Photovoltaic Power Forecasting Using Multiscale-Model-Based Machine Learning Techniques

Manel Marweni 1, Mansour Hajji 2, Majdi Mansouri 3,* and Mohamed Fouazi Mimouni 1

1 Laboratory of Automatic Electrical Systems and Environment, National Engineering School of Monastir, University of Monastir, Monastir 5000, Tunisia; marwenimanel1@gmail.com (M.M.); mfouzimim@gmail.com (M.F.M.)
2 Research Unit Advanced Materials and Nanotechnologies, Higher Institute of Applied Sciences and Technology of Kasserine, Kairouan University, Kasserine 1200, Tunisia; hajji_mansour@yahoo.fr
3 Electrical and Computer Engineering Program, Texas A&M University at Qatar, Doha 23874, Qatar
* Correspondence: majdi.mansouri@qatar.tamu.edu

Abstract: The majority of energy sources being used today are traditional types. These sources are limited in nature and quantity. Additionally, they are continuously diminishing as global energy consumption increases as a result of population growth and industrial expansion. Their compensation is made from clean energy and renewable energy. Renewable energy is strongly dependent on climatic conditions; therefore, an aspect of energy management is needed, which is essential in distribution systems, because it enables us to calculate the precise energy used by the load as well as by its many components. It also helps us understand how much energy is required and its origin. The energy management aspect contains two main phases: forecasting and optimization. In this study, we are focused on the forecasting level using intelligent machine learning (ML) techniques. To ensure better energy management, it is very important to predict the production of renewable energy over a wide time period. In our work, several cases are proposed in order to predict the temperature, the irradiance, and the power produced by a PV system. The proposed approach is validated by an experimental procedure and a real database for a PV system. The big data from the sensors are noisy, which pose a major problem for forecasting. To reduce the impact of noise, we applied the multiscale strategy. To evaluate this strategy, we used different performance criteria, such as mean error (ME), mean absolute error (MAE), root mean square error (RMSE), nRMSE and the coefficient of determination ($R^2$). The obtained experimental results show good performance with lower error. Indeed, they achieved an error for nRMSE criteria between 0.01 and 0.37.

Keywords: photovoltaic (PV); energy management (EM); forecasting; stand-alone PV system

1. Introduction

Electricity is a significant factor in a country’s development in terms of economy and technology. The growing dependence on energy is resulting from a considerable increase in electricity production all over the world. The primary source of energy production in the world is conventional fossil fuels. These sources are on the point of extinction, and they are one of the key sources of gas emissions that result in extreme global warming [1]. As a result, in the last few years, attention has increasingly been switched from fossil fuels to renewable energy (RE) sources for electricity generation. Amongst these RE sources, the solar energy is considered as one of the promising sources [2]. Looking for new clean energy sources is not the only way to reduce the impact on energy consumption, but one of the most important ways to address this problem is to optimise energy use [3]. Therefore, the aim of designing a sustainable energy production system is to make smart energy use. Accordingly, the interest in energy management is growing. The energy management systems (EMSs) is a great choice for smart energy use and the improvement of company competitiveness, especially in the case of industry and building applications. In order to
improve the generation and distribution of energy, modelling and forecasting of energy has become an extremely important challenge in EM.

In general, EM contains two principal steps, including the forecasting and optimization. In this paper, we are focused on the step of forecasting. The increased generation of renewable energy, optimal operation of storage devices, and more adaptable EMS techniques can all result in lower operating costs for power systems. In this regard, this article focuses on solar energy output and irradiance predictions for the best energy management (EM). The two primary categories for forecasting power models are typically physical and statistical models. The statistical approach is based on historical data and how those data statistically relate to forecasts made by meteorologists as well as SCADA (supervisory control and data acquisition) measurements [4]. Moreover, numerical weather prediction (NWP) data can be employed with the statistical forecasting methodology. Forecasts for weather variables such as wind speed, temperature, precipitation, humidity, pressure, and solar irradiation are introduced by NWP models. The physical method [5], however, focuses on including physical information in the model, such as topography data and energy source attributes. In addition, there are many power forecasting time horizon methodologies, which are often divided into several basic scales. A number of recent studies address the issue as interest in predicting solar power has grown. Many of these take into account predictions of the world’s radiation, which is effectively the same issue as forecasting solar power.

In [6], they developed a variety of state-of-the-art probabilistic models for forecasting solar irradiance. They investigated the use of post-hoc calibration techniques for ensuring well-calibrated probabilistic predictions. Four probabilistic models that output a probability distribution over the outcome space instead of a point prediction were discussed: a Gaussian process regression model, a neural network with uncertainty based on a dropout variation (dropout neural network), a neural network whose predictions parameterized a Gaussian distribution optimized to maximize likelihood (variational neural network), and [7] present natural gradient boosting (NGBoost) for probabilistic prediction. They used CRPS and maximum likelihood (MLE) as indicators to evaluate probabilistic prediction. In addition, an ANN inference system was used in the study by [8] to forecast electricity usage and guarantee an ideal EM. Therefore, the paper [9] gives a brief development of computational models based on machine learning (ML), support vector machine (SVM), and Gaussian process regression (GPR) techniques to estimate the PV panel power. The results showed that Matern 5/2 GPR offered the top performance among the offered ML algorithms, with RMSE, MAE, and R2 values of 7.967, 5.302, and 0.98, respectively. This validates the Matern 5/2 GPR model’s excellent dependability and accuracy. Authors in [10] examined how several environmental factors, including irradiance, relative humidity, ambient temperature, wind speed, PV surface temperature, and collected dust, affected the output power of the PV panel. An internal PV system’s calibration of a number of sensors was reported. To anticipate the PV system’s hourly power output, a number of multiple regression models and ANN-based prediction models were developed and tested, and they yielded root mean square errors (RMSEs) of 2.1436, 6.1555, and 5.5351, respectively. The ANN models, with all the features and features selected using correlation feature selection (CFS) and relief feature selection (Relieff) approaches, were shown to accurately forecast PV output power. The goal of this study [11] is to compare the effectiveness of several machine learning models, such as artificial neural networks (ANNs), support vector regression (SVR), and regression trees (RTs), with varying hyper-parameters and variables, in forecasting the power output of PV systems. The ANN beat the other models in a comparison of ideally planned models, where it achieved the lowest prediction mean absolute percentage error (MAPE) and normalised root mean square error (nRMSE) of 0.6% and 0.76%, respectively. Because the nRMSE findings were 1.13% and 1.33%, respectively, the prediction capabilities of the SVR and RT models can be deemed similar in this scenario. Finally, all of the models outperformed the PM in terms of relative improvement, because the skill score (SS) ranged from 86 to 100.
In recent years, there has been an increase in interest in the subject of machine learning techniques for prediction [12]. Many engineering, natural science, and social science applications require precise forecasting models (FM), which are a key component. A forecast is created using data from prior efforts and research into potential future manifestations of identified characteristics [13]. The essential premise of predicting is that the future will in some way resemble the patterns or distribution of the past. The secret to making good forecasts is finding the patterns or hidden information in historical data. Energy forecasting has drawn interest from the development of artificial intelligence (AI) and machine learning (ML) approaches. For more than three decades, AI/ML algorithms have been used for energy forecasting [14,15]. Because of the development of computing technologies, the discipline of AI/ML has recently made more progress. For example, deep learning [16,17], reinforcement learning [18], and transfer learning [19] are some of the advanced AI/ML approaches that have been implemented in energy forecasting. Compared to the suggested forecasting method, this study represents the state-of-the-art in PV power forecasting techniques established over the recent years.

The aim of this paper is to predict the irradiance and PV power using stand-alone system to exploit the ability of each model and show the effectiveness of ML techniques under different operating cases using intervals of nine days (one minute) utilizing real data from the NOAA Surface Radiation (SURFRAD) network [6]. The techniques considered in this study are artificial neural network (ANN), Gaussian process regression (GPR) and boosting trees. Many sources of failure can affect sensor data, including hardware noise, environmental influences, hardware imperfections, and hardware noise from external sources. In order to avoid this problem, we discuss a multiscale data representation to denoise data. The latter shows us its effectiveness and impact for data cleaning and improved predicting performance.

The mean error (ME), mean absolute error (MAE), root mean squared error (RMSE), nRMSE, and coefficient of determination ($R^2$) were used to assess the results of each technique. The findings demonstrate that the prediction models have exceptional predicting accuracy and received very top status. The rest of paper is structured as follows: Section 2 presents the description of the Energy Management by giving its structure. A brief description of the proposed ML techniques, presenting a definition of the most well known performance criteria are given in Section 3. Section 4 presents and discusses the results, giving a statement of the source information. Section 5 concludes the paper with a discussion of potential new ideas.

2. Energy Management

The primary management goal is to maintain a balance between electricity production and demand by giving renewable energies more of a priority and taking into account system constraints.

The structure of a management system for renewable energy is shown in Figure 1. The energy management system (EMS) deals with energy control, management, maintenance, and consumption issues to help with the upkeep and repair of electrical equipment in a factory, farm, or even an entire city [20,21]. It may check on the equipment’s operational status and rapidly enhance management in general. An effective management technique can lower expenses and increase the life of electrical equipment. The EMS can also alert management staff to start looking for replacements for old, high-energy-demanding equipment. The system can instantly send out an alarm in the event of equipment malfunctions or other conditions, thereby allowing management employees to monitor and maintain the system and reduce losses to a minimum. In the renewable energy management system, the EMS can connect monitoring stations, management, and control centers distributed throughout the site using programmed control system technology, network communication technology, and database technology. This enables data collection, storage, processing, statistics, query and analysis, and even data monitoring and diagnosis.
In Figure 1, it is shown that, in order to accomplish the objectives of energy monitoring and efficient management, the EMS distributes renewable-energy-generated power in accordance with a power projections made using the forecasting model. Energy consumption per unit is decreased and economic and energy efficiency are considerably increased by centralized monitoring and the efficient administration of energy data. Hence, this paper focuses on the forecasting model, which is a vital component of the EMS.

**Figure 1.** Structure of an energy management system (EMS).

### 3. Methodology and Techniques

This section describes information sources and the data clarification approach and shows how to build machine learning techniques to predict solar irradiance.

To predict the power of the field, we applied several machine learning techniques such as: ANN, GPR, LASSO, and boosting trees.

#### 3.1. Artificial Neural Network

Artificial neural networks (ANNs) are a subset of machine learning and artificial intelligence (AI) that emulate human learning, memory, and relationship-finding abilities using computer programs. The technology known as artificial neural networks was developed after research into the brain and nervous system [22].

Artificial neural networks are mentioned as one of the time series prediction methods in [23] due to their high adaptability and ability to address difficult, nonlinear situations. The input layer, hidden layer, and output layer are the three layers that make up the ANN structure. Each layer is made up of a sets of nodes, where the hidden layer processes the information after it has been processed by the input layer, and the output layer then sends the network response. The input layer’s number of neurons is equal to the number of inputs. The fundamental equation for a neuron is

$$y = f \left( \sum (w_i x_i + b) \right)$$

where $y$ is the output, $f$ is the activation function, $w_i$ and $x_i$ are the weight and input, respectively, $b$ is the bias term, and $\sum$ is the sum all inputs.

Ref. [24] asserts that a simple ANN architecture yields better predictions than a complicated ANN structure. Additionally, a signal with weight $w_{ij}$ connects every two neurons in a layer of increasing complexity. Each neuron sends the information to the neurons in the layer below after processing it using an activation function.
Figure 2 illustrates an ANN’s basic structure. It involves neural connections. Each connection is allocated a weight, which is a numerical value. Neuron $i$’s output in the hidden layer is $h_i$ [25].

$$h_i = \sigma \left( \sum_{j=1}^{N} W_{ij} x_j + T_{hid}^i \right)$$  \hspace{1cm} (2)

where

- $\sigma()$ is the activation function.
- $N$ is the total number of input neurons.
- $W_{ij}$ are the weights.
- $x_j$ are the inputs to the input neurons.
- $T_{hid}^i$ are the threshold terms of the hidden neurons.

Given that it is a non-linear function that can be distinguished, the sigmoid activation function is the one that is employed the most commonly [26]. This function is a logistic function that has a range of 0 to 1 and has the following expression:

$$f = \frac{1}{1 + \exp(x)}$$  \hspace{1cm} (3)

### 3.2. Gaussian Process Regression (GPR)

Gaussian process regression (GPR) is a flexible nonparametric Bayesian model that allows a prior probability distribution to be constructed over functions directly [27,28]. It is commonly used for non linear problems. Due to their exceptional generalization capabilities, GPs represent one of the most significant Bayesian discriminative kernel learning approaches, where the interpolated values are modeled by a Gaussian process controlled by prior covariance [29]. The best linear unbiased prediction of the values is produced by the GPR by including the proper prior assumptions [30].

The GP offers accurate distributions for all possible functions using the training dataset. In light of this, the number of variables in a GP is limitless and grows as more training datasets are added. A Gaussian process of the mean function $n(x)$ and kernel function $t(x, x')$ is what is known as a GPR [9].

$$F(x) = GP(n(x), t(x, x'))$$  \hspace{1cm} (4)

where $n(x)$ represents the central tendency of $F$. The test input $x$ and test output $Y$ values are related as follows:

$$Y = F(x) + \epsilon$$  \hspace{1cm} (5)

where $\epsilon$ denotes the independent noise term. It is enased by a distribution with a mean of 0 and a variance of $\epsilon_m$ and is defined as:

$$\epsilon = D(0, \sigma^2_m)$$  \hspace{1cm} (6)

The marginal likelihood for the sample dataset is provided by:
\[ H(y|f) = D(y|f, \sigma^2_m f) \]  
(7)

here

\[ Y = [Y_1, Y_2, ..., Y_n]^T \]  
(8)

\[ f = [f(x_1), f(x_2), ..., f(x_m)]^T \]  
(9)

The distribution of the predicted dataset is provided by:

\[ H(Y_s|x, x', Y) = D(\mu_s, \sigma^2_s) \]  
(10)

\[ \mu_s = k_s M (k_s M + \sigma^2_m)^{-1} Y \]  
(11)

\[ \sigma^2_s = k_s s - k_s M (k_s M + \sigma^2_m)^{-1} k_s M \]  
(12)

where

- \( \mu_s \) signifies the GP posterior mean’s average value.
- \( \sigma^2_s \) displays the prediction’s covariance matrix.
- \( k_s M \) is the covariance matrix that associated the training and the test data.
- \( x' \) is the training data.
- \( J \) represents the \( M \times M \) matrix.

The GP prefers functions that correctly smooth and explain trained data. The smoothing functional property allows for excellent generalizations [31].

### 3.3. Least Absolute Shrinkage and Selection Operator (LASSO)

For computing effective model descriptions of nonlinear systems, the least absolute shrinkage and selection operator (LASSO) method is examined [32].

The least absolute shrinkage and selection operator (LASSO) [33] is a well-known high dimensional data analysis technique that may be used for biomarker data, because it can perform regularization and variable selection at the same time. This can increase the precision of predictions and the readability of the results [34]. This approach reduces the residual sum of squares for a linear regression model, provided that the total absolute value of the coefficients is smaller than a tuning parameter [33] Penalized binomial logistic regression has previously been discussed in detail [35]. For \( i = 1, \ldots, N \), let \( x_i = (x_{i1}, \ldots, x_{ip})^T \) signify \( p \)-predictors for \( N \) observations. Assume that the binary logistic regression model’s answers can have values \( G = 1, 2 \). Then,

\[
\Pr(G = 1|x) = \frac{1}{1 + e^{-\beta_0 + x_i^T \beta}}, \quad \Pr(G = 2|x) = \frac{1}{1 + e^{\beta_0 + x_i^T \beta}}
\]  
(13)

where \( \beta_0 \) is the intercept, and \( \beta = (\beta_1, \beta_2, \ldots, \beta_p)^T \) is a \( p \)-vector of regression parameters. This suggests

\[
\log \frac{\Pr(G = 1|x)}{\Pr(G = 2|x)} = \beta_0 + x_i^T \beta.
\]

Then, to identify parameter values to minimize, the LASSO method is applied, which is shown as follows:

\[
\left[ \frac{1}{N} \sum_{i=1}^{N} y_i \times (\beta_0 + x_i^T \beta) - \log(1 + e^{\beta_0 + x_i^T \beta}) \right] + \lambda \sum_{j=1}^{p} |\beta_j|
\]  
(14)

where, \( \lambda \sum_{j=1}^{p} |\beta_j| \) denotes the LASSO’s penalty function.

It has been demonstrated that the LASSO approach does not always offer reliable variable selection. Even when the coefficients are big, the LASSO penalises all of them equally. In contrast, the AL penalises coefficients differently by using adaptive weights [36].

The AL employs a weighted fine
\[ \lambda \sum_{j=1}^{p} w_j |\beta_j|, \quad \text{(15)} \]

where
\[ w_j = \frac{1}{|\hat{\beta}_j|} |\hat{\beta}_j| \quad \text{(16)} \]

These represent the most likely estimate, and \( v > 0 \). The weighted penalty will make it possible for variables with larger coefficients to obtain smaller penalties, thereby potentially leading to a more reliable outcome.

3.4. Boosting Trees

Boosting was first used in the machine learning community after AdaBoost was released [37]. The fundamental concept is to average out the predictions from a number of weak classifiers (high PE) to create a strong classifier (low PE) [38]. Let us say we want to approximate the answer \( y \) by the function \( f(x) \) given the predictors \( x \) and the response \( y \). Usually, to estimate \( f(x) \), we first give the functional form of \( f(x) \) along with a loss function \( L(y, f(x)) \).

The linear model is defined with \( f(x) = x\beta \), where \( \beta \) is a matrix of parameters, and the squared error loss function is the most well-known form of \( f \), i.e., \( L(y, f(x)) = (yf(x))^2 \).

3.5. Performance Evaluation Metrics

The prediction performance of the suggested models was assessed using distinct statistical evaluation criteria. In this paper we are focused on the mean absolute error (MAE), which is computed by expressing the mean absolute deviation of the difference between the expected values and the actual values; the root mean square Error (RMSE), in which it is calculated by using the estimation errors’ standard deviation as its input; and finally the coefficient of determination, which it is determined by using a metric that represents the strength of the linear relationship between the predicted values of the modeling techniques and the actual values. Their expressions are respectively the following:

\[
\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |o_i - p_i| \quad \text{(17)}
\]

\[
\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (o_i - p_i)^2}{N}} \quad \text{(18)}
\]

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (o_i - p_i)^2}{\sum_{i=1}^{N} (o_i - \bar{o})^2} \quad \text{(19)}
\]

where
- \( o_i \) is the observation’s actual value.
- \( p_i \) is predicted value of of the observation.
- \( \bar{o} \) is the average of the actual observation values.
- \( N \) is the number of samples utilized for the statistical evaluation criteria.

These errors are frequently standardized, especially for the RMSE; as a standard, the mean value is typically employed, however, various definitions can be obtained.

\[
n\text{RMSE} = \sqrt{\left(\frac{1}{N} \times \sum_{i=1}^{N} (o_i - p_i)^2\right)} \quad \text{(20)}
\]
where \( \bar{y} \) is the mean value of \( p \). The index of agreement (d), which is normalized between 0 and 1, and the correlation coefficient R (Pearson Coefficient) are two more indices that can be employed.

4. Simulation Results

4.1. System Description

A single PV array block was composed of (2) parallel strings with (12) modules connected in series on each string. The stand-alone or off-grid PV was considered in this study, where the PV system is independent of the utility grid. We studied this type of system by supplying with a load based on the climatic for the purpose of extracting the power.

The “basic” single diode model typically serves as an illustration of how a PV cell operates. This simplified model is frequently employed for failure investigations, performance analysis, and stability analysis [39]. Figure 3 displays the equivalent circuit of a PV cell.

![Figure 3. Equivalent circuits of one-diode model for a PV cell.](image)

This model is composed of five parameters and it contains:

- \( I_{ph} \): a current source which represents the irradiance received by the cell.
- A diode for modeling the PN junction of the cell.
- \( R_s \): the series resistance representing the resistivity of the material through which the cell is made.
- \( R_{sh} \): the parallel resistance, which represents all the paths crossed by the leakage current, either in parallel with the leakage current, in parallel with the cell, or at the edge of the cell.

The current supplied by the cell is given by the following equation:

\[
I = I_{ph} - I_d - I_{shunt} \tag{21}
\]

where

\[
I_{shunt} = \left( \frac{V + R_s I}{R_{sh}} \right) \tag{22}
\]

As we said before, this model is also called “implicit model with five parameters” which are: \( I_{ph}, I_s, a, R_s, R_{sh} \).

Where

- \( I_{ph} \) is the photo current, which is proportional to the irradiance received by the cell.
- \( I_s \) is the saturation current of the diode.
- \( a \) is the ideality factor of the diode at (1 to 2).
- \( V_t = N_s K_b T_c / q \) is the thermal voltage as a function of the number of cells in series in the PV module, the cell temperature, Boltzmann’s constant, and the charge of the electron.

The photo current depends linearly on the global irradiance \( G \) and the temperature \( T \), and it is found in the following equation:

\[
I_{ph} = \frac{G}{G_n} \left( I_{ph,n} + K_s \Delta T \right) \tag{23}
\]
\[ \Delta T = T - T_n \]  
\[ T_c = T_n + \left( 0.2 \left( \frac{G}{G_n} \right) \right) \]

where

- \( T = T_c + 273.15 \)
- \( I_{ph,n} \) is the photo current at the standard test condition, which corresponds to the short-circuit current of the PV module given by the manufacturer.
- \( K_i \) is the temperature coefficient of the short circuit current.
- \( T \) is the temperature of the PV cell in kelvin.
- \( T_n \) is the cell temperature at standard test condition [25 °C].
- \( G \) is the solar radiation received by the PV cell [W/m²].
- \( G_n \) is the solar radiation at standard test conditions [1000 W/m²].
- \( T_a \) is the ambient temperature in °C.
- \( T_c \) is the temperature of the PV cell in °C.

The well-known equation for the saturation current of the diode is given by:

\[ I_s = I_{s,n} \left( \frac{T}{T_n} \right)^3 \exp \left( \frac{qE_g}{aK_b} \left( \frac{1}{T_n} - \frac{1}{T} \right) \right) \]  
\[ E_g \] is the the band energy of the semiconductor.  
\( I_{s,n} \) is the the nominal saturation current at (STC).

An improved equation to describe the saturation current that considers the variation of temperature is given by:

\[ I_s = \frac{(I_{sc,n} + K_v \times \Delta T)}{\exp[(V_{sc,n} + K_v \times \Delta T) \cdot aVt] - 1} \]

where the constant \( K_v \) is the temperature coefficient of the open circuit voltage.

These equations were implemented on the Simulink that followed the architecture in Figure 4. In order to evaluate the model (Figure 4), simulation tests were performed using the manufacturer settings [39].

4.2. Data Collection

MATLAB software (R2021b) was used in this work to train machine learning techniques. A case study was conducted using a dataset obtained from the NOAA’s Surface Radiation (SURFRAD) network, which is composed of seven stations from the U.S. [6] to assess the effectiveness of the proposed method, which includes nine days of data (from 1 July 2017 to 9 July 2017) from the Goodwin Creek, Mississippi (GWN) station. The initial photoelectric data included \( 9 \times 660 = 5940 \) observations from 8 am to 6 pm; the choice of this time interval was made at the time that there was an energy production and a power generation, which were gathered at 1-min intervals. A collection of data inputs was employed in this experiment. In fact, we fed the studied system with the irradiance and the temperature, which were processed in the database detailed above to extract the power. Then, for each input pair (irradiance, temperature), we determined the power produced. Therefore, from a real database, we determined a database for the power supplied by the PV system. Figure 5 shows us the curves of the temperature, irradiance, and the power measured on one day. The following figure shows us, moreover, that the variation of the irradiance was not very big compared to the variation of the power, and the variation of the temperature was light. We can conclude that if the temperature is not very important, the power will be the image of the irradiance. However, the variation in the power was multiplied to the value of the irradiance (between 120 and 800); that is why we predicted the latter and we made a prediction of the power, since it affects the robustness of the techniques. This expresses the importance of the power compared to the irradiance.
To accomplish the objectives of energy monitoring and efficient management, the EMS distributes renewable energy produced power in response to a power projection made...
using the forecasting model. The energy consumption per unit is decreased, and economic and energy efficiency are considerably increased by the centralised monitoring and efficient administration of energy data. As a result, the forecasting model is a key component of the EMS. Because of this, the EMS can predict future energy output with accuracy.

The best option for quick and effective energy reduction is energy management. In order to maximise energy conservation, our major goal is to analyse the techniques of consuming energy and to improve generation and utilization. For optimal energy usage, both the generating and utilisation sides must be considered. The most interesting goal for this paper is to ensure a better management, to reach an optimal energy, to reduce the energy consumption, and to obtain higher accuracy. We proposed four different cases by adding time intervals in each case to obtain better management.

In the current work, different predictions horizon lengths were used to improve the energy management as well as to evaluate the forecasting performance of the proposed techniques. To achieve these goals, four cases were addressed as follows:

- **Case 1**: During this case, one day was examined. A total of 8 hours (h) and 48 minutes (min) were used for training to predict 2 h and 12 min.
- **Case 2**: In this case, two days were considered (1320 observations) for the training phase to predict the next day.
- **Case 3**: Four days (2640 observations) were examined for training to predict the following two days.
- **Case 4**: We continued increasing the training samples to obtain larger prediction horizon lengths. In such a way, six days (3960 observations) were intended to perform the training phase to predict three successive days.

The selection of the forecasting horizon length was based on the significant variation of the irradiance, either during the hours of one day or through successive days, in which the irradiance underwent several peaks with totally different behavior. These considerable variations are clearly shown in the following curves illustrated in Figures 6–9, especially in the three latter.
For example, in Figure 9, which represents Case 4, we see the variation in irradiance and the power over the time, such as in (Figure 9a). The three successive predicted days did have not the same variation in irradiance; every day had its behavior. We can see a big variation in the first forecasted day with a significant irradiance. On the second predicted day, we can see that the values of irradiance decreased compared to the first day; then, they started to increase a little with a totally different variation in the irradiance in comparison with the previous day. We can say here that the irradiance was a bit weak with another weak behavior. For the third day, we found another different pattern than the two previous ones, with weak variation peaks.

As a result, the error was practically low, but we attempted to diminish it and enhance the obtained results to show better energy management; we used the multiscale model to denoise the raw data. By using this model, the cases used show us that the errors were decreased even more and that the proposed techniques remained generally robust.

4.3. Discussion

The previous section briefly described the proposed approaches, as well as the use and the need of the system and the studied cases. According to the latter, we can discuss the following results. Figure 6 displays the predicted values of the proposed ANN model that was used to estimate the irradiance, as well as a time graph showing the actual values over the first day (1 July 2017). Additionally, it was reported that the actual and predicted values indicated almost the same trend. In addition, this figure shows the average daily irradiance and power, where the error of the irradiance and power was higher, as shown in Figure 6a,b.

Figures 7–9 show us the prediction of the irradiance and the power for the three other cases. It can be concluded here that the error was lower than the first case, so we conclude that these cases have good prediction properties.

In addition, the effectiveness of the proposed ANN model was evaluated against four trained and tested machine learning methods, including LASSO, GPR, and boosting trees for the time period of 1 July through 9 July. The ME, MAE, RMSE, NRMSE, and $R^2$ values were used to compare the effectiveness of these strategies. The following table displays the findings. The mentioned evaluation indicators were used to evaluate the performance of
the NN and other models in order to assess the accuracy of various prediction models. The performance of the persistence models for raw data and multiscale model, including NN, GPR, LASSO, and boosting trees, is shown in Tables 1 and 2 for four distinct prediction scales. For Case 1, Table 1 show us that the different criteria of GPR, which are ME, RMSE, nRMSE, MAE, and $R^2$, were, respectively, 18.40, 19.10, 0.11, 18.40, and 1.00 in raw data, as shown in the same table that the predicted values in multiscale were improved with blue 43.70, 43.40, 45.45, 43.70, and 0% for the five different criteria to achieve the following respective values of 10.36, 10.81, 0.06, 10.36, and 1.00. This approach had the good performance. For the second case, the GPR model had the best values where, in the raw data, it achieved 7.85, 13.30, 0.06, 10.53, and 1.00 for the different evaluation criteria ME, RMSE, nRMSE, MAE, and $R^2$, respectively. Correspondingly, these values were enhanced with 87.64, 83.83, 83.33, 85.94, and 0% when we used the multiscale, so the obtained values were 0.97, 2.15, 0.01, 1.48, and 1.00, respectively.

For Case 3, when we increased the data more than the previous case, we can see that the GPR technique had the best values of performance criteria for all different performances, which were 1.75, 22.87, 0.05, 18.98, and 1.00 in the raw data. When using the multiscale, these values were 56.57, 81.99, 80, 83.14, and 0% to obtain 0.76, 4.12, 0.01, 3.20, and 1.00, respectively, for the ME, RMSE, nRMSE, MAE, and $R^2$. By increasing the data in the last case to achieve six days for the training and three days for testing to show the performance of our models, we can see that the ME, RMSE, nRMSE, MAE, and $R^2$ of the GPR model were, respectively 0.55, 21.87, 0.06, 17.87, and 1.00; after using the multiscale, this technique achieved 0.13, 3.71, 0.01, 2.92, and 1.00 respectively. This model was enhanced by a percentage of 76.36, 83.04, 83.33, 83.66 and 0%, respectively. We can say that the multiscale had an especially good impact for this technique. For the four prediction irradiance scales of Case 1, Case 2, Case 3, and Case 4, these findings demonstrate that multiscale representation has good impact for enhancing the performance criteria of the different models.

We selected four prediction scales for Cases 1, 2, 3, and 4 to test the performance of the power prediction model. As shown in Table 2, the LASSO model had the best values for raw data, and the multiscale model performed best for the first case. In the second case, we can see that the GPR technique had the minimum error values for the two applications. Moving to the third case, we can see that the multiscale model had a good impact for enhancing the NN method. The last case showed us that the NN technique outperformed other prediction models for the raw data and for using the multiscale model.

<table>
<thead>
<tr>
<th>Cases</th>
<th>Methods</th>
<th>Raw Data</th>
<th>Multiscale</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 1</td>
<td>NN</td>
<td>44.60</td>
<td>18.40</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>18.40</td>
<td>18.40</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>24.39</td>
<td>24.39</td>
</tr>
<tr>
<td></td>
<td>Boosting Trees</td>
<td>63.82</td>
<td>80.19</td>
</tr>
<tr>
<td>Case 2</td>
<td>NN</td>
<td>7.32</td>
<td>7.85</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>18.40</td>
<td>18.40</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>24.39</td>
<td>24.39</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>63.82</td>
<td>80.19</td>
</tr>
<tr>
<td>Case 3</td>
<td>NN</td>
<td>1.88</td>
<td>1.75</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>18.40</td>
<td>18.40</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>24.39</td>
<td>24.39</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>63.82</td>
<td>80.19</td>
</tr>
<tr>
<td>Case 4</td>
<td>NN</td>
<td>0.39</td>
<td>0.55</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>18.40</td>
<td>18.40</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>24.39</td>
<td>24.39</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>63.82</td>
<td>80.19</td>
</tr>
</tbody>
</table>
Table 2. Performance evaluation of Predicted Power under different cases.

<table>
<thead>
<tr>
<th>Cases</th>
<th>Methods</th>
<th>Global Performances</th>
<th>Raw Data</th>
<th>Multiscale</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>ME</td>
<td>RMSE</td>
<td>nRMSE</td>
</tr>
<tr>
<td>Case 1</td>
<td>NN</td>
<td>41.75</td>
<td>48.01</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>11.92</td>
<td>11.92</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>7.31</td>
<td>9.75</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>30.40</td>
<td>35.69</td>
<td>0.48</td>
</tr>
<tr>
<td>Case 2</td>
<td>NN</td>
<td>10.74</td>
<td>26.49</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>10.29</td>
<td>21.01</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>23.36</td>
<td>42.55</td>
<td>0.50</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>11.66</td>
<td>28.62</td>
<td>0.24</td>
</tr>
<tr>
<td>Case 3</td>
<td>NN</td>
<td>0.18</td>
<td>39.17</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>0.27</td>
<td>36.90</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>76.52</td>
<td>97.80</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>1.54</td>
<td>45.14</td>
<td>0.14</td>
</tr>
<tr>
<td>Case 4</td>
<td>NN</td>
<td>0.41</td>
<td>37.48</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>GPR</td>
<td>0.93</td>
<td>39.97</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>LASSO</td>
<td>61.35</td>
<td>87.55</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>Boosting trees</td>
<td>1.26</td>
<td>49.13</td>
<td>0.18</td>
</tr>
</tbody>
</table>

5. Conclusions

In this study, the use of multiscale machine learning techniques (i.e., ANN, GPR, LASSO, and boosting trees) to process irradiance and power forecasting has been demonstrated using historical real data from the Goodwin Creek, Mississippi (GWN) station in the United States. Given that measured process data are generally tainted by errors (noise) that hide the significant features in data and degrade the effectiveness of the forecasting approach, the multiscale representation data has been proven to be an effective tool of data analysis and feature extraction owing to its capacity to present useful separations of features. Accordingly, the goal of the developed approach was to apply the multiscale representation data to further improve the efficiency of applied ML irradiance and power forecasting techniques. In order to evaluate the proposed approach, four cases were studied that varied from one day to nine (1 July 2017 to 9 July 2017) using the mean error (ME), mean absolute error (MAE), root mean square error (RMSE), nRMSE, and coefficient of determination \(R^2\) criteria. The obtained results proved the efficiency of the proposed approach with nRMSE values (for example) between 0.01 and 0.37 for predicted irradiance, and nRMSE values between 0.02 and 0.71 for predicted power.

Author Contributions: Methodology, M.M. (Majdi Mansouri); software, M.M. (Manel Marweni); validation, M.M. (Manel Marweni); investigation, M.H. and M.F.M.; writing—original draft preparation, M.M. (Mael Marweni); writing—review and editing, M.H. and M.M. (Majdi Mansouri); supervision, M.H., M.M. (Majdi Mansouri) and M.F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Qatar National Library through the Qatar National Research Fund (QNRF) Research Grant.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, MM, upon reasonable request.

Acknowledgments: Open access funding provided by the Qatar National Library.

Conflicts of Interest: The authors declare no conflict of interest.

References

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.