



**HAL**  
open science

# Photovoltaic power prediction for solar micro-grid optimal control

Sonja Kallio, Monica Siroux

► **To cite this version:**

Sonja Kallio, Monica Siroux. Photovoltaic power prediction for solar micro-grid optimal control. 2022 9th International Conference on Power and Energy Systems Engineering (CPESE 2022), Doshisha University, Kyoto, Japan, 9–11 September 2022, Sep 2022, Kyoto, Japan. pp.594-601, 10.1016/j.egyvr.2022.11.081 . hal-03991561

**HAL Id: hal-03991561**

**<https://hal.science/hal-03991561v1>**

Submitted on 15 Feb 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



2022 9th International Conference on Power and Energy Systems Engineering (CPSE 2022),  
Doshisha University, Kyoto, Japan, 9–11 September 2022

# Photovoltaic power prediction for solar micro-grid optimal control

Sonja Kallio, Monica Siroux\*

*INSA Strasbourg ICUBE, University of Strasbourg, Strasbourg, France*

Received 17 October 2022; accepted 6 November 2022

Available online xxxx

## Abstract

In a solar micro-grid, a hybrid renewable energy system generates electricity for a building's onsite use. The battery storage and the main power grid connection are used to facilitate the matching between the demand and production. To control energy flows optimally, an accurate day-ahead prediction of the photovoltaic (PV) panels output is required. However, this is a challenging task due to the fluctuating nature of solar radiation availability. The accuracy of the prediction is influenced by the modelling method and input parameters. In this study, the measured power and weather data is gathered from an experimental installation of PV panels to predict PV output for a 24-hours horizon in 15 min intervals. The multiple linear regression (MLR) and artificial neural network (ANN) methods are considered in the prediction modelling and compared using performance indicators. The micro-inverter technology is used to gather the individual PV panel output in addition to the overall system output. The results show that the modelling methods have different accuracy performances and the ANN model built with the individual PV output data results in the highest accuracy. Utilizing the micro-inverter technology leads to an advantage of having more accurate PV prediction for the control purpose.

© 2022 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Peer-review under responsibility of the scientific committee of the 9th International Conference on Power and Energy Systems Engineering, CPSE, 2022.

**Keywords:** Solar energy; Photovoltaics; Prediction model; Multiple linear regression; Artificial neural network; Machine learning

## 1. Introduction

Solar energy is an inexhaustible source of renewable energy. Additionally, it is free of charge and any emissions. However, the availability of solar energy is strongly fluctuating and depends on the changing weather conditions, time of the day and location. That creates significant uncertainty in the power output of solar systems. The photovoltaic panel is a silent device without moving parts and can convert sunlight into electricity through the photovoltaic effect.

The electrical efficiency of the crystalline silicon PV panels varies from 11% to 22% [1]. An increasing amount of distributed PV installations in the building sector enables building owners to act as a prosumer by generating and storing their own electricity onsite or selling it to the grid [2]. A solar micro-grid can be used to generate profit

\* Corresponding author.

*E-mail address:* [monica.siroux@insa-strasbourg.fr](mailto:monica.siroux@insa-strasbourg.fr) (M. Siroux).

**Nomenclature**

GHI	global horizontal irradiation, W/m <sup>2</sup>
GTI	global tilted irradiation, W/m <sup>2</sup>
P	power, W
p	pressure, hPa
RH	relative humidity, %
T	temperature, °C
WS	wind speed, m/s

for the building owner, and through the optimal control of hybrid renewable energy systems reviewed in [3], the profit and self-consumption can be maximized. However, PV power output has uncontrollable and abrupt nature. Thus, an accurate forecast of the output is required for optimal energy management in the solar micro-grid as well as in the other levels of the electric grid [4].

In the literature, different PV power prediction methods have been proposed with different data inputs. The most used methods are deterministic, machine learning (ML) and hybrid techniques [5,6]. The deterministic methods are based on the physical model of the PV to predict the power output, and the artificial intelligence-based ML methods, such as multiple linear regression (MLR), artificial neural network (ANN) and support vector machine (SVM), aim to form a relationship between input and output without any further knowledge of the system [5]. Kalogirou [7] conducted a comprehensive review on the ANN method in renewable energy systems considering PV systems. Zazoum [8] compared different ML techniques for PV power prediction. He concluded that the ML methods are suitable and fast to predict the PV power of any system.

The scientific literature has aimed to improve the accuracy of the PV power forecasting models in terms of different prediction algorithms and input data. Different sets of weather and PV operation parameters have been investigated to increase the accuracy [9]. However, to the best of the author's knowledge, there is no work focusing on the different ML methods using as an input parameter the PV power data measured by the individual micro-inverters to increase the accuracy of the day-ahead power prediction. In this study, the MLR and ANN prediction models are built for each PV panel and the overall system using the input data measured from the micro-inverter and aggregated system level. The study aims to reveal if using the data of the micro-inverter level can increase the prediction accuracy of the whole PV system.

## 2. Methodology

A comparative study of different PV power prediction models is conducted. First, the experimental set-up of PV panels is presented with data gathering methods. Next, the monitored data is pre-processed and used to create different data-driven models for the PV power prediction. The first method is a multiple linear regression (MLR) with interaction terms which is compared to a method called Artificial Neural Network (ANN). Both models are regression models, and the monitored data from the PV installation and weather station are used to train and validate the models in the MATLAB environment. The training data was collected from 27.9.2021 to 31.3.2022 and includes large variation of different weather conditions. The different prediction models are built by using as input the weather data and power output data from the micro-inverter and aggregated system level.

### 2.1. Experimental set-up

The installation of four PV panels in Fig. 1 was realized on the rooftop of the INSA ICUBE Laboratory in Strasbourg, France (48°N and 7°E) in September 2021. Strasbourg is classified to have an oceanic and semi-continental climate according to the Köppen climate classification [10]. Each PERC (Passivated Emitter & Rear Contact) monocrystalline PV panel has a peak power of 300 W, and micro-inverters are used to convert each panel's DC output to AC power individually. The energy production is maximized by a Maximum Power Point Tracking (MPPT) of each micro-inverter. Two of the panels have a tilt angle of 7°, and the rest is set to an 18° angle. The PV panels were installed on the available roof area which has a clear face with the sun during a part of the day. The

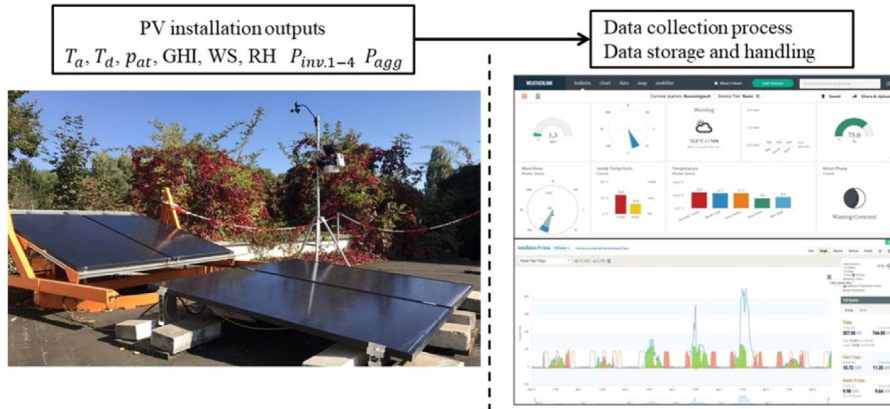


Fig. 1. The PV installation with the weather station and data collection process.

shading from the surrounding buildings during the day cannot be avoided. However, the micro-inverter technology enables the independent operation of each panel and partial shading has an impact only on the shaded panels. Using the micro-inverter technology has also several other advantages over traditional inverters discussed in [11].

The four micro-inverters are connected to a gateway which is used to communicate measured system data to the web-based monitoring and analysis software provided by Enphase company [12]. The power output of each panel ( $P_{inv,1-4}$ ) and the whole system ( $P_{agg}$ ) is measured every 15 min and uploaded into Excel file from the online software. The installation includes also Lithium-Ion AC battery storage to store the PV production if not used directly in the building. The optimal battery use is managed by the gateway.

In addition to the PV panels, the installation includes a weather station to measure the local weather parameters global horizontal irradiance (GHI), ambient temperature ( $T_a$ ), dew point temperature ( $T_d$ ), pressure ( $p_{at}$ ), wind speed (WS) and relative humidity (RH) in 15 min intervals. A data logger is used to transfer the weather data to a computer for further analysis.

### 2.2. Correlation analysis

The PV power production correlates with the surrounding weather conditions. The Pearson correlation analysis was used to find out the correlation between the PV power output and individual weather variables. Generally, this analysis method reveals how strong are the linear relationship between the different variables in a given data set. The analysis shows the correlation coefficients that vary from  $-1$  to  $1$ . The magnitude of the absolute value of the coefficient indicates the strength of proportionality between two variables. This indicator is used to select the most significant variables for the prediction model. The high number of parameters can result in lower accuracy of the model.

### 2.3. Multiple linear regression model

In this study, the measured historical data of both the PV power output and weather variables were used as an input to create a multiple regression model with interaction effects. The interaction effect occurs when the effect of the independent weather variable on the PV production changes depending on the other weather variables. This interaction is presented as a product of the independent weather variables. The regression model with interaction effects is an extension of the general regression presented as follows [5]:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \varepsilon \tag{1}$$

where  $\beta_0$  is the intercept and  $x$  represents each independent weather variables. The other  $\beta$  parameters present the slope coefficient of the variable. All  $\beta$  parameters are defined in the model creation process to minimize the error  $\varepsilon$ . The interaction effects are added to the general form of the multiple regression model as follows:

$$\hat{y} = \beta_0 + \beta_1x_1 + \beta_2x_2 + \beta_3x_1x_2 \dots + \beta_nx_n + \beta_lx_nx_m + \varepsilon \tag{2}$$

where  $x_1x_2$  is the interaction between two variables.

## 2.4. Artificial neural network model

The ANN algorithms are commonly used to solve time series prediction tasks because they have the ability to learn the behaviour of variables from historical measurements, such as PV power output and weather data in terms of the PV prediction modelling [13]. The ANN is based on supervised learning, and the prediction performance increases with the amount of historical data used in the training of the model. In addition to the training dataset, a part of the historical dataset is used only for the test and validation of the model to show the generalization capability of the prediction model. The learning or training process is based on continuously improving the weights and biases within the neural network to reach the desired output, which minimizes MSE in Eq. (2) (or RMSE in Eq. (3)) between the actual and predicted output. This is performed by using a learning algorithm, such as gradient descent, error back-propagation or Levenberg–Marquardt [13]. The back-propagation algorithms are the most used and powerful in terms of learning [14]. In this study, the MATLAB software is used to train the ANN model and the limited-memory Broyden–Fletcher–Goldfarb–Shanno quasi-Newton (LBFGS) optimization algorithm is used to minimize the MSE [15].

The number of inputs to the ANN model indicates the “neurons” in the input layer presented in Fig. 2. The layer structure includes the input layer, a certain number of hidden layers with hidden neurons and the output layer. In the case of PV power prediction, the input layer includes the selected weather variables ( $x_1, x_2, x_3 \dots, x_n$ ), and the output is the power prediction. Generally, the ANN model can include any number of hidden layers and neurons. However, the accuracy of the model depends on the right number of layers and neurons. In this study, the Multi-Layer-Feed-Forward Neural Network (MLFFNN) structure is used in which the information flows only in one direction from the input to the output layer. The outputs of the previous layer are the inputs for each neuron on the next layer. The inputs of each neuron are multiplied by weights ( $w_1, w_2, w_3 \dots, w_n$ ) and combined using a weighted linear combination [9]. Before the outputs of the neurons are ready, a nonlinear activation function  $f_{ac}$ , such as Hyperbolic tangent or Sigmoid function, is applied to the results of the layer. In this study, the Rectified linear unit function is used to perform a threshold operation on the input values to set the values less than zero to be zero [15].

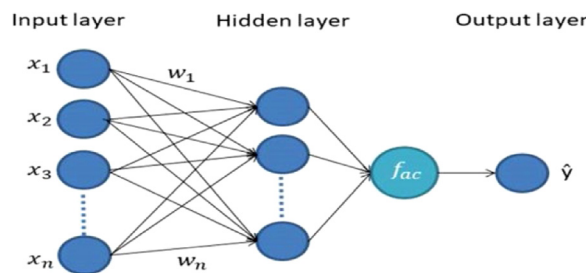


Fig. 2. The structure of a simple ANN model with one hidden layer.

After training, the ANN is a quick method to provide predictions for complex non-linear problems and it can reach good accuracy also with smaller data set used in the training and validation [9].

## 2.5. Model performance indicators

To conduct a comparative study between different models, certain performance indicators are used to evaluate the accuracy of the prediction models. In addition to a simple absolute error between the actual and predicted value, the following commonly used performance indicators for the forecasting models are used in this study: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Square Error (RMSE) and Coefficient of Determination (COD) [9,16].

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

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

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (5)$$

$$COD = R^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2} \quad (6)$$

where  $n$  is the number of observations,  $\bar{y}$  is the mean value of the observations,  $y$  and  $\hat{y}$  are the actual and predicted values of the PV power, respectively.

In the regression analysis, the COD is used to present how close the prediction model output is compared to the observed data and it results in values between zero and 1. To evaluate the overfitting of the prediction model, the COD requires the other indicators alongside it. In the evaluating process, the COD should be closer to 1 and the other indicators closer to zero [9].

### 3. Results and discussion

In this section, the performance results of the MLR and ANN model are presented. The training data was collected from the PV installation between September 2021 and March 2022. In addition to the performance indicators, a representative day in April was selected and the prediction was computed and compared to the actual power output of the system.

#### 3.1. Correlation coefficient analysis

To evaluate the significance of the individual weather variables to the PV power, a correlation coefficient analysis was conducted. The results of the analysis were used to select the most relevant weather variables for the modelling. The weather variables under the investigation were the ambient temperature  $T_a$ , relative humidity RH, dew point temperature  $T_d$ , wind speed WS, atmospheric pressure  $p_{at}$ , global horizontal irradiance GHI and global tilted irradiation GTI for both panel arrays in the tilt angle of  $7^\circ$  and  $18^\circ$ . The GTI values were calculated based on the measured GHI, number of a day, latitude and tilt angle of the panels. The PV power output  $P$  was recorded per panel and overall system. The correlation coefficients between the weather variables and PV power output are presented in Table 1.

**Table 1.** The correlation coefficients.

Variable	Correlation with PV power
Ambient temperature	0.27
Relative humidity	−0.39
Dew point temperature	0.01
Wind speed	0.15
Atmospheric pressure	0.11
GHI	0.94
GTI_up	0.94
GTI_down	0.94

The highest correlation was recognized between solar irradiation on the horizontal and tilted surface (GHI and GTI) and the PV power. The coefficient for each solar irradiation variable was 0.94. The second highest significance was shown by the RH followed by the ambient temperature. The impact of the RH was negative, which decreased PV power when the RH increased. This was caused by the increased amount of water vapour particles in the air reducing the solar insolation such as dust. The dew point temperature  $T_d$  showed the lowest significance by the value of 0.01 and was excluded from the modelling. The other variables in Table 1 were included in the prediction modelling by MLR and ANN.

### 3.2. Multiple linear regression model performance

To predict the PV power, an MLR model was trained in the MATLAB software environment. The performance indicators presented in Section 2.5 were used to evaluate the accuracy of the model.

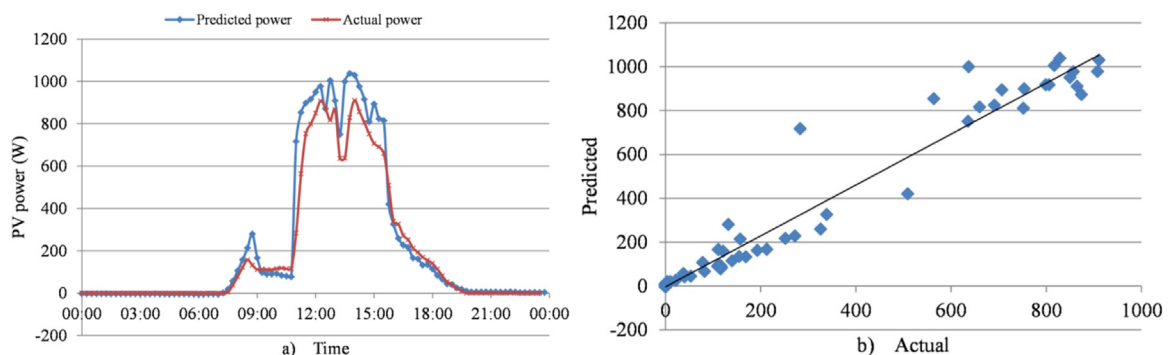
First, the regression model was trained with the weather data and the aggregated PV power output of the system. Next, the PV power output data of each micro-inverter was used to predict the output of each panel. The predictions of each panel were summed up to have a prediction of the whole system. The performance of the prediction models generated by the aggregated PV output data and by the individual micro-inverter data was compared.

The overall performance results in Table 2 show that the MLR, which was based on the micro-inverter data, showed better performance with 13.86 MAE, 1438 MSE, 37.92 RMSE and 0.938 COD than the model based on the aggregated PV power data with 16.32 MAE, 2160 MSE, 46.48 RMSE and 0.907 COD. The results revealed that a more accurate prediction model was achieved if the micro-inverter data was used first to model each panel output and then summed up to have PV power prediction for the whole system. This method can help to predict more accurately the PV power of the system if all the panels are not positioned equally to the sun during the day, for example, if some shading occurs.

**Table 2.** The overall performance comparison of the models based on the different methods and PV data input.

Model		Performance indicator			
		MAE	MSE	RMSE	COD (R <sup>2</sup> )
Multiple regression	Overall	16.32	2161	46.485	0.9074
	Micro-inv.	13.86	1438	37.919	0.9376
ANN	Overall	13.34	1517	38.96	0.935
	Micro-inv.	9.87	930	30.497	0.9596

The MLR model based on the micro-inverter data was used to predict the PV power for the representative day in April, shown in Fig. 3a. This day was not used in the training dataset. Fig. 3b presents the correlation between the actual and predicted values. In this case, the model was able to perform as follows: 43.4 MAE, 7552 MSE, 86.9 RMSE and 0.91 COD.



**Fig. 3.** (a) The predicted and actual PV power using the micro-inverter based MRL model. (b) Actual vs. predicted PV power correlation with the linear trendline from the MRL model.

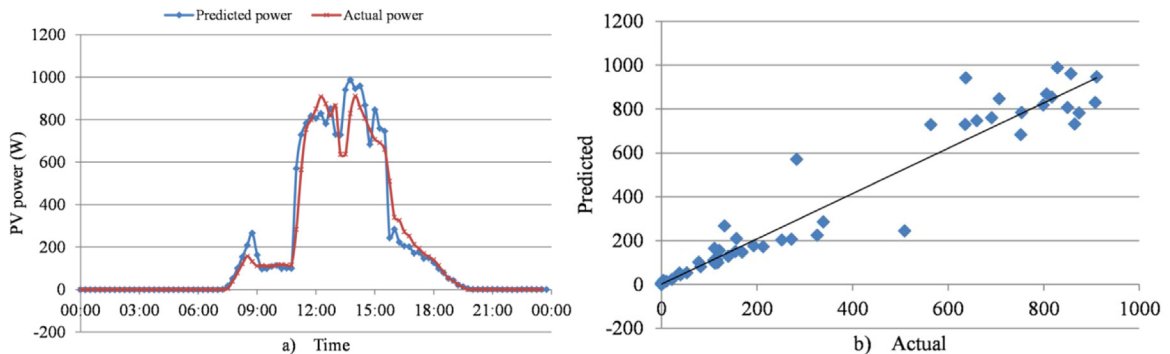
### 3.3. Artificial neural network model performance

The same dataset, as for the MLR model, was used to train an ANN model. Again both aggregated and micro-inverter based PV power data were used to train the model to see if the micro-inverter data can be used to have a more accurate prediction model for the installation.

The overall performance results in Table 2 show that the ANN model, which was based on the micro-inverter data, showed again better performance with 9.87 MAE, 930 MSE, 30.497 RMSE and 0.9596 COD than the model

based on the aggregated PV power data with 13.34 MAE, 1517 MSE, 38.96 RMSE and 0.935 COD. However, in both cases, the ANN model showed the better overall performance than the MLR models.

The ANN model based on the micro-inverter data was used to predict the PV power for the representative day in April, shown in Fig. 4a. Fig. 4b presents the correlation between the actual and predicted values. The model was able to perform as follows: 32.26 MAE, 4632 MSE, 68.06 RMSE and 0.9471 COD.



**Fig. 4.** (a) The predicted and actual PV power using the micro-inverter based ANN model. (b) Actual vs. predicted PV power correlation with the linear trendline from the ANN model.

#### 4. Conclusion

In this paper, the PV panel installation using micro-inverter technology was realized in INSA ICUBE Laboratory in Strasbourg to build a PV power prediction model using two different methods: multiple linear regression and artificial neural network. The weather station was used to collect the weather data. The correlation analysis was used to select the most significant weather parameters correlating with the PV power output. The parameters were the ambient temperature, relative humidity, pressure, wind speed and solar irradiation.

In addition to the two different modelling methods, the prediction models were built using two different approaches to collect the PV power output data from the installation. First, the aggregated PV power was used in the modelling process. Next, the PV power data, detected by each individual micro-inverter, was used to model each PV panel output to get a more accurate model of the whole system.

The results showed that the overall performance of the ANN models was better than the MLR models resulting in lower MAE, RMSE and MSE and higher COD. Additionally, the results revealed that building a model of each panel by using the PV power data from the micro-inverters resulted in more accurate prediction models. Due to this, it is recommended to use the micro-inverter technology in the PV installations to facilitate the prediction of the system output power.

For future work, the PV prediction model based on the ANN and micro-inverter data will be enhanced by collecting more data from the installation. The comparison between different models will be extended with other methods, such as Support Vector Machine, Fuzzy prediction and Regression Tree. Additionally, the prediction model will be used in the model predictive control of a hybrid renewable energy system.

#### Declaration of competing interest

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

#### Data availability

The authors do not have permission to share data.

#### Acknowledgements

The authors would like to thank Interreg V Rhin supérieur, France for their support and funding of this research.



## References

- [1] Wang J, Han Z, Guan Z. Hybrid solar-assisted combined cooling, heating, and power systems: A review. *Renew Sustain Energy Rev* 2020;133:110256.
- [2] Barbu M, Darie G, Siroux M. Analysis of a residential photovoltaic-thermal (PVT) system in two similar climate conditions. *Energies* 2019;12:1–18.
- [3] Kallio S, Siroux M. Hybrid renewable energy systems based on micro-cogeneration. *Energy Rep* 2022;8:762–9.
- [4] Gupta P, Singh R. PV power forecasting based on data-driven models: a review. 2021;14:1733–1755. <https://doi.org/10.1080/19397038.2021.1986590>.
- [5] Abubakar Mas'ud A. Comparison of three machine learning models for the prediction of hourly PV output power in Saudi Arabia. *Ain Shams Eng J* 2022;13:101648.
- [6] Theocharides S, Theristis M, Makrides G, et al. Comparative analysis of machine learning models for day-ahead photovoltaic power production forecasting. *Energies* 2021;14:1081, 2021, Vol 14, Page 1081.
- [7] Kalogirou SA. Artificial neural networks in renewable energy systems applications: a review. *Renew Sustain Energy Rev* 2001;5:373–401.
- [8] Zazoum B. Solar photovoltaic power prediction using different machine learning methods. *Energy Rep* 2022;8:19–25.
- [9] Alshafeey M, Csáki C. Evaluating neural network and linear regression photovoltaic power forecasting models based on different input methods. *Energy Rep* 2021;7:7601–14.
- [10] Kallio S, Siroux M. Energy analysis and exergy optimization of photovoltaic-thermal collector. *Energies* 2020;13:5106.
- [11] Mundada AS, Nilsiam Y, Pearce JM. A review of technical requirements for plug-and-play solar photovoltaic microinverter systems in the United States. *Sol Energy* 2016;135:455–70.
- [12] Enphase. 2022, <https://enphase.com/fr-fr> (accessed 8 2022).
- [13] Dolaro A, Grimaccia F, Leva S, et al. Comparison of training approaches for photovoltaic forecasts by means of machine learning. *Appl Sci* 2018;8:228, 2018, Vol 8, Page 228.
- [14] Kalogirou SA. *Solar energy engineering*. Elsevier; 2009.
- [15] Matlab. Regression neural network. 2022, <https://fr.mathworks.com/help/stats/regressionneuralnetwork.html> (accessed 5 2022).
- [16] Nageem R, Jayabarathi R. Predicting the power output of a grid-connected solar panel using multi-input support vector regression. *Procedia Comput Sci* 2017;115:723–30.