*Research article*

# Forecasting energy production of a PV system connected by using NARX neural network model

**Marwa M. Ibrahim, Amr A. Elfeky and Amal El Berry\***

Mechanical Engineering Department, Engineering and Renewable Energy Research Institute, National Research Centre (NRC), Cairo 12622, Egypt

**\* Correspondence:** Email: amalelberry@yahoo.com; Tel: +20233371615.

**Abstract:** Applying artificial neural network techniques to forecast the electricity production of photovoltaic (PV) power plants is a novel concept. A reliable analytical model for calculating the energy output of a grid-connected solar plant is very difficult to establish because of hourly, daily, and seasonal variations in climate. The current study estimated and predicted the energy production of a connected PV system that was installed in Cairo, Egypt (30.13° N and 31.40 ° E) using an artificial neural network. Four seasons' worth of data (summer, autumn, winter, and spring) were methodically assessed using information from the climate database. The parameters that had an impact on the electrical data of PV modules included meteorological and irradiation variables, energy output, and the user's needs used to verify the NARX feedback neural networks. Prediction performance metrics were obtained, such as the correlation coefficient (R) and root mean square error (RMSE). The observed correlation coefficient ranged from 99% to 100%, indicating that the expected results are verified, while the mean error fluctuates very little.

**Nomenclature:** ANN: Artificial neural networks; DiffHor: horizontal diffuse irradiation (kWh/m²); E_Solar: Energy supplied to the user from solar (kWh); E_User: Energy needs of the user (kWh); Earray: Effective energy at the array output(kWh); EFrGrid: Energy from the Grid (kWh); GlobEff: Effective global irradiation on the collectors (kWh/m²); GlobHor: Horizontal global irradiation (kWh/m²); GlobInc: Incident global irradiation in the collector plane (kWh/m²); MSE: Mean squared error; NARX: Nonlinear autoregressive network with exogenous input; PnomPV: STC installed power; PR:

Performance ratio; PV: Photovoltaic; R: Correlation coefficient (R); $R^2_{adj}$: Adjusted R-squared; RMSE: Root mean square error; STC: Standard test condition; T_Amb: Ambient temperature (°C)

## 1. Introduction

Because of hourly, daily, and monthly variations in climate, it is difficult to find a credible analytical model for determining the energy production of a grid-connected solar plant. PV system's energy production will be estimated and predicted utilizing accurate methods, and some previous related studies will be described.

A power inverter device connects the solar array or panels in a grid-connected photovoltaic (PV) system to the utility grid, enabling them to run in parallel with the electric utility grid. An energy-producing solar power system that is linked to the utility grid is called a grid-connected PV system. Solar panels, a power conditioning unit, one or more inverters, and grid connection equipment are components of a grid-connected photovoltaic system [1]. When everything is operating properly, the grid-connected PV system feeds electricity not needed by the related load into the utility grid [2]. Nonetheless, in recent years, there has been a considerable increase in the number of solar-powered homes connected to the local electrical grid. These grid-connected PV systems may be able to meet most, if not all, of daily power demand with solar panels while remaining linked to the local electrical grid network at night [3].

A grid-connected PV system allows homes and businesses to use solar energy for all or part of their energy needs while still receiving power from the standard electrical mains grid at night or on cloudy, rainy days, giving them the best of both worlds. In grid-connected PV systems, power is transferred between the main grid and the sun in accordance with the actual demand for electricity [4]. In a grid-connected PV system, sometimes referred to as a "grid-tied" or "on-grid" system, the PV solar panels or array can supply electricity back into the grid through an electrical connection to the local main power grid [5]. The main advantages of a grid-connected PV system are its simplicity of use, relatively low maintenance and operation costs, and reduced electricity bills. The disadvantage is that, in order to generate the required amount of additional power, a sufficient number of solar panels must be installed [6]. Since grid-tied systems feed solar energy back into the grid, most grid-connected designs can be created without expensive backup batteries [7]. In addition, there is no need to determine the size of the solar panels or calculate solar energy consumption, because this type of PV system is permanently connected to the grid. This implies a wide range of possibilities, ranging from a system as tiny as 1.0 kWh on the roof to nearly eliminating your electricity expenditures with a floor-mounted array [8].

The most common machine learning strategy used is the artificial neural network (ANN) technique (e.g., [9]). Neural networks can be used to simulate, forecast, and optimize the performance of engineering systems, such as renewable energy systems. ANNs are commonly used in the real world to save money and time when solving complex nonlinear engineering problems [10]. Neural networks are algorithms based on the structure and operation of the human brain. The structure is divided into three layers: input, concealed, and output. Each layer is connected to the one below by a network of nodes, or neurons. Weights are computed iteratively throughout the training phase and utilized to define the connection between neurons. During the training phase, networks are randomly started, and learning takes place by altering weights until a specific criterion is reached [11,12], to balance supply and demand, anticipate fuel production and electrical power supply [13], and evaluate the electrical load for demand or energy consumption (building, transportation, and industrial use) [14,15]. Ghenai

et al. [16] used the ANFIS technique to produce highly accurate and short-term energy consumption projections for educational institutions. The forecasting algorithm was evaluated using historical data with extremely short temporal spans (0.5–4 hours ahead). The predictive algorithm did a good job projecting the building's future energy use.

Grid-connected PV systems vary in size, from small rooftop solar power systems for homes and businesses to enormous solar power plants for utilities. When the conditions are right, the grid-connected PV system sends any excess electricity to the utility grid by the linked load. Most consumers' needs can be met by large-scale systems [17–19]. Using historical data and learning, the NARX network model can be utilized to achieve multistep forward prediction. The NARX neural network architecture is made up of input, hidden, and output. Different input information is received by the input layer from the concealed layer, which could be made up of one or more layers. The objective of the current study is to forecast the power output of a grid-connected PV by using the NARX neural network model, including meteorological data and irradiation variables, energy output, and user's needs.

## 2. Materials and methods

### 2.1. Weather data from the site of the illustrated grid-connected PV system

The system under investigation is located in Cairo, northeast Egypt, on the eastern coast of the Nile River, approximately 500 miles (800 kilometers) downstream of the Aswan High Dam. The climate is mild-to-hot for the majority of the year, with summer temperatures reaching 34 °C (92 °F) and winter temperatures reaching 18 °C (65 °F). Temperatures are significantly warmer than in Central Europe, rarely dropping below 20 °C. March and April can be windy, resulting in sandstorms. In July and August, temperatures soar, with daily highs frequently reaching >30 °C (approximately 100 °F). Daylight duration in the summer is approximately 4 hours longer than in the winter. In June, daylight can exceed 14 hours. Winter has the longest nights, the opposite being true in the southern hemisphere. In December, the night in Cairo lasts over 14 hours, the day being two hours shorter. Table 1 depicts the times for sunrise, sunset, total hours of daylight, and solar noon throughout the year.

**Table 1.** Approximate sunshine and sunset duration in Cairo (https://www.worlddata.info/) in 2024.

| Month | Sunrise | Sunset | Hours of daylight | Solar noon |
|---|---|---|---|---|
| January | 06:50 am | 05:17 pm | 10:28 h | 12:04 pm |
| February | 06:34 am | 05:44 pm | 11:10 h | 12:09 pm |
| March | 06:03 am | 06:04 pm | 12:02 h | 12:03 pm |
| April | 05:26 am | 06:23 pm | 12:57 h | 11:54 am |
| May | 05:59 am | 07:42 pm | 13:43 h | 12:51 pm |
| June | 05:52 am | 07:59 pm | 14:07 h | 12:55 pm |
| July | 06:02 am | 07:59 pm | 13:56 h | 01:01 pm |

| Month | Sunrise | Sunset | Hours of daylight | Solar noon |
|---|---|---|---|---|
| August | 06:21 am | 07:37 pm | 13:16 h | 12:59 pm |
| September | 06:38 am | 07:01 pm | 12:23 h | 12:50 pm |
| October | 06:56 am | 06:25 pm | 11:29 h | 12:40 pm |
| November | 06:19 am | 05:00 pm | 10:41 h | 11:39 am |
| December | 06:42 am | 04:58 pm | 10:16 h | 11:50 am |

In June, daylight lasts approximately 14 hours. This suggests that visual obstruction occurs approximately every 2:37 hours every day. In December, on the other hand, daylight lasts a little over 10 hours, with visual obstruction occurring every 3:10 hours.

## 2.2. Characteristics of design 545-kW grid-connected PV array model

The investigated solar power facility is located in Cairo (30.13° N and 31.40° E). The optimal tilt angle of the solar cell allows the solar panels to attain the best energy conversion efficiency. Array orientation is a crucial component of any PV performance model computation, and is typically classified as either tracked or fixed. The orientation with a fixed tilt is immovable; on the other hand, the orientation of a tracked array moves over time in order to reduce the angle of incidence between the array and the sun. The way trackers move categorizes them into various types. A fixed tilt array orientation is defined by its azimuth angle [θ_ (T, Array)] and tilt angle (θ_T) [20]. Table 2 displays the characteristics of the grid-connected PV array model as a result of the PVsyst simulation.

The array azimuth angle [θ_ (T, Array)] and array tilt angle (θ_T) for a fixed tilt array orientation range from 10 to 80 by step 10. This is one of the most important aspects to consider when determining a PV system's efficiency. The monthly average irradiation in a is optimum in tilted plane in general, throughout most of the months of the year, from March to October. An optimal tilt angle allows the solar module to obtain more energy than a horizontal plane. The PV support system type is fixed and monocrystalline with a total power of 545 Wp as seen in Table 2. The total number of solar cells is 770 modules. The total solar array power in standard conditions is 420 kWp at 25 °C while the total capacity is 400 kWp at operating conditions at 50 °C.

The performance ratio is the ratio of energy successfully produced (used) to energy that would be produced if the system operated continuously at its nominal STC efficiency. In most grid-connected systems, the available energy is denoted as E_Grid. The potential energy produced under STC conditions is indeed equal to "GlobInc × PnomPV", "PnomPV" being the STC installed power. The performance ratio is determined in Eq (1) as (PVsyst guide):

$$PR = E\_Grid \, / \, (GlobInc \times PnomPV) \tag{1}$$

**Table 2.** Design characteristics of the grid-connected PV array model.

| Location | | | System information | | | | | | User's need | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | PV array | | Inverters | | | | Daily household consumers | |
| Cairo, Egypt | Latitude | 30.13 °N | No. of modules | Pnom total | No. of units | Pnom total | Pnom ratio | | Seasonal modules | |
| | Longitude | 31.40 °E | 770 modules | 420 kWp | 35 units | 350 kWac | 1.199 | | Average | 11.0 kWh/day |
| Grid-connected system | | | | | | | | | | |
| PV field orientation, fixed plane, tilt ($\theta_T$)/azimuth (0) | | 10/0 | 20/0 | 30/0 | 40/0 | 50/0 | 60/0 | | 70/0 | 80/0 |
| PV module | | | | | | | | | | |
| | | Unit nom power | | Number of PV module | | Normal (STC) | | | OptimizerArray | |
| 545-Wp Monocrystalline | | 545 Wp | | 770 modules | | 420 kWp | | | 70 string x11 in series | |

*At reference conditions, the maximum power 13.04 A and temperature coefficient 5.6 mA/°C or 0.04%/°C*

Details of housed consumer seasonal modulation, average = 11.0 kWh/day

| Summer (Jun–Aug) | Autumn (Sep–Nov) | Winter (Dec–Feb) | Spring (Mar–May) |
|---|---|---|---|
| 12642 | 10142 | 10942 | 10142 |
| kWp: The kilowatt peak (maximum value to be achieved) | kWh (kilowatt hour yield achieved PV) | | |

## 2.3. Artificial neural network (ANN) approach

When predicting the power output of a plant that can be compared to the recorded power trend, the artificial neural network (ANN) technique is quite useful. ANNs can deal with complex system modeling, prediction, and optimization, being widely used in energy and renewable energy systems. In [21], authors proposed an application that includes modeling, simulation, sizing, control, and diagnosis of diverse energy systems, such as grid-connected hybrid PV systems. The main objective of this study is to examine if multilayer networks are suitable for modeling and forecasting the electricity generated by a (545 wp) panel grid-connected solar plant that is erected on a rooftop near Cairo, Egypt. To this end, models will be created and examined by employing the nonlinear autoregressive exogenous model (NARX), a nonlinear autoregressive model with exogenous inputs, for time series modeling. The model will analyze if the driving (exogenous) series, or the externally determined series that drives the series of interest, is related to both its current and previous values, as well as the series' historical values. Furthermore, the model includes an error element that represents the difficulty of accurately predicting the current value of the time series without knowledge of other factors. The NARX network model is generated in MATLAB/Simulink.

### 2.3.1. Time-series NARX feedback neural

The nonlinear autoregressive network with exogenous input (NARX) is a recurrent neural network widely utilized in time series applications. The NARX model is made up of two main parts:

autoregressive (AR) and exogenous inputs (X). The X component simulates the effect of external influences on the time series (parameters that influence the electrical data of PV modules include GlobHor, DiffHor, T_Amb, GlobInc, and GlobEff), whereas the AR section depicts the temporal correlations between past and current time series values, forecasting future values based on past observations. NARX maps the input and output using a multilayer perceptron with a time delay unit and output feedback in the input. This model is effective for modeling and anticipating grid-connected PV behavior by forecasting grid output energy, which is the user's requirement. This is achieved by combining the powers of the electrical grid and a PV system via electronic inverters. The NARX model is advantageous because it can represent nonlinearity while considering feedback signals and external variables. It is based on the nonlinear autoregressive model, which is widely used in time series modeling.

### 2.3.2. Defining equation for the NARX model

NARX solutions are more accurate than others, but this solution is only used if previous values of y(t) are unavailable when deployed. Input data is a 96 × 5 matrix representing dynamic data, with 96 timesteps and 5 entries. Target data is a 96 × 5 matrix that represents dynamic data, with 96 timesteps with 5 elements, as in [22]. The NARX model can be expressed by the following equation, where y(t) is the expected output value, u(t) is the input variable, and $n_y$ and $n_u$ are the input and output time delays, respectively as shown in Eq (2):

$$y(t) = f\ (y\ (t{-}1),\ y(t{-}2),\ y(t-n_y),\ u(t-1),\ u(t-2),\ u(t-n_u)) \tag{2}$$

Equation (2) can be simplified in Eq (3):

$$y(t) = f\ [y\ (t-1);\ u\ (t-1)] \tag{3}$$

Where the input and output regressor vectors are u $(t-1)$ and $(t-1)$, respectively. A typical perceptron network, which has multiple layers, can be used to estimate the usually unknown nonlinear mapping function $f$. We then discuss the resulting connectionist design using a NARX network. A two-hidden-layer NARX network is depicted in Figure 1, where the historical values of the output signal are used to regress the subsequent value y(t) of the dependent output signal, using the historical values of an independent PV energy output input signal. A feedforward neural network may be used to approximate the function $f$ to implement the NARX model. The resulting network, which uses a two-layer feedforward network for approximation, is shown in the diagram below. Furthermore, the implementation described in [23] supports a vector ARX model with multidimensional input and output.
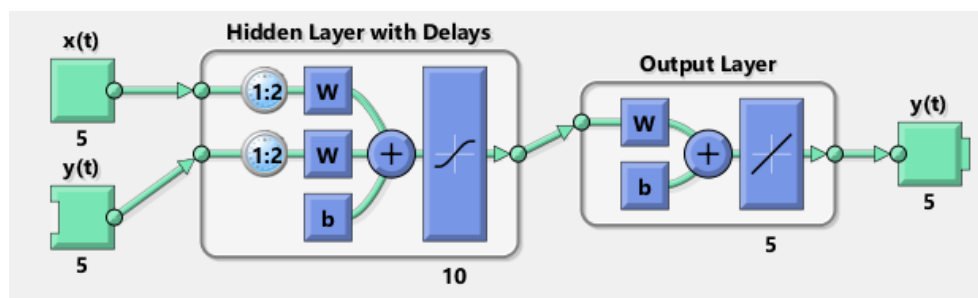


**Figure 1.** Two-hidden-layer NARX network.

There are several applications for the NARX network, namely as a predictor to determine the next value of the input signal. In this study, the accuracy of the NARX model in calculating the grid-connected PV system is measured using three standard metrics. The trained network outcomes in training, testing, and total data regression stages are determined by the root mean square error (RMSE), correlation coefficient (R), and adjusted $R^2$ provided in Eqs (4) and (5). The superlatively trained NARX model is indicated by an RMSE value close to 0 and an adjusted $R^2$ value close to 1.

Figure 1 depicts the model developed by the dynamic nonlinear autoregressive system with external exogenous input (NARX). By minimizing the output function through output feedback to the input, it is possible to forecast the future value of the output y(t) based on the past values of y(t) and x(t). The median squared difference between the target variable's actual and anticipated values is measured by the mean squared error (MSE). Lower MSE values indicate a better performance; a value of 0 denotes flawless prediction. The number of data points (the target variable's actual value for the ith data point) is given by yi, and ŷi indicates the target variable's forecasted value for the ith data point [24].

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(Yi - \hat{Y}i\right)^2 \tag{4}$$

where:
MSE = Mean square error
n = Number of data points
Yi = Observed values
$\hat{Y}i$ = Predicted values

The square root of the mean squared error is called the root mean squared error (RMSE). It calculates the residuals' standard deviation; the average squared difference between the expected and actual numbers is taken as the square root.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Yi - \hat{Y}i\right)^2} \tag{5}$$

The coefficient of determination, often known as R-squared, represents the fraction of the variance in the dependent variable that can be explained by the linear regression model. It is a scale-free score, which means that regardless of the values, R-squared as expressed in Eq (6) will always be lower than 1.

$$R^2 = 1 - \frac{\sum(Yi - \hat{Y})^2}{\sum(Yi - \bar{y})^2} \tag{6}$$

The adjusted R-squared is a modified in Eq (7) form of R-squared that accounts for the number of independent variables in the model. It is always less than or equal to $R^2$. In the following formula, n represents the number of observations in the data, and k represents the number of independent variables.

$$R_{adj}^2 = 1 - \left[\frac{(1-R^2)(n-1)}{n-k-1}\right] \tag{7}$$

Prediction is a type of dynamic filtering in which previous values from one or more time series are used to forecast future values. Nonlinear filtering and prediction are performed using dynamic neural networks with tapped delay lines. Predictive models are also employed in system identification (or dynamic modeling), which involves creating dynamic models of physical systems.

These dynamic models are useful for analyzing, simulating, monitoring, and controlling a wide range of systems, including PV solar system behaviors.

### 2.3.3. Data processing for NARX ingestion

Professionals use data to generate well-informed predictions every day, everywhere. Meteorologists forecast future weather by utilizing historical meteorological data, similarly to the characteristics that affect the future energy output of PV modules. Time series data are used to make predictions, sometimes referred to as "time-stamped data", which is a group of data points on a particular topic where each value is assigned a time period. In addition to external parameters, the model can accept feedback from the output based on changes in meteorological data over a predetermined period of time. After a certain period of time, the model can accept feedback from the output as well as external variables. Furthermore, ambient temperature can affect GlobHor, DiffHor, T_Amb, GlobInc, and GlobEff, all of which effectively modify cell temperature.

Even small variations in the outside temperature can result in a large buildup of heat from the annual DC energy generated by the PV array, the annual AC energy added to the grid, and idle conditions. The data is organized into a table using MATLAB and Excel, with columns including historical power production data from solar PV plants, hour angles, zenith angles, and weather variables. Power and meteorological data must be synchronized with respect to sunrise and sunset in order for them to be correct. Since the missing data contains significant information that could affect the model's performance, it is necessary to take it into consideration. The MATLAB Spline interpolation method is used in this study to fill in missing data since it is superior to linear, nearest neighbor, and shape-preserving methods in terms of smoothness [25].

### 2.3.4. Validation and test training

Test training and validation were completed in the following way: Three sets of input and target vectors were randomly selected; 70% was utilized for training, 15% was used to assess the network's generalization and prevent overfitting, and the remaining 15% was used as an independent test of network generalization.

The training process was repeated multiple times using different algorithms (Bayesian regularization, scaled conjugate gradient, and Levenberg-Marquardt) until the optimal result was obtained. The loop was closed for multistep prediction tests, and simulation based on the accuracy parameters of the Bayesian regularization after training was finished. Although the Bayesian regularization process usually takes longer, it can produce good generalizations for challenging, tiny, or noisy datasets. Training halts in accordance with regularization or adaptive weight minimization. The average squared difference between the targets and the outputs is called the mean squared error. It is better to have lower values. Zero indicates the absence of an error. The correlation between targets and outputs is measured by R values. A relationship is said to be close when the R value is 1, and random when it is 0.

## 3. Results and discussion

### 3.1. Results of PVsyst simulation to optimize 545-wp grid-connected PV array setting

Figure 2 displays the optimum tilt angle in terms of incident global irradiation in the collector plane GlobInc (kWh/m²).
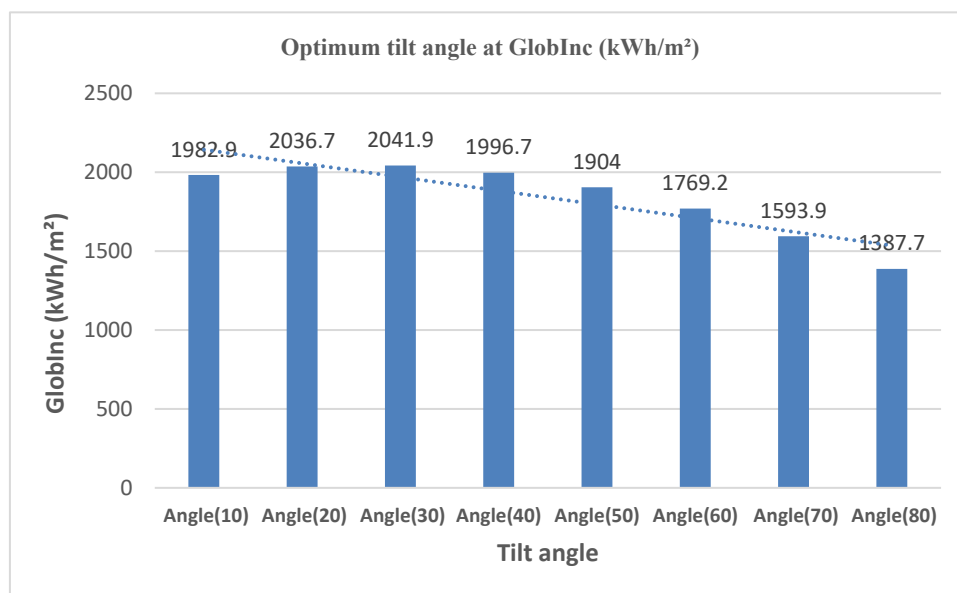


**Figure 2.** Optimum tilt angle in terms of incident global irradiation in the collector plane GlobInc (kWh/m²).

The variables horizontal global irradiation [GlobHor (kWh/m²)], horizontal diffuse irradiation [DiffHor (kWh/m²)], ambient temperature [T_Amb (°C)], incident global irradiation in the collector plane [GlobInc (kWh/m²)], and GlobEff (kWh/m²) without any optical corrections are effective globally; after all, optical losses are included in the balances and primary outcomes of the grid-connected PV setting using PVsyst software. The monthly values at PV field orientation, fixed plane, and tilt ($\theta\_T$) varied from 10 to 80 by step 10. These computed values were acquired for all the variables indicated in the main results and balances. The collecting plane's incident global irradiation [GlobInc (kWh/m²)] is at its optimal when tilted to 30º, as Figure 2 illustrates. From the PVsyst simulation of a grid-connected system, annual average values are possible for temperature, efficiency, and sums of irradiance and energy. For the study site, the annual global irradiance on the horizontal plane is 1882.4 kWh/m², while the annual global incident energy on the collector without optical adjustments and effective global irradiance after optical losses are 2041.9 and 1998.2 kWh/m², respectively. With this effective irradiance, the PV array generates 756.137 MWh of DC energy each year and injects 744.350 MWh of AC energy into the grid.

The 545 wp Si-mono photovoltaic system produces 747.062 MWh of energy per year. The second parameter, the specific annual production per installed kWp, is 1780 kWh/kWp/year. The third parameter, the annual average performance ratio (PR), is 87.2%, while the efficiency, which is provided by the manufacturer, equals 21.11%.

## 3.2. Results of the generated NARX network model

The NARX approach based on ANN was employed for the main simulation results for predicting the energy output and continuous use. Five main parameters were assessed: the total effective energy at the array output [Earray (kWh)], energy needs of the user [User (kWh)], energy supplied to the user from solar [E_Solar (kWh)], energy injected into the grid [E_Grid (kWh)] and energy from the grid EFrGrid (kWh), produced from the 545 wp Si-mono photovoltaic system on an annual basis, which is stated as produced energy. In addition to these variables, the DC energy generated by the Si-mono photovoltaic array, the energy injected into the grid while accounting for electrical component losses, and the efficiency of the photovoltaic array were also calculated. Regression, training state, and validation performance are shown in Figures 3–5, respectively. In addition to providing the estimation linear equation and regression value between targets and predicted values, Figure 3 displays the regression between target outputs and predicted values.
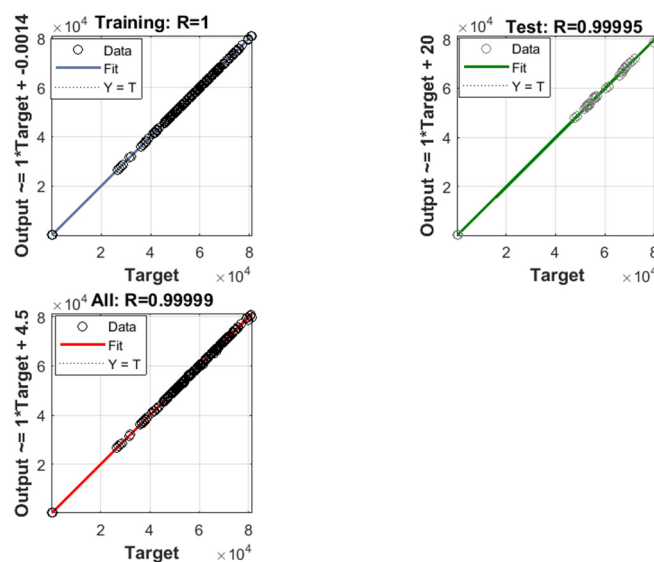


**Figure 3.** Datasets of linear regressions: training data regression, test data regression, and total data regression.

All dataset's errors are analyzed using linear regression (training, validation, and testing). The values between the specified targets and the obtained output are displayed in Figure 3. The ideal circumstances are training data with 100% regression effectiveness. This outcome can be attributed to the quantity of data used for the training—70% of the total—that was used. Regarding the efficacy of regression, there is also a great deal of proximity to the other data groups. The BEST_NARX network was calculated and presented with the correlation coefficient values. Figure 3 shows that the NARX artificial neural network evolved as a result of simulation which is represent datasets of linear regressions. The linear regression model's variables' capacity to explain the variability in the dependent variable was measured by R-squared and adjusted R-squared. The R-squared value varied from 0.99 to 1, with Figure 3 indicating that it always rises with the addition of independent variables, which could cause our model to include redundant variables. Nonetheless, the adjusted R-squared resolves this issue if necessary.

Figure 4 displays the neural network's training states at the epoch when the goal is attained. The neural network's optimal validation performance is depicted in Figure 5, which also displays the training, validation, and test curves for each case's goal as well as the total number of epochs (training iterations) at which the objective has been met. It displays the number of validation checks performed throughout the neural network training epochs, as well as the gradient and weight changes. The training gain, or MU, regulates how much the weights change throughout each iteration. These training process outputs show that the NARX neural network is operating successfully.
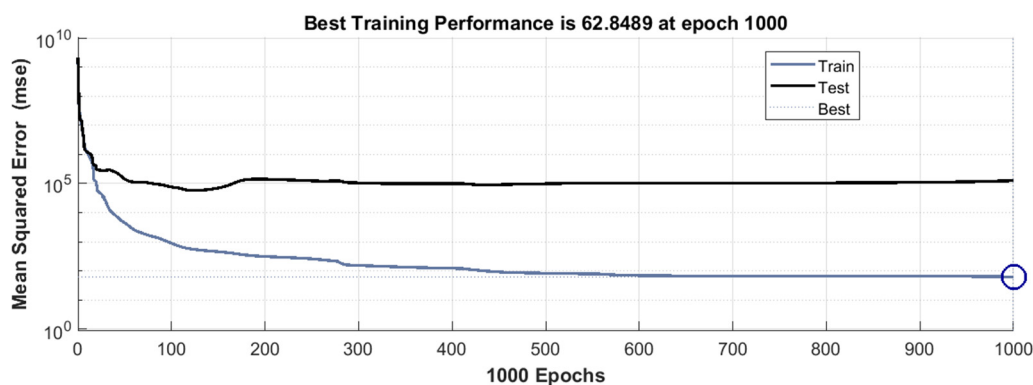


**Figure 4.** Development of training and testing in relation to the number of iterations used, as well as the MSE obtained (performance).
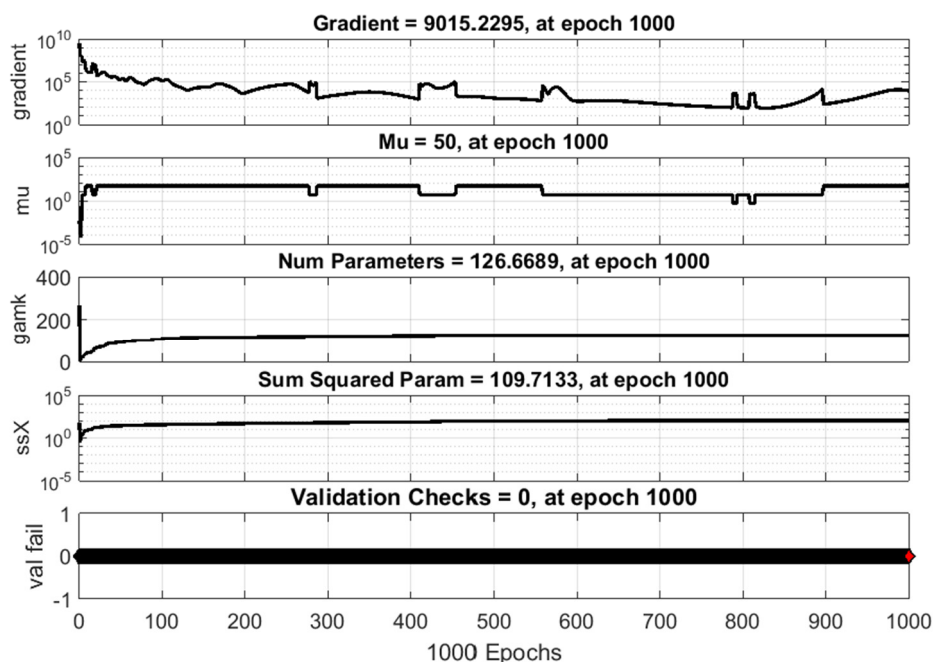


**Figure 5.** Training state.

The NARX network was parameterized and then trained using the Bayesian regularization method. The algorithm was stopped after 1000 iterations of the 1,000 available epochs, and the

network's performance did not change when it was tested using validation data. As a result, the best training performance is shown in Figure 4, which illustrates how the mean squared error (MSE) varied along the training and testing curves during the training epochs. The graph indicates that the epoch 1000, where the training performance is recorded at 62.8489, yielded the best training performance.

Figure 5 illustrates, throughout algorithm execution, how mistakes in training, validation, and test data exhibit similar patterns.

Figure 5 displays the training, validation, and test curves for the goal set, as well as the number of epochs (training iterations) completed to achieve the target. It shows how the gradient and weights change, as well as the number of validation checks performed throughout each epoch of the neural network training process. MU = 50, at approximately 1000, is the training gain that governs the weight change between iterations. The training process's outputs indicate a successful state for the NARX neural network.

Figure 6 specifies the histogram of the error of outputs against the target timeseries dataset, intended to predict output energy and user needs, which were used as exogenous variables in the BEST_NARX network training algorithm.
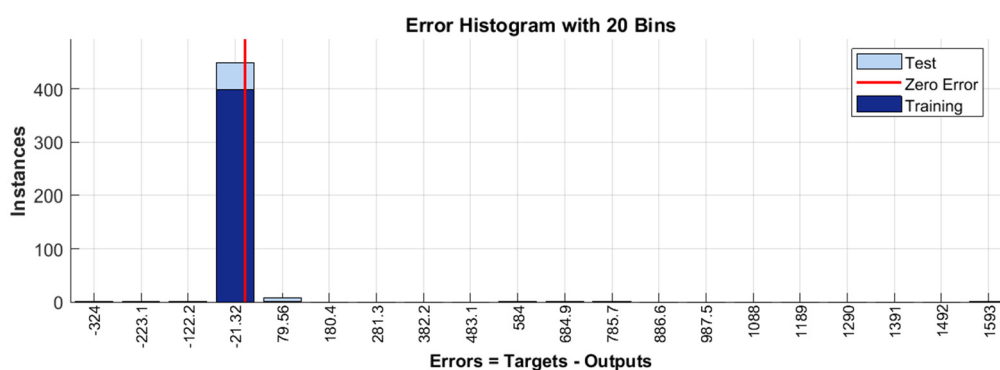


**Figure 6.** Histogram of the error of outputs against target.
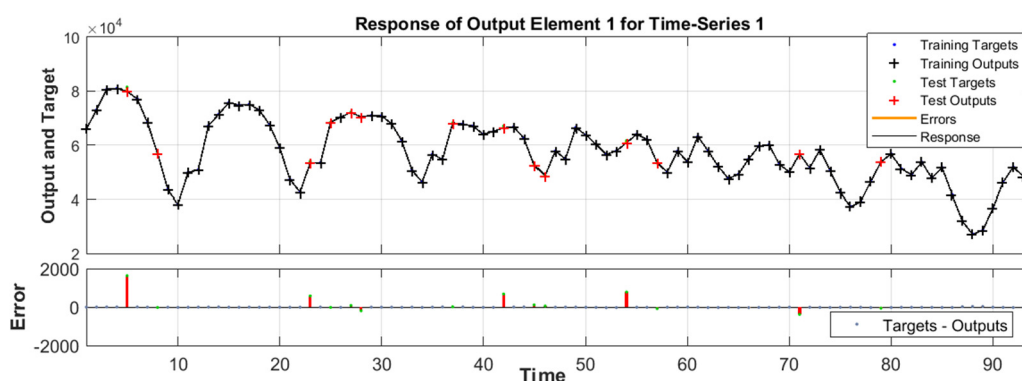


**Figure 7.** Temporary response obtained and error with respect to the objective (time series response).

The results are computed and shown in Figure 6. Upon analyzing this figure, it can be observed that the majority of the errors fall within the value of –21.3; furthermore, the error histogram indicates

that most mistakes are centered around a relatively small value which is equal to –21.3, keeping the dataset's faults within bounds. The training set's data ratio is found to be higher on the center line; this behavior is consistent when analyzing the data that is farthest from the null error.

Figure 7 indicates temporary response obtained and error with respect to the objective (time series response).

The response displays the features of the discrepancy between the NARX's response and the current signal in the grid feeder when power from the PV system is present. The discrepancies between the target values and those obtained from the training datasets are very close to zero, as can be seen from the errors' graph in Figure 7; this indicates that the largest errors occur when the waveform of the signal behaves in a way that is quick and abrupt. Finally, the errors' correlation plots with regard to time and inputs are shown.

Figure 8 shows the error autocorrelation function corresponding to the BEST_NARX network.
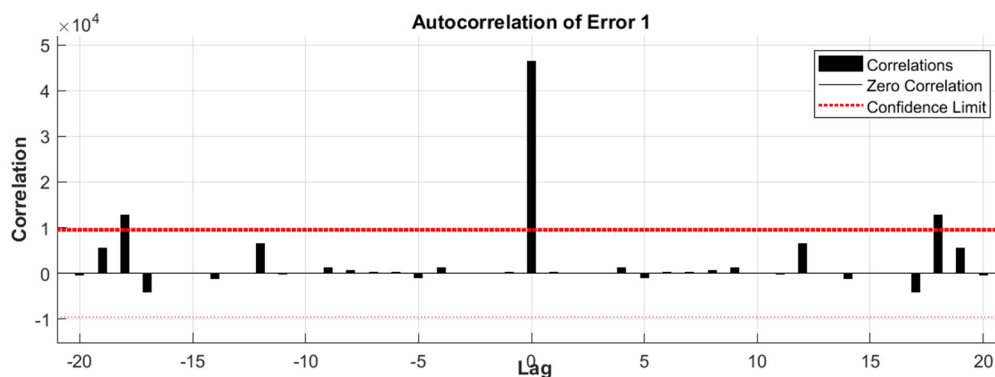


**Figure 8.** Error-time autocorrelation.

Figure 8 illustrates how this network's predicting errors functions. It can be observed that all of the correlations—aside from the one with zero lag—fall inside the intended confidence intervals surrounding zero. This plot illustrates the relationship between the forecasting errors of this network over time, and upon analysis, it can be shown that all correlations fall within the acceptable confidence bounds around zero, with the exception of the zero-lag correlation. Error-time autocorrelation shows how adequate the training is; the center correlation (MSE) with zero value is bigger, while the remainder are within the predicted confidence bounds.

The current study was compared with the NARX model of [19], and the anticipated values were generated by the regression model, provided in Eqs 5 and 6. The predictable values and the NARX model results have a very good relationship. The $R^2$ values of the equations generated by the regression model for the current study ranged from 0.99 to 1. As a comparison, in [19], the values of $R^2$ ranged from 0.9446 to 0.9724. As a result, the regression model was demonstrated to be successful for estimates in both studies.

## 4. Conclusions

As solar energy becomes more prevalent in power generation, forecasting power output from PV power plants is necessary for energy trading, plant optimization, and operational planning.

Furthermore, ambient temperature can affect GlobHor, DiffHor, T_Amb, GlobInc, and GlobEff. PVsyst simulation was used to optimize the 545-wp grid-connected PV array setting. These computed values were acquired for all the variables indicated in the main results and balances. The collecting plane's incident global irradiation [GlobInc (kWh/m$^2$)] is at its optimal when PV field orientation, fixed plane, tilt (/Azimuth (0)) is 30º. The regression coefficients between each input parameter and the solar PV output power data determined the input data combination.

The results of the generated NARX network model confirmed that:

➢ The accuracy to which a linear regression model fits a dataset is measured by both RMSE and R-squared. While R-squared indicates how well the predictor variables can explain the variation in the response variable, the RMSE indicates how well a regression model can predict the value of a response variable in absolute terms. The findings showed that the R-squared value varied from 0.99 to 1.

➢ The NARX model's performance is evaluated based on statistical factors. The epoch 1000, where the training performance is recorded at 62.8489, yielded the best training performance.

➢ Gradient and weights change, as well as the number of validation checks performed throughout each epoch of the neural network training process. MU = 50 at approximately 1000 is the training gain that governs the weight change between iterations. The training process's outputs specify an effective status for the NARX neural network.

➢ The majority of the errors fall within are centered around a relatively small value equal to −21.3.

➢ The response displays the features of the discrepancy between the NARX's response and the current signal in the grid feeder when power from the PV system is present. The discrepancies between the target values and those obtained from the training datasets are very close to zero.

With the exception of the correlation with zero lag, all of the correlations in the network's predicting errors function are related throughout time and, upon closer inspection, are found to lie inside the intended confidence intervals surrounding zero. Analysis of the relationship between this network's forecasting errors over time reveals that, with the exception of the zero-lag correlation, all correlations fall inside the allowable confidence ranges around zero. Error-time autocorrelation indicates how well the training went; the remaining values are within the expected confidence ranges, but the center correlation (MSE) with a zero value is larger.

## Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

## Acknowledgments

## Conflict of interest

The authors declare that they have no conflicts of interest to report regarding the present study.

## Author contributions

The authors confirm contribution to the paper as follows: study conception and design: Amal El Berry, Marwa M. Ibrahim; data collection: Amal El Berry, A. A. Elfeky; analysis and interpretation of results: Amal El Berry, A. A. Elfeky and Marwa M. Ibrahim; draft manuscript preparation: Amal El Berry, Marwa M. Ibrahim. All authors reviewed the results and approved the final version of the manuscript.

## References

1. Abo-Khalil AG, Sayed K, Radwan A, et al. (2023) Analysis of the PV system sizing and economic feasibility study in a grid-connected PV system. *Case Studies Thermal Eng* 45: 102903. https://doi.org/10.1016/j.csite.2023.102903

2. Satpathy R, Pamuru V (2021) Chapter 9—Grid-connected solar PV power systems. *Sol PV Power*, 365–433. https://doi.org/10.1016/B978-0-12-817626-9.00009-5

3. Qazi S (2017) Chapter 5—Fixed standalone PV systems for disaster relief and remote areas. *Standalone Photovoltaic (PV) Syst Disaster Relief Remote Areas*, 139–175. https://doi.org/10.1016/B978-0-12-803022-6.00005-8

4. Huang P, Lovati M, Zhang XX, et al. (2019) Transforming a residential building cluster into electricity prosumers in Sweden: Optimal design of a coupled PV-heat pump-thermal storage-electric vehicle system. *Appl Energy* 255: 113864. https://doi.org/10.1016/j.apenergy.2019.113864.

5. Boscaino V, Ditta V, Marsala G, et al. (2024) Grid-connected photovoltaic inverters: Grid codes, topologies and control techniques. *Renewable Sustainable Energy Rev* 189: 113903. https://doi.org/10.1016/j.rser.2023.113903

6. Numbi B, Malinga S (2017) Optimal energy cost and economic analysis of a residential grid-interactive solar PV system-case of eThekwini municipality in South Africa. *Appl Energy* 186: 28–45. https://doi.org/10.1016/j.apenergy.2016.10.048

7. Guliyev M, Rustamova S, Makhmudova V (2023) The modern status and prospects for further development in the Australian energy sector: Transformation, external economic relations, investment climate. *Energy Policy J* 26: 47–64. https://doi.org/10.33223/epj/166436

8. Nakach I, Mouhat O, Shamass R, et al. (2023) Review of strategies for sustainable energy in Morocco. *Energy Policy J* 26: 65–104. https://doi.org/10.33223/epj/163373

9. Olabi AG, Abdelkareem MA, Semeraro C, et al. (2023) Artificial neural networks applications in partially shaded PV systems. *Thermal Sci Eng Progress* 37: 101612. https://doi.org/10.1016/j.tsep.2022.101612

10. Al-Othman A, Tawalbeh M, Martis R, et al. (2022) Artificial intelligence and numerical models in hybrid renewable energy systems with fuel cells: Advances and prospects. *Energy Convers Manage* 253: 115154. https://doi.org/10.1016/j.enconman.2021.115154

11. Elsheikh A, Sharshir S, Abd Elaziz M, et al. (2019) Modeling of solar energy systems using artificial neural network: A comprehensive review. *Sol Energy* 180: 622–39. https://doi.org/10.1016/j.solener.2019.01.037

12. Moreira M, Balestrassi P, Paiva A, et al. (2021) Design of experiments using artificial neural network ensemble for photovoltaic generation forecasting. *Renewable Sustainable Energy Rev* 135: 110450. https://doi.org/10.1016/j.rser.2020.110450

13. Ledmaoui Y, El Fahli A, Chehri A, et al. (2023) Monitoring solar energy production based on internet of things with artificial neural networks forecasting. *Proc Comput Sci* 225: 88–97. https://doi.org/10.1016/j.procs.2023.09.095

14. Liu R, Chen T, Sun G, et al. (2022) Short-term probabilistic building load forecasting based on feature integrated artificial intelligent approach. *Electr Power Syst Res* 206: 107802. https://doi.org/10.1016/j.epsr.2022.107802

15. Xu L, Hu MM, Fan C (2022) Probabilistic electrical load forecasting for buildings using Bayesian deep neural networks. *J Build Eng* 46: 103853. https://doi.org/10.1016/j.jobe.2021.103853

16. Chaouki G, Al-Mufti OAA, Al-Isawi OAM, et al. (2022) Short-term building electrical load forecasting using adaptive neuro-fuzzy inference system (ANFIS). *J Build Eng* 52: 104323. https://doi.org/10.1016/j.jobe.2022.104323

17. Irshad A, Nqibullah K, Elkholy M, et al. (2024) Techno-economic evaluation and comparison of the optimal PV/wind and grid hybrid system with horizontal and vertical axis wind turbines. *Energy Convers Manage: X* 23: 100638. https://doi.org/10.1016/j.ecmx.2024.100638

18. Irshad A, Ludin G, Ludin S, et al. (2024) Integration and performance analysis of optimal large-scale hybrid PV and pump hydro storage system based upon floating PV for practical application. *Energy Convers Manage: X* 22: 100599. https://doi.org/10.1016/j.ecmx.2024.100599

19. Nkuryngoma O, Seljuk SB (2021) Solar power plant generation forecasting using NARX neural network model: A case study. *Int J Energy Appl Technol* 8: 80–92. https://doi.org/10.31593/ijeat.870088

20. Jiménez-Castillo G, Muñoz-Rodríguez FJ, Martinez-Calahorro AJ, et al. (2020) Impacts of array orientation and tilt angles for photovoltaic self-sufficiency and self-consumption indices in olive mills in Spain. *Electronics* 9: 348. https://doi.org/10.3390/electronics9020348

21. Rahman MM, Shakeri M, Tiong SK, et al. (2021) Prospective methodologies in hybrid renewable energy systems for energy prediction using artificial neural networks. *Sustainability* 13: 2393. https://doi.org/10.3390/su13042393

22. Liu Q, Chen W, Huo H, et al. (2020) An optimal NARX neural network identification model for a magnetorheological damper with force-distortion behavior. *Front Mater* 7: 10. http://dx.doi.org/10.3389/fmats.2020.00010

23. Menezes JMP, Barreto GA (2008) Long-term time series prediction with the NARX network: An empirical evaluation. *Neurocomputing* 71: 3335–3343. http://dx.doi.org/10.1016/j.neucom.2008.01.030

24. Berggren JL, Folkerts M, Knorr WR, et al. (2024) Mathematics: References & edit history. *Encyclopedia Britannica*. Available from: https://www.britannica.com/science/mathematics.

25. Haykin S (1998) Neural networks: A comprehensive foundation. 2nd edition, Prentice Hall, PTR Upper Saddle River, NJ, USA. Available from: https://dl.acm.org/doi/book/10.5555/521706.