International Journal of Engineering Inventions
e-ISSN: 2278-7461, p-ISSN: 2319-6491
Volume 14, Issue 2 [February. 2025] PP: 82-89

Adaptive approach of Multilayer Perceptron Neural Network for Global Irradiation Prediction: The Case of Madagascar

Nomenjanahary Tantely Rakotondraibe, Solofo Hery Rakotoniaina, Harlin Samuel Andriatsihoarana
Doctoral School in Engineering Sciences, Techniques and Innovation, University of Antananarivo, Madagascar
Corresponding Author: dee.meh.sth@gmail.com

ABSTRACT:
This research focuses on adapting Multilayer Perceptron (MLP) neural networks for the prediction of global solar irradiation in Madagascar, emphasizing the integration of temporal variables as explanatory factors. Madagascar, endowed with substantial solar potential but facing limited electrification, requires innovative approaches to harness this resource effectively. Accurate solar irradiation forecasting is essential to optimize photovoltaic system management and support the energy transition. In this study, four years of global solar irradiation data collected via Solcast were utilized. After rigorous preprocessing, including data normalization and reformulation of temporal variables using trigonometric bases (sine and cosine), an MLP model was trained and tested. The dataset was split into 70% for training and 30% for testing to evaluate the model's robustness across three strategic sites: Ambatolampy, Betainomby, and Ankatso. The results demonstrate a significant improvement in the MLP model's performance with the integration of temporal variables. Mean Absolute Percentage Error (MAPE) and Normalized Root Mean Square Error (nRMSE) were notably reduced, while the coefficient of determination (R^2) reached high levels, exceeding 0.91 in several cases. These performances, particularly pronounced for medium-term prediction horizons (1 hour) and the short-term prediction horizon (5 minutes), highlight the importance of tailoring models to local geographic and temporal specificities. This study offers promising perspectives for generalizing the methodology and developing operational tools for solar plant management and energy planning.

Keywords: *Solar irradiation forecasting, Multilayer Perceptron, temporal variables, meteorological and geographical factors.*

Date of Submission: 07-02-2025 Date of acceptance: 18-02-2025

I. INTRODUCTION

The development of renewable energies, particularly solar energy, represents a strategic priority for Madagascar, a country characterized by significant solar potential and increasing electricity demand [1]. With a low electrification rate and high dependence on fossil fuels [2], effectively harnessing solar energy offers a viable solution to meet the nation's energy needs while promoting a transition toward a sustainable energy system.

However, the variability of solar irradiation, influenced by meteorological and geographical factors, poses a major challenge for the optimal integration of this resource into the power grid [3]. Accurate solar irradiation forecasting is therefore essential to:

- Improve the management and planning of photovoltaic systems;
- Reduce uncertainties related to the balance between energy supply and demand;
- Optimize the performance of solar power plants connected to the grid.

This article is part of ongoing research on global solar irradiation prediction for Madagascar. The primary objective is to develop and evaluate a precise and robust prediction model for global solar irradiation tailored to Madagascar’s specific context. Using a Multilayer Perceptron (MLP), an advanced architecture of artificial neural networks, this study utilizes meteorological data enriched with parameters, including temporal resolutions down to 5 minutes.

A distinctive feature of this research is the integration of innovative temporal parameters into the neural network inputs, such as seasonal, daily, and hourly indicators, which are rarely considered in previous studies. This approach aims to capture Madagascar's unique geographical and climatic specificities while maximizing prediction accuracy.

This work contributes to improving modeling methodologies in renewable energy by envisioning direct applications to optimize solar energy integration into Madagascar's electrical grids and supporting the country's energy transition. To achieve this, three strategic sites were selected: **Ambatolampy** (19∘29′24.78″S, 47∘26′38.07″E): This site hosts a 40 MW solar power plant interconnected with the Antananarivo grid. It is a critical example of large-scale solar energy integration, enabling analysis of real-world performance and variability impacts on production. **Betainomby** (18°10'9.97"S, 49°23'13.15"E): Located in Toamasina, this 2 MW plant is connected to the coastal region's grid in Eastern Madagascar. Despite its smaller size, it provides insights into solar energy integration in areas with variable sunlight and weather conditions. The site illustrates the challenges of interconnection in regions distant from the main grid. **Ankatso Laboratory** (18°54'56.69"S, 47°33'34.08"E): This site with a capacity of 3kWp, at the solar laboratory of the Polytechnic School Research Center at the University of Antananarivo, is pivotal for high-quality data collection due to its advanced measurement equipment and its role in renewable energy academic research.

These sites offer a comprehensive base to test theoretical hypotheses and predictive models in a controlled setting. They enable a thorough analysis of solar irradiation across varied contexts, from large interconnected infrastructures to smaller research-focused installations.

II. METHODOLOGY

Data used

The data used and considered as measurement data in this study were obtained from **Solcast** [4], a specialized platform providing high-resolution satellite-based meteorological data available in real-time and for short-term intervals (ranging from hourly to 5-minute steps).

The parameters used are: G_H Global Horizontal irradiation [Wh/m^2]: total solar energy received per unit area. η : Cloud opacity [%] is the cloud cover can attenuate solar irradiation. P Atmospheric pressure [Pa], influences the propagation of solar radiation in the atmosphere. P_r : Precipitation [mm] can reduce solar irradiation by creating clouds or humidity in the air. Rh : Relative humidity [%] is an important factor for determining the air's capacity to absorb moisture and its impact on the diffusion of solar rays. w_S: Wind speed [m/s], is an indirect factor affecting evaporation, and therefore cloud cover and humidity. Zenith θ_z , is the parameters related to the position of the sun, influencing the angle of incidence of solar radiation. T: Temperature [°].

These data enable the construction of time series, which form the foundation for our prediction model. Given the lack of meteorological stations in our regions, the use of satellite data serves as an effective solution to address this gap, providing precise information on local climatic conditions. In our case, we had access to four years of data, spanning from January 1, 2020, to December 31, 2023.

For the simulation, we used Matlab® software and its neural network toolbox.

Data preprocessing

To ensure accurate predictions and minimize the impact of low irradiation values, the data were filtered by removing observations recorded before 07:00:00 and after 17:00:00. This timeframe corresponds to periods of low irradiation where variations are not significant for modeling. Consequently, only 11 hours per day were retained, consistent with the reality of tropical solar cycles in Madagascar. This preprocessing step reduces noise in the data and optimizes model training, as recommended in several similar studies [5].

Data normalization

Normalization was a crucial step to enhance the performance of the Multi-Layer Perceptron (MLP) model. This process prepared the global solar irradiation data to be compatible with the neural network by ensuring a uniform scale for all input variables, including measured global irradiation, exogenous meteorological variables, and temporal variables.

Data normalization is performed for each dataset of global irradiation measurements. The formula applied to normalize the global irradiation $y(t)_{(norm)}$, is as follows [6]:

$$y(t)_{(norm)} = \frac{y(t) - y(t)_{min}}{y(t)_{max} - y(t)_{min}} \tag{1}$$

Where $y(t)$ represents the value of global irradiation at time t, $y(t)_{max}$ et $y(t)_{min}$ are respectively the minimum and maximum values observed in the global irradiation data series;

This normalization eliminates biases caused by differing scales of input data, which is particularly beneficial for the MLP model, as it can be sensitive to magnitude differences among input variables. Additionally, it ensures greater stability during training, accelerates convergence, and enhances prediction quality.

Training and testing phases

The data used for training and testing were sourced from four years of Solcast records. To evaluate the robustness and generalization ability of the model, the dataset was randomly split into two subsets: 70% of the data were used for training the model, while 30% were reserved for testing its performance. This partitioning ensures that the model does not overfit the training data and provides a more reliable assessment of its ability to predict solar irradiation for unobserved periods.

Data denormalization

After the model generates predictions on the normalized data, it is necessary to denormalize the results to restore them to their original scale. The formula for denormalizing the predicted global solar irradiation $y(t)_{(denorm)}$ is as follows:

$$y(t)_{(denorm)} = \left[y(t)_{(norm)} * (y(t)_{max} - y(t)_{min})\right] + y(t)_{min} \tag{2}$$

Thus, $y(t)_{(denorm)}$ represents the predicted value of global solar irradiation on the original scale, enabling direct comparison with the measured data.

Selection of the MLP model (Multi-Layer Perceptron)

The MLP was chosen as the prediction model for several reasons: ***Non-linear flexibility:*** Unlike traditional linear models such as persistence or ARMA (Autoregressive Moving Average), the MLP can capture non-linear relationships between input variables and outputs [7]**.** ***Empirical validation***: Numerous studies have demonstrated the effectiveness of MLPs for time series modeling in various fields, including solar irradiation prediction [8, 9]. ***Simplicity and popularity***: Although many neural network architectures are available, the MLP remains widely used due to its ease of implementation and efficiency in prediction tasks [10].

A classical MLP network with one hidden layer and one output layer was employed. The mathematical relationship can be expressed as follows:

$$\hat{y}(t) = f\left(\sum_{j=1}^{H} w_j^{(2)}\, g\left(\sum_{i=1}^{I} w_{ij}^{(1)}\, x_i(t) + b_j^{(1)}\right) + b^{(2)}\right) \tag{3}$$

$x_i(t)$: The inputs at time *t*, including past values of irradiation and exogenous variables (meteorological and temporal), H: The number of neurons in the hidden layer, $w_{ij}^{(1)}$ and $w_j^{(2)}$: The biases associated with the hidden layer and the output layer, respectively, $b_j^{(1)}$ et $b^{(2)}$: the associated biases, g and f : activation functions in the hidden layer and the output layer, respectively.

The input variables include: the endogenous variables, which are the past values of global solar irradiation, and the exogenous variables, which consist of meteorological parameters provided by Solcast and temporal parameters such as hour, day, and month, transformed into sinusoidal and cosinusoidal components to capture periodicity [11].

The temporal variables are defined as:

$$\begin{aligned} Var\_sin_t &= sin\left(2\pi\, \frac{t}{\tau}\right), \\ Var\_cos_t &= cos\left(2\pi\, \frac{t}{\tau}\right), \end{aligned} \tag{4}$$

t : Time (hour, day, or month)

τ : Period, τ=11 hours for a day, τ= 365 days for a year and τ=12 months in a year

These temporal variables were incorporated to transform raw temporal information (hour, month, day of the year) into features that can be utilized by the Multilayer Perceptron (MLP) model. The idea is to represent this information in a way that preserves its cyclicity while enabling the model to extract insights effectively.

Prediction Error Calculation

In this study, several metrics were used to evaluate the performance of the MLP model in predicting global solar irradiation. These metrics include the **nRMSE** (Normalized Root Mean Square Error) with a 95% Confidence Interval **(CI95)**, **MAE** (Mean Absolute Error), **MAPE** (Mean Absolute Percentage Error) and **R²** (Coefficient of Determination) [12].

$$nRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}}{\bar{y}} \times 100 \tag{5}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (6)$$

$$MAPE = \frac{\frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\bar{y}} \times 100 \quad (7)$$

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

$$IC95 = \pm Z\alpha/_2 \cdot \frac{\sigma_{\hat{y}}}{\sqrt{n}} \quad (9)$$

y_i: measured value, $\hat{y}_i$: predicted value, n: number of observation or data points and $\bar{y}$: the mean of the y_i $Z\alpha/_2$ is the quantile of the standard normal distribution for a confidence level of 1−α ($Z\alpha/_2$=1.96 for a 95% confidence level where α=0.05, significance level of 5%) and $\sigma_{\hat{y}}$ is the standard deviation of the prediction errors [12].

III. RESULTS AND DISCUSSION

Table 1: Data splitting

Data	Period
• Total dataset used	01-Jan-2020 07:00:00 à 31-Dec-2023 17:00:00
• Training and validation 70%	01-Jan-2020 07:00:00 à 19-Oct-2022 14:00:00
• Test 30%	19-Oct-2022 14:05:00 à 31-Dec-2023 17:00:00

The four years of measurement data were divided into two distinct sets: 70% for training and validation, corresponding to the period from January 1, 2020, at 7:00 AM to October 19, 2022, at 2:00 PM, and 30% for testing, covering the period from October 19, 2022, at 2:05 PM to December 31, 2023, at 5:00 PM. This division ensures that the model is trained on a broad dataset while being tested over a sufficiently representative period to evaluate its robustness and generalization capacity.

Table 2: Choice of the MLP network architecture, Ambatolampy site, horizon m+5

Neuron in hidden layer	nRMSE	R^2	MAE	MAPE
1	4.68%	0.9933	12.8725 Wh/m^2	2.57%
2	4.61%	0.9935	12.8812 Wh/m^2	2.57%
3	4.51%	0.9938	12.6247 Wh/m^2	2.52%
4	4.50%	0.9938	12.6755 Wh/m^2	2.53%
5	4.49%	0.9939	12.5684 Wh/m^2	2.51%
6	4.46%	0.9939	12.5363 Wh/m^2	2.51%
7	4.38%	0.9942	12.5079 Wh/m^2	2.50%
8	4.35%	0.9942	12.5585 Wh/m^2	2.51%
9	4.36%	0.9942	12.5113 Wh/m^2	2.50%
10	**4.34%**	**0.9943**	**12.3785 Wh/m^2**	**2.47%**
11	4.35%	0.9942	12.5046 Wh/m^2	2.50%
12	4.46%	0.9939	12.5541 Wh/m^2	2.51%
13	4.35%	0.9942	12.4214 Wh/m^2	2.49%
14	4.35%	0.9942	12.5520 Wh/m^2	2.51%
15	4.35%	0.9942	12.4731 Wh/m^2	2.49%
16	4.35%	0.9942	12.5398 Wh/m^2	2.51%
17	4.35%	0.9942	12.5221 Wh/m^2	2.50%
18	4.35%	0.9942	12.5279 Wh/m^2	2.50%
20	4.35%	0.9942	12.5215 Wh/m^2	2.50%

According to Table 2, a number of 10 neurons in the hidden layer was selected in this study, as this configuration yields the best error values across the different tests.

Table 3: Prediction horizon: *h+1* (one hour)

Sites	Ambatolampy			Betainomby			Laboratory		
Errors	MAPE	nRMSE ±CI95	R^2	MAPE	nRMSE ±CI95	R^2	MAPE	nRMSE ±CI95	R^2
MLP without temporal variables	19.03%	25.25% ±3.34	0.8437	21.56%	27.25% ±3.60	0.7956	17.63%	23.53% ±3.31	0.8483
MLP with temporal variables	**13.46%**	**18.58% ±2.44**	**0.9166**	**12.15%**	**16.85% ±2.20**	**0.9228**	**13.26%**	**18.33% ±2.56**	**0.9098**

Table 4: Prediction horizon: *m+5* (5 minutes)

Sites	Ambatolampy			Betainomby			Laboratory		
Errors	MAPE	nRMSE ±CI95	R^2	MAPE	nRMSE ±CI95	R^2	MAPE	nRMSE ±CI95	R^2
MLP without temporal variables	2.79%	4.74% ±0.20	0.9931	2.75%	4.54% ±0.19	0.9928	2.95%	5.28% ±0.24	0.9902
MLP with temporal variables	**2.49%**	**4.34% ±0.18**	**0.9943**	**2.13%**	**3.98% ±0.17**	**0.9946**	**2.75%**	**4.95% ±0.22**	**0.9914**

These results compare the performance of an MLP (Multilayer Perceptron) model with and without the integration of temporal variables for prediction horizons of h+1 (medium-term horizon, 1 hour) and m+5 (short-term horizon, 5 minutes). The analysis shows that integrating temporal variables significantly enhances the MLP model's performance in predicting solar irradiation in Madagascar, with particularly notable improvements for the medium-term prediction horizon (h+1, 1 hour).

For example, for Table 3, at Betainomby, the MAPE decreases substantially from 21.56% to 12.15%, and the nRMSE decreases from 27.25% to 16.85%, while the coefficient of determination (R^2) increases from 0.7956 to 0.9228. Similar improvements are observed at Ambatolampy, where R^2 increases from 0.8437 to 0.9166, and at the Ankatso Laboratory, where the MAPE decreases from 17.63% to 13.26%. These results highlight the model's improved ability to capture hourly variations in solar irradiation due to the inclusion of temporal variables.

For the short-term prediction horizon (m+5, 5 minutes), Table 4, the model's overall performance is already high, with R^2 values close to 1 across all sites. At Betainomby, integrating temporal variables further reduces the MAPE from 2.75% to 2.13% and the nRMSE from 4.54% to 3.98%, confirming an almost perfect correlation between predictions and measurements. These results demonstrate the model's capability to effectively handle short-term predictions, where data variability is less pronounced.

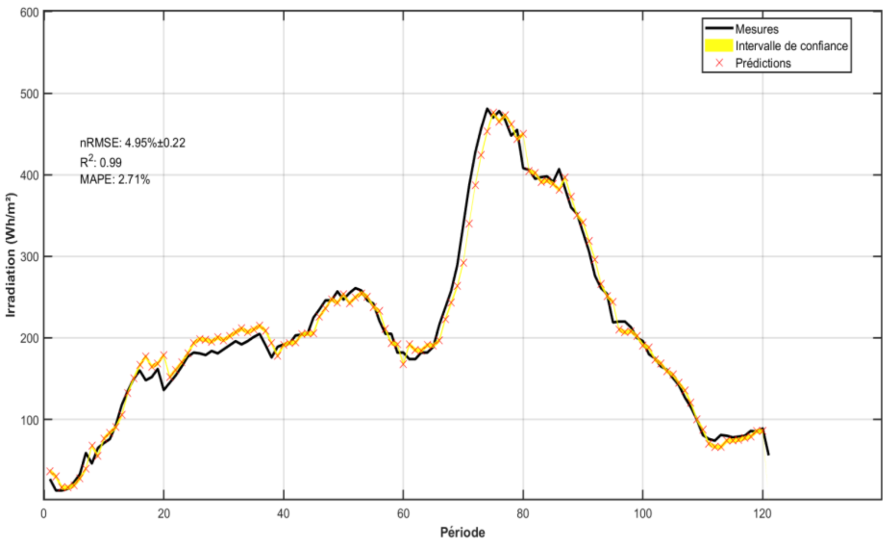

Fig.1 Prediction for the Laboratory site on January 26, 2023, from 07:00:00 to 17:00:00 (Period m+5)

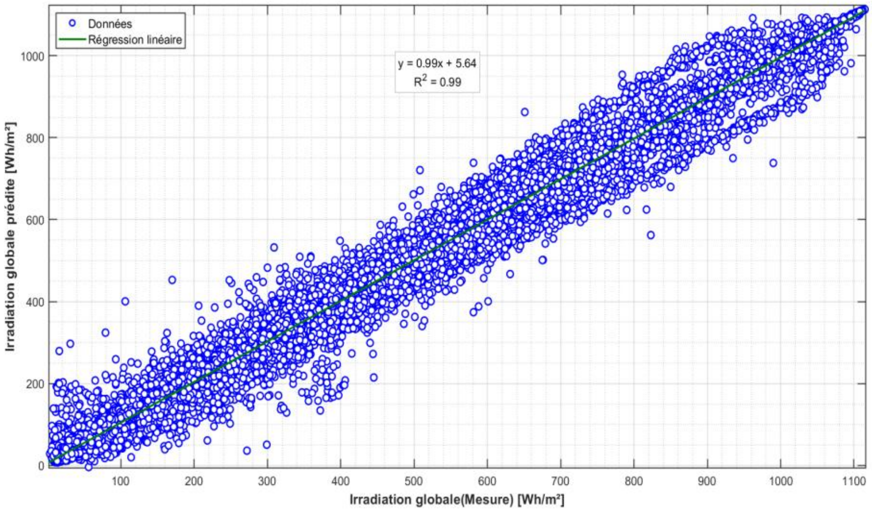

Fig.2 Linear regression curve between measured global irradiation and predicted global irradiation for Laboratory Site, y=0.99x+5.64, R^2=0.99

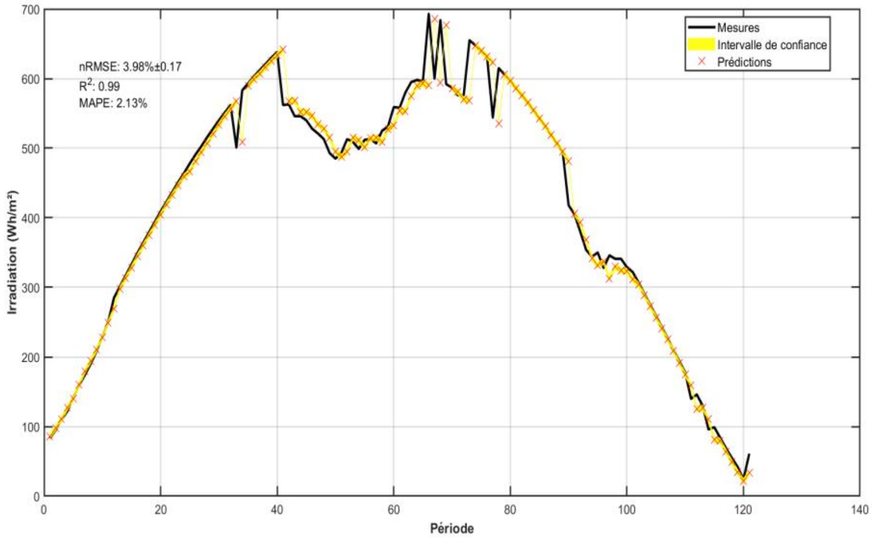

Fig.3 Prediction for Betainomby site for the Period from July 2, 2023, 07:00:00 to 17:00:00 (Period m+5)

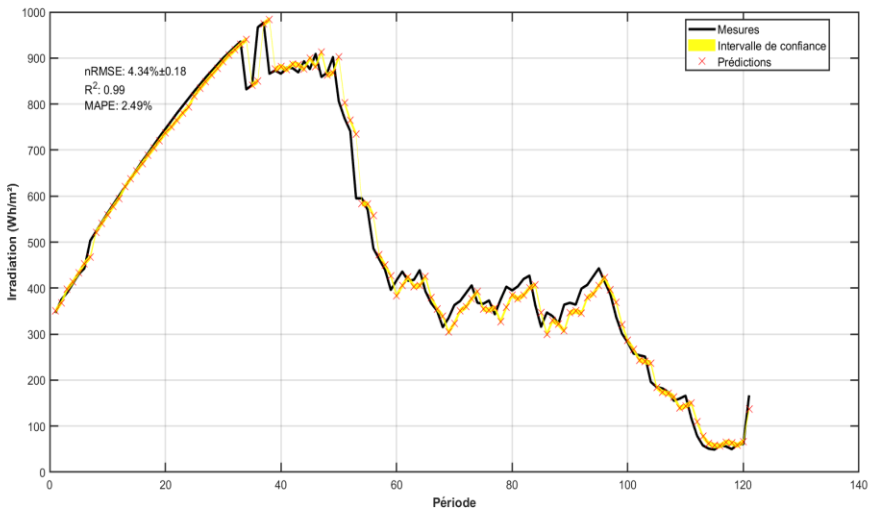

Fig.4 Prediction for Ambatolampy site for the Period from November 30, 2023, 07:00:00 to 17:00:00 (Period m+5)

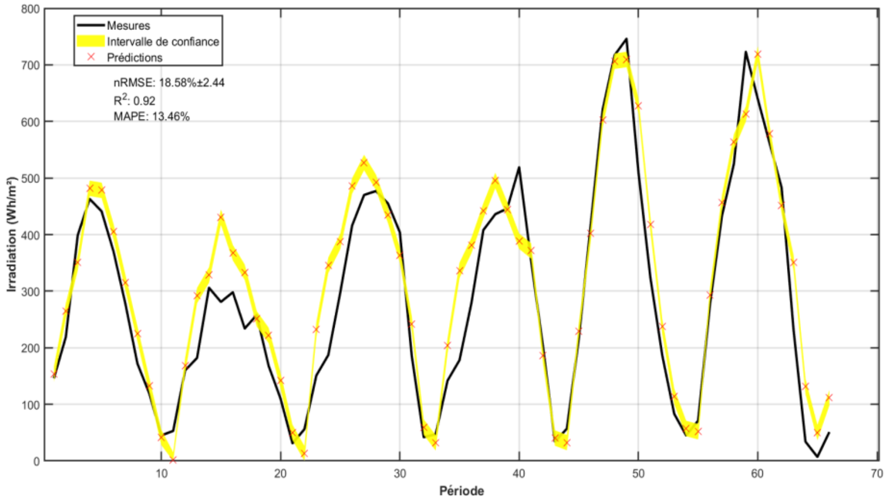

Fig.5 Prediction for Ambatolampy site for the period from July 7, 2023, 07:00:00 to July 12, 2023, 17:00:00 (Period h+1)

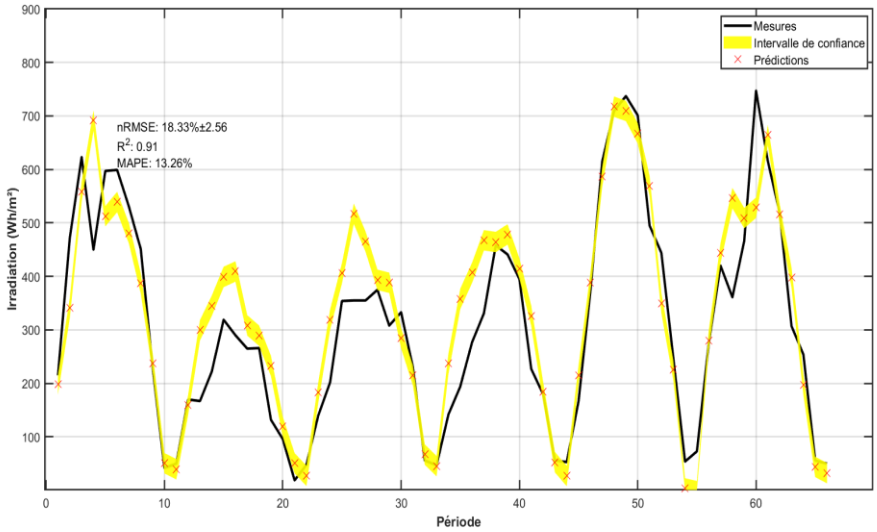

Fig.6 Prediction for Laboratoire site for the period from July 7, 2023, 07:00:00 to July 12, 2023, 17:00:00 (Period h+1)

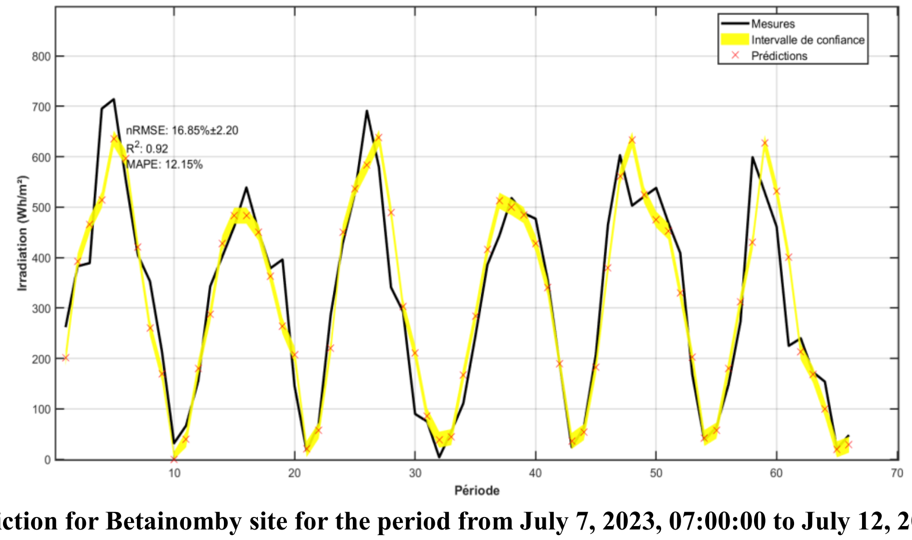

Fig.7 Prediction for Betainomby site for the period from July 7, 2023, 07:00:00 to July 12, 2023, 17:00:00 (Period h+1)

IV. CONCLUSION

The results obtained demonstrate that integrating temporal variables into the MLP model significantly enhances the accuracy of solar irradiation predictions for the three sites. This approach enables better capture of the country's specific temporal and geographical variations, making the models more robust and better suited to local conditions. The observed performance improvements, particularly in Betainomby for the medium-term horizon, confirm the importance of incorporating site-specific data for optimal modeling.

To further this research, several directions can be explored:

- Geographical Generalization: Extend the analysis to other regions of Madagascar or areas with similar climatic characteristics to evaluate the generalizability of the results;
- Practical Applications: Develop decision-support tools for solar plant management and energy planning, leveraging the improved predictions of the model.

REFERENCES

[1]. World Bank. [2018] October 10 “The Force of the Sun: Madagascar Embarks on Renewable Energy Production”.
[2]. Economic development board of Madagascar, [2018] “Madagascar, l‘île aux reserves d‘énergie”.
[3]. Mahamat Adoum, A., Gabriel Jean Philippe, T., Djicknoum, D., Amadou Seidou, M. [2020] “Impact of the Intermittency of Photovoltaic power plants on the Frequency Management: case of the Senegalese electricity grid”.
[4]. Solcast, a DNV company [2023], https://toolkit.solcast.com.au/
[5]. Voyant, C., Muselli, M. Paoli, C., Nivet Marie, L. [2012] “Numerical weather prediction (NWP) and hybrid ARMA/ANN model to predict global radiation” Energy, 39 (1), pp: 341-355.
[6]. Zhiqiang, W., Qingyun, S., PengTao, Z., Junlin, Z., [2020] “Correct normalization Matters: Understanding the effect of normalization on deep neural network Models for click-through rate prediction” arXiv preprint arXiv:2006.12753.
[7]. Brownlee, J. [2020] “Deep Learning for Time Series Forecasting: predict the future with MLPs, CNNs, LSTMs in Python”, Edition v1.9: pp.3-5.
[8]. Paoli, C., Voyant, C., Muselli, M., Nivet Marie, L. [2009] “Solar radiation forecasting using ad-hoc time series preprocessing and neural networks”. arXiv preprint arXiv:0906.0311.
[9]. Fathima, T. A., Nedumpozhimana, V., Lee, Y. H., & Dev, S. [2021] “Embedding Cyclical Information in Solar Irradiance Forecasting”. arXiv preprint arXiv:2110.09761.
[10]. Tang, P., & Zhang, W. [2024] “PDMLP: Patch-based Decomposed MLP for Long-Term Time Series Forecasting”. arXiv preprint arXiv:2405.13575.
[11]. Walter, A., Shewhart, Samuel, S. Wliks [2000] “Fourier Analysis of time series”. 2 Fitting sinusoids, pp: 9-24.
[12]. Wackerly, D., Mendenhall, W., & Scheaffer, R. L. [2008] “Mathematical Statistics with Applications (7th ed.)”. Cengage Learning. Chapter 8, pp: 406-423.