Research Article

The Importance of Distance between Photovoltaic Power Stations for Clear Accuracy of Short-Term Photovoltaic Power Forecasting

Abdelhakim El hendouzi,1 Abdennaser Bourouhou,2 and Omar Ansari3

1Lab Research in Electrical Engineering, National School of Computer Science and Systems Analysis and Higher Normal School of Technical Education, Avenue of the Royal Army, Madinat Al Irfane, District Riad, Rabat 100100, Morocco
2Lab Research in Electrical Engineering, Higher Normal School of Technical Education, Avenue of the Royal Army, Madinat Al Irfane, District Riad, Rabat 100100, Morocco
3Lab Research in Mechanical Engineering, Higher Normal School of Technical Education, Avenue of the Royal Army, Madinat Al Irfane, District Riad, Rabat 100100, Morocco

Correspondence should be addressed to Abdelhakim El hendouzi; abdelhakim.elhendouzi@um5s.net.ma

Received 27 November 2019; Revised 18 February 2020; Accepted 2 March 2020; Published 10 April 2020

Guest Editor: Minh Quan Duong

Copyright © 2020 Abdelhakim El hendouzi et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The current research paper deals with the worldwide problem of photovoltaic (PV) power forecasting by this innovative contribution in short-term PV power forecasting time horizon based on classification methods and nonlinear autoregressive with exogenous input (NARX) neural network model. In the meantime, the weather data and PV installation parameters are collected through the data acquisition systems installed beside the three PV systems. At the same time, the PV systems are located in Morocco country, respectively, the 2 kWp PV installation placed at the Higher Normal School of Technical Education (ENSET) in Rabat city, the 3 kWp PV system set at Nouasseur Casablanca city, and the 60 kWp PV installation also based in Rabat city. The multisite modelling approach, meanwhile, is deployed for establishing the flawless short-term PV power forecasting models. As a result, the implementation of different models highlights their achievements in short-term PV power forecasting modelling. Consequently, the comparative study between the benchmarking model and the forecasting methods showed that the forecasting techniques used in this study outperform the smart persistence model not only in terms of normalized root mean square error (nRMSE) and normalized mean absolute error (nMAE) but also in terms of the skill score technique applied to assess the short-term PV power forecasting models.

1. Introduction

The reports by Renewables 2017 Global Status and International Energy Agency (IEA) confirmed that the solar PV power has grown tremendously which implied many economic and social benefits. The cumulative solar PV capacity, meanwhile, reached 398 GW which generated over 460 TWh and represented around 2% of global power energy [1]. However, the penetration of renewable energy particularly the solar PV remains trivial in comparison to the fuel and coal-fired power plants due to numerous technical and economic challenges. In this case, the need for high penetration of solar PV in power systems is chronic and required. The solar PV, meanwhile, depends on the weather parameters and the location of PV installation, which are unpredictable and affect the daily solar energy generation. However, in the case of solar PV grid-tied, the poor electrical grids cannot support this source of energy. For that reason, the strong penetration of solar PV energy in the global energy mix has driven the thinking to next generation of electrical power grids and the renovation of most existence electrical grids to host the new mode of solar PV and
guaranteeing its integration. In this case, the need to smart energy management systems (SEMS) that incorporate the forecasting methods of solar PV power is an important key to overcome many trials of renewable energy challenges and allow them (especially the PV power) the flexibility in terms of control and monitoring. Moreover, the forecasting methods can help the integration of natural and sustainable energy resources and encourage the adoption of recent energy systems such as microgrids, which are smart small microgenerations based on microsources including the renewable energy. The microgrids, meanwhile, request advanced techniques of control and forecasting to overcome the effect of solar PV variability. Also, the use of short-term PV power forecasting algorithms can support the integration of solar PV in microgrids by providing the profile of PV power for next 24 hours, which can aid the control flexibility, regulation, monitoring, and dispatching of microgrids. In addition, further advantages of PV power forecasting such as the economic returns which they are clear in the planning of energy generation, which supported by the demand forecasting. However, the electricity cost optimization.

The current research article states the need for accurate short-term PV power forecasting due to its positive effect on scheduling. They can help the energy market operators escape the potential penalties due to the eccentricities between the planned and produced energies. This research study, meanwhile, suggests the best design of PV power forecasting model, which consisted on selecting the right future time horizon that means the choice of time between the current time and the needed future time, choosing the right forecasting resolution, and selecting the suitable forecasting approach. The time horizons mostly considered by the literature include very short term that starts from some seconds and ends in few minutes also includes the “time scale starts from several days to several months. In addition, the spatial horizon is also needed for PV power forecasting, which can display the total of space foreseen by the forecasting model; this forecasting horizon begins from one site to regional areas, also called regional forecast or multisite forecast [2]. Moreover, the need for forecasting approaches is primal for forecasting modelling. In the meantime, the survey of literature showed that the PV power forecasting is possible by using the direct and indirect techniques. The direct technique, meanwhile, resides on counting directly the amount of PV power in a future time horizon; also, the experts recommend the techniques of artificial intelligence and machine learning for short-term PV power forecasting. The indirect technique or solar irradiation forecasting consists on transforming the solar irradiation forecasting through a PV model to the PV power [3]. The literature review also recognized three main approaches for PV power forecasting modelling [4]. They include the physical approach based on real model of PV installation, the physical model, the rental of equipment, etc. The statistical approach includes methods belonging to the artificial intelligence, data mining, and machine learning. The hybrid approach is a new approach that gathers the techniques of different approaches or considers the collaboration between techniques of the same approach.

Certainly, other approaches used for PV power forecasting include the time series, regressive, and probabilistic methods.

In the meantime, the literature review showed some related research articles in short-term PV power forecasting topic based on artificial neural networks and classification methods. In this case, the review article by Inman et al. [5] showed successful applications of solar forecasting methods and other theories related to the PV power resources and forecasting. The focus of this paper, meanwhile, is about the comparative study established between the artificial neural networks and K-nearest neighbors (KNNs), which both considered methods of artificial intelligence. The review article by Voyant et al. [6] presented a list of machine learning methods including the K-NN method which is considered as the groundwork for this current research paper. The research article by Zamo et al. [7] presented a set of PV power forecasting methods called (PEARP). In the meantime, the focus of this study is about the use of data provided by 28 PV power plants, which encourage the use of data provided by multiple sites for feeding the forecasting models such as the case of this current research paper. The research study by Almenacid et al. [8] proposed multilayer perceptron neural network for forecasting the global irradiance and air temperature, alongside with NAR neural network that is used for calculating the PV power; however, the use of NAR did not take into account the effect of outputs on the forecasting results; nevertheless, the present research paper proposed NARX instead of NAR. The research article by Chu et al. [9] presented three smart models for reforecasting PV power; the models included the KNN; nevertheless, in their study, they did not combine the KNN method with any algorithm of similarity and thus it is very clear in their results. This present research paper is also inspired by the studies conducted by Li et al. and Gigoni et al. [10, 11] which presented some useful methods for data normalization and assessments for error minimization measured between the forecasted and real PV powers.

The context of this research paper is the contribution to resolve the dilemma of short-term PV power forecasting by the application of similarity algorithm (SA) with the KNN method and NARX neural network model applied to three different sites with varied sizes and distinct geographical locations. The forecasting model, meanwhile, consisted on choosing the right variables that fit more the pattern of PV power and then the use of artificial intelligence methods. The SA method, meanwhile, calculates the distance between the weather variables and PV system parameters. In addition, the KNN, which is a straightforward method, is used for short-term PV power forecasting with NARX neural network. The main goal of this research article, meanwhile, is contributing to the short-term PV power forecasting modelling. Also, this research article highlights the effect of distance between the PV power installations on short-term PV power forecasting by answering the need for optimal number of variables that fit more the PV power [12]. Moreover, the smart persistence model is used in this study as the benchmarking model of PV power forecasting.
The body of this research article contains the outline of PV system at Rabat ENSET School, which consists of a profound study of DC and AC installation, alongside the model of PV system. In the meantime, the highlight of equipment that used for measuring the weather and PV system parameter data. Moreover, the presentation of solar PV power forecasting methods explains the process of forecasting modelling as well as the contrast on useful equations and models used in this topic. Lastly, the demonstration of results and the perspectives are presented.

2. Outline of PV System at Rabat ENSET School

The purpose of this section is building the PV model of the PV system located at ENSET School. In this case, the PV model supports the PV power forecasting modelling since it allows the complete understanding of the PV system and the knowledge of important PV parameters. With the aim to facilitate the study of PV system, this research article considers the separation of DC and AC parts. In the meantime, the modelling of PV system starts from the study of the PV system location and the DC and AC materials. The PV system at ENSET School, meanwhile, availed for lighting. However, the extra power incorporates the electrical grid since the grid-tied inverter is used for the integration service as illustrated in Figures 1 and 2.

2.1. Overview of Geographical Characteristics of ENSET School Location. The ENSET School located in Rabat of Morocco benefits from an extraordinary site, which is most of the time sunny. For further information about the system location, Table 1 shows the geographical coordinates of the ENSET School site. This location holds a PV system of 2 kWp established by eight (8) PV panels installed in south facing as illustrated in Figures 3(a) and 3(b).

The ENSET PV system equipped by eight PV panels has the identical electrical features as shown in Table 2.

2.2. PV System Electrical Characteristics

2.2.1. PV System DC Parameters. The DC component of the PV system includes a metallic structure designed for eight PV modules, a DC junction box that contains the DC circuit breakers, and the electrical cables as shown in Figures 3(a) and 3(b). In the meantime, the serial configuration of PV panels is adopted with the aim to answer the input voltage required by the inverter. Furthermore, Table 3 provides further data about the PV array installed at the ENSET ground.

In this study, the single diode model is used for PV modelling. The implementation and simulation of the PV model, meanwhile, based on MATLAB software are shown in Figure 4. In addition, the PVSYST software used to simulate the voltage at the maximum power point tracking (MPPT) of the PV array, which is designed to provide the essential input voltage of the inverter, is illustrated in Figure 5.

2.2.2. PV System AC Parameters. The second part of PV system modelling regards the AC kit, which covers an inverter, the AC junction box that contains the AC circuit breakers, and the electrical cables. Furthermore, the inverter used to convert the PV power provided by the PV array as well as to order the device of power limitation is shown in Figure 6. Table 4, meanwhile, provides further data about the AC part.
2.2.3. Balances and Main Results. This part presents the main analysis of the energy produced by the whole PV system of the ENSET School over the year as shown in Figures 7 and 8. The energy over year, meanwhile, remains unpredictable and depends on the location of PV system and the days and the season of the year. Although, in Morocco, the daily energy for December and January months is less than other periods because this two months belong to the winter season. For further analysis in terms of losses of the energy provided by the ENSET PV system, the chart in Figure 9 illustrates the estimation of the PV system losses over the year, which is 10.3%.

Table 2: The electrical characteristics of PV modules of the ENSET PV system.

<table>
<thead>
<tr>
<th>Module PV</th>
<th>Si-poly</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturer</td>
<td>Solar World</td>
</tr>
<tr>
<td>Model</td>
<td>SUNMODULE PLUS SW 250 POLY</td>
</tr>
<tr>
<td>Power at maximum point power pmpp(W)</td>
<td>250</td>
</tr>
<tr>
<td>Open circuit voltage $V_{oc}$ (V)</td>
<td>37.6</td>
</tr>
<tr>
<td>Voltage at maximum point power $V_{mpp}$ (V)</td>
<td>30.5</td>
</tr>
<tr>
<td>Open circuit current $I_{oc}$ (A)</td>
<td>8.81</td>
</tr>
<tr>
<td>Current at maximum point power $I_{mpp}$ (A)</td>
<td>8.27</td>
</tr>
<tr>
<td>Module efficiency (%)</td>
<td>14.91%</td>
</tr>
</tbody>
</table>

Table 3: The DC parameters of the ENSET PV system.

<table>
<thead>
<tr>
<th>PV modules in series</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>PV modules in parallel</td>
<td>N/A</td>
</tr>
<tr>
<td>Metallic structure</td>
<td>8</td>
</tr>
<tr>
<td>Tilt °</td>
<td>26</td>
</tr>
<tr>
<td>PV array area (m²)</td>
<td>13.4</td>
</tr>
<tr>
<td>PV module area (m²)</td>
<td>11.7</td>
</tr>
<tr>
<td>Array global power at nominal conditions (STC) kWc</td>
<td>2</td>
</tr>
<tr>
<td>Array global power at operating conditions (50°C) kWp</td>
<td>2.13</td>
</tr>
<tr>
<td>Array operating characteristics (50°C) $U_{mpp}$ (V)</td>
<td>256</td>
</tr>
<tr>
<td>Array operating characteristics (50°C) $I_{mpp}$ (A)</td>
<td>8</td>
</tr>
</tbody>
</table>

2.2.3. Balances and Main Results. This part presents the main analysis of the energy produced by the whole PV system of the ENSET School over the year as shown in Figures 7 and 8. The energy over year, meanwhile, remains unpredictable and depends on the location of PV system and the days and the season of the year. Although, in Morocco, the daily energy for December and January months is less than other periods because this two months belong to the winter season. For further analysis in terms of losses of the energy provided by the ENSET PV system, the chart in Figure 9 illustrates the estimation of the PV system losses over the year, which is 10.3%.

2.3. Overview of the Monitoring Gear of the ENSET School PV System. The PV system at ENSET School includes a sensor network that embeds sensors of ambient temperature, module temperature, wind speed, and solar irradiance. The sensors can provide the data through RS485/422 cables or...
Ethernet mode to the central data acquisition called solar log, which is able to communicate the data through the website and android application. The solar log’s target, meanwhile, is ensuring the management and monitoring of the ENSET PV system including the visualization, optimization, and management process of self-consumption and grid-tied PV system. In the meantime, this equipment achieves the drop of power generation. At the same time, this equipment can ensure the limitation of reactive current through an installed external box.

However, with the aim of establishing the database for forecasting modelling, the website as well as the USB device used to export the data needs from the solar log in Excel files is illustrated by the diagram in Figure 10.

![Figure 6: The energy distribution from the inverter, which depends on PV array voltage and current; in this case, the operating voltage of the inverter corresponds to 175–560 V.](image)

**Table 4: Recaps of simulation of ENSET’s AC PV system.**

<table>
<thead>
<tr>
<th>Inverter</th>
<th>SMA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>Sunny Boy 2000 HF</td>
</tr>
<tr>
<td>Unit nominal power kW AC</td>
<td>2</td>
</tr>
<tr>
<td>Operating voltage (V)</td>
<td>175–560</td>
</tr>
<tr>
<td>Total nominal power kw AC</td>
<td>2</td>
</tr>
<tr>
<td>Produced energy kWh/year</td>
<td>3483</td>
</tr>
<tr>
<td>Specific production kWh/kWp/year</td>
<td>1741</td>
</tr>
<tr>
<td>Performance ratio PR (%)</td>
<td>90, 90%</td>
</tr>
</tbody>
</table>

Ethernet mode to the central data acquisition called solar log, which is able to communicate the data through the website and android application. The solar log’s target, meanwhile, is ensuring the management and monitoring of the ENSET PV system including the visualization, optimization, and management process of self-consumption and grid-tied PV system. In the meantime, this equipment achieves the drop of power generation. At the same time, this equipment can ensure the limitation of reactive current through an installed external box.

However, with the aim of establishing the database for forecasting modelling, the website as well as the USB device used to export the data needs from the solar log in Excel files is illustrated by the diagram in Figure 10.

### 3. Data and Methods of Solar Photovoltaic Power Forecasting

#### 3.1. Data Normalization

The standardization or data normalization corresponds to the process of scaling the data through mathematical equations. The data normalization, meanwhile, is worthwhile for the flawlessness of PV power forecasting models, for example, neural networks perform better when inputs have the appropriate scale. The Z-score is known as the best used normalization method, which corresponds to the number of standard deviations from the mean. Furthermore, the mean and standard deviation are used to calculate the Z-score of data.

#### 3.1.1. Mean

Equation (1) provides the mathematical form for finding the mean value of a specified variable.

\[
V_m = \frac{1}{n} \sum_{j=1}^{n} V_j, \quad (1)
\]

where \(V_m\) is the mean value of the vector parameters \(V\), \(n\) is the number of elements in \(V\), and \(V_j\) is the \(j^{th}\) element in \(V\).

#### 3.1.2. Standard Deviation

The standard deviation provides the spreading of values. In addition, it is mostly practical to set the range of data [13]. In the case of total population, equation (2) provides the mathematical structure for calculating the standard deviation of a specified variable. Moreover, when the data are a sample, equation (3) is preferred.

\[
\text{std}_1 = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (V_j - V_m)^2}, \quad (2)
\]

\[
\text{std}_2 = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (V_j - V_m)^2}, \quad (3)
\]

where \(V_j\) is the \(j^{th}\) element in vector \(V\) and \(V_m\) is the mean value of \(V\).

#### 3.1.3. Z-Score

Equation (4) is used to calculate the Z-score value of a stated variable.

\[
Z_{\text{score}} = \frac{V - V_m}{\text{std}}, \quad (4)
\]
Figure 7: Assessment of ENSET PV array efficiency based on PVSYST software for each individual month of the year.

Loss diagram for "Enset_Rabat_2kWc"-Year

- Horizontal global irradiation
- Global incident in coll. plane
- IAM factor on global
- Effective irradiance on collectors
- PV conversion
- Array nominal energy (at STC effic.)
- Array losses (Temp, Mod. Qual., Mism., Res.)
- Array virtual energy at MPP
- Available Energy at Inverter Output
- Energy injected into grid

Figure 8: Assessment of ENSET PV array efficiency based on PVSYST software for each individual month of the year.

Figure 9: The loss diagram of the whole year for the ENSET PV system.
$V$ is the vector of original values, $V_m$ is the mean value of $V$, and std is the standard deviation of the data.

3.2. Physical Model of Photovoltaic Power Forecasting. The physical model of PV power forecasting is the most common one, which is based on the data measurement from both PV systems and weather stations [14].

3.2.1. Mathematical Expression of PV Power. The PV power produced by solar PV panels can be predictable by using a mathematical equation [15] as shown in the following equation:

$$P_{pv} = \xi \cdot S_m \cdot N \cdot I_{rr}, \quad (5)$$

where $N$ is the number of solar PV panels, $S_m$ is the surface of a solar PV module, $I_{rr}$ is the solar radiation on the plan of PV module, and $\xi$ is the instantaneous performance of the solar PV panel. The expression of $\xi$ is shown in the following equation:

$$\xi = \xi_r \cdot (1 - \Gamma_{tstc} \cdot (\Gamma_{md} - \Gamma_{stc})), \quad (6)$$

where $\xi_r$ is the reference efficiency of the PV module under STC conditions, $\Gamma_{tstc}$ is the temperature coefficient under STC conditions, which is a value given by the manufacturer, and $\Gamma_{md}$ and $\Gamma_{stc}$ are, respectively, the temperatures of the module at STC conditions and under any conditions.

3.2.2. Expression of Energy Irradiance. The effective irradiance received by the PV cell can be calculated by using the following equation:

$$I_{rr} = I_{eff} + \nabla d (I_{dc} + I_{rd}), \quad (7)$$

where $I_{eff}$ is the efficient irradiance, $I_{dc}$ is the energy irradiance diffused by the sky, $I_{rd}$ is the energy irradiance reflected from the ground, and $\nabla d$ is the fraction of diffuse.

3.2.3. Mathematical Expression of SANDIA Cell Temperature Model. The Sandia Energy first proposed the SANDIA cell temperature model as a part of the Sandia PV system performance model [16]. Equation (8) meanwhile, calculates the temperature of PV module in degrees Celsius (C).

$$\Gamma_{md} = I_{rr} \cdot e^{a+b \cdot ws} + \Gamma_{at}, \quad (8)$$

where $a$ and $b$ are the temperature coefficients, $ws$ is the wind speed, and $\Gamma_{at}$ is the ambient temperature.

In the meantime, the temperature of PV cell in degrees Celsius is available by using

$$\Gamma_{celpv} = \Gamma_{md} + \frac{I_{rr}}{1000} \alpha_t, \quad (9)$$

where $\alpha_t$ is a coefficient of temperature.

3.3. Correlation of Calculated and Measured Data. The PV power measured from the PV system is totally different from the PV power calculated by a PV model according to [17]. In the meantime, the PV power calculated by a PV model is equal to the PV power measured multiplied by a coefficient $\lambda_{pow}$ as shown in the following equation:

$$P_{cal} = \lambda_{pow} \cdot P_{real}. \quad (10)$$

The similar analysis above is practical for calculating the temperatures as shown in the following equation:

$$\Gamma_{md} = \lambda_{temp} \cdot \Gamma_{real}. \quad (11)$$

3.4. Artificial Neural Networks (ANNs). Machine learning (ML) and artificial intelligence are considered as the advanced techniques since they allow the easy classification and forecasting of data. The ML includes the ANNs. Their objective is asking the machine to classify the data by splitting, but sometimes the error revealed the presence of misclassified data. Therefore, the aim of error function is getting the information about the misclassified data and also modelling the
error. The error function meanwhile is either discrete which is convenient for the classification problems or continuous which is suitable for the optimization problems. In the meantime, the gradient descent is applicable for minimization criteria. The probability is also used to describe the error function since the product of probabilities defines this function. Therefore, the existence of times of probabilities revealed that there is no error, and the maximum likelihood is determined. Moreover, the error function is guilty for the choice of right activation function in the ANNs modelling.

In this paper, the NARX neural network is used with feedback or closed loop architecture with time delay for both external inputs and feed-forwards from outputs as shown in Figure 11. The closed loop also called parallel architecture, meanwhile, is convenient for multistep PV power forecasting [18]. In this item, the NARX neural network with two layers is applied for short-term PV power forecasting model as well as used to contribute in the flawlessness of PV power forecasting.

3.5. Classification of Weather and PV System Parameters.

The process of classification of weather variables and PV system parameters is useful to distinct the variables that fit more the pattern of measured PV power and to classify them according to their importance as primary, secondary, etc. This process realized through computer algorithms is based on mathematical equations such as the Euclidian distance, root mean square Euclidian distance difference, and weighted hybrid distance. The Euclidian distance, meanwhile, is used to calculate the distance between the elements of PV power vector and the elements of other variables. In the meantime, the root mean square Euclidian distance differences are used to cut the prime variable and the weighted hybrid distance is used to compute the rank of other external variables. In this paper, the classification rank is six variables.

Equation (12) is practical for calculating the Euclidian distance between the elements of PV power vector for days \( d \) and \( d + 1 \).

\[
D(v_j, d, d') = \sqrt{\sum_{i=1}^{r} \left( p_i^{(d)} - p_i^{(d')} \right)^2}. \tag{12}
\]

Furthermore, equation (13) is useful to compute the Euclidian distance between the elements of other variables for days \( d \) and \( d + 1 \).

\[
D(v_j, d, d') = \sqrt{\sum_{i=1}^{r} \left( v_{ij}^{d} - v_{ij}^{d'} \right)^2}. \tag{13}
\]

However, to cut the prime external variable, the algorithm should organize the variables that fit more the PV power pattern by calculating the root mean square Euclidian distance differences (\( \mathcal{G} \)) as shown in the following equation:

\[
\mathcal{G}(P, v_j) = \sqrt{\frac{\sum_{d=2}^{n} \sum_{d'=3}^{n} (D(P, d, d') - D(v_j, d, d'))^2}{(1/2)n \cdot (n - 1)}}. \tag{14}
\]

where \((1/2)n \cdot (n - 1)\) corresponds to the distance size and \( n \) is the variable dimension.

In addition, to discover the rank of other external variables, the weighted hybrid distance (\( \zeta \)) and the root mean square weighted hybrid distance differences (\( \zeta_i \)) are calculated for the Euclidian distance of the prime external variable and the Euclidian distance of other variables for days \( d \) and \( d + 1 \) as shown in equations (15) and (16).

\[
\zeta(v_j, d, d') = (1 - \gamma_{v_j})D(R_{v_1}, d, d') + \gamma_{v_j}D(v_j, d, d'). \tag{15}
\]

\[
\zeta(R_{v_2}, d, d') = (1 - \gamma_{R_{v_2}})D(R_{v_1}, d, d') + \gamma_{R_{v_2}}D(R_{v_2}, d, d'). \tag{16}
\]

where \( R_{v_j} \) is the \( j \)th label of the variable found by the similarity algorithm and \( \gamma_{v_j} \) is a coefficient whose value is chosen as the smallest. In the perspective of this article, the algorithm of similarity is applicable to the data provided by the three sites that are previously discussed.

3.6. Approximate Method for Information Extraction from the Data. The question there is how one can sort the forecasted day from forecasting result. The extraction of the forecasted day is going to be hard when big data are available. In the meantime, equations (17) and (18) below are practical and helpful for finding the corresponding day and month in the forecasting result. Equation (17) is developed to sort the forecasted day from big data and finds its label in months of year.

\[
\Delta_f = \frac{f_d}{N_{dy}}, \tag{17}
\]

where \( \Delta_f \) is the month of year, \( f_d \) is the forecasted day, and \( N_{dy} = 30.42167 \) corresponds to the average of 365 days of year. The result given by equation (17) is often with a comma where the decimal part corresponds to the month and the fractional part corresponds to the day of month.

Furthermore, equation (18) is practical to find the sorting of similar days to the forecast day.

\[
\Delta_s = \frac{S_d}{N_{dy}}, \tag{18}
\]

where \( \Delta_s \) is the month of year. The outcome given by this equation has often a decimal part that corresponds to the month, and the fractional part corresponds to the day of month.

3.6.1. Example. The detected day by the algorithm of similarity is \( f_d = 324 \) and \( \Delta_f = 10.652 \). Therefore, the result corresponds to the month October and the fractional part 0.652 timed by the coefficient \( N_{dy} \) matches the day 20. Finally, the forecast day corresponds to 20 October.

3.7. Error Metrics of PV Power Forecasting. The mean absolute error (MAE) and root mean square error (RMSE) are mostly relevant and practical methods for assessing the
accuracy of PV power forecasting models [3]. The MAE, meanwhile, is used for finding the steady distance between the real and the outputted values from forecast models. Therefore, the MAE is appropriately practical for estimating the persistent forecast errors, whereas the RMSE deals with severely large errors in square order. Equations (19) and (20) show, respectively, the structure of MAE and RMSE.

\[
\text{MAE} = \frac{1}{n} \sum_{j=1}^{n} |P_{\text{for}} - P_{\text{real}}|, \tag{19}
\]

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{j=1}^{n} (P_{\text{for}} - P_{\text{real}})^2}, \tag{20}
\]

where \(P_{\text{for}}\) is the forecasted PV power and \(P_{\text{real}}\) is the measured PV power.

However, to compare results generated from forecast models, the skill score \(\Delta_f\) is the most practical method [19]. Equation (21) shows the structure of skill score technique.

\[
\Delta_{sk} = 1 - \frac{M_{f_x}}{M_{f_z}}, \tag{21}
\]

where \(M_{f_x}\) corresponds to the result from the forecasting model \(j\) and \(M_{f_z}\) corresponds to the result from the model \(j+1\).

4. Results and Discussion

4.1. Simulation Results. This research paper provides the best results based on simulation of two kinds of model that belong to two different areas of artificial intelligence modelling. The first part of simulation concerns the results of artificial neural network application, specifically the use of NARX neural network model, and the second part of simulation concerns the application of classification methods, specifically the use of K-nearest neighbors with similarity algorithm to forecast the short-term PV power.

4.1.1. Weather and PV System Data. The building of database is the key process of forecasting modelling, and it has taken more time than expected. In the case of PV power forecasting study, two kinds of data can be initiated which are the weather data from meteorological stations and the PV system data measured directly from PV systems. In addition, for the subject of this research paper, the data used to feed the PV power forecasting models correspond to three different locations characterized by size dissimilarity. In the meantime, the PV system data at ENSET School provide 2247 hours of weather and PV system data; besides, the Casablanca and Rabat sites provide 8760 hours of weather and PV system data. Furthermore, the PVGIS which is a platform of weather data from the European Commission is used to provide other meteorological data from 2007 to 2016 for each site which they also used to feed the forecasting process as shown in Table 5 [20]. The platforms Excel, R, and MATLAB were used to create the database and implementing different forecasting models.

4.1.2. NARX Neural Network for PV Power Forecasting. The NARX neural network model is applicable for both weather and PV system data provided by the locations described above. In the meantime, the implementation of NARX pursues the separation or multimodel approach that means the implementation of NARX model for each individual site.

(1) NARX Forecasting Model for 2kWp PV Power Station. The implementation of NARX forecasting model on ENSET School PV system and PVGIS weather data for the chosen days 25, 26, and 27 February shows satisfactory results that are clear by the PV power curves as presented in Figure 12. The NARX model of the ENSET PV system, meanwhile, contains three (3) hidden neurons.

(2) NARX Forecasting Model for 3kWp PV Power Station. The implementation of NARX model of PV system on Casablanca (3kWp) PV system shows satisfactory results for the chosen days 27, 28, and 29 November, which are clear by the PV power curves as shown in Figure 13. In this path, the NARX neural network model contains eleven (11) hidden neurons.

(3) NARX Forecasting Model for 60kWp PV Power Station. The NARX neural network model applied to forecast the quantity of PV power (60kWp) of another site located at Rabat city illustrates satisfactory results for the chosen days 27, 28, and 29 November as shown by the PV power curves in Figure 14. In the meantime, the NARX model contains fourteen (14) hidden neurons.
forecasting model, meanwhile, revealed perfect results in terms of skill scores in comparison with the smart persistence model as shown in Table 6. The use of straightforward ANNs such as the NARX model for PV power forecasting shows excellent results. Nevertheless, the process of forecasting by NARX model takes more time such as the time allowed to the data preparation and standardization. In addition, the presence of some slighted outliers stresses the NARX neural network models and reduces their efficiency, which drives the thinking to other methods.

### 4.1.3. Similarity Algorithm and KNN for PV Power Forecasting

The first part of this approach concerns the algorithm of similarity that is based on root mean squared difference distances $I$, which is used to detect the similar days to the forecast day. The variable with the lowest $I$ is ordered as the prime external variable. The other variables are classed regarding their calculated root mean square weighted hybrid difference distances $\zeta$. The second part uses the KNN model to forecast the short-term PV power.

In addition, the proposed forecasting process uses the same data shown in Table 5 for feeding the models of different locations. In the meantime, this research paper considers six external variables chosen by the similarity algorithm for each individual site as presented, respectively, in Table 7 for the 2kWp PV station, Table 8 for the 3kWp PV station, and Table 9 for the 60kWp PV station. Furthermore, the simulation results for both classification and forecasting noticeably present best results as shown, respectively, in Figure 15 for the 20 February forecasting day of the 2 kWp PV system, Figure 16 for the 26 September forecasting day of the 3 kWp PV system, and Figure 17 for the 7 July forecasting day of the 60 kWp PV system. The result summary of the similarity algorithm is shown in Table 10.

Therefore, the KNN forecasting model presents satisfactory results in terms of skill scores in comparison with the
### Table 6: Summary of accuracy assessment of NARX neural network forecasting model for each location.

<table>
<thead>
<tr>
<th>Metrics/Locations</th>
<th>Rabat 60kWp</th>
<th>Rabat 2kWp</th>
<th>Casablanca 3kWp</th>
</tr>
</thead>
<tbody>
<tr>
<td>RMSE\textsubscript{NARX} 27 (%)</td>
<td>10.145</td>
<td>—</td>
<td>11.178</td>
</tr>
<tr>
<td>RMSEPERSISTENCE 27 (%)</td>
<td>23.339</td>
<td>—</td>
<td>22.887</td>
</tr>
<tr>
<td>$\Delta_{\text{skRMSE}}$ (%)</td>
<td>0.565</td>
<td>—</td>
<td>0.511</td>
</tr>
<tr>
<td>MAE \textsubscript{NARX} 27 (%)</td>
<td>5.130</td>
<td>—</td>
<td>8.333</td>
</tr>
<tr>
<td>MAEPERSISTENCE 27 (%)</td>
<td>15.642</td>
<td>—</td>
<td>15.448</td>
</tr>
<tr>
<td>$\Delta_{\text{skMAE}}$ (%)</td>
<td>0.672</td>
<td>—</td>
<td>0.460</td>
</tr>
<tr>
<td>RMSE\textsubscript{NARX} 28 (%)</td>
<td>5.948</td>
<td>—</td>
<td>6.570</td>
</tr>
<tr>
<td>RMSEPERSISTENCE 28 (%)</td>
<td>15.570</td>
<td>—</td>
<td>15.098</td>
</tr>
<tr>
<td>$\Delta_{\text{skRMSE}}$ (%)</td>
<td>0.617</td>
<td>—</td>
<td>0.564</td>
</tr>
<tr>
<td>MAE \textsubscript{NARX} 29 (%)</td>
<td>3.508</td>
<td>—</td>
<td>3.593</td>
</tr>
<tr>
<td>MAEPERSISTENCE 29 (%)</td>
<td>15.642</td>
<td>—</td>
<td>9.041</td>
</tr>
<tr>
<td>$\Delta_{\text{skMAE}}$ (%)</td>
<td>0.672</td>
<td>—</td>
<td>0.460</td>
</tr>
<tr>
<td>RMSE\textsubscript{NARX} 25 (%)</td>
<td>—</td>
<td>10.043</td>
<td>—</td>
</tr>
<tr>
<td>RMSEPERSISTENCE 25 (%)</td>
<td>—</td>
<td>75.119</td>
<td>—</td>
</tr>
<tr>
<td>$\Delta_{\text{skRMSE}}$ (%)</td>
<td>—</td>
<td>0.930</td>
<td>—</td>
</tr>
<tr>
<td>MAE \textsubscript{NARX} 26 (%)</td>
<td>—</td>
<td>5.491</td>
<td>—</td>
</tr>
<tr>
<td>MAEPERSISTENCE 26 (%)</td>
<td>—</td>
<td>79.015</td>
<td>—</td>
</tr>
<tr>
<td>$\Delta_{\text{skMAE}}$ (%)</td>
<td>—</td>
<td>0.314</td>
<td>—</td>
</tr>
<tr>
<td>RMSE\textsubscript{NARX} 27 (%)</td>
<td>—</td>
<td>8.945</td>
<td>—</td>
</tr>
<tr>
<td>RMSEPERSISTENCE 27 (%)</td>
<td>—</td>
<td>11.212</td>
<td>—</td>
</tr>
<tr>
<td>$\Delta_{\text{skRMSE}}$ (%)</td>
<td>—</td>
<td>0.202</td>
<td>—</td>
</tr>
<tr>
<td>MAE \textsubscript{NARX} 28 (%)</td>
<td>—</td>
<td>3.593</td>
<td>—</td>
</tr>
<tr>
<td>MAEPERSISTENCE 28 (%)</td>
<td>—</td>
<td>6.688</td>
<td>—</td>
</tr>
<tr>
<td>$\Delta_{\text{skMAE}}$ (%)</td>
<td>—</td>
<td>0.044</td>
<td>—</td>
</tr>
</tbody>
</table>

### Table 7: Classification of variables according to $\mathcal{I}$ and $\zeta_A$ for 2kWp PV power station.

<table>
<thead>
<tr>
<th>$R_{v_j}$</th>
<th>$v_j$</th>
<th>$\gamma$</th>
<th>$\mathcal{I}$</th>
<th>$\zeta_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{v_{t1}}$</td>
<td>Module temperature (ENSET)</td>
<td>—</td>
<td>0.005508720545916</td>
<td>—</td>
</tr>
<tr>
<td>$R_{v_{t2}}$</td>
<td>Relative humidity (R76)**</td>
<td>$10^{-3}$</td>
<td>—</td>
<td>0.019802673656454</td>
</tr>
<tr>
<td>$R_{v_{t3}}$</td>
<td>Wind direction (R76)</td>
<td>$10^{-6}$</td>
<td>—</td>
<td>0.019802837275625</td>
</tr>
<tr>
<td>$R_{v_{t4}}$</td>
<td>Ambient temperature (R76)</td>
<td>$10^{-6}$</td>
<td>—</td>
<td>0.019802846771643</td>
</tr>
<tr>
<td>$R_{v_{t5}}$</td>
<td>Direct normal irradiance (R76)</td>
<td>$10^{-9}$</td>
<td>—</td>
<td>0.019802847479204</td>
</tr>
</tbody>
</table>


### Table 8: Classification of variables according to $\mathcal{I}$ and $\zeta_A$ for 3kWp PV power station.

<table>
<thead>
<tr>
<th>$R_{v_j}$</th>
<th>$v_j$</th>
<th>$\gamma$</th>
<th>$\mathcal{I}$</th>
<th>$\zeta_A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{v_{t1}}$</td>
<td>Ambient temperature (NC)*</td>
<td>—</td>
<td>0.005508720545916</td>
<td>—</td>
</tr>
<tr>
<td>$R_{v_{t2}}$</td>
<td>Module temperature (NC)</td>
<td>$10^{-3}$</td>
<td>—</td>
<td>0.001322625296966</td>
</tr>
<tr>
<td>$R_{v_{t3}}$</td>
<td>Relative humidity (C76)**</td>
<td>$10^{-3}$</td>
<td>—</td>
<td>0.001322639903851</td>
</tr>
<tr>
<td>$R_{v_{t4}}$</td>
<td>Wind speed (C76)</td>
<td>$10^{-7}$</td>
<td>—</td>
<td>0.001322641151184</td>
</tr>
<tr>
<td>$R_{v_{t5}}$</td>
<td>Direct normal irradiance (C76)</td>
<td>$10^{-8}$</td>
<td>—</td>
<td>0.001322641234634</td>
</tr>
<tr>
<td>$R_{v_{t6}}$</td>
<td>Ambient temperature (C76)</td>
<td>$10^{-9}$</td>
<td>—</td>
<td>0.001322641242248</td>
</tr>
</tbody>
</table>

**C76: weather data from 2007 to 2016 of Casablanca. **NC: data from Nouasseur location at Casablanca.
4.2. Discussion and Outlines. In the first case, the contribution in short-term PV power forecasting through the employment of classification techniques and artificial neural networks was negligible. However, in the second case, the contribution was substantial, as evidenced by the high accuracy rates achieved.

4.3. Conclusion. The proposed approach demonstrated significant improvements in PV power forecasting, particularly for short-term predictions. The integration of classification techniques and artificial neural networks provided a robust framework for enhancing the accuracy of PV power forecasts, which is crucial for the efficient management of grid connectivity and renewable energy integration.

4.4. Future Work. Further research is needed to explore the impact of various weather conditions on PV power generation and to develop more sophisticated models for real-time forecasting. The integration of advanced machine learning algorithms and big data analytics could potentially lead to even more accurate predictions, thereby optimizing the grid’s response to variations in renewable energy production.


Table 9: Classification of variables according to $\zeta$ and $\zeta_\Delta$ for 60kWp PV power station.

<table>
<thead>
<tr>
<th>$R_{ij}$</th>
<th>$v_j$</th>
<th>$\gamma$</th>
<th>$\zeta$</th>
<th>$\zeta_\Delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{v1}$</td>
<td>Ambient temperature (Rabat)**</td>
<td>0.0007698674850908259</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>$R_{v2}$</td>
<td>Module temperature (Rabat)</td>
<td>$10^{-5}$</td>
<td>—</td>
<td>0.001365317131326</td>
</tr>
<tr>
<td>$R_{v3}$</td>
<td>Relative humidity (R76)*</td>
<td>$10^{-6}$</td>
<td>—</td>
<td>0.001365332792013</td>
</tr>
<tr>
<td>$R_{v4}$</td>
<td>Wind speed (R76)</td>
<td>$10^{-7}$</td>
<td>—</td>
<td>0.00136534030451</td>
</tr>
<tr>
<td>$R_{v5}$</td>
<td>Direct normal irradiance (R76)</td>
<td>$10^{-8}$</td>
<td>—</td>
<td>0.00136534123127</td>
</tr>
<tr>
<td>$R_{v6}$</td>
<td>Wind direction (R76)</td>
<td>$10^{-9}$</td>
<td>—</td>
<td>0.00136534131037</td>
</tr>
</tbody>
</table>


---

**Figure 15:** Similarity algorithm combined with KNN for forecasting the PV power of 20 February.

**Figure 16:** Similarity algorithm combined with KNN for forecasting the PV power of 26 September.

The persistence model as shown in Table 11. The $K$ is chosen equal to one (1) since in this simulation, just one day is detected as similar to the forecast day.
networks. Although, the both approaches are belonging to the artificial intelligence and machine learning. In the meantime, the KNN with similarity algorithm and NARX neural network models is established for each individual PV power system described in the aforementioned sections. The forecasting system showed satisfactory results due to the use of similarity algorithm for selecting the significant variables, which means the classification of variables that fit more pattern of PV power. Hence, this process of selection decreases the time of modelling.

Moreover, the application of KNN method combined with the similarity algorithm revealed perfect results in comparison to the benchmarking model as well as the application of NARX neural network for short-term PV power forecasting. The NARX neural network, meanwhile, is a robust and powerful model since it takes into account the effect of outputs that feed-forwarded to inputs (see Figure 11). Nevertheless, it needs huge size of data that are used particularly for training, testing, and validation. For that reason, the NARX model is applied to the overall data of each individual location.

Consequently, the classification methods showed perfect results in terms of modelling simplicity in comparison to the artificial neural network models that suffer from the overfitting and memorization problems, even though data normalization performs well. In short, this research article recommends the process of similarity algorithm associated with KNN as the flawless short-term PV power forecasting model.

In the second case, the research article has proven the effect of distance between PV systems on short-term PV power forecasting modelling. The examination of results, meanwhile, has shown that the similarity algorithm must be employed to the weather and PV system parameters of each individual site even if the sites belong to the same geographical location (e.g., in this study, Rabat city covers two PV systems, 2 kWp and 60 kWp, respectively). However, the distance between PV system locations is an important parameter. Therefore, this confirmation is very significant when there is a need for multisite PV power forecasting modelling.

### 5. Conclusions

As a conclusion, firstly, this research article shows best results from the use of NARX and KNN methods. Therefore,
this present research article recommends the practice of classification techniques such as KNN combined with similarity algorithm for flawless short-term PV power forecasting. Furthermore, the optimization of forecasting modelling by selecting the optimal parameters is required since the choice of optimal variables that fit more the pattern of PV power can lead to forecasting error minimization and improving the forecasting accuracy. Secondly, this research article presents the effect of distance between PV power installations on PV power forecasting process. Therefore, this new parameter needs more studies and developments to show its real effect on forecasting models.

For future work, the advanced neural networks with optimization methods will be able to give a further solution to the dilemma of short-term PV power forecasting, as well as the consideration of other variables and parameters.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References