doctoral Dissertation, Université Mohamed Khider–Biskra, 2021.

[18] B. Selahdja, "The geographical distribution of prehistoric sites in the region of Biskra," *Journal El-Baheth in Human and Social Sciences*, vol. 15, no. 1, pp. 135-147, 2024.

[19] A. D. Hidaoui, A. K. Kerroumi, and D. Benatiallah, "Development of a geographic information system to support decision-making in the agricultural sector," Doctoral Dissertation, Universite Ahmed Draia-Adrar, 2023.

[20] D. Benatiallah, K. Bouchouicha, A. Benatiallah, A. Harrouz, and B. Nasri, "Forecasting of solar radiation using an empirical model," *Algerian Journal of Renewable Energy and Sustainable Development*, vol. 1, no. 2, pp. 212-219, 2019.

[21] N. Aoun and K. Bouchouicha, "Estimating daily global solar radiation by day of the year in Algeria," *The European Physical Journal Plus*, vol. 132, pp. 1-12, 2017. https://doi.org/10.1140/epjp/i2017-11495-7

[22] A. Khosravi, R. Koury, L. Machado, and J. Pabon, "Prediction of wind speed and wind direction using artificial neural network, support vector regression and adaptive neuro-fuzzy inference system," *Sustainable Energy Technologies and Assessments*, vol. 25, pp. 146-160, 2018. http://doi.org/10.1016/j.seta.2018.01.001

[23] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *International Journal of Data Mining & Knowledge Management Process*, vol. 5, no. 2, p. 1, 2015. https://doi.org/10.5121/ijdkp.2015.5201

[24] G. López, F. Batlles, and J. Tovar-Pescador, "Selection of input parameters to model direct solar irradiance by using artificial neural networks," *Energy*, vol. 30, no. 9, pp. 1675-1684, 2005. http://doi.org/10.1016/j.energy.2004.04.035
[25] S. Pereira, P. Canhoto, and R. Salgado, "Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data," *Energy and AI*, vol. 15, p. 100314, 2024. https://doi.org/10.1016/j.egyai.2023.100314
[26] B. Güzel, O. Sevli, and E. Okatan, "Forecasting solar radiation based on meteorological data using machine learning techniques: A case study of Isparta," *International Journal of Engineering Research and Development*, vol. 15, no. 2, pp. 704-713, 2023. https://doi.org/ 10.29137/umagd.1268055