Energy Reports 9 (2023) 6267–6294

Contents lists available at ScienceDirect

Energy Reports

journal homepage: www.elsevier.com/locate/egyr

Research paper

A user-friendly and accurate machine learning tool for the evaluation of the worldwide yearly photovoltaic electricity production

Domenico Mazzeo [a,*], Sonia Leva [a], Nicoletta Matera [b], Karolos J. Kontoleon [c], Shaik Saboor [d], Behrouz Pirouz [b], Mohamed R. Elkadeem [e,f]

[a] Department of Energy, Politecnico di Milano, IT-20156, Milan, Italy
[b] Independent Researcher, IT-87036, Rende (CS), Italy
[c] Department of Civil Engineering, Aristotle University of Thessaloniki (A.U.Th.), University Campus, GR-54124, Thessaloniki, Greece
[d] School of Mechanical Engineering, Vellore Institute of Technology (VIT), IN-632014, Vellore, Tamil Nadu, India
[e] Electrical Power and Machines Engineering Department, Faculty of Engineering, Tanta University, EG-31521, Tanta, Egypt
[f] Interdisciplinary Research Center for Renewable Energy and Power Systems (IRC-REPS), King Fahd University of Petroleum and Minerals (KFUPM), SA-31261, Dhahran, Saudi Arabia

ARTICLE INFO

Article history:
Received 17 January 2023
Received in revised form 26 April 2023
Accepted 22 May 2023
Available online xxxx

Keywords:
Photovoltaic module
Electricity
Machine learning
Artificial neural network
PV forecasting
Validation

ABSTRACT

While traditional methods for modelling the thermal and electrical behaviour of photovoltaic (PV) modules rely on analytical and empirical techniques, machine learning is gaining interest as a way to reduce the time, expertise, and tools required by designers or experts while maintaining high accuracy and reliability. This research presents a data-driven machine learning tool based on artificial neural networks (ANNs) that can forecast yearly PV electricity directly at the optimal PV inclination angle without geographic restrictions and is valid for a wide range of electrical characteristics of PV modules. Additionally, empirical correlations were developed to easily determine the optimal PV inclination angle worldwide. The ANN algorithm, developed in Matlab, systematically and quantitatively summarizes the behaviour of eight PV modules in 48 worldwide climatic conditions. The algorithm's applicability and robustness were proven by considering two different PV modules in the same 48 locations. Yearly climatic variables and electrical/thermal PV module parameters serve as input training data. The yearly PV electricity is derived using dynamic simulations in the TRNSYS environment, which is a simulation program primarily and extensively used in the fields of renewable energy engineering and building simulation for passive as well as active solar design. Multiple performance metrics validate that the ANN-based machine learning tool demonstrates high reliability and accuracy in the PV energy production forecasting for all weather conditions and PV module characteristics. In particular, by using 20 neurons, the highest value of R-square of 0.9797 and the lowest values of the root mean square error and coefficient of variance of 14.67 kWh and 3.8%, respectively, were obtained in the training phase. This high accuracy was confirmed in the ANN validation phase considering other PV modules. An R-square of 0.9218 and values of the root mean square error and coefficient of variance of 31.95 kWh and 7.8%, respectively, were obtained.

The results demonstrate the algorithm's vast potential to enhance the worldwide diffusion and economic growth of solar energy, aligned with the seventh sustainable development goal.

1. Introduction

1.1. Context

Clean or renewable energies are those that can be generated simultaneously with their consumption. In contrast to non-renewable sources such as oil, coal, and gas, renewable energy sources (RESs) are clean and inexhaustible. They offer alternative energy options to traditional fossil fuels, with many of them not releasing harmful substances into the atmosphere that could negatively impact the climate. Despite their potential benefits, renewable sources are characterized by irregularity, which poses a significant challenge at the industrial level. However, given their potential to combat climate change, the promotion of RESs has become a critical and pressing issue. Consequently, several directives have been issued to promote the use of RESs and decrease the European Union's (EU's) reliance on imported fossil fuels.

* Corresponding author.
E-mail address: domenico.mazzeo@polimi.it (D. Mazzeo).

https://doi.org/10.1016/j.egyr.2023.05.221

At a global level, there is the agreement, called the Kyoto Protocol, signed in December 1997 during the Conference of the Parties in Kyoto (COP3) with objectives of improvement of energy efficiency, promoting sustainable agriculture and reducing emissions in the transport sector (Babiker et al., 2000). The fight against climate change, which is the most important current problem in the environmental field, forms the core of the agreement. The general objectives of the Kyoto Protocol are the improvement of energy efficiency, promotion of sustainable agriculture, and reduction of emissions in the transport sector. In particular, the Kyoto Protocol has its main objective of reducing greenhouse gas emissions by all signatory countries, which also have the task of drawing up a national system of annual monitoring of greenhouse gas emissions. Subsequently, in December 2015, the first global climate agreement was adopted by 195 countries at the Paris Climate Conference (COP21). The main goal of this agreement is to keep the global average temperature increase below 2 °C above pre-industrial levels. In addition, the signatory countries submitted national global climate action plans, which, however, were not sufficient to achieve the goal set. The goal can be achieved in the future with the commitment of the countries, which are obliged to meet every 5 years to discuss climate actions by drafting national plans to achieve the goal. Following the U.S. Green New Deal (Galvin and Healy, 2020) — a package of measures proposed, in 2019, by the U.S. Legislature to address climate change — Europe has its own "green deal". On January 14, 2020, the EU Parliament approved a massive investment plan aimed at transforming Europe into a "zero climate impact" country by 2050 (Laurent, 2020). The European Green New Deal aims to decarbonize the energy sector across the continent, renovate buildings, support industry in a green economy process and make the transport system cleaner. All EU countries will receive a financial aid package to kick-start the transition: total investments will amount to around 1000 billion euros over ten years. In addition, several funds will be activated, which are necessary for member states to start the economic, productive, and labour reconversion.

RESs such as hydroelectric energy, solar energy, wind energy, marine or tidal energy and geothermal energy are commonly considered eco-friendly energies. In particular, solar energy is clean, renewable thermal or electrical energy produced by directly harnessing the energy radiated from the Sun to the Earth. Indeed, the amount of solar energy that reaches the Earth's soil is enormous, about ten thousand times greater than all the energy used by humanity as a whole. Photovoltaic (PV) modules use light energy (photons) from the sun to generate electricity through the PV effect. Each module is evaluated by its DC power output under standard test conditions (STC). The power typically ranges from 100 W to 365 W. The efficiency of a module determines the area of a module given the same power rating: a 230 W module with 8% efficiency has twice the area of a 230 W module with 16% efficiency. Some commercially available solar modules exceed 24% efficiency. A PV system typically includes an array of PV modules, an inverter, an energy storage battery, a charge controller, interconnect wiring, circuit breakers, fuses, disconnect switches, voltage meters, and optionally a solar tracking mechanism. The equipment is carefully selected to optimize production, and energy storage, and reduce power loss during power transmission and DC to AC conversion. The efficiency of solar panels can be calculated based on the panels' MPP (maximum power point) value. Inverters convert DC power to AC power by performing the maximum power point tracking (MPPT) process: the solar inverter samples the output power (IV curve) from the solar cell and applies the appropriate resistance (load) to the solar cells to obtain maximum energy. A comprehensive review of PV theory and applications was developed by Tiwari et al. (2011), while a simple explicit model for determining the I–V characteristic of different PV module technologies was developed by Boutana et al. (2017).

Tools to predict PV module energy production under real conditions are essential for selecting different PV modules and suitable sites. Despite the availability of many PV models, the development of a method to evaluate the PV energy performance shared by researchers is still an open problem. Previous models can be classified into physical-based models, based on the full I–V curve, or empirical-based models solely based on the maximum power point (MPP) (de la Parra et al., 2017). Usually, the physical models are more complex and used by researchers, while the empirical ones are easier, quicker, require few input data, and are more indicated for designers or policy-makers. The empirical methods are based on fitting equations that summarize the PV behaviour. A recent review summarized the most known algebraic forms that express the temperature dependence of solar electrical efficiency, solar power, and power/energy rating methods (Skoplaki and Palyvos, 2009). These methods are characterized by a black-box approach since the PV efficiency or electrical power as a function of cell/module operating temperature and essential environmental variables are obtained by means of linear or nonlinear multivariable regression equations. According to de la Parra et al. (2017), empirical models with just three independent parameters suffice to accurately describe the relationship between PV array performance and operating conditions. Also, Artificial neural networks (ANNs) fall into this class of models since, after a training process, they learn the PV response as a function of key input parameters with a black-box approach. They can be trained to forecast the PV output for any weather condition and PV cell type with a black-box approach. Recently, ANN or artificial intelligence approaches are, in general, becoming common in different energy and climate fields, such as smart energy management (Li et al., 2023), the thermal analysis of green roofs (Mazzeo et al., 2023), precipitation rate predictions (Ghazikhani et al., 2022), the smart framework for supplying biogas energy (Shahsavar et al., 2021), the design of clean energy community with hybrid renewable systems (Mazzeo et al., 2021), the forecasting of the electrical energy demand for an online monitoring system (Ghadami et al., 2021), a novel regenerator design for a caloric cycle (Kang and Elbel, 2023), the prediction of the specific heat capacity of hybrid nanofluid (Seawram et al., 2022), selecting the most appropriate locations of the offshore wind farms (Marin et al., 2022), load and price forecasting in power systems (Alhendi et al., 2023), predicting the performance of solar collectors (Du et al., 2022) and proton exchange membrane fuel cell power and voltage prediction (Wilberforce and Biswas, 2022).

With large amounts of data available from solar stations, different artificial intelligence techniques are used to calculate, predict and forecast solar radiation energy (Sudharshan et al., 2022). Karamirad et al. (2013) adopted ANNs to predict PV panel behaviours under realistic weather conditions, comparing the results with experimental data and analytical four and five-parameter models of PV modules. They optimized the topology of a multilayer perceptron ANN model in which the input layer consists of three neurons (total irradiation, air temperature, and module voltage), and the output layer contains one neuron (module current). According to Ogliari et al. deterministic models for the day ahead PV output power forecast, based on electric equivalent circuits with three and five parameters, were compared with a hybrid method based on ANNs by employing real data measured for one year in an existing PV plant located at SolarTechlab in Milan (Ogliari et al., 2017). The proposed ANN model can be used to test the maximum power point tracking under real weather conditions. Ye et al. designed a five-layer multi-perceptron ANN, which includes one input layer, three

hidden layers, and one output layer, for the estimation of the power output of a photovoltaic panel (Ye et al., 2022). All the previous ANNs for PV forecasting are not a general value and were not tested for other localities, namely climatic conditions and PV module characteristics. Only very recently, an hourly ANN for the forecasting of PV electricity at any latitude using ANNs was proposed (Matera et al., 2023a); however, a similar approach was never applied from a yearly point of view. Yearly analysis is very important in the design phase of PV systems both in terms of size and in terms of the best site selection. As far as the authors are aware, no previous work has used historical global climate data to accurately predict annual PV energy for a very wide range of variations in key PV electrical and thermal parameters.

For the first time, this paper aims to overcome the issues of classical approaches for the determination of PV cell performance using complex I–V curve modelling by proposing an easy, reliable, quick, and accurate data-driven machine learning model employing the ANN model for the direct calculation of the yearly energy produced by the PV system with any electrical characteristics located in any worldwide locality. The ANN algorithm is very useful for designers and is proposed as an alternative to existing global simple energy methods or empirical correlations. The yearly input and output data used to train the ANN are derived from hourly simulated data summarized in yearly terms. The training database was obtained with an extensive parametric dynamic simulation of eight different PV modules located in 48 worldwide localities. The other two modules were used to develop a verification test of the accuracy of the ANN tool in the calculation of the yearly PV electricity generated in different world climates.

The development of this ANN will allow researchers and designers to directly predict yearly PV energy by using as inputs only some yearly weather variables and electrical and thermal characteristics of the PV module and without solving any electric circuit and, hence, non-linear equations to extract parameters in the reference conditions. In this way, no hourly simulation is required and the result in terms of annual PV energy will be very close to the yearly value that would have been obtained from an hourly simulation.

The rest of the paper is structured as follows. Section 2 describes the ANN forecasting algorithm, accuracy metrics and PV models used. Section 3 presents all electrical and thermal PV characteristics of the module, weather data of the 48 localities considered, optimization of the inclination angle worldwide and input and outputs used in the ANN training process. Simulation results and the calculated accuracy metrics are presented in Section 4 for the ANN training and validation process. The final remarks, conclusions, limitations and future outlook are presented in Section 5.

2. Materials and methods

A forecasting ANN machine learning model for the yearly PV electricity estimation was developed by using as training data yearly input and output data summarizing hourly weather variables and electrical power deriving from dynamic simulations of ten PV modules located in 48 worldwide localities. The proposed model ANN accuracy was verified on 2 further PV modules. In the following sections, the mathematical models used in this research work for the ANN training and hourly PV simulation are analyzed.

2.1. Artificial neural networks

ANNs are machine learning algorithms inspired by the brain's biological functioning, simulating the behaviour of neurons. ANNs

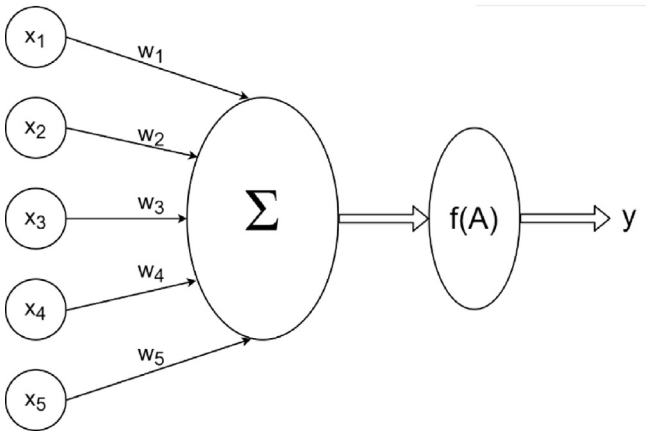

Fig. 1. ANN graphical representation.

are utilized for solving artificial intelligence engineering problems (Ahmad et al., 2014; Cong et al., 2013; Ghritlahre and Prasad, 2018; Liu et al., 2021; Hoang et al., 2021), and other sectors' dilemmas such as biomedicine and data mining (Wang and Huang, 2021; Gong, 2021).

ANNs are mathematical models capable of processing the incoming information at the ANN nodes, called neurons. ANN consists of: (i) neural nodes connected through links; (ii) weights associated with the connections; and (iii) activation functions. In particular, the signals to be processed pass between the various neurons through communication links to which a weight is associated, which usually multiplies the transmitted signal. Each neuron generates the output by applying an activation function to the weighted sum of the inputs. The state of activation of each neuron depends on the input of the neuron itself. The neurons use the activation function to create the signal that is subsequently sent via the communication links to the other neurons. Graphically, the neuron is represented with a circle; instead, the connections between the neurons are represented with oriented arrows indicating the direction of the information flow (see Fig. 1).

The artificial neuron is characterized by a set of synapses (connections), and each synapse is represented by a weight w, which can be positive or negative. The initial operation of the artificial neuron is to calculate an activation function $f(A)$ by carrying out a weighted sum of the input signals x_j with the relative weights w_{ij}. The neuron excitation threshold value s_i, which increases or decreases the input of the activation function according to its positivity or negativity, must be subtracted from this weighted sum (see Eq. (1)).

$$y(x) = f(A) = f\left(\sum_{j=1}^{n} x_j\, w_{ij} - s_i\right) \tag{1}$$

where $y(x)$ represents the output of the neuron that, in turn, represents the input for the other neurons.

The type of activation function used determines the neuron's response. In the literature, different types of activation functions are employed. The most used are:

1. Step activation function (Heaviside function)
   The step activation function $f(A)$ of Eq. (2) assumes value 1 if the weighted sum is greater than the threshold value s; otherwise, it assumes value 0.

$$f(A) = \begin{cases} 1 & \text{if } A \geq s \\ 0 & \text{else} \end{cases} \tag{2}$$

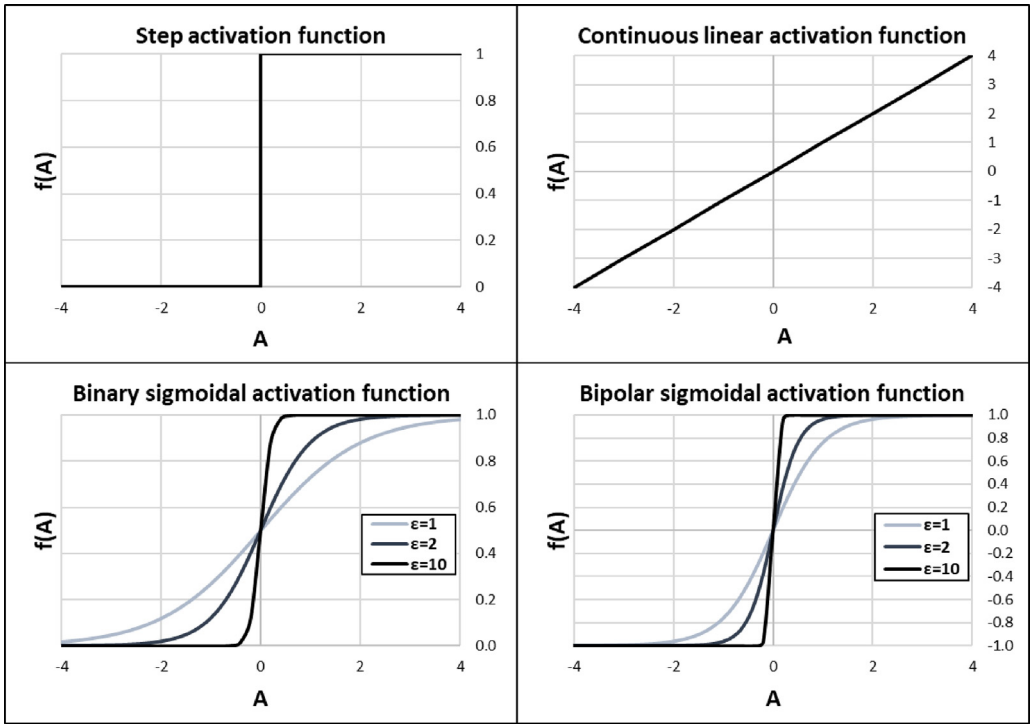

Fig. 2. Most employed activation functions.

2. Sign step activation function
   The sign activation function of Eq. (3) assumes value 1 if it is greater than the threshold value, otherwise, it assumes value -1.

$$f(A) = \begin{cases} 1 & \text{if } A \geq s \\ -1 & \text{else} \end{cases} \tag{3}$$

3. Continuous linear activation function
   The continuous linear activation function of Eq. (4) is directly proportional to the weighted sum of the input signals. When $\varepsilon = 1, f(A)$ is equal to this weighted sum.

$$f(A) = \varepsilon A \tag{4}$$

   The previous activation functions are all continuous. This allows the transmission of signals of gradual intensity and makes them similar to biological neurons. In addition, there are also widely used non-linear continuous activation functions.
4. Binary sigmoidal activation function
   The binary sigmoidal activation function of Eq. (5) is an increasing function and varies in the interval [0, 1].

$$f(A) = \frac{1}{1 + e^{-\varepsilon A}} \tag{5}$$

   where ε indicates the slope of the function. When $\varepsilon = 1$ the Log-sigmoid activation function is obtained.
5. Bipolar sigmoidal activation function
   The bipolar sigmoidal activation function of Eq. (6) varies in the interval [1, -1] and uses the hyperbolic tangent:

$$f(A) = \tanh(\varepsilon A) \tag{6}$$

   where ε represents the slope of the function. When $\varepsilon = 1$, the hyperbolic tangent sigmoid transfer function of Eq. (7) is obtained.

$$f(A) = \frac{2}{1 + e^{-2A}} - 1 \tag{7}$$

Fig. 2 shows all the activation functions described.

Different ANN architecture types are available based on the number of input and output neurons and based on their connection. The neurons are organized in layers. In particular, neurons belonging to the same layer present similar behaviour. It is necessary to mention that the input nodes are not considered layers since they are not involved in processing. According to Kim (2017), the ANN can be classified according to its architecture in: (i) single-layer feedforward ANNs, composed of a layer of input neurons and a layer of output neurons, in which the propagation of the signal occurs only in one direction, from the input layer to the output layer and, therefore, it is strongly acyclic or feedforward; (ii) multilayer feedforward ANNs, characterized by the presence of one or more layers of hidden nodes (hidden layers) placed between input nodes and output nodes, in which the operation takes place only in one direction from input to output and the final efficiency of the model is greater than the single-layer ANN; and (iii) recurrent ANNs, characterized by at least one feedback or a counter-reaction cycle, in which there is a level of neurons that sends the output signals back to the input and the learning capacity is enhanced.

To build an effective ANN model, it is necessary to carry out an ANN training phase in which the model receives a series of input data, apparently unrelated to each other, and a series of outputs, to learn the relationship between input and output. These data are numerically or experimentally analyzed and elaborated by experts and the data produced constitute the model outputs.

In the training phase, the ANN learns the relationships existing between the inputs, which are the data collected, and the outputs, which are the evaluations made by the experts. The training phase aims to provide an output solution even when different input data are provided. In particular, the ANN training phase can be based on three different algorithms: supervised, unsupervised, and reinforcement. Supervised learning is used to solve classification and regression problems, and the objective is to learn the relationship between inputs and outputs given as training data. Then the ANN generalizes this relationship by processing correct

outputs even when it receives different data inputs. In particular, the ANN increases the weights that provide a correct solution and decreases those that provide an incorrect solution.

The validation phase represents the last step and aims to validate the final model using input data different from those used in the training phase to verify that the ANN has learned to generalize the model. If the verification is not satisfied, it is necessary to return to the training phase, while if the solution provided by the ANN is compatible with that provided by experts, the model is used with various hardware and software. The ANN validation phase quickly provides the results obtained with the new data, even if the ANN training phase was slow.

Several ANN learning methods were developed to speed up the learning process. The most used training algorithms by an ANN are the Error back-propagation algorithm and the Levenberg–Marquardt algorithm. The first one is a decreasing gradient method that optimizes the value of the weights minimizing the total square error between the ANN output (output) and the desired output (target) (Kim, 2017). The Levenberg–Marquardt algorithm is widely used in ANNs because it is the most stable and fastest method in finding the solution, while the method of back-propagation of the error is valid in the case of ANNs composed of a large number of hidden layers and neurons. The Levenberg–Marquardt algorithm refers to the Newton approximation method and the Hessian matrix to assign values to weights (Levenberg-Marquardt backpropagation; Hagan et al., 1996; Hagan and Menhaj, 1994). It is based on both the Newton method and the gradient descent method and is an iterative regression technique used to solve multi-variable nonlinear problems. To better understand the Levenberg–Marquardt algorithm, the Descending gradient and Newton methods underlying this algorithm are summarily described.

• In the descending gradient method, the weights w are changed at step $k+1$ using Eq. (8).

$$w_{k+1} = w_k - \alpha g_k \tag{8}$$

in which the constant α is called the learning rate and g is the gradient with a negative sign, which is the first derivative of the function sum of the square of the errors. The problem with this method is the difficulty of choosing the learning rate α, from which the speed with which the function converges to the minimum strongly depends.

• In the Newton method, the weights w are updated at step $k+1$ with Eq. (9):

$$w_{k+1} = w_k - H_k^{-1} g_k \tag{9}$$

in which H is the Hessian matrix of the second derivatives of the sum function of the quadratic error with respect to the weights. The introduction of the Hessian matrix allows the learning rate α to be adjusted at each step k, while α is fixed in the previous method. To calculate the Hessian matrix, the Gauss–Newton method is used, which modifies the weights w at step $k+1$ with Eq. (10).

$$w_{k+1} = w_k - \left(J_k^T J_k\right)^{-1} J_k e_k \tag{10}$$

where J^T is the transposed Jacobian matrix which approximates the Hessian matrix using the prime derivatives and not the secondary derivatives, so the Hessian matrix is written via the Jacobian matrix as:

$$H_k = J_k^T J_k \tag{11}$$

while the gradient g is:

$$g = J^T e \tag{12}$$

where e is the ANN error vector.

In this way, the Gauss–Newton method solves the problem of slowness in finding the minimum of the sum function of the square of the errors. However, in the case of complex problems may diverge from the solution. Instead, the Levenberg–Marquardt method modifies the weights w at step $k+1$ with the equation:

$$w_{k+1} = w_k - \left(J_k^T J_k + \mu I\right)^{-1} J_k e_k \tag{13}$$

Compared to the Gauss–Newton method, this method introduces the parameter μ, called the damping coefficient, and the identity matrix I. The damping coefficient is introduced to prevent the algorithm from diverging from the minimum error function. Parameter μ varies according to the following criteria: is reduced when a step in the algorithm leads to a better value of the error function, such that convergence to the minimum value is accelerated; is increased when the value of the error function moves away from the minimum so that the variation of the weights is reduced to find a better value of the error function. The iterative Levenberg–Marquardt algorithm steps used during the ANN training are: (i) starting the ANN training with random values of the weights w_k; (ii) calculation of the errors, sum function of the square of the errors and Jacobian matrix; (iii) modification of the values of the weights w_k using a random value of the damping parameter μ; (iv) recalculation of the errors E and sum function of the square of the errors with the new values of the weights; (v) if the error is decreased ($E_{k-1} < E_k$), then the damping parameter μ is divided by 10 such that the speed of convergence is increased, while if the error has increased ($E_{k-1} > E_k$), then the damping parameter μ is multiplied by 10 to decrease the speed of convergence; (vi) return to step (ii). The epochs indicate the number of times the ANN is trained with the training set. The training algorithm is interrupted when the sum function of the square of the errors reaches a minimum value set as a threshold or when this function starts to increase on the percentage of data of the "validation test".

2.1.1. Advantages and limitations of artificial neural networks

The ANN model presents a series of characteristics that allows their use in different areas. This model presents three fundamental properties that are: flexibility, robustness, and generalization. These features make the model resistant, i.e., able to provide an answer even if some connections are eliminated or if there is noise; moreover, they make it flexible, i.e. able to be used for different objectives, and general as once trained it can provide a correct answer in output even if new data are inserted in input.

Therefore, ANNs have several advantages:

- They work in parallel, which allows them to process a multitude of data in a short amount of time. From this comes a good tolerance to failures and noise working in parallel, the ANN can provide an adequate response even if some units do not work well or if the inputs are inaccurate, however decreasing the performance.
- The ANN can adapt to any changes as it learns autonomously through experience and training.
- Data processing is distributed in such a way that many elements work on the same task.

The ANN model also has some flaws:

- It works in a black-box approach, i.e. it is not possible to understand the method used by the model to arrive at the exact output. It receives the data in input and provides in the output the correct solution but does not allow the examination of the various stages of elaboration of the process.

- An initial training phase is necessary to fix the weights of the neurons, and this requires time, attention and a good experience as it is necessary to provide the model with a lot of input data.
- The solutions provided in the output are not completely correct but have a margin of error.

2.1.2. Implementation of artificial neural networks in Matlab environment

The Neural net fitting tool of Matlab software was used for the ANN design (Neural net fitting tool). The classical steps to create an ANN are:

1. ANN training data collection and importation: collection of the model training input and output data and other input data to use the created ANN;
2. ANN creation and configuration: choice of the number of hidden layers and neurons;
3. ANN parameters initialization: initialization of the weights and biases;
4. ANN training, testing, and validation: the dataset is randomly divided into three groups so that, usually, 70% of the data are used for the ANN training, 15% of the data for the ANN testing, and the remaining 15% for the data for the ANN validation; for a resulting low accuracy, the ANN can be re-trained by varying the number of neurons or by importing a larger dataset;
5. ANN application: use of the ANN created to test its accuracy for new dataset considering accuracy metrics.

In this study, a feed-forward ANN with one hidden layer with a sigmoidal activation function and an output layer with a linear activation function was considered. The training algorithm employed is the Levenberg–Marquardt backward propagation algorithm.

Typically, using a higher number of neurons and layers requires more computation, and time and tends to overuse data but allows the ANN to solve more complex problems. By increasing the number of neurons present in the hidden layer, a risk of adapting the ANN in a perfect way to the data used (overfitting), and losing the capacity of ANN generalization can be produced. In other words, a high number of neurons allows the resolution of a complex problem optimally, but there is the risk of creating an ANN suitable only for the data used and not a general ANN that can also be used with different data. One of the main goals of ANN creation is the design of the ANN architecture. For this reason, the ANN was trained by varying the number of neurons from 2 to 20 with a step of 2 neurons.

The training stops automatically when the generalization stops improving, as indicated by an increase in the mean square error (MSE) of the validation samples. The most important parameters are the performance, the magnitude of the performance gradient, and the number of validation checks. In particular, the latter two are used to terminate the ANN training. The value of the gradient becomes very small as the training reaches the minimum MSE. The training is terminated when the magnitude of the gradient is less than 1×10^{-7}. The validation checks number represents the number of subsequent iterations so that performance does not decrease; by default, this number is set to six. In addition, the training process is stopped when the error rate of untrained (validation) data continuously increases for more than six epochs. Also the damping parameter μ of the Levenberg–Marquardt algorithm is monitored during the training process. At each iteration, the value of μ guides the optimization process since: when the error decreases quickly, then a smaller value of μ is used; on the contrary, when the decrease of the error is insufficient, then it is necessary to increase the μ value. In general, μ is decreased with each successful iteration, i.e., when the performance function decreases and is increased only when an iteration would improve the performance function. In this way, the performance function is always reduced with each iteration.

The overfitting can be observed when the error starts to increase on the validation dataset. The training process stops after six consecutive increases in the validation error, and the best performance is taken in the epoch with the lowest validation dataset error. In particular, the performance plot can be considered valid only if the final mean square error is small, the error of the testing dataset and the error of the validation dataset have similar characteristics, or no significant iterations occurred at the time when the validation performance reached the limit.

Usually, the validation and testing curves by increasing the epochs are very similar. When the test curve increases significantly before the validation curve increases, it means that overfitting of the data occurred and, therefore, the number of neurons should be decreased. Another graph to visualize the training process is the error histogram which shows the ANN error distribution. It indicates the outliers, which are data where the fit is significantly worse than in most of the data. Finally, the data should fall along a 45° inclined line, where the ANN outputs are equal to the targets, for a perfect regression plot between the ANN output and target for training, testing, and validation datasets.

2.1.3. Artificial neural network accuracy and validation

The number of neurons is chosen based on some accuracy metrics to optimize the ANN reliability and accuracy. Furthermore, the same metrics can be used to validate the ANN when further input datasets, external to the training dataset, are employed to determine the output.

The most common metrics are reported in Table 1 (Matera et al., 2023a; Elsheikh et al., 2019), where t_i, t_m, $t_{\max}$ and $t_{\min}$ are the ith, mean, maximum and minimum values of the target output obtained from the TRNSYS simulations, y_i and y_m are the ith and mean values of the output predicted by the ANN and N is the total number of comparisons.

Additional accuracy metrics are used to evaluate ANN accuracy:

• the minimum E_{min}, maximum E_{max}, mean E_m and standard deviation E_{sd} values of the error, the difference between the target t and output y.

$$E_{min} = \min\left(t_i - y_i\right);$$
$$E_{max} = \max\left(t_i - y_i\right);$$
$$E_m = \frac{\sum(t_i - y_i)}{N}; \tag{14}$$
$$E_{sd} = \sqrt{\frac{\sum\left(E_i - E_m\right)^2}{N}}$$

Finally, Pearson's correlation coefficient ρ_{XY}, defined as the covariance of input X and output Y variables divided by the product of their standard deviations σ_X and σ_Y, measures their linear correlation and always has a value between -1 and 1. A value of 1 determines a perfect linear relationship between X and Y perfectly, a value of -1 implies an inversely linear proportion between X and Y, while a value of 0 implies that there is no linear correlation between the variables. This coefficient can be used to identify the most correlated inputs with the output.

$$\rho_{XY} = \frac{\operatorname{cov}(X, Y)}{\sigma_X \sigma_Y} \tag{15}$$

Table 1
Accuracy metrics name, equation, variation range and optimal value (Matera et al., 2023a; Elsheikh et al., 2019).

Accuracy metric	Formula	Range and optimal value
Mean square error (MSE)	$MSE = \frac{\sum_{i=1}^{N}(t_i - y_i)^2}{N}$	$(\mathbf{0}, +\infty)$
Mean absolute error (MAE)	$MAE = \frac{\sum_{i=1}^{N}\lvert t_i - y_i\rvert}{N}$	$(\mathbf{0}, +\infty)$
Root mean square error (RMSE)	$RMSE = \sqrt{MSE}$	$(\mathbf{0}, +\infty)$
Coefficient of variance (COV)	$COV = \frac{RMSE}{\frac{\sum_{i=1}^{N} y_i^2}{N}} \cdot 100$	$(\mathbf{0}, +\infty)$
Correlation coefficient (CC)	$CC = R = \frac{N\sum_{i=1}^{N} t_i y_i - \left(\sum_{i=1}^{N} t_i\right)\left(\sum_{i=1}^{N} y_i\right)}{\left[N\left(\sum_{i=1}^{N} t_i^2\right) - \left(\sum_{i=1}^{N} t_i\right)^2\right]\left[N\left(\sum_{i=1}^{N} y_i^2\right) - \left(\sum_{i=1}^{N} y_i\right)^2\right]}$	$(-\infty, \mathbf{1})$
R-square (R^2)	$R^2 = CC^2$	$(0, \mathbf{1})$
Coefficient of determination (COD)	$COD = \frac{\left[\sum_{i=1}^{N}(t_i - t_m)(y_i - y_m)\right]^2}{\sum_{i=1}^{N}(t_i - t_m)^2 \sum_{i=1}^{N}(y_i - y_m)^2}$	$(0, +\infty)$
Efficiency coefficient (EC)	$EC = 1 - \frac{\sum_{i=1}^{N}(t_i - y_i)^2}{\sum_{i=1}^{N}(t_i - t_m)^2}$	$(-\infty, \mathbf{1})$
Overall index of model performance (OIMP)	$OIMP = \frac{1}{2}\left[1 - \left(\frac{RMSE}{t_{max} - t_{min}}\right) + EC\right]$	$(-\infty, \mathbf{1})$
Coefficient of residual mass (CRM)	$CRM = \frac{\sum_{i=1}^{N} t_i - \sum_{i=1}^{N} y_i}{\sum_{i=1}^{N} t_i}$	$(-\infty, +\infty)$ **0**

2.2. Photovoltaic mathematical modelling

Determining the energy produced by a PV system for any weather conditions is one of the most essential steps in designing and verifying a PV system's performance. Most of the forecasting models for PV electricity are based on a global simple energy method or complex method according to the resolution of the equivalent electric circuit representing a PV cell.

Regarding the global simple energy methods, the input data related to a specific day are employed as a reference in representing the entire month. All days of the month in question are considered identical to this single representative day. The simplified energy methods use the statistical concept of the usability of solar radiation. Those commonly used are the Siegel method and the Clark method, applicable both for the verification and for the general design of the systems.

With respect to Siegel, it is valid if the electricity produced is simultaneously absorbed by the load, assumed constant (Siegel et al., 1981; Mazzeo et al.). This hypothesis is true in the case of grid-connected systems, in which any electrical power produced in excess, compared to that absorbed by the load, is transferred to the electricity grid. The method estimates the monthly average daily yield of the PV field based on the monthly average daily irradiation and air temperature values. Also, it can be applied for evaluating the monthly average daily system performance for a constant 24 hr-per-day load with a battery of specified capacity. While the Clark method considers the variable monthly average hourly profiles of the electrical load and uses monthly average hourly irradiation and air temperature in the average monthly days (Clark et al., 1984). The method is based on radiation statistics and utilizability, in addition, it can account for variability in the electrical demand as well as for the variability in solar radiation.

The modelling of the physical behaviour of the PV cell is essential to correctly evaluate the electrical energy that can be produced for a reliable estimation of the economic return of the investment. The most widely used complex methods are based on the possibility of describing the characteristic curve (I–V) of the PV cell analytically as a function of the absorbed irradiance and cell temperature. The physical behaviour of the cell varies substantially as a function of solar irradiation, electrical load and operating cell temperature. This last parameter is strongly variable in relation to the convective thermal exchange with the external environment. It is also sensitive to wind speed and direction, external air temperature and surrounding surfaces. The operating cell temperature is a function of the radiative heat exchange (and therefore of the transmission coefficient of the cover glass and the absorption coefficient of the semiconductor). It is essential to fully understand the variations in operating cell temperature as it directly influences the electrical parameters of the generation system.

Various analytical and numeric methods, of varying complexity, were proposed for describing the behaviour of a PV cell (I–V characteristic curve) for specific temperature and radiation conditions. Many studies have focused on the modelling of PV modules and have developed electric models with different levels of complexity. These models differ mainly in the number of diodes, shunt resistance (infinite or finite), ideality factor (fixed or variable), and numerical methods used to determine unknown parameters. Most of the electrical models available in the literature, considering single diode and two diodes electrical models based on 3 to 7 parameters, are described in detail by Tossa et al. (2014). Another useful review on PV modelling and simulation was outlined by Chin et al. (2015), deepening the concepts behind the main models of PV cells and highlighting their respective advantages and drawbacks. Similarly, Humada et al. (2016) comprehensively reviewed the foremost issues of the methods of the extraction of PV cell parameters for single-diode and double-diode models. The main task is the model parameter extraction based on the number of PV parameters (Lun et al., 2013; Siddiqui and Abido, 2013; Ma et al., 2014; Ali et al., 2016; Chen et al., 2016; Wu et al., 2018; Nunes et al., 2019; Xu and Wang, 2017; Yu et al., 2019, 2018; Mares et al., 2015; Humada et al., 2020; Xu et al., 2014; Orioli and Di Gangi, 2013). Recently, a novel algorithm was proposed to identify the unknown parameters for different PV models including the static PV models (single-diode and double-diode) and dynamic PV model (Elaziz et al., 2021).

Determining the PV parameters of the model requires solving nonlinear equations obtained from the resolution of the electrical circuit. After choosing an electrical model, another major problem lies in the method of extracting these parameters; it is difficult to obtain the optimal parameters of these models analytically. For this purpose, many implicit, explicit, or iterative algorithms were developed based on experimental, numerical, and optimization techniques or combinations of these (Tossa et al., 2014). This obstacle is often the reason why the two-diode model is not used, even though it might be the most accurate. Some authors have proposed and used an equivalent circuit with a pair of

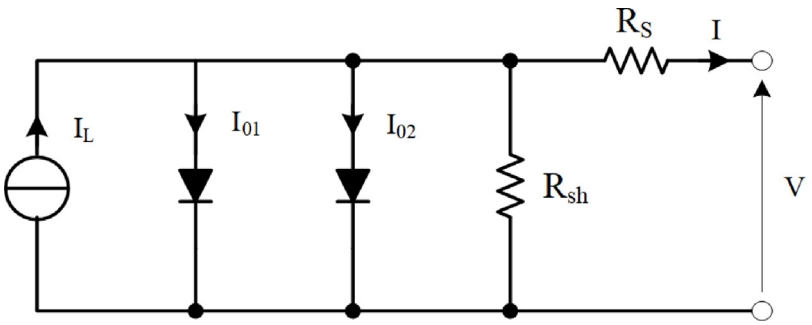

Fig. 3. Two-diode equivalent electrical circuit.

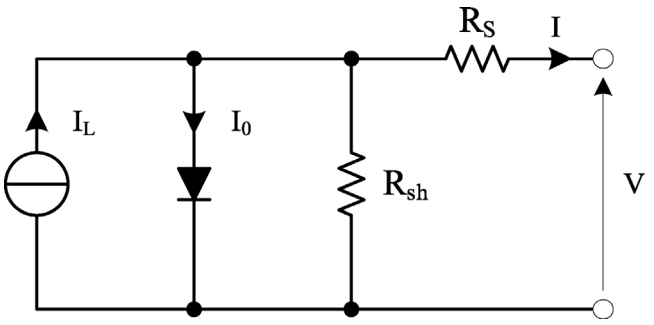

Fig. 4. Equivalent electric circuit with one diode (5 parameters).

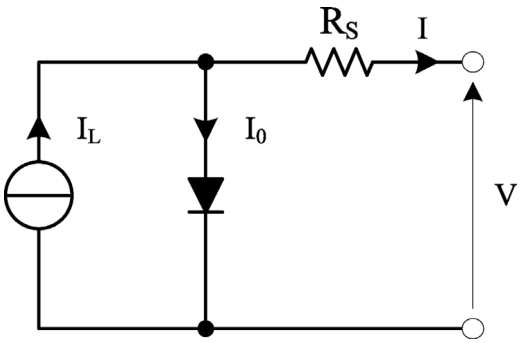

Fig. 5. Equivalent electrical circuit of a diode (4 parameters).

diodes, a current generator, and two resistors (see Fig. 3). The inverse saturation currents of the two diodes (I_{01}, I_{02}) and the two diode ideality factors (n_1, n_2) depend on the properties of the two diodes. Such a model involves the determination of seven parameters, namely the light current (I_L), diode ideality factors (n_1, n_2), series resistance (R_s) and shunt resistance (R_p).

The equation in an implicit form related to the equivalent circuit in Fig. 3 is as follows:

$$I = I_L - I_{01}\left(e^{\left(\frac{V+I\,R_s}{n_1 V_t}\right)} - 1\right) - I_{02}\left(e^{\left(\frac{V+I\,R_s}{n_2 V_t}\right)} - 1\right) - \frac{V + I\,R_s}{R_{sh}} \tag{16}$$

Where:

I_L is the light current;

I_{01}, I_{02} are the inverse saturation current of the two diodes;

n_1, n_2 are the two diode quality factors;

R_s is the series resistance: it represents the set of resistances due to both the cell material and contact resistance between the metal grid and the crystal surface;

R_{sh} is the shunt resistance: this resistance is due to current leakage in the junction, and it depends almost exclusively on the method used to make the junction;

$V_t = \frac{K_B T_c}{q}$ which is defined as thermal voltage;

q is the charge of the electron $q = 1.602 \times 10^{-19}$ C;

K_B is the Boltzmann constant $K_B = 1.381 \times 10^{-23}\ \frac{J}{K}$; and

T_c is the cell temperature.

The solution of the previous equation exists and is mathematically determinable; however, the implicit form and the presence of two exponential elements make the calculation of the seven parameters of the equivalent circuit very complex, and also the resolution methods in the literature are unreliable as they can quickly diverge towards incongruent solutions. For these reasons, other authors have preferred to use, in the development of their models, an equivalent electrical circuit representing a 5-parameter model, in which there is only one diode, in addition to the current generator and the two resistors (see Fig. 4).

The equation in an implicit form related to the equivalent circuit in Fig. 4 is as follows:

$$I = I_L - I_0\left(e^{\left(\frac{V+I\,R_s}{n\,V_t}\right)} - 1\right) - \frac{V + I\,R_s}{R_{sh}} \tag{17}$$

its resolution is simpler since it involves the determination of five parameters (I_L, I_0, n, R_s, R_{sh}).

In general, it is possible to state that the problem of modelling the electrical behaviour of a PV cell was addressed, in most of the models in the literature, using the simplified one-diode model that provides for the determination of five parameters, as this model manages to fully describe the characteristic curve (I–V), under standard conditions, of most modern PV panels. This happens because the curve (I–V) of modern panels is particularly squared as they are characterized by very small values of series resistance R_s and very high values of shunt resistance R_{sh}.

A further reasonable simplification can be applied to eliminate one of the five unknown parameters on the same type of one-diode model. The simplification that is generally adopted considers a shunt resistance of infinite value ($R_{sh} \to \infty$); this assumption is particularly correct for crystalline silicon modules. In this case, the model requires the determination of four parameters (I_L, I_0, n, R_s), so the computational cost is further reduced (see Fig. 5). The equation in an implicit form related to the equivalent circuit in Fig. 5 is as follows:

$$I = I_L - I_0\left(e^{\left(\frac{V+I\,R_s}{n\,V_t}\right)} - 1\right) \tag{18}$$

Three physical models (corresponding to three-, four-, and five-parameter equivalent electric circuits) and two thermal models were compared by Dolara et al. (2015).

For this purpose, ten monocrystalline and eight polycrystalline modules were calibrated and tested in SolarTechLab at Politecnico di Milano. The degree of complexity of the model will determine which of the methods is most reliable for determining the parameters involved in the mathematical expression of the model. In

general, these models can be divided into two groups: numerical methods that require powerful mathematical means and iterative methods to solve the implicit nonlinear equation associated with the PV device, and analytical methods that introduce a series of simplifications and approximations that lead to a simpler solution without introducing significant errors in the results (Tossa et al., 2014). To estimate the parameters of electrical models, the Levenberg–Marquardt method has been demonstrated in many studies (Tossa et al., 2014). To solve these models based on a system of exponential equations, known data are required. However, manufacturers provide only limited PV panel operating data, such as open-circuit voltage (V_{oc}), short-circuit current (I_{sc}), maximum power current (I_{mp}) and maximum power voltage (V_{mp}), open-circuit voltage temperature coefficient, short-circuit current temperature coefficient (β_{voc} and α_{1sc}, respectively), and nominal cell operating temperature (NOCT). Furthermore, these data are available only under standard conditions, where the irradiance is 1000 W/m^2, and the cell temperature (T_c) is 25 °C (except for NOCT, which is determined at 800 W/m^2 and an ambient temperature of 20 °C). These conditions result in high electrical power but are rarely encountered in actual operations. Indeed, solar system designers need reliable means of predicting the power output from a PV panel under all conditions to decide whether or not to adopt this technology. Ideally, a PV panel should always operate at a voltage that produces maximum power. This is possible, approximately, by using a maximum electrical power tracker (MPPT). Without an MPPT, the PV panel operates at a point on the I–V characteristic curve that coincides with the I–V characteristic of the load. But the commercially available PV module sold in the market, very often, has not been tested under natural environmental conditions, and its performance is lower than measured in standard condition tests (STC) given by the manufacturer (Tossa et al., 2014, 2016). Moreover, the energy production is sometimes different for modules of different technologies with the same maximum power measured under STC conditions. Module performance is closely related to intrinsic characteristics such as absorption, strength, and construction process. The latter are generally affected (but not in the same way) by environmental conditions, temperature, and solar irradiance (solar power and spectrum) (Merten et al., 2008). Therefore, it is difficult but essential to know how to choose the technology that provides the best tradeoff between cost and actual performance of a module in a natural environment for a given site. One of the major changes in this choice lies in the mathematical model used to predict PV module performance accurately under real operating conditions. The choice of the electrical model can be based on several criteria: computational speed, PV technology, accuracy, or be guided by the analysis of the statistical errors of each electrical parameter. Thus, the latter approach can be used as a fast and accurate means of decision-making, not only for choosing the best electrical model to estimate the energy production of a given PV technology but also for choosing the best PV technology suitable for the climatic and environmental characteristics of a given site. It can be argued that there is no clear relationship between PV technology, environmental conditions, and the electrical model. In addition, modelling a given module requires comparing the results with different modules. Finally, suitable PV models to analyze the effects of partial shading must be considered (Piccoli et al., 2019).

In this work, the electrical power produced by a PV module was obtained by employing a 5-parameter model based on the equivalent electric circuit of Fig. 4 and Eq. (17).

Under the actual operating conditions, by changing the absorbed solar radiation G_T and the PV cell temperature T_c, the following parameters are updated as a function of the parameters in the reference conditions:

$$I_L = I_{L,\mathrm{ref}} \frac{G_T}{G_{T,\mathrm{ref}}} \tag{19}$$

$$\frac{I_0}{I_{0,\mathrm{ref}}} = \left(\frac{T_c}{T_{c,\mathrm{ref}}}\right)^3 \tag{20}$$

The procedure used to determine the 5 parameters in the reference and actual conditions is developed by Fry (1999). This procedure is implemented in Type 94 of TRNSYS 17 software (University of Wisconsin, 2012).

TRNSYS (http://www.trnsys.com/) is a simulation software used to simulate the behaviour of transient systems such as energy simulation and building simulation with respect to solar design and thermal and electrical energy performance; therefore, it is mostly used in energy/environmental engineering, such as in Mazzeo et al. (2021), Baglivo et al. (2020), Herdem et al. (2020) and Matera et al. (2023b).

Type 94 includes several options, one of which is the mathematical model that is used to predict electrical performance. The type can use the four-parameter model for single-crystal or polycrystalline PV modules or the five-parameter model for amorphous or thin-film PV modules. The difference between the two mathematical models lies in the fact that the four-parameter model assumes that the slope of the current–voltage curve is zero under short-circuit conditions, while the five-parameter model assumes a finite negative slope in the voltage-current curve under short-circuit conditions.

Another option that type 94 allows is whether or not the simulation should call the "incidence angle modifier" correlation, which considers the increase in reflective losses when radiation is incident on the module at wide angles. Type 94 also considers an optional angle-of-incidence modifier correlation for calculating the change in PV module surface reflectance as a function of the angle of incidence of solar radiation.

Mathematically, type 94 uses an iterative search routine to calculate the equivalent circuit characteristics, to calculate the values of $I_{L,ref}$, $I_{0,ref}$, $R_{s,ref}$, R_{sh} and a_{ref}. The parameter values under operating conditions are obtained by updating the I_L and I_0 values as a function of absorbed solar radiation and cell temperature, respectively. The latter is calculated using the nominal operating cell temperature ($NOCT$). In this way, the characteristic curve is updated at each instant as a function of cell temperature and absorbed solar radiation. In addition, the absorbed solar energy is evaluated considering the incidence angle modifier IAM. Finally, the electrical PV power output $P_{pv}(t)$ is calculated at the maximum power point of the characteristic curve by the following equation:

$$P_{pv}(t) = I_{mp}(t)\, V_{mp}(t) \tag{21}$$

The adopted model type uses temperature data from standard NOCT measurements to calculate the module temperature T_c at each time step. The NOCT temperature $T_{c,NOCT}$ is the operating temperature of the module at a wind speed of 1 m/s, with no electrical load, and at a certain ambient temperature and specific solar radiation. The values for solar radiation $G_{T,NOCT}$ and ambient temperature $T_{a,NOCT}$ are usually 800 W/m^2 and 20°C. The module temperature T_c is calculated with Eq. (22).

$$T_c = T_a + \frac{\left(1 - \frac{\eta_c}{\tau\alpha}\right)}{\frac{G_T \tau\alpha}{U_L}} \tag{22}$$

η_c is the conversion efficiency of the module, U_L is the array thermal loss coefficient and $\tau\alpha$ is the module transmittance–absorptance product.

3. Yearly photovoltaic artificial neural network

Herein, this analysis aims to train and validate an ANN capable of predicting the yearly energy produced by a PV module in a

Table 2
PV module parameters at reference conditions (Canadian Solar Inc., 2023; Jakson Group, 2023; LG Electronics, 2023; Mitsubishi Electric US Inc., 2023; Panasonic Corporation, 2023; Wuxi Suntech Power Co., 2023; Trina Solar Co., 2023; ENF Solar Ltd., 2023; Vikram Solar Limited, 2023).

Name	Cells	V_{oc} (V)	I_{sc} (A)	$V_{mp,ref}$ (V)	$I_{mp,ref}$ (A)	NOCT (K)	A (m^2)	μ_{Isc} (%/°C)	β_{Voc} (%/°C)	$P_{pv,n}$ (W)	η_m (%)
Training											
CanadianSolar290	60	38.50	9.72	31.60	9.18	43.0	1.64	0.0500	−0.290	290.09	17.72
Jakson250	72	44.50	7.45	35.90	6.97	47.0	1.62	0.0400	−0.320	250.22	12.88
LG300	60	40.10	9.65	32.90	9.15	45.0	1.64	0.0300	−0.280	301.04	16.23
MLU250	120	37.60	8.79	31.00	8.08	45.7	1.66	0.0560	−0.350	250.48	15.10
Panasonic330	96	69.70	6.07	58.00	5.70	44.0	1.67	0.0340	−0.164	330.60	19.70
Suntech250	60	37.40	8.63	30.70	8.15	45.0	1.63	0.0500	−0.340	250.21	15.40
TallMax320	72	45.80	9.10	37.10	8.63	44.0	1.94	0.0500	−0.320	320.17	16.50
TP250	60	37.30	8.71	30.20	8.30	47.0	1.67	0.0442	−0.290	250.66	15.00
Validation											
CHSM250	60	38.19	8.65	30.30	8.27	43.0	1.64	0.0520	−0.344	250.58	15.20
VikramSolar320	72	45.96	9.03	37.65	8.50	45.0	1.92	0.0520	−0.310	320.03	16.67

given location by using as inputs some yearly variables identifying the climate and electrical and thermal characteristics of the PV module.

A direct prediction of yearly PV energy can be made with this ANN using only a few yearly weather variables, as well as electrical and thermal characteristics of the PV module as inputs, while avoiding solving electric circuit equations and, therefore, non-linear equations for determining parameters. As a result, no hourly simulation is needed since the resulting ANN PV energy is very close to what would have been calculated from an hourly simulation.

The ANN was trained by considering 8 PV modules, different in terms of electrical and thermal behaviour, and 48 localities uniformly placed around the world and characterized by high, intermediate, and low levels of solar radiation available, as well as very cold, cold, medium, hot and very hot climates. The ANN was validated on two other PV modules with different electrical characteristics and included in the range of electrical parameters of the 8 PV modules.

3.1. Photovoltaic modules

Eight PV modules with different electrical characteristics were considered to create the ANN starting from input and output extracted from the TRNSYS software. Two other PV modules (CHSM250 and VikramSolar320), having electrical characteristics in the range of values of the previous eight modules were chosen as validation case studies by comparing the yearly electrical energy obtained from the ANN and the target electrical energy deriving from the TRNSYS simulation. Overall, ten PV modules were considered whose electrical and thermal characteristics at reference conditions are reported in Table 2.

The main PV module parameters at reference conditions are: the open-circuit voltage $V_{oc,ref}$; the short-circuit current $I_{sc,ref}$; the voltage at the maximum power point $V_{mp,ref}$; the current at the point of maximum power I_{mp}; the nominal operating cell temperature NOCT; the area of the PV module A; the temperature coefficient of the short-circuit current μ_{Isc}; the temperature coefficient of the open-circuit voltage β_{Voc}; the nominal power of the PV module P_n obtained from the product between voltage and current at the point of maximum power; the efficiency η_m. Reference conditions mean a temperature of 25 °C and solar radiation of 1000 W/m^2. These parameters are supplied by the manufacturer (Canadian Solar Inc., 2023; Jakson Group, 2023; LG Electronics, 2023; Mitsubishi Electric US Inc., 2023; Panasonic Corporation, 2023; Wuxi Suntech Power Co., 2023; Trina Solar Co., 2023; ENF Solar Ltd., 2023; Zhejiang Chint New Energy Development Co. - Astronergy Solar Inc., 2023; Vikram Solar Limited, 2023) and are required by Type 94 in TRNSYS to determine the power produced by using the 5-parameter model.

3.2. Climatic data

To make the ANNs as general as possible and adaptable to any locality, 48 localities belonging to different climate groups of the Köppen classification and characterized by very different climates were considered (Kottek et al., 2006; Rubel and Kottek, 2010; Arnfield, 2020). In particular, two different localities for each Koppen climate subgroup were considered, except for the Cfb subgroup which has four localities. The 48 locations considered in this work are reported in Fig. 6. The hourly values of air temperature and horizontal total solar radiation in a typical year *tm2* file are known for each locality from the TRNSYS library.

A preliminary analysis was carried out to maximize the yearly PV electrical energy produced worldwide and to identify the optimal inclination angle for each locality and PV module. Recently, a novel data-driven approach proved that power-based optimization provides a different tilt angle than conventional irradiance-based optimization because of the trade-off between irradiance and efficiency (Ye et al., 2022).

For this purpose, the hourly power produced by varying the inclination angle was calculated through an analysis carried out with the TRNSYS software. A parametric analysis was done by changing the inclination angle from −90° to 90° with a step of 2° and using as surface azimuth 0° or 180° in North and South hemisphere localities, respectively. Dynamic simulations performed have determined the hourly electrical power produced varying the inclination angle, and for each locality, 91 different series of hourly electrical powers were printed in one year. Overall, for each PV module, 4368 (91 angles × 48 localities) Excel sheets were printed containing the hourly electrical power at varying angles of inclination. The hourly data were used to determine the yearly electrical energy produced for each location, given by the sum of the hourly powers. The optimum angle of inclination for a given location and PV module is that which leads to the maximum energy. Fig. 7 illustrates the yearly electrical energy produced by varying the PV inclination angle in all 48 localities for the Jakson250 module.

The optimal inclination angle depends on the locality and, as expected, is negative in the North hemisphere localities and positive in the South hemisphere localities. Further, the optimal inclination angle and the relative highest yearly energy generated E_g are reported for each locality and PV module in Table 3.

In the training phase, the PV modules were considered placed at the optimal angle of each specific locality. In this way, the ANNs can predict the PV hourly performance worldwide directly on the optimal inclination angle. For each PV module, the relationship between the optimal PV inclination angle and the latitude of the 48 locations considered was derived. Empirical equations of the optimal angles as a function of the latitude were obtained for

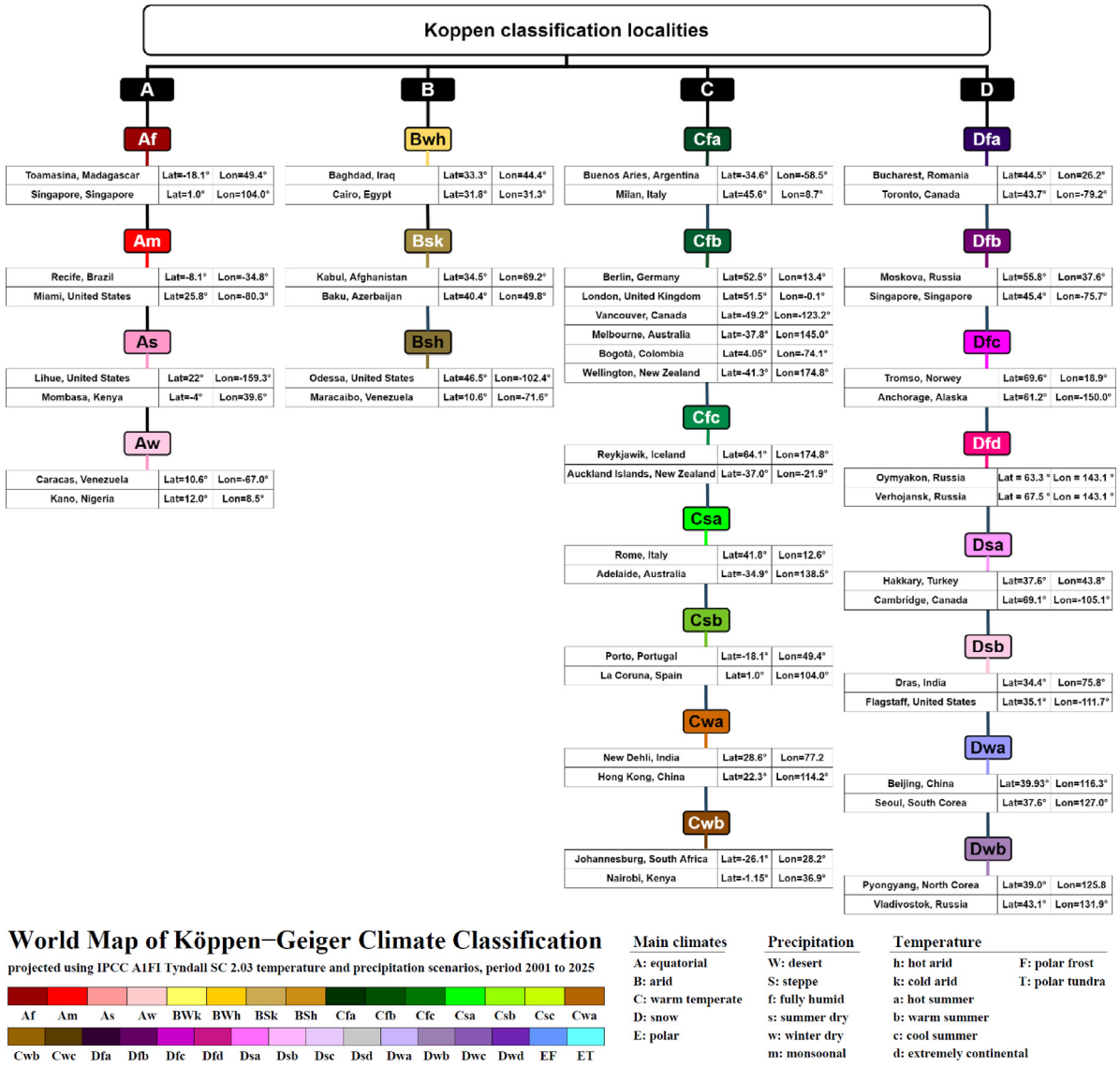

Fig. 6. Localities belonging to the training and validation set along with the Koppen climate classification.

each PV module, as shown in Fig. 8. These equations are useful to calculate the optimal inclination angle for any locality worldwide.

As shown by Table 3 and by the regression curves, the latitude strongly influences the optimal inclination angle value, while the PV module typology has a very slight impact.

3.3. Input and output data for the ANN training

To train the yearly ANN, the electrical characteristics of 8 PV modules and the environmental characteristics of all 48 localities were considered. The ANN created was validated by considering two other PV modules with different electrical characteristics in the 48 localities.

The ANN training input data matrix is composed of:

- The yearly horizontal total solar energy obtained as the sum of the hourly solar radiation of a typical year $E_{s,y}$ [Wh/m^2];
- The yearly average external air temperature obtained as an average of the hourly temperatures of a typical year $T_{ea,y}$ [°C];
- Number of individual cells in a PV module N_{cs} [–];
- Open-circuit voltage at reference conditions $V_{oc,ref}$ [V];
- Short-circuit current at reference conditions $I_{sc,ref}$ [A];
- Voltage at maximum power point along IV curve at reference conditions $V_{mp,ref}$ [V];
- Current at maximum power point along IV curve at reference conditions $I_{mp,ref}$ [A];
- Nominal operating cell temperature *NOCT* [°C];
- Temperature coefficient of short-circuit current μ_{Isc} [%/°C];
- Temperature coefficient of open-circuit voltage β_{Voc} [%/°C].

The ANN training output data matrix is composed of:

- Yearly produced solar PV energy at the optimal inclination angle $E_{pv,y}$ (Wh) calculated as the sum of the hourly electrical power values obtained with the TRNSYS software for each locality and each PV module.

The training input matrix of the ANN consists of 10 parameters for each locality and PV module. The input vector is composed of as many rows as there are parameters and as many columns as there are localities and PV modules. Overall, an input matrix of 10 rows and 384 columns (48 localities × 8 PV modules) is used for the ANN training phase. The training output vector of the ANN

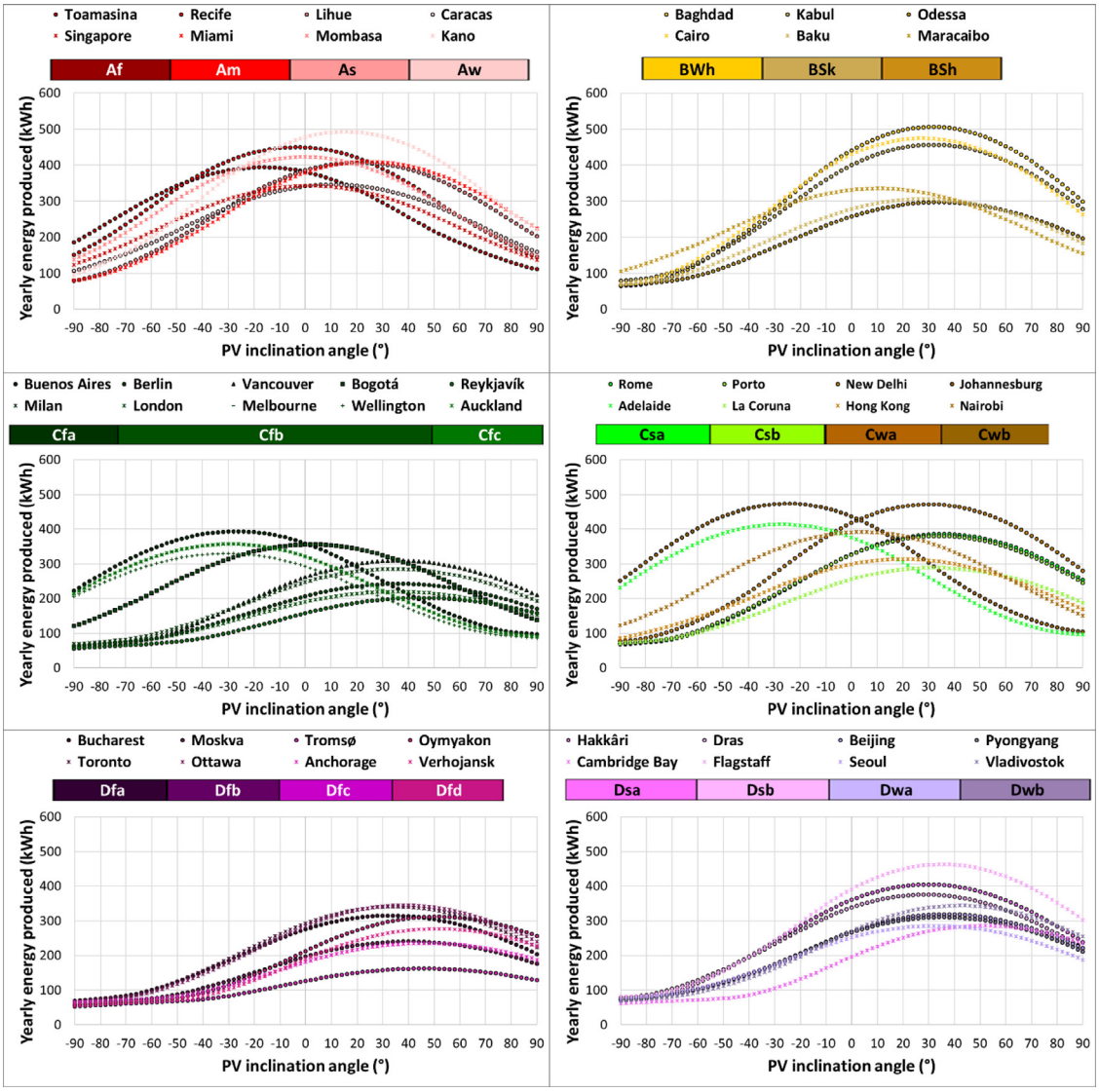

Fig. 7. Produced yearly energy by varying the PV inclination angle in all 48 localities for the Jakson250 module.

is made up of the yearly electrical energy produced by each of the 8 PV modules in each of the 48 localities. Overall, an output matrix of 1 row and 384 columns (48 localities × 8 PV modules) is used for the ANN training phase. The yearly horizontal solar energy $E_{s,y}$ and average external air temperature $T_{ea,y}$ for the 48 localities considered are reported in Fig. 9.

Fig. 9 highlights the extremely variable climatic conditions considered with external air temperatures variable between −20 °C and 30 °C and solar radiation between 500 kWh/m^2 and 2500 kWh/m^2. This makes the database and ANN created very general. In addition, the yearly produced solar PV energy at the optimal inclination angle $E_{pv,y}$ obtained with the TRNSYS software for each PV module by varying the locality is illustrated in Fig. 10.

Fig. 10 shows that the yearly produced PV energy regularly increases for a locality with a higher yearly horizontal total solar energy for all PV modules. The lowest electricity production is observed for the MLU250 module variable between 148.55 kWh and 419.37 kWh from the locality with the lowest yearly solar energy available to the locality with the highest yearly solar energy available, while the highest one for the Panasonic330, variable between 211.67 kWh and 683.56 kWh.

4. Results and discussion

4.1. Training phase

The input matrix of size [10 × 384] and output matrix of size [1 × 384] was imported into Matlab and the Neural net fitting tool was used for the ANN training purpose. The dataset was divided into three groups: 268 (70%) combinations of parameters used for training, 58 (15%) combinations of parameters used for the validation, and 58 (15%) combinations of parameters of the data used for testing. The Levenberg–Marquardt algorithm was chosen as the learning method, and the sigmoid transfer function and the linear transfer function were used, respectively, in the hidden layer and output layer. In addition, an ANN optimization analysis was carried out that consisted of evaluating the accuracy metrics described in Section 2.1.3., by varying the number of neurons in the hidden layer from 2 to 20 with a step of 2 neurons. The results of this analysis are reported in Figs. 11 and 12.

Fig. 13 shows the boxplots of the error and relative error on the yearly PV energy evaluation by means of the ANNs by varying the number of neurons.

Table 3
Optimal inclination angle and the relative highest yearly energy generated for each locality and PV module.

	CanadianSolar290		Jakson250		LG300		MLU250		Panasonic330		SunTech250		TallMax320		TP250		CHSM250		VikramSolar320	
	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)	β (°)	E_g (kWh)
Toamasina	−16	453.8	−16	393.9	−16	463.8	−16	320.9	−16	530.7	−16	392.3	−16	504.6	−16	390.2	−16	398.7	−16	504.5
Singapore	2	394.7	2	342.7	2	402.4	2	270.2	2	463.2	2	341.2	2	439.4	2	339.0	2	347.5	2	439.2
Recife, Pernambuco	−4	520.0	−4	450.1	−4	526.9	−4	352.6	−4	612.2	−4	450.3	−4	579.2	−4	446.1	−4	457.8	−4	579.4
Miami, Florida	24	471.7	24	409.2	24	481.3	26	331.0	24	552.3	24	407.9	24	524.6	24	405.4	24	414.6	24	524.6
Lihue, Hawaii	20	468.5	20	406.7	20	478.5	10	274.0	20	548.3	20	405.1	20	521.0	10	342.1	20	411.8	20	521.0
Mombasa	0	487.6	0	422.6	0	495.8	16	388.0	0	572.5	0	421.9	0	542.7	16	488.9	0	428.9	0	542.8
Caracas	10	398.2	10	345.8	10	406.2	20	329.7	10	467.2	10	344.2	10	443.2	20	402.8	10	350.5	10	443.1
Kano	16	570.0	16	492.8	16	577.2	0	336.2	16	671.2	16	493.8	16	634.8	0	418.7	16	501.5	16	635.2
Baghdad	32	584.9	32	506.5	32	597.3	32	419.4	32	683.6	32	506.1	32	649.8	32	502.9	32	512.8	32	650.0
Cairo	28	548.0	28	475.1	28	559.6	28	392.4	28	640.4	28	474.1	28	609.0	28	471.3	28	480.9	28	609.1
Kabul	32	526.2	32	457.0	32	544.4	34	260.1	32	608.3	32	454.2	32	582.6	34	295.0	32	459.2	32	582.5
Baku	28	352.4	28	306.6	28	364.2	10	259.9	28	409.1	28	303.9	28	390.8	10	331.9	28	308.5	28	390.5
Odessa, Texas	34	341.6	34	297.6	34	355.7	34	403.3	34	394.2	34	294.1	34	378.1	32	454.1	34	298.3	34	377.7
Maracaibo	10	386.6	10	335.5	10	393.3	28	259.3	10	455.0	10	334.3	10	430.7	28	303.8	10	340.8	10	430.6
Buenos Aires	−26	452.6	−26	393.4	−28	466.4	36	331.6	−26	525.5	−26	390.7	−26	502.1	36	382.3	−26	396.3	−26	501.8
Milan	34	327.7	34	285.5	34	340.5	−28	353.0	34	378.9	34	282.2	34	363.0	−28	410.1	34	286.5	34	362.6
Berlin	36	277.4	36	242.0	36	289.6	34	331.3	36	319.9	36	238.6	36	307.0	34	376.9	36	242.2	36	306.6
London	36	251.1	36	219.2	36	262.7	32	251.5	36	289.1	36	215.8	36	277.8	32	286.2	36	219.1	36	277.3
Vancouver, British Columbia	36	355.2	36	309.6	36	370.6	−28	335.3	36	408.9	36	305.9	36	392.9	−26	390.1	36	309.8	36	392.5
Melbourne, Victoria	−28	409.4	−28	356.5	−28	424.2	36	246.0	−28	473.6	−28	353.1	−28	453.7	34	282.8	−28	358.0	−26	453.3
Bogotá, Cundinamarca	2	408.6	2	355.8	2	424.1	32	377.1	2	471.9	2	352.2	2	452.5	30	467.5	2	357.0	2	452.1
Wellington	−30	378.3	−30	329.6	−30	393.3	20	254.5	−30	436.5	−30	325.9	−30	418.8	20	310.5	−30	330.4	−30	418.4
Reykjavík	44	230.8	44	201.7	44	244.3	38	210.7	44	263.3	44	197.9	44	254.4	36	239.6	44	200.4	44	253.9
Auckland Islands	−28	409.9	−30	357.0	−30	425.0	36	191.9	−28	473.9	−30	353.4	−30	454.2	36	217.0	−28	358.4	−28	453.8
Rome	36	443.2	36	385.5	36	457.7	36	276.0	36	513.8	36	382.5	36	491.4	36	307.1	36	387.8	36	491.2
Adelaide	−28	475.8	−28	413.4	−28	490.0	−30	309.4	−26	552.6	−28	410.9	−28	527.8	−28	353.4	−28	416.6	−28	527.6
Porto	34	436.7	34	380.0	34	452.0	44	185.3	34	505.2	34	376.8	34	483.9	44	200.0	34	381.7	34	483.6
La Coruna	32	331.5	32	288.9	32	344.7	−30	311.0	32	382.7	32	285.5	32	367.1	−28	353.8	32	289.7	32	366.6
New Delhi	30	544.6	30	471.4	30	553.2	−26	409.9	30	639.5	30	471.5	30	606.0	−24	469.9	30	478.7	30	606.3
Hong Kong	20	360.9	20	313.7	20	370.0	4	327.4	20	421.8	20	311.7	20	401.2	4	388.2	20	317.1	20	401.0
Johannesburg	−24	545.0	−24	473.3	−24	561.2	2	311.7	−24	632.3	−24	471.0	−24	604.4	2	352.8	−24	476.7	−24	604.3
Nairobi	4	450.9	4	391.7	4	462.9	−32	289.6	4	525.3	4	389.6	4	500.8	−30	326.7	4	395.4	4	500.6
Bucharest	32	361.9	32	315.1	32	375.9	34	273.4	32	418.3	32	311.8	32	400.8	32	312.4	32	316.2	32	400.4
Toronto, Ontario	34	390.9	34	340.7	34	408.9	36	306.0	34	449.1	34	336.4	34	431.9	34	338.1	34	340.4	34	431.4
Moskva	40	275.1	40	240.1	40	289.2	38	279.6	40	315.5	40	236.3	40	303.8	36	316.3	40	239.4	40	303.3
Ottawa, Ontario	38	395.4	38	344.9	38	415.1	34	250.4	38	453.3	38	340.0	38	436.6	32	283.0	38	343.9	38	436.0
Tromsø	46	185.6	46	162.2	46	196.8	30	355.6	46	211.7	46	159.0	46	204.5	30	401.6	46	161.1	46	204.1
Anchorage, Alaska	46	268.8	46	234.9	46	284.5	52	287.5	46	306.9	46	230.6	46	296.3	52	285.1	46	233.4	46	295.8
Oymyakon, Sakha Republic	52	356.8	52	311.9	52	387.2	42	215.4	50	400.9	52	304.3	52	390.1	40	238.0	50	305.9	52	389.2
Verhojansk, Sakha Republic	52	316.5	52	276.6	52	340.4	40	312.8	52	357.3	52	270.6	52	347.0	38	342.3	52	272.5	52	346.2
Hakkâri	30	465.2	30	404.4	30	482.3	36	276.9	30	537.6	30	401.3	30	515.0	36	308.0	30	406.0	30	514.8
Cambridge Bay, Nunavut	52	326.8	52	286.1	52	354.8	44	320.0	52	366.1	52	279.1	52	357.3	42	341.6	52	280.2	52	356.5
Dras	28	430.7	28	375.5	28	453.8	30	349.6	28	491.9	28	370.2	30	474.8	28	373.2	28	373.6	28	474.2
Flagstaff, Arizona	36	531.1	36	462.1	36	553.1	36	418.9	36	610.8	36	457.9	36	587.1	36	459.2	36	462.5	36	586.8
Beijing	36	366.2	36	319.1	36	381.4	46	148.6	36	422.5	36	315.3	36	405.3	46	160.8	36	319.7	36	404.8
Seoul	32	327.5	32	285.6	32	341.9	48	215.6	32	377.4	32	281.8	32	362.3	46	233.0	32	285.8	32	361.8
Pyongyang	36	356.1	36	310.6	36	372.9	52	305.8	36	409.4	36	306.3	36	393.6	52	310.8	36	310.3	36	393.1
Vladivostok	42	394.0	42	344.1	42	416.2	52	266.9	42	449.5	42	338.4	42	434.3	52	275.3	42	341.9	42	433.6

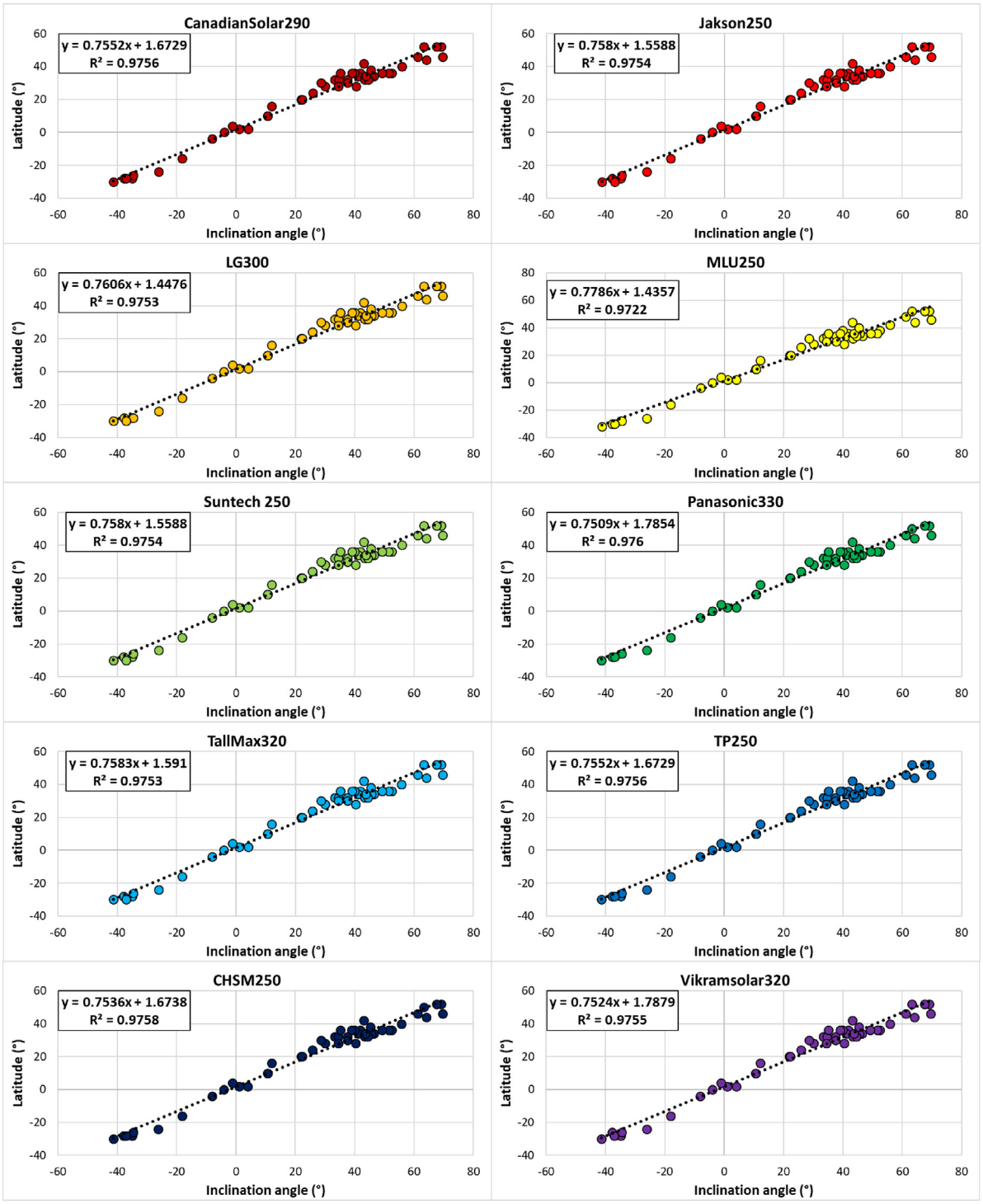

Fig. 8. Optimal PV inclination angle as a function of the latitude for the ten PV modules.

An overall analysis of all metrics and graphs shows that by increasing the number of neurons from 2 to 10, a sensible improvement of the ANN accuracy can be observed. Instead, above 10 neurons the accuracy remains rather stable. This is evident from Fig. 11 on the top, where R^2, R and OIMP increase and MSE and RMSE decrease with more evidence from 2 to 10 neurons. Similar behaviour can be observed for the COD in Fig. 12 on the top; however, the COD growing trend is developed until 20 neurons with an intermediate reduction between 14 and 18 neurons. The absolute error obtained by using 2 neurons has a range of −80000 kWh to 70000 kWh, as highlighted in Fig. 11 below. This range narrows as the number of neurons increases, becoming minimal and equal to ≈100000 kWh ($E_{max} \approx 40000$ kWh and $E_{min} \approx -60000$ kWh) for 14 neurons. The average and the standard deviation of the absolute error are stable around about 1000 kWh and 150000 kWh, respectively. This is also highlighted in Fig. 12 at the top, where the COV only decrease from 4.8% to 3.8% by increasing the number of neurons from 2 to 20. All the ANNs, on average, slightly underestimate the yearly PV energy as highlighted by the negative CRM in Fig. 12 at the bottom. Only with 14 neurons, on average, the ANN slightly overestimates. By using 20 neurons, the highest R, R^2, OIMP and COD of 0.9898,

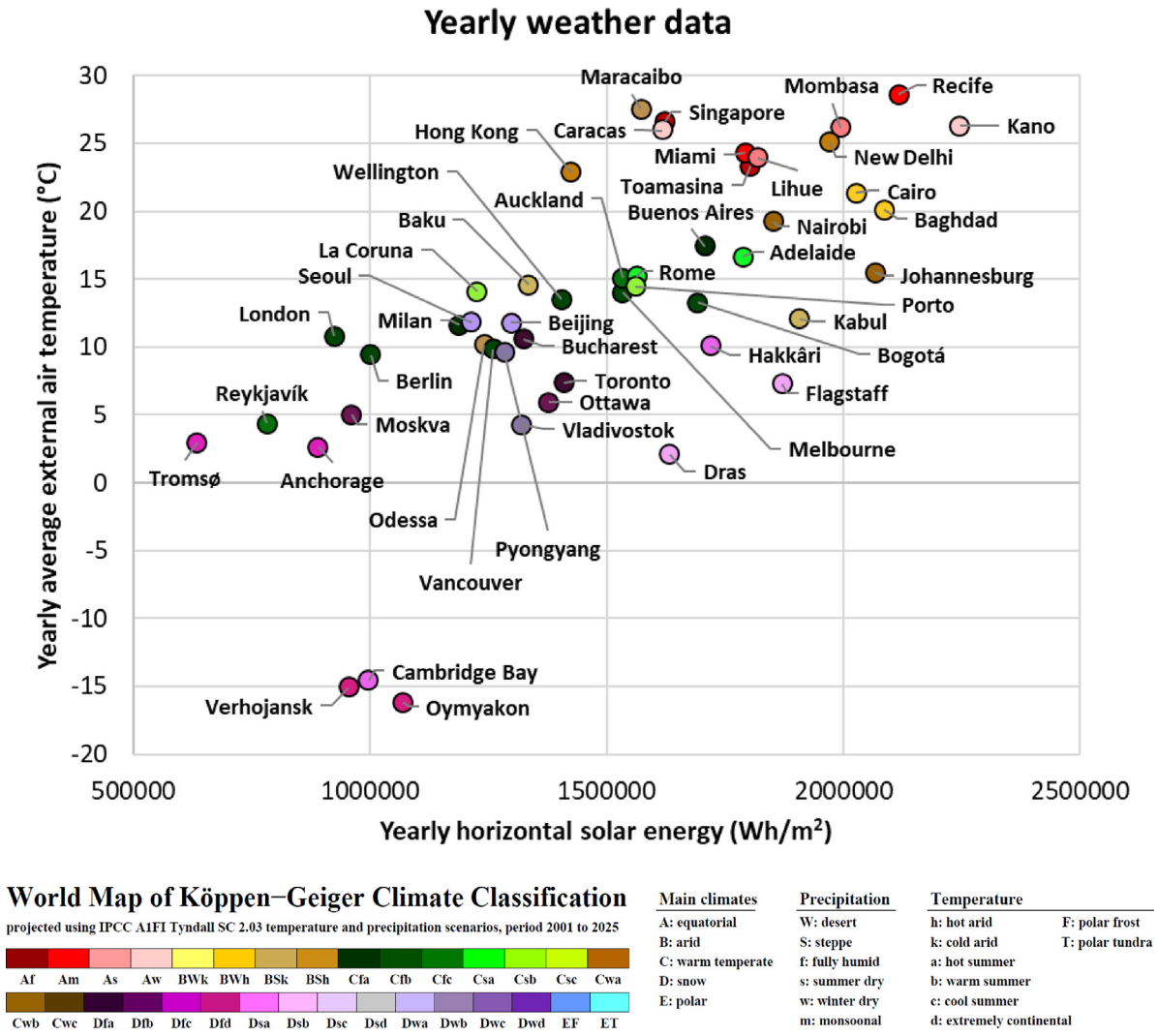

Fig. 9. Yearly horizontal solar energy and average external air temperature for the 48 localities.

0.9797, 0.9761 and 0.9809, respectively, and the lowest RMSE, MSE, and COV of 14.67 kWh, 215.30 MWh and 3.8%, respectively, are obtained. In addition, the CRM is almost zero and the error boxplot is the least wide and is characterized by the mean value closest to zero. For all these reasons, the ANN with 20 neurons was considered the optimal one and is proposed as reliable and accurate for the yearly PV energy forecasting.

All details related to the ANN architecture and the training, testing and validation phases are reported in Fig. 14, which shows the screen of the Neural net fitting tool after the end of the learning process. In the figure, also the number of epochs required in the training phase can be seen. Fig. 15 highlights the mean square error trend for the training, testing, and validation phases on the left and the trends of the gradient, damping factor, and validation checks on the right by increasing the number of epochs.

The mean square error of training, testing, and validation curves significantly decrease in the first 8 epochs. The validation and testing curves slightly increase after the 8th epoch, while the training curve slightly decreases. For this reason, the learning process required 14 epochs and was terminated after six validation checks. The results obtained after 8 epochs were considered the best ones since by increasing the epochs the generalization stops improving and the overfitting can be observed. The validation and test curves, representing the mean square error as a function of the epochs, are very close to the 8th epoch. The Levenberg–Marquardt backpropagation algorithm at the 8th epoch provides a gradient of 6.39×10^7 and a damping parameter μ of 1×10^6. The gradient decreases during the entire process, while the damping parameter first increases rapidly and then very slowly.

The selected optimal ANN architecture is very accurate, for all data groups' training, testing, and validation, as demonstrated by the high frequency of errors very close to zero (see Fig. 16) and by the almost perfect regression between the target and ANN output on the bisector line as shown in Fig. 17.

Fig. 16 shows that the frequency of outliers is very low for all data groups, demonstrating the highest reliability in reproducing accurate results. In addition, Fig. 17 reveals the very high R of the different regressions that are between 0.985 for the validation data and 0.991 for the training data. Overall, R is 0.990 for the entire dataset. Besides, Table 4 shows the Pearson correlation coefficient ρ_{xy} between the ANN output and different ANN inputs.

By analysing Table 4, it can be deduced that the yearly horizontal total solar energy $E_{s,y}$ is the most correlated input to the output with a ρ_{xy} value of 0.81, representing an almost perfect linear dependency between the variables. Also the yearly average external air temperature has a linear dependency with the output; however, with a lower ρ_{xy} value. The same considerations can also be done for the temperature coefficient of open-circuit voltage μ_{Voc}, the open-circuit voltage at reference conditions $V_{oc,ref}$ and the voltage at maximum power point at

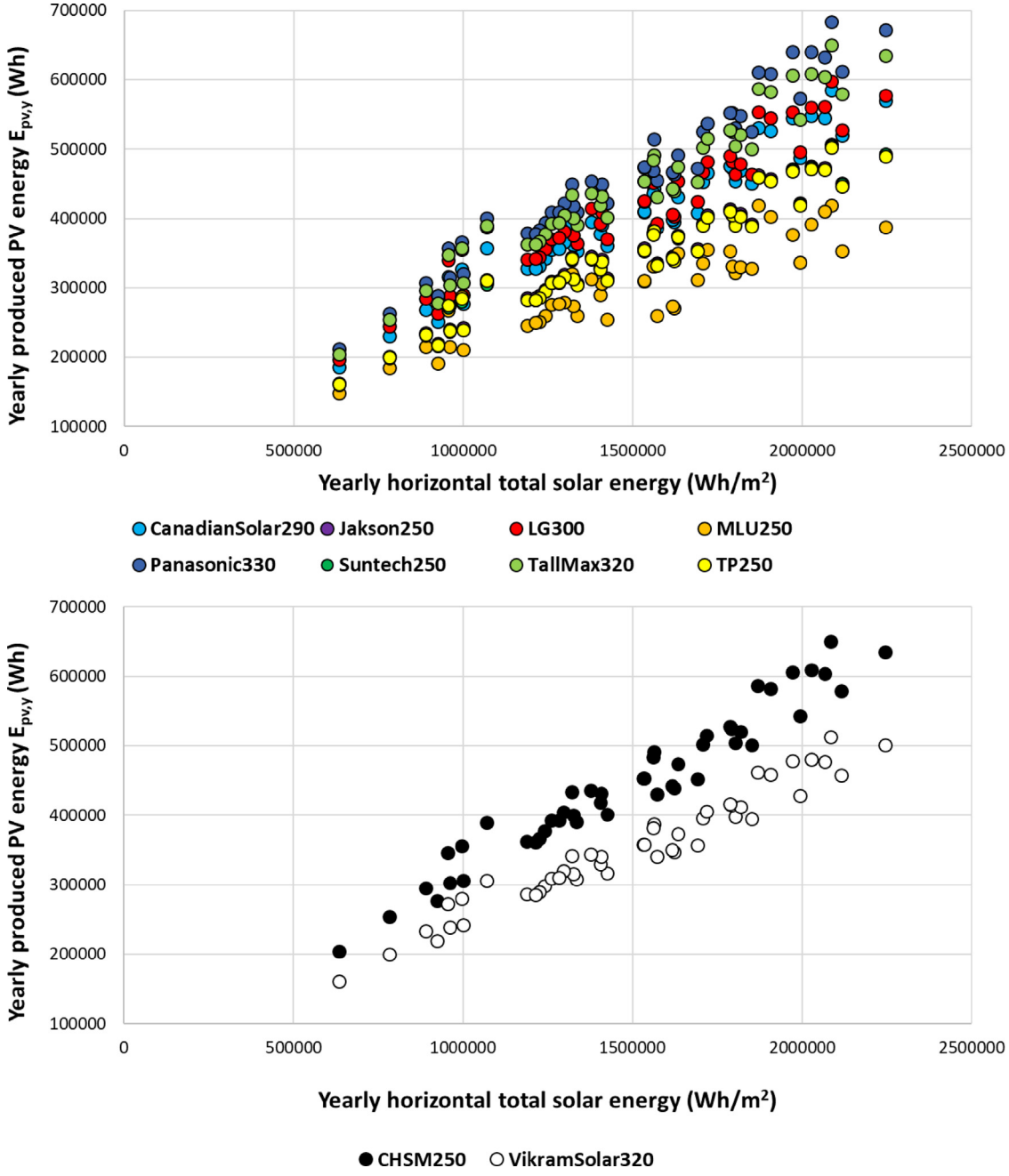

Fig. 10. Yearly produced solar PV energy at the optimal inclination angle $E_{pv,y}$ obtained with the TRNSYS software for each PV module by varying the locality. Top: Training PV modules; Bottom: Validation PV modules.

Table 4
Pearson correlation coefficient between the ANN output and inputs.

Variable	Pearson correlation coefficient ρ_{xy}
$E_{pv,y}$ (Wh)	1.00
$E_{s,y}$ (Wh/m^2)	0.81
$T_{ea,y}$ (°C)	0.43
N_{cs} (–)	−0.10
$V_{oc,ref}$ (V)	0.35
$I_{sc,ref}$ (A)	−0.11
$V_{mp,ref}$ (V)	0.35
$I_{mp,ref}$ (A)	−0.09
NOCT (°C)	−0.35
μ_{Isc} (%/°C)	−0.27
μ_{Voc} (%/°C)	0.37

reference conditions $V_{mp,ref}$. Instead, an increase of the nominal operating cell temperature NOCT or temperature coefficient of short-circuit current μ_{Isc} leads to a decrease in the output, denoting an inversely proportional correlation. Finally, there is no significant linear dependency between the output and the number of individual cells N_{cs} in a PV module and the current at maximum power point at reference conditions $I_{mp,ref}$.

Fig. 18 highlights the linear regression between the output (yearly PV electricity $E_{pv,y}$) and the most two linearly correlated inputs, namely the yearly horizontal total solar energy $E_{s,y}$ and the yearly average external air temperature $T_{ea,y}$.

In particular, an R^2 of 0.66 characterizes the regression between annual total horizontal solar energy and annual PV electricity. Taking into account the wide range of weather conditions in the locations considered and the different electrical and thermal characteristics of the PV modules selected, this value can be considered relatively high.

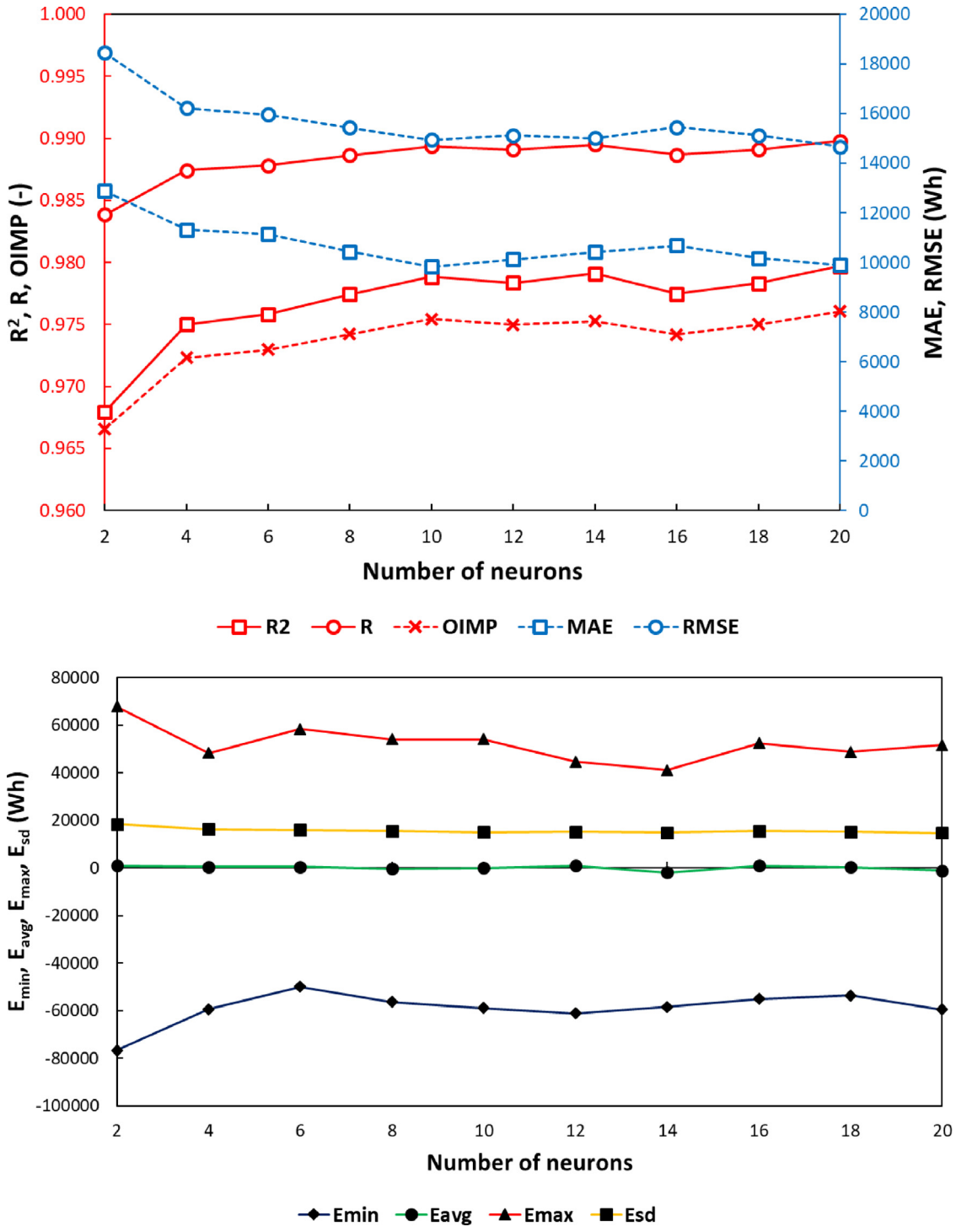

Fig. 11. Accuracy metrics by changing the number of neurons. Top: R^2, R OIMP, MAE and RMSE; Bottom: E_{min}, E_{avg}, E_{max} and E_{sd}.

4.2. Validation phase

The ANN trained in the previous section was validated separately on two new PV modules with electrical characteristics in the range of the PV modules used in the training phase. The Matlab algorithm code used for this purpose is reported in Appendix. A new input vector of size [10 × 48] for each PV module was loaded to simulate the yearly produced PV energy. The input vector contains: the yearly horizontal total solar energy $E_{s,y}$ for each of the 48 localities in the first row, the yearly average external air temperature $T_{ea,y}$ for each of the 48 localities in the second line and, in the remaining 8 rows, there are the electrical parameters of the PV module considered, which are constant for a specific PV module. By launching the Matlab algorithm code with this new input data, the vectors of the yearly energies produced by the CHSM250 and VikramSolar320 PV modules were obtained for each of the 48 localities. To assess the ANN accuracy, these values were compared with the yearly energy values obtained from the TRNSYS software for the 48 localities, as reported in the regression curves of Fig. 19.

The linear regression shows that there is an almost perfect correlation between the yearly energy calculated with TRNSYS

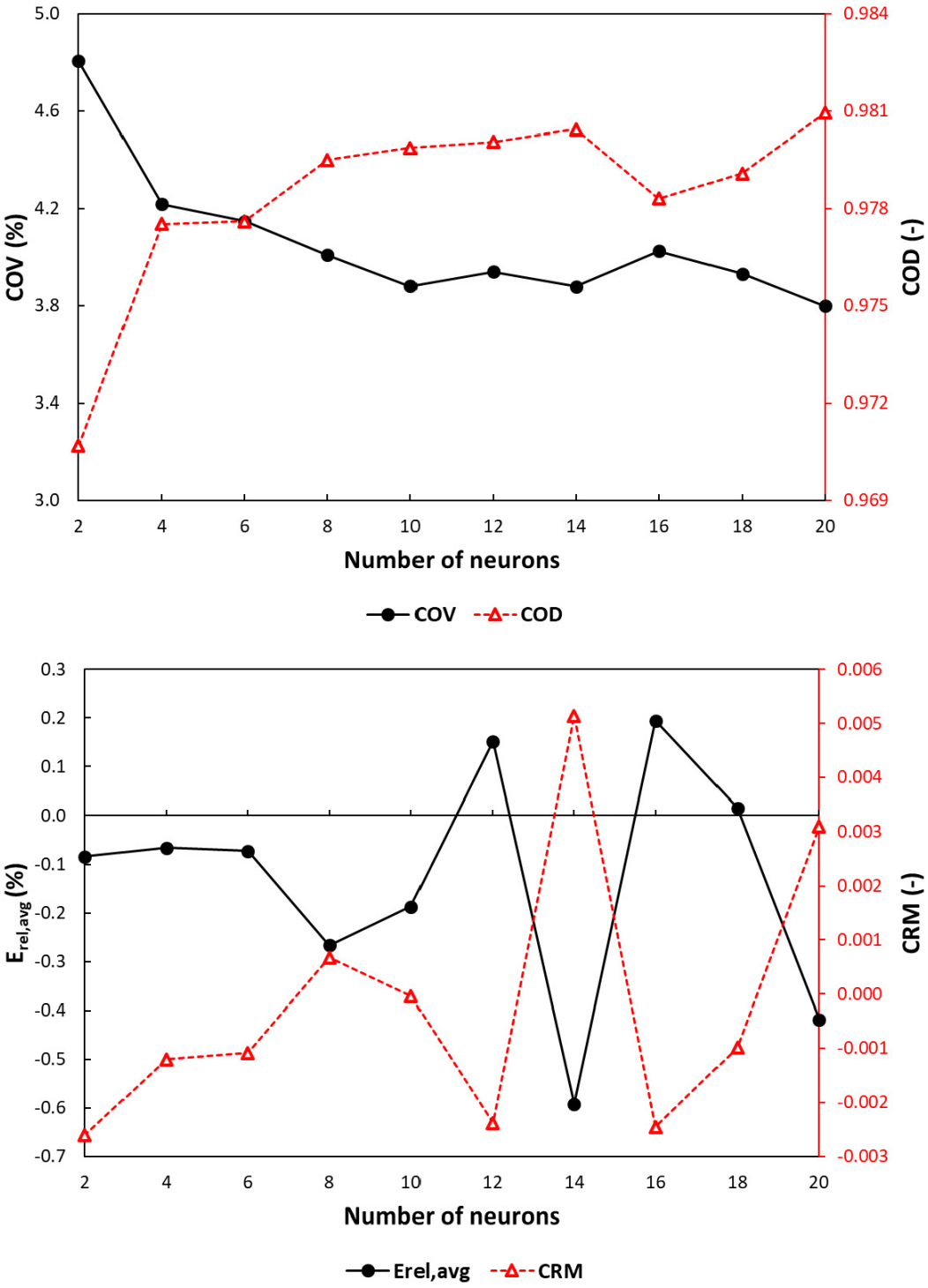

Fig. 12. Accuracy metrics by changing the number of neurons. Top: COV and COD; Bottom: $E_{\text{rel,avg}}$ and CRM.

and that obtained with the ANN for all 48 localities considered. The R2 is in the range of 0.96 ÷ 0.97 for both PV modules, while the angular coefficients of the regression curves are 0.7834 and 0.8989, respectively, for the VikramSolar320 and CHSM250 PV modules. The high accuracy of the ANN in the validation phase is also confirmed by the values of accuracy metrics obtained for the two PV modules individually and globally listed in Table 5.

Globally, despite the ANN was not trained with data related to the VikramSolar320 and CHSM250 PV modules, the accuracy metrics are only slightly worse than those obtained in the training phase;

Fig. 20 shows the comparison between the yearly energies produced by the two PV modules obtained with the TRNSYS software and with the ANN for the 48 localities; moreover, the relative percentage error $E_{rel,avg}$, calculated as the average of all percentage relative error obtained in each locality, is also reported. The frequency distribution of the relative error is reported for the two PV modules in Fig. 21.

The obtained figures demonstrate that the ANN also works perfectly with these new PV modules, as highlighted by the comparison of the target and ANN output for the different localities. The relative error ranges between −19.5% (New Delhi) and

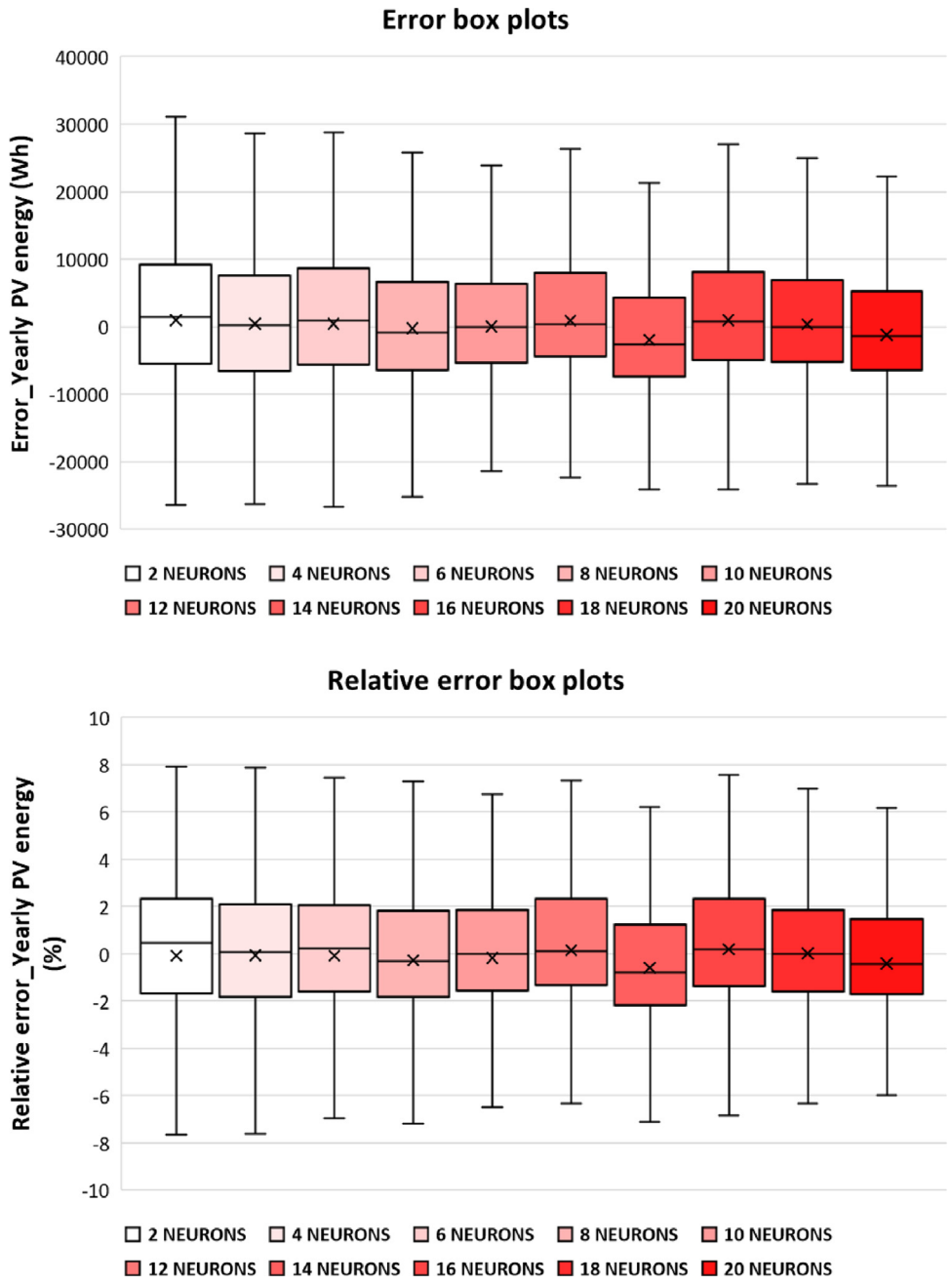

Fig. 13. Boxplots of the error, at the top, and relative error, at the bottom, on the yearly PV energy evaluation by means of the ANNs by varying the number of neurons.

4.8% (in Johannesburg) for the CHSM250 and between −13.3% (Hakkâri) and 13.4% (in Johannesburg) for the VikramSolar320. The relative error distribution highlights that the ANN mainly overestimates the yearly PV electricity for the CHSM250 module characterized by a lower module power. By increasing the PV module power, the relative distribution is almost perfectly symmetrical around zero, as exhibited for the VikramSolar module. Finally, relative errors between −7.5% and 7.5% are highlighted for 50% of localities in the case of the CHSM250 PV module and 81% of localities in the case of the VikramSolar320 PV module.

5. Conclusions

Recently, ANN-based machine learning models employing models have proved their capabilities as a precious prediction tool. In this research work, the ANN model for the prediction of yearly PV electricity directly on the optimal PV inclination angle was trained and validated. For the inclination angle, some empirical equations were obtained as a function of the latitude. Meteorological and PV module parameters were used in the TRNSYS environment to generate the training dataset. The ANN model requires as inputs only a small number of yearly weather data and PV module parameters and was successfully trained in Matlab software for 8 PV modules in 48 localities and validated for two other PV modules. The final numerical results of the ANN model showed considerable accuracy compared to those obtained with dynamic simulations performed with physical-based PV analytical models. In the training phase, the performance metrics of the developed ANN model revealed values of R, R^2 and OIMP of 0.9898, 0.9797 and 0.9761, respectively, and RMSE, MSE, and COV of 14.67 kWh, 215.3 mWh and 3.8%, respectively. In addition, the CRM is almost zero.

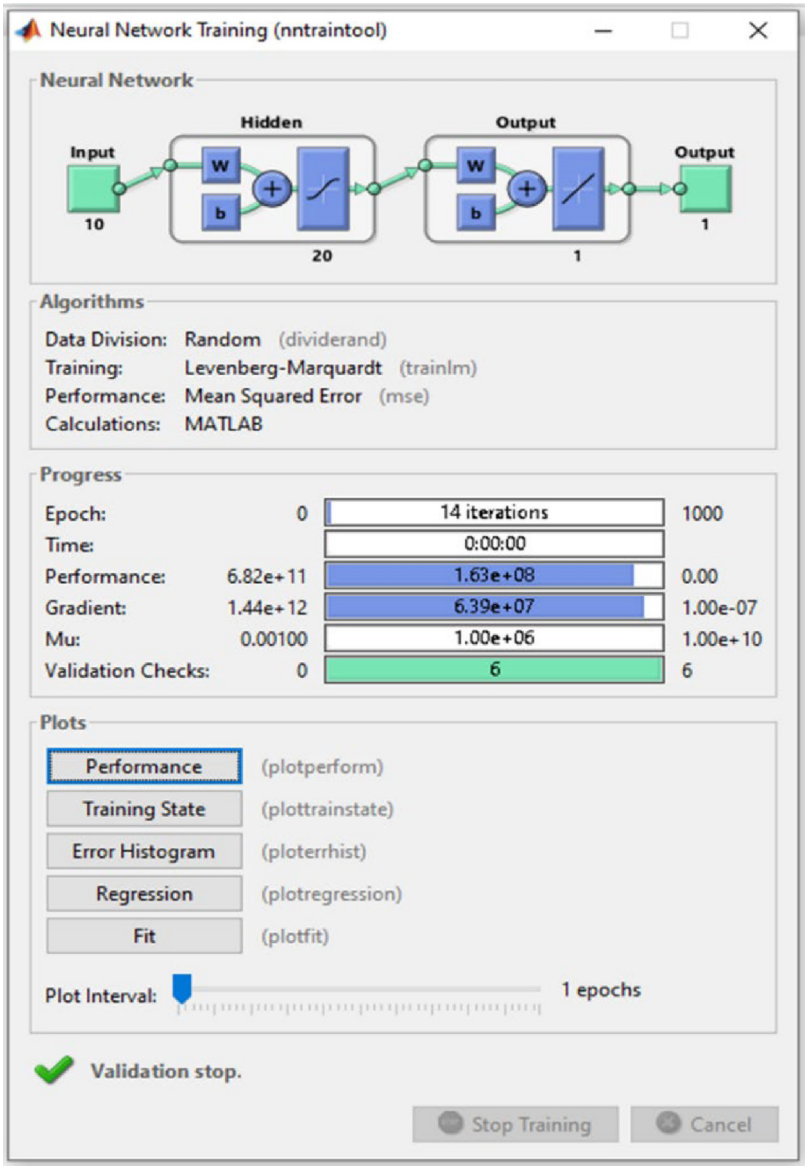

Fig. 14. Screen of the Neural net fitting tool after the end of the training process.

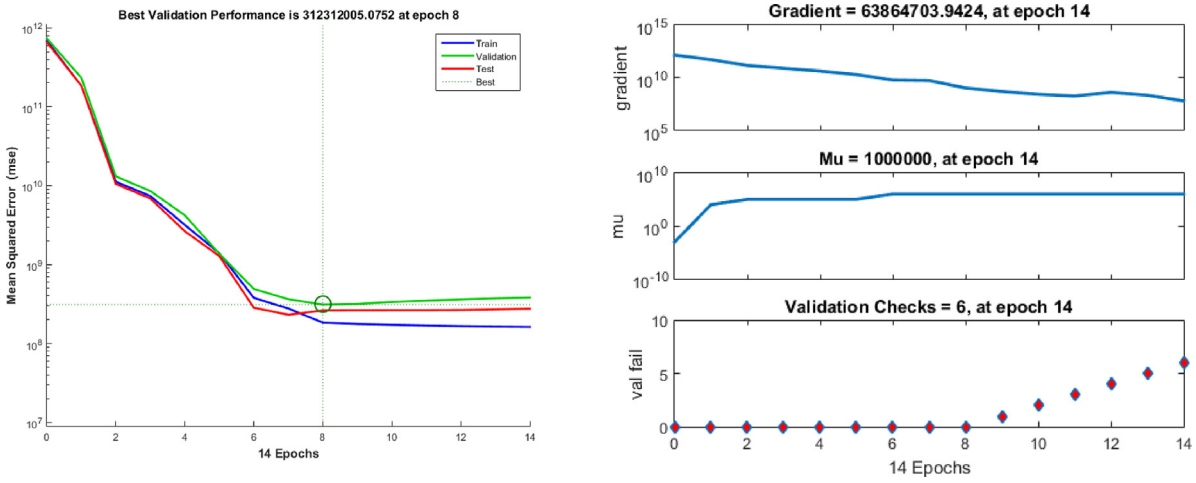

Fig. 15. Mean square error trend for the training, testing and validation phases on the left and trends of the gradient, damping factor and validation checks on the right by increasing the number of epochs.

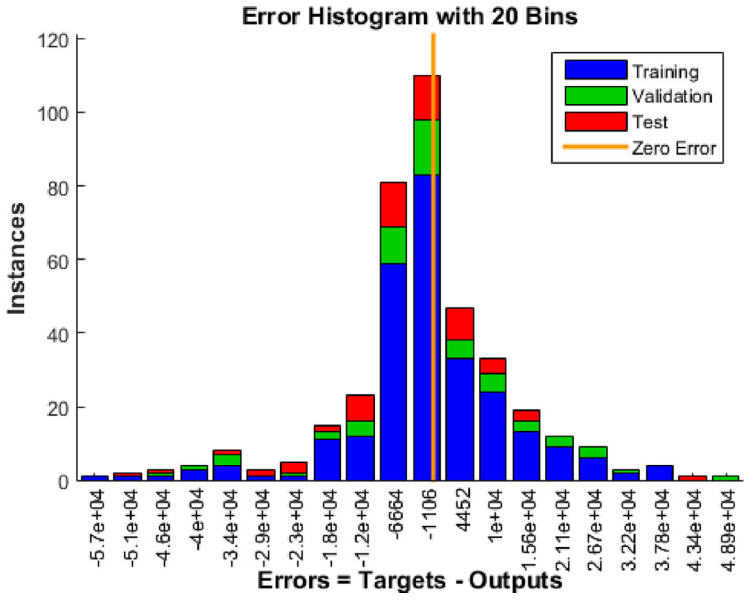

Fig. 16. Error histogram for the training, testing, validation and all data.

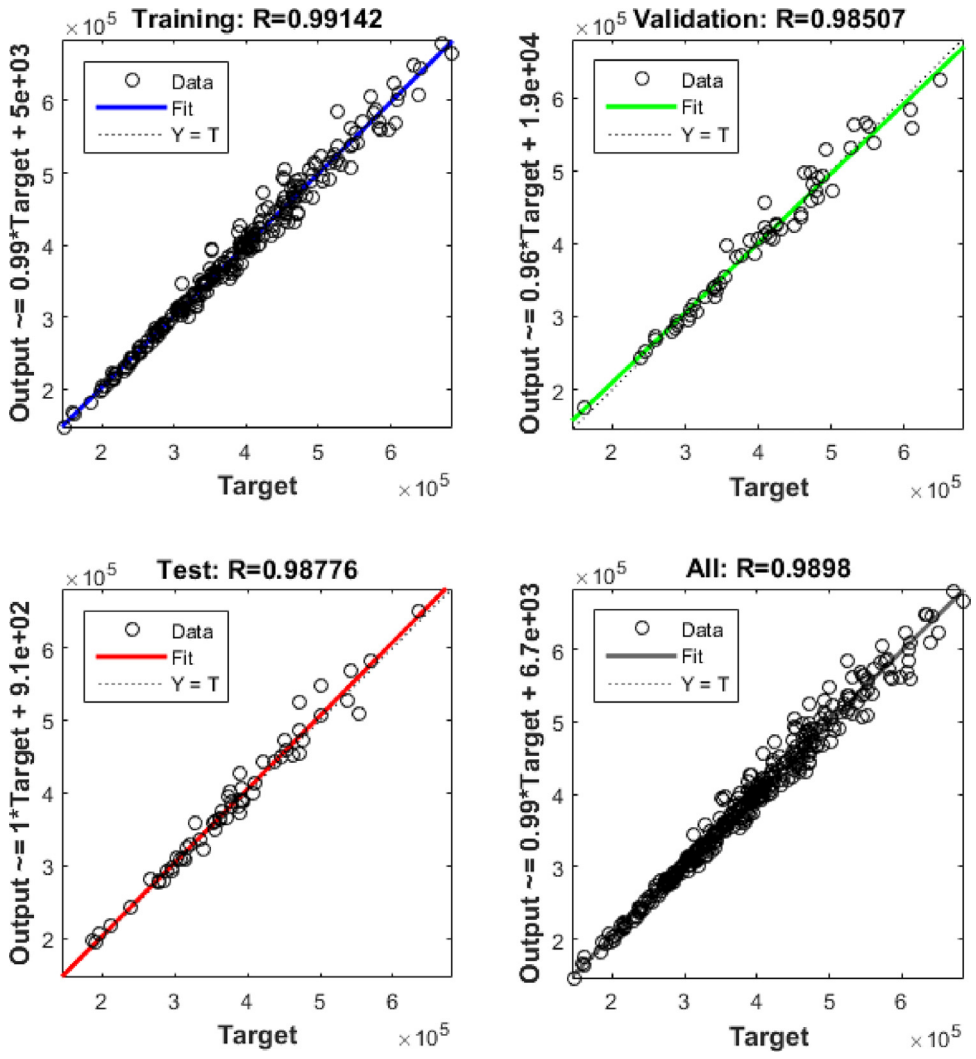

Fig. 17. Regression between the ANN output and the target for the training, testing, validation and all data.

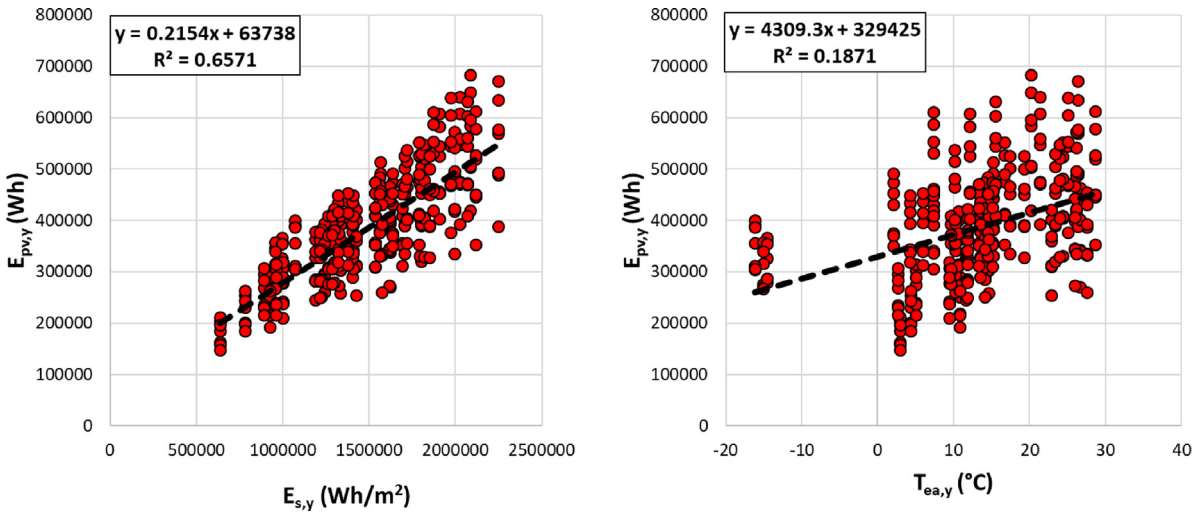

Fig. 18. Linear regression between the output and yearly horizontal total solar energy on the left, and between the output and yearly average external air temperature on the right.

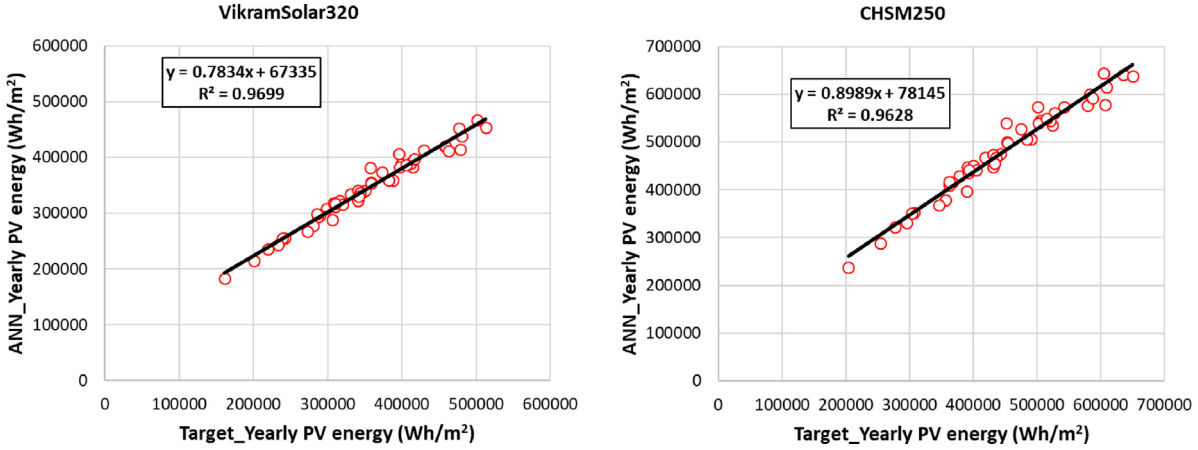

Fig. 19. Linear regression between the ANN output and the target for the VikramSolar 320 PV module on the left and for the CHSM250 on the right.

Table 5
Obtained accuracy metrics for CHSM250 and VikramSolar320 PV modules and globally.

		Global	VikramSolar320	CHSM250
MAE	(kWh)	25.98	17.10	34.85
R^2	(-)	0.9218	0.9699	0.9628
MSE	(MWh)	1020.72	513.82	1527.62
RMSE	(kWh)	31.95	22.67	39.08
EC	(–)	0.9052	0.9218	0.8551
CRM	(–)	0.0301	−0.0254	0.0740
COV	(%)	7.7693	6.6037	8.1566
R	(–)	0.9601	0.9848	0.9812
COD	(–)	0.9310	0.9699	0.9628
OIMP	(–)	0.9199	0.9287	0.8837
E_{max}	(kWh)	63.99	63.99	28.92
E_{min}	(kWh)	−88.20	−25.22	−88.20
E_{avg}	(kWh)	−12.03	8.96	−33.02
E_{sd}	(kWh)	29.75	21.04	21.13
$E_{rel,avg}$	(%)	−3.44	1.45	−8.34

Instead, in the validation phase, R^2 and RMSE obtained are 0.9218 and 31.95 kWh. The highest relative error found is −19.5% in New Delhi and relative errors between −7.5% and 7.5% are obtained for 50% of localities in the case of the first PV module with a nominal power of 250 W and for 81% of localities in the case of the second PV module of 320 W.

This emphasizes that the ANN model is almost accurate and versatile as compared to other complex models to predict yearly PV electricity while requiring few weather and PV input data, lower user knowledge and expertise and execution time.

Future research and PV sizing can be directly performed with the ANN proposed using only a few yearly weather variables, as well as electrical and thermal characteristics of the PV module as inputs, while avoiding solving electric circuit equations and, therefore, non-linear equations for determining parameters. As a result, no hourly simulation is needed since the resulting ANN PV energy is very close to what would have been calculated from an hourly simulation.

This study has two limitations:

- The most important limitation is that the 5-parameter model was used for all types of considered PV modules. Such an assumption must be made to make uniform all simulations, make the ANN of global value and reduce time spent to identify the best model to be used for each type of PV module. The selected 5-parameter represents a trade-off between computation cost and accuracy.
- Another limitation is that the yearly weather parameters are summarized with their average values; no input data are provided to ANN to take into account the variability of climatic conditions in time.

For future research, we recommend the following tasks: (i) use of an ANN with a more advanced algorithm to further improve the prediction accuracy; (ii) proposal of other optimized ANNs using data from simulations performed by other PV models with 4, 6 or 7 parameters; (iii) evaluation of the effect of introducing the

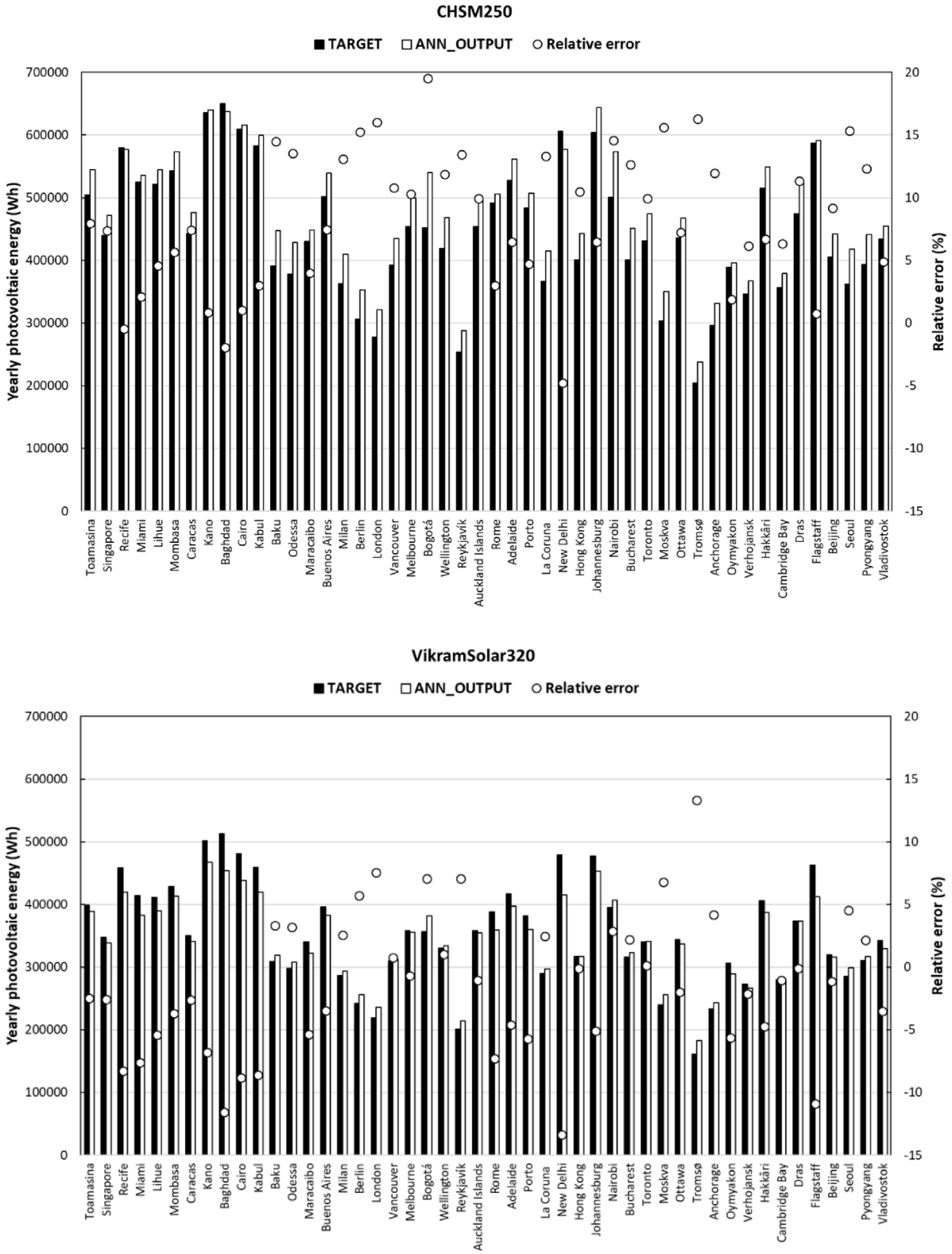

Fig. 20. Yearly electrical energies produced by the CHSM250 and VikramSolar320 PV modules obtained with the TRNSYS software and with the ANN for the 48 localities.

standard deviation of weather variables as an additional input in the training data set on the accuracy of the ANN.

As a further development of this research, it should be taken into account that statistical analysis of the results indicated that the yearly horizontal total solar energy and yearly average air temperature had the most significant correlation with the PV output. For this reason, these two variables may be used as the only input variables in future analysis to make the ANN further simple and more user-friendly.

Given the demonstrated tool flexibility and accuracy for different weather conditions, the use of the ANN-based machine learning algorithm proposed can be very beneficial in future investment decision-making to develop incentive plans for PV economic growth, taking into account the ongoing climate change.

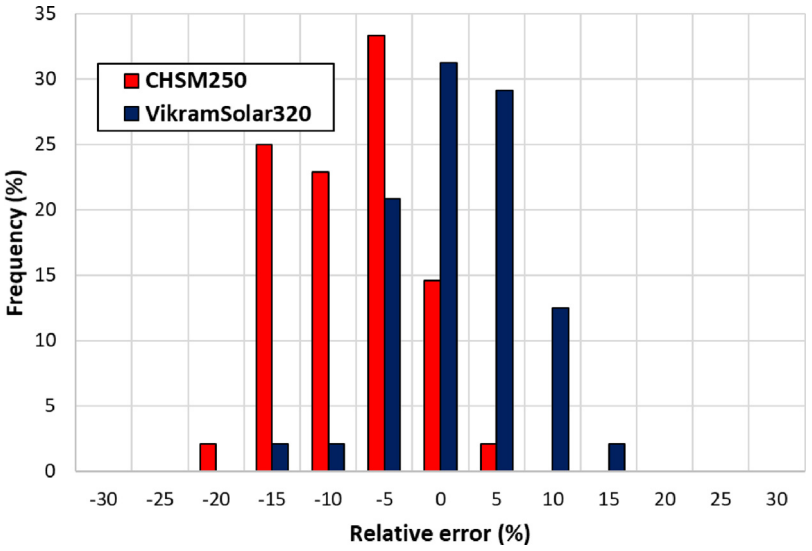

Fig. 21. Frequency distribution of the relative error of the ANN for the CHSM250 and VikramSolar320 PV modules.

Abbreviations	
ANN	Artificial neural network
CC	Correlation coefficient
COD	Coefficient of determination
COP21	Paris climate conference
COP3	Conference of the Parties in Kyoto
COV	Coefficient of variance
CRM	Coefficient of residual mass
EC	Efficiency coefficient
EU	European Union
MAE	Mean absolute error
MPPT	Maximum electrical power tracker
MSE	Mean square error
NOCT	Nominal cell operating temperature
OIMP	Overall index of model performance
PV	Photovoltaic
RES	Renewable energy source
RMSE	Root mean square error
STC	Standard condition tests
TRNSYS	Transient system simulation tool

Symbols	
I_{01}, I_{02}	Inverse saturation currents of diodes 1 and 2
I_L	Light current
n_1, n_2	Ideality factors of diodes 1 and 2
R_s	Series resistance
R_p	Shunt resistance
q	Charge of the electron
K_B	Boltzmann constant
T_c	Photovoltaic cell temperature
V	Voltage
V_{oc}	Open-circuit voltage of the PV module
I_{sc}	Short-circuit current of the PV module
β_{Voc}	Open-circuit voltage temperature coefficient
Uc	Array thermal loss coefficient
η_c	Conversion efficiency of the module
$\tau\alpha$	Module transmittance-absorptance product
α_{Isc}	Short-circuit current temperature coefficient
$f(A)$	Activation function
x_j	Input signal number to artificial neurons i
w_{ij}	Weight value of synapse connecting artificial neurons i and j
s_i	Threshold value of artificial neuron i
$y(x)$	Output of the neuron that represents the weighted sum of the input Signals x_j with the relative weights w_{ij}
k	$k-th$ step in the training process
ε	Parameter of the activation function
g	Gradient
α	Learning rate
H	Hessian matrix
J	Jacobian matrix
t_i	$i-th$ value of the target output obtained from the TRNSYS simulations
t_m	Mean value of the target output obtained from the TRNSYS simulations
t_{max}	Maximum value of the target output obtained from the TRNSYS simulations
$t_{\min}$	Minimum value of the target output obtained from the TRNSYS simulations
y_i	$i-th$ value of the output predicted by the ANN
y_m	Mean value of the output predicted by the ANN
N	Total number of training data
e	Error vector
μ	Damping coefficient
I	Identity matrix
E	Error value
E_{min}	Minimum error
E_{max}	Maximum error
E_{sd}	Standard deviation of error
$E_{rel,avg}$	Relative average error
ρ_{XY}	Pearson's correlation coefficient of input X and output Y variables
σ_X	Standard deviation of input variable X
σ_Y	Standard deviation of output variable Y
G_T	Solar radiation
IAM	Incidence angle modifier
P_{pv}	Electrical power output of the PV module
P_n	Nominal power of the PV module
I_{mp}	Current at maximum power point of the PV module
V_{mp}	Voltage at the maximum power point of the PV module
$T_{c,NOCT}$	Operating cell temperature at standard NOCT measurements
$G_{T,NOCT}$	Solar radiation at standard NOCT measurements
$T_{a,NOCT}$	Ambient temperature at standard NOCT measurements
μ_{sc}	Temperature coefficient of the short-circuit current
A	Area of the PV module

η_m	Efficiency of the PV module
$E_{pv,y}$	Yearly produced solar PV energy at the optimal inclination angle
$E_{s,y}$	Yearly horizontal total solar energy
$T_{ea,y}$	Yearly average external air temperature
N_{cs}	Number of individual cells in a PV module
$V_{mp,ref}$	Voltage at maximum power point along IV curve at reference conditions
$I_{mp,ref}$	Current at maximum power point along IV curve at reference conditions

Declaration of competing interest

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

Data availability

Data will be made available on request.

Acknowledgements

This study was partly carried out within the Agritech National Research Center and received partial funding from the European Union Next-GenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) — MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.4 — D.D. 1032 17/06/2022, CN00000022). This manuscript reflects only the authors' views and opinions, neither the European Union nor the European Commission can be considered responsible for them.

Appendix. Matlab algorithm code of the ANN for the yearly produced PV energy

```
function [y1] = myNeuralNetworkFunction(x1)
%MYNEURALNETWORKFUNCTION neural network simulation function.
%
% Generated by Neural Network Toolbox function genFunction, 05-May-2021 12:38:41.
%
% [y1] = myNeuralNetworkFunction(x1) takes these arguments:
%   x = 10xQ matrix, input #1
% and returns:
%   y = 1xQ matrix, output #1
% where Q is the number of samples.
%#ok<*RPMT0>

% ===== NEURAL NETWORK CONSTANTS =====

% Input 1
x1_step1_xoffset = [634301.871035932;-16.1517351598173;60;37.3;6.07;30.2;5.7;43;0.03;-0.35];
x1_step1_gain = [1.24110736568843e-06;0.0447027838964855;0.0333333333333333;0.0617283950617284;0.547945205479452;0.0719424460431655;0.574712643678161;0.5;76.9230769230769;10.752688172043];
x1_step1_ymin = -1;

% Layer 1
b1 = [-3.3106233360505319;2.0215782736193648;-1.6196156203080219;1.1079039707661795;-1.163829783034684;0.7692985355237903 5;0.9224265084126147 5;-0.55066419953101153;-0.29312075185926117;0.15421029440801823;0.88658421587163661;-0.68065138634336786;0.21679859910885299;0.12098895155292323;-2.1059690261634141;0.45115933175230249;-1.0249936465060878;-2.179046112509694;1.9500568908151776;1.7573249369962363];
IW1_1 = [0.25135343666320231 2.2200369794238464 -0.19103769153788885 0.037326464511691027
0.60542815015084128 0.40707667750529336 -0.96058272254164689 -0.12272477295596733
0.93381823673876274 0.44939190910529137;0.86764108986960575 1.2970695756417814
0.87089465069664218 -1.0608652309697024 1.2566082125554907 -0.65598538372516102
0.29871410959519834 0.89747212337103521 -0.31386819191117621 -0.21822164124207646;-
0.10437802752198724 -0.54947254730402151 -1.0843146614639683 0.0029777701415117132
0.20116344090210511 0.89191151128012658 0.37937404186382101 -0.53991450111499029
0.49514521265557676 -0.98333551255040141;-0.44423238561897088 -0.11061193353266725
0.52793661756133869 -0.64647521497400873 -0.58568815112715811 0.66518892473956504
0.31563658679532458 0.071513247961091125 0.058876450037957065
0.35475451245762135;0.7094086954741643 -0.063757919375982103 -0.6959028599266196
0.44818509032187942 1.3011451048000413 -0.013548464164914049 0.11525551204271212
0.51094181374420256 -1.1174294416443449 0.29867332047670786;0.62665519660505065
0.34625628630894878 0.2547412153542345 -1.0786876950883473 0.83259999373666937 -
0.96934475046733337 1.1182698101631328 0.53916376072493577 0.1123086347442325 -
0.27604016550039384;0.98111665604953413 0.30553616243291887 0.27891438169663241
0.59608396972012057 -0.14920765316533288 -1.3003540214725178 0.93571116753032801 -
0.18077355705416215 -0.49347958865713082 -1.2093669623445502;-0.35179034946790744 -
0.31694144405536162 0.44301507417371377 0.15191907229694887 0.22291153768515373
0.53631780136373663 1.1955562167221128 0.23643528345228693 1.2830870633239366
1.2956426619096841;1.0790099171973491 -0.10609339208730839 1.020941090245278 -
1.1876419638389351 1.1973930177166108 -0.5988026353659901 -0.10199054462262155 -
0.15408212067568022 -1.0007414357830271 -0.11071855482712892;0.89581664127446547 -
0.010715430758994386 1.0787350593517566 -0.12614489282585087 0.27249375518594404
1.2728996946264639 -0.12757646424372279 0.076848889000360632 0.51188550114200493 -
0.2393515027333292;0.83228932171418069 -0.48015320185155919 -0.13675500656659595
0.10734236352061971 -0.04727379406860463 0.64216847588508097 0.041905541304288217
0.72991042411430862 0.8917695710600626 0.21494349049023134;-0.36537761684045256
0.25097326231285177 -0.25621317552407502 0.79417118986404067 -0.052404360141884349 -
0.21007310245571589 0.93273972827000162 1.0985751457478141 -0.13518197817109320 -
```

```
0.4878835635914639;2.7042369667441841 -2.1064212208589521 -1.7387405918876451
0.45084454670506457 -0.17527307314355889 1.0334140303554937 0.44494273732555878
0.8299859660998522 -0.71282865510809268 -0.133552097485306;-0.048798639798191309 -
0.34534051530330134 2.1187634870938452 0.36826720799025681 -0.061418503964738726
0.86038449449585441 -0.89360778760204207 0.24608703069376225 0.76731026530812141
0.32594841287290771;-1.4891246830536549 0.33783387573257811 -0.21149678125979274
0.71616506426802984 -0.4443574096567342 0.014803327920995127 0.55405565065047524 -
0.9351631934320549 0.6430296681290798 0.28765457431199387;0.31043077615429249
0.41892911449177672 -0.59383263530084907 -1.2830019226360581 1.0513527754609555
0.30486932630923885 -0.24419401920980149 0.94577868773534113 0.4801573638607865 -
0.63750714124681884;-0.15053411774558223 -0.66765994929904404 -0.067866713645861235 -
0.1044149825536257 -0.43685888847194376 2.0541290854164016 -0.71823653834123957 -
0.091751663041859707 0.50198434795107982 0.86955121115290335;0.36938160088707223 -
0.23455921226982823 -0.49729550558785574 -0.78239045766494153 0.13787840104210849
0.43024231604349811 0.19103534586256229 -0.32882654128382699 -0.7377567108319969
0.73251302583261024;1.024215609989495 0.3649003190784943 -0.12953246499724322 -
0.42308870940465054 -0.1885252338175851 0.56528603007770173 0.99665209361916718
0.65694678679928054 -0.24510622818581226 0.92773202605048144;2.2827124011547877 -
0.8460362940482169 -1.0643574937551594 0.68919855913987571 -0.9346250763534405
0.35751307070952326 0.79105235188852929 -0.13418476653392167 0.11333222474878721
0.88657844961273591];

% Layer 2
b2 = -0.90570906796381223;
LW2_1 = [-0.92474945084102955 0.55755738734068649 -0.34220509648738562 -0.38713274115950858
0.38341924943928479 -0.021170184863182329 0.36895025642055007 0.21118506440417223 -
0.31546163262736177 0.59332482931739561 0.31649596768521393 -0.59471051232018701 -
0.023308418082055915 -0.55438883896478997 -0.17528297592529721 -0.33850756669250626
0.33235223305491401 0.39811091159645529 -0.066902382063219856 0.1645680475649029];

% Output 1
y1_step1_ymin = -1;
y1_step1_gain = 3.73826012562395e-06;
y1_step1_xoffset = 148551.724680299;

% ===== SIMULATION ========

% Dimensions
Q = size(x1,2); % samples

% Input 1
xp1 = mapminmax_apply(x1,x1_step1_gain,x1_step1_xoffset,x1_step1_ymin);

% Layer 1
a1 = tansig_apply(repmat(b1,1,Q) + IW1_1*xp1);

% Layer 2
a2 = repmat(b2,1,Q) + LW2_1*a1;

% Output 1
y1 = mapminmax_reverse(a2,y1_step1_gain,y1_step1_xoffset,y1_step1_ymin);
end

% ===== MODULE FUNCTIONS ========

% Map Minimum and Maximum Input Processing Function
function y = mapminmax_apply(x,settings_gain,settings_xoffset,settings_ymin)
y = bsxfun(@minus,x,settings_xoffset);
y = bsxfun(@times,y,settings_gain);
y = bsxfun(@plus,y,settings_ymin);
end
% Sigmoid Symmetric Transfer Function
function a = tansig_apply(n)
a = 2 ./ (1 + exp(-2*n)) - 1;
end

% Map Minimum and Maximum Output Reverse-Processing Function
function x = mapminmax_reverse(y,settings_gain,settings_xoffset,settings_ymin)
x = bsxfun(@minus,y,settings_ymin);
x = bsxfun(@rdivide,x,settings_gain);
x = bsxfun(@plus,x,settings_xoffset);
end
```

References

Ahmad, A.S., Hassan, M.Y., Abdullah, M.P., Rahman, H.A., Hussin, F., Abdullah, H., Saidur, R., 2014. A review on applications of ANN and SVM for building electrical energy consumption forecasting. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 33, 102–109. http://dx.doi.org/10.1016/j.rser.2014.01.069.

Alhendi, A., Saad Al-Sumaiti, A., Marzband, M., Kumar, R., Zaki Diab, A.A., 2023. Short-term load and price forecasting using artificial neural network with enhanced Markov chain for ISO new England. Energy Rep. (ISSN: 2352-4847) 9, 4799–4815. http://dx.doi.org/10.1016/j.egyr.2023.03.116.

Ali, E.E., El-Hameed, M.A., El-Fergany, A.A., El-Arini, M.M., 2016. Parameter extraction of photovoltaic generating units using multi-verse optimizer. Sustain. Energy Technol. Assess. (ISSN: 2213-1388) 17, 68–76. http://dx.doi.org/10.1016/j.seta.2016.08.004.

Arnfield, A.J., 2020. Köppen climate classification. In: EncyclopæDia Britannica. Encyclopædia Britannica, Inc., https://www.britannica.com/science/Koppen-climate-classification.

Babiker, M., Reilly, J.M., Jacoby, H.D., 2000. The Kyoto protocol and developing countries. Energy Policy (ISSN: 0301-4215) 28 (8), 525–536. http://dx.doi.org/10.1016/S0301-4215(00)00033-1.

Baglivo, C., Mazzeo, D., Panico, S., Bonuso, S., Matera, N., Congedo, P.M., Oliveti, G., 2020. Complete greenhouse dynamic simulation tool to assess the crop thermal well-being and energy needs. Appl. Therm. Eng. (ISSN: 1359-4311) 179, 115698. http://dx.doi.org/10.1016/j.applthermaleng.2020.115698.

Boutana, N., Mellit, A., Haddad, S., Rabhi, A., Massi Pavan, A., 2017. An explicit I-V model for photovoltaic module technologies. Energy Convers. Manage. (ISSN: 0196-8904) 138, 400–412. http://dx.doi.org/10.1016/j.enconman.2017.02.016.

Canadian Solar Inc., 2023. Guelph ON, Canada. https://www.csisolar.com. (Accessed 21 April 2023).

Chen, Z., Wu, L., Lin, P., Wu, Y., Cheng, S., 2016. Parameters identification of photovoltaic models using hybrid adaptive Nelder-Mead simplex algorithm based on eagle strategy. Appl. Energy (ISSN: 0306-2619) 182, 47–57. http://dx.doi.org/10.1016/j.apenergy.2016.08.083.

Chin, V.J., Salam, Z., Ishaque, K., 2015. Cell modelling and model parameters estimation techniques for photovoltaic simulator application: A review. Appl. Energy (ISSN: 0306-2619) 154, 500–519. http://dx.doi.org/10.1016/j.apenergy.2015.05.035.

Clark, D.R., Klein, S.A., Beckman, W.A., 1984. A method for estimating the performance of photovoltaic systems. Sol. Energy (ISSN: 0038-092X) 33 (6), 551–555. http://dx.doi.org/10.1016/0038-092X(84)90010-0.

Cong, T., Su, G., Qiu, S., Tian, Wenxi, 2013. Applications of ANNs in flow and heat transfer problems in nuclear engineering: A review work. Prog. Nucl. Energy (ISSN: 0149-1970) 62, 54–71. http://dx.doi.org/10.1016/j.pnucene.2012.09.003.

Dolara, A., Leva, S., Manzolini, G., 2015. Comparison of different physical models for PV power output prediction. Sol. Energy (ISSN: 0038-092X) 119, 83–99. http://dx.doi.org/10.1016/j.solener.2015.06.017.

Du, B., Lund, P.D., Wang, J., 2022. Improving the accuracy of predicting the performance of solar collectors through clustering analysis with artificial neural network models. Energy Rep. (ISSN: 2352-4847) 8, 3970–3981. http://dx.doi.org/10.1016/j.egyr.2022.03.013.

Elaziz, M.A., Thanikanti, S.B., Ibrahim, I.A., Lu, S., Nastasi, B., Alotaibi, M.A., Md A. Hossain, Yousri, D., 2021. Enhanced marine predators algorithm for identifying static and dynamic photovoltaic models parameters. Energy Convers. Manage. (ISSN: 0196-8904) 236, 113971. http://dx.doi.org/10.1016/j.enconman.2021.113971.

Elsheikh, A.H., Sharshir, S.W., Elaziz, M.A., Kabeel, A.E., Guilan, W., Haiou, Z., 2019. Modeling of solar energy systems using artificial neural network: A comprehensive review. Sol. Energy (ISSN: 0038-092X) 180, 622–639. http://dx.doi.org/10.1016/j.solener.2019.01.037.

ENF Solar Ltd., 2023. https://www.enfsolar.com. (Accessed 21 April 2023),

Fry, B.A., 1999. Simulation of Grid-Tied Building Integrated Photovoltaic Systems (M.S. thesis). Solar Energy Laboratory, University of Wisconsin, Madison.

Galvin, R., Healy, N., 2020. The Green new deal in the United States: What it is and how to pay for it. Energy Res. Soc. Sci. (ISSN: 2214-6296) 67, 101529. http://dx.doi.org/10.1016/j.erss.2020.101529.

Ghadami, N., Gheibi, M., Kian, Z., Faramarz, M.G., Naghedi, R., Eftekhari, M., Fathollahi-Fard, A.M., Dulebenets, M.A., Tian, G., 2021. Implementation of solar energy in smart cities using an integration of artificial neural network, photovoltaic system and classical Delphi methods. Sustainable Cities Soc. (ISSN: 2210-6707) 74, 103149. http://dx.doi.org/10.1016/j.scs.2021.103149.

Ghazikhani, A., Babaeian, I., Gheibi, M., Hajiaghaei-Keshteli, M., Fathollahi-Fard, A.M., 2022. A smart post-processing system for forecasting the climate precipitation based on machine learning computations. Sustainability 14 (11), 6624. http://dx.doi.org/10.3390/su14116624.

Ghritlahre, H.K., Prasad, R.K., 2018. Application of ANN technique to predict the performance of solar collector systems — A review. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 84, 75–88. http://dx.doi.org/10.1016/j.rser.2018.01.001.

Gong, J., 2021. Network data mining based on artificial intelligence inference engine. Microprocess. Microsyst. (ISSN: 0141-9331) 82, 103794. http://dx.doi.org/10.1016/j.micpro.2020.103794.

Hagan, M.T., Demuth, H.B., Beale, M.H., 1996. Neural Network Design. PWS Publishing, Boston, MA.

Hagan, M.T., Menhaj, M.B., 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Netw. (ISSN: 1941-0093) 5 (6), 989–993. http://dx.doi.org/10.1109/72.329697.

Herdem, M.S., Mazzeo, D., Matera, N., Wen, J.Z., Nathwani, J., Hong, Z., 2020. Simulation and modeling of a combined biomass gasification-solar photovoltaic hydrogen production system for methanol synthesis via carbon dioxide hydrogenation. Energy Convers. Manage. (ISSN: 0196-8904) 219, 113045. http://dx.doi.org/10.1016/j.enconman.2020.113045.

Hoang, S.A.T., Nižetić, S., Ong, H.C., Tarelko, W., Pham, V.V., Le, T.H., Chau, M.Q., Nguyen, X.P., 2021. A review on application of artificial neural network (ANN) for performance and emission characteristics of diesel engine fueled with biodiesel-based fuels. Sustain. Energy Technol. Assess. (ISSN: 2213-1388) 47, 101416. http://dx.doi.org/10.1016/j.seta.2021.101416.

Humada, A.M., Darweesh, S.Y., Mohammed, K.G., Kamil, M., Mohammed, S.F., Kasim, N.K., Tahseen, T.A., Awad, O.I., Mekhilef, S., 2020. Modeling of PV system and parameter extraction based on experimental data: Review and investigation. Sol. Energy (ISSN: 0038-092X) 199, 742–760. http://dx.doi.org/10.1016/j.solener.2020.02.068.

Humada, A.M., Hojabri, M., Mekhilef, S., Hamada, H.M., 2016. Solar cell parameters extraction based on single and double-diode models: A review. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 56, 494–509. http://dx.doi.org/10.1016/j.rser.2015.11.051.

Jakson Group, 2023. Noida, Uttar Pradesh, India. https://www.jakson.com. (Accessed 21 April 2023).

Kang, M., Elbel, S., 2023. Novel regenerator design for caloric cycles using artificial neural network — Genetic algorithm method and additive manufacturing. Energy Rep. (ISSN: 2352-4847) 9, 4257–4274. http://dx.doi.org/10.1016/j.egyr.2023.03.031.

Karamirad, M., Omid, M., Alimardani, R., Mousazadeh, H., Heidari, S.N., 2013. ANN based simulation and experimental verification of analytical four- and five-parameters models of PV modules. Simul. Model. Pract. Theory (ISSN: 1569-190X) 34, 86–98. http://dx.doi.org/10.1016/j.simpat.2013.02.001.

Kim, P., 2017. Matlab deep learning, with machine learning. Neural Netw. Artif. Intell. 130 (21), http://dx.doi.org/10.1007/978-1-4842-2845-6, eBook,

Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Koppen-Geiger climate classification updated. Meteorol. Z. 15 (3), 259–263. http://dx.doi.org/10.1127/0941-2948/2006/0130.

Laurent, É., 2020. The European Green deal: Bring back the new. OFCE Policy Brief 63, 1–10, https://spire.sciencespo.fr/hdl:/2441/5h72jqj51784nakihclqujf0ke/resources/ofce-policy-brief-63.pdf.

Levenberg-Marquardt backpropagation, 0000. Deep Learning Toolbox. ©COPYRIGHT 1992–2004 by The MathWorks, Inc., Natick, Massachusetts, United State, https://www.mathworks.com/help/deeplearning/ref/trainlm.html.

LG Electronics, 2023. Seoul, korea. https://www.lg.com. (Accessed 21 April 2023).

Li, J., Herdem, M.S., Nathwani, J., Wen, J.Z., 2023. Methods and applications for artificial intelligence, big data, Internet of Things, and blockchain in smart energy management. Energy AI (ISSN: 2666-5468) 11, 100208. http://dx.doi.org/10.1016/j.egyai.2022.100208.

Liu, X., Tian, S., Tao, F., Yu, W., 2021. A review of artificial neural networks in the constitutive modeling of composite materials. Composites B (ISSN: 1359-8368) 224, 109152. http://dx.doi.org/10.1016/j.compositesb.2021.109152.

Lun, Shu-xian., Du, Cun-jiao., Guo, Ting-ting., Wang, Shuo, Sang, Jing-shu., Li, Jia-pei., 2013. A new explicit I–V model of a solar cell based on Taylor's series expansion. Sol. Energy (ISSN: 0038-092X) 94, 221–232. http://dx.doi.org/10.1016/j.solener.2013.04.013.

Ma, T., Yang, H., Lu, L., 2014. Development of a model to simulate the performance characteristics of crystalline silicon photovoltaic modules/strings/arrays. Sol. Energy (ISSN: 0038-092X) 100, 31–41. http://dx.doi.org/10.1016/j.solener.2013.12.003.

Mares, O., Paulescu, M., Badescu, V., 2015. A simple but accurate procedure for solving the five-parameter model. Energy Convers. Manage. (ISSN: 0196-8904) 105, 139–148. http://dx.doi.org/10.1016/j.enconman.2015.07.046.

Marin, R., Ciortan, S., Rusu, E., 2022. A novel method based on artificial neural networks for selecting the most appropriate locations of the offshore wind farms. Energy Rep. (ISSN: 2352-4847) 8 (16), 408–413. http://dx.doi.org/10.1016/j.egyr.2022.10.248.

Matera, N., Mazzeo, D., Baglivo, C., Congedo, P.M., 2023a. Hourly forecasting of the photovoltaic electricity at any latitude using a network of artificial neural networks. Sustain. Energy Technol. Assess. (ISSN: 2213-1388) 57, 103197. http://dx.doi.org/10.1016/j.seta.2023.103197.

Matera, N., Mazzeo, D., Congedo, P.M., 2023b. Energy-sustainable hospitals: Integration of a novel compound parabolic concentrator system with two storage tanks for domestic hot water production at high and low temperatures. Appl. Therm. Eng. (ISSN: 1359-4311) 221, 119773. http://dx.doi.org/10.1016/j.applthermaleng.2022.119773.

Mazzeo, D., Matera, N., Bevilacqua, P., Arcuri, N., Energy and economic analysis of solar photovoltaic plants located at the university of Calabria. Int. J. Heat Technol.. (ISSN: 0392-8764) 33 (4), 41-50. https://doi.org/10.18280/ijht.330406.

Mazzeo, D., Matera, N., Peri, G., Scaccianoce, G., 2023. Forecasting green roofs' potential in improving building thermal performance and mitigating urban heat island in the Mediterranean area: An artificial intelligence-based approach. Appl. Therm. Eng. (ISSN: 1359-4311) 222, 119879. http://dx.doi.org/10.1016/j.applthermaleng.2022.119879.

Mazzeo, D., S.Herdem, M., Matera, N., Bonini, M., Wen, J.Z., Nathwani, J., Oliveti, G., 2021. Artificial intelligence application for the performance prediction of a clean energy community. Energy (ISSN: 0360-5442) 232, 120999. http://dx.doi.org/10.1016/j.energy.2021.120999.

Merten, J., Sicot, L., Delesse, Y., Guérin de Montgareuil, A., 2008. Outdoor evaluation of the energy production of different module technologies. In: 23rd European Photovoltaic Solar Energy Conference and Exhibition, 1-5 September 2008. Valencia, Spain, (ISSN: 3-936338-24-8) pp. 2841–2845. http://dx.doi.org/10.4229/23rdEUPVSEC2008-4AV.3.23.

Mitsubishi Electric US Inc., 2023. California, United States. https://www.mitsubishielectricsolar.com. (Accessed 21 April 2023).

Neural net fitting tool, 0000. Deep Learning Toolbox, ©COPYRIGHT 1992–2004 by The MathWorks, Inc., Natick, Massachusetts, United State. https://www.mathworks.com/help/deeplearning/ref/nftool.html.

Nunes, H.G.G., Pombo, J.A.N., Bento, P.M.R., Mariano, S.J.P.S., Calado, M.R.A., 2019. Collaborative swarm intelligence to estimate PV parameters. Energy Convers. Manage. (ISSN: 0196-8904) 185, 866–890. http://dx.doi.org/10.1016/j.enconman.2019.02.003.

Ogliari, E., Dolara, A., Manzolini, G., Leva, S., 2017. Physical and hybrid methods comparison for the day ahead PV output power forecast. Renew. Energy (ISSN: 0960-1481) 113, 11–21. http://dx.doi.org/10.1016/j.renene.2017.05.063.

Orioli, A., Di Gangi, A., 2013. A procedure to calculate the five-parameter model of crystalline silicon photovoltaic modules on the basis of the tabular performance data. Appl. Energy (ISSN: 0306-2619) 102, 1160–1177. http://dx.doi.org/10.1016/j.apenergy.2012.06.036.

Panasonic Corporation, 2023. Osaka, Japan. https://na.panasonic.com. (Accessed 21 April 2023).

de la Parra, I., Muñoz, M., Lorenzo, E., García, M., Marcos, J., Martínez-Moreno, F., 2017. PV performance modelling: A review in the light of quality assurance for large PV plants. Renew. Sustain. Energy Rev. (ISSN: 1364-0321) 78, 780–797. http://dx.doi.org/10.1016/j.rser.2017.04.080.

Piccoli, E., Dama, A., Dolara, A., Leva, S., 2019. Experimental validation of a model for PV systems under partial shading for building integrated applications. Sol. Energy (ISSN: 0038-092X) 183, 356–370. http://dx.doi.org/10.1016/j.solener.2019.03.015.

Rubel, F., Kottek, M., 2010. Observed and projected climate shifts 1901–2100 depicted by world maps of the köppen-geiger climate classification. Meteorol. Z. 19, 135–141, http://koeppen-geiger.vu-wien.ac.at/shifts.htm.

Seawram, S., Nimmanterdwong, P., Sema, T., Piemjaiswang, R., Chalermsinsuwan, B., 2022. Specific heat capacity prediction of hybrid nanofluid using artificial neural network and its heat transfer application. Energy Rep. (ISSN: 2352-4847) 8 (16), 8–15. http://dx.doi.org/10.1016/j.egyr.2022.10.238.

Shahsavar, M.M., Akrami, M., Gheibi, M., Kavianpour, B., Fathollahi-Fard, A.M., Behzadian, K., 2021. Constructing a smart framework for supplying the biogas energy in green buildings using an integration of response surface methodology, artificial intelligence and petri net modelling. Energy Convers. Manage. (ISSN: 0196-8904) 248, 114794. http://dx.doi.org/10.1016/j.enconman.2021.114794.

Siddiqui, M.U., Abido, M., 2013. Parameter estimation for five- and seven-parameter photovoltaic electrical models using evolutionary algorithms. Appl. Soft Comput. (ISSN: 1568-4946) 13 (12), 4608–4621. http://dx.doi.org/10.1016/j.asoc.2013.07.005.

Siegel, M.D., Klein, S.A., Beckman, W.A., 1981. A simplified method for estimating the monthly-average performance of photovoltaic systems. Sol. Energy (ISSN: 0038-092X) 26 (5), 413–418. http://dx.doi.org/10.1016/0038-092X(81)90220-6.

Skoplaki, E., Palyvos, J.A., 2009. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. Sol. Energy (ISSN: 0038-092X) 83 (5), 614–624. http://dx.doi.org/10.1016/j.solener.2008.10.008.

Sudharshan, K., Naveen, C., Vishnuram, P., Kasagani, D.V.S.K.R., Nastasi, B., 2022. Systematic review on impact of different irradiance forecasting techniques for solar energy prediction. Energies 15 (17), 6267. http://dx.doi.org/10.3390/en15176267.

Tiwari, G.N., Mishra, R.K., Solanki, S.C., 2011. Photovoltaic modules and their applications: A review on thermal modelling. Appl. Energy (ISSN: 0306-2619) 88 (7), 2287–2304. http://dx.doi.org/10.1016/j.apenergy.2011.01.005.

Tossa, A.K., Soro, Y.M., Azoumah, Y., Yamegueu, D., 2014. A new approach to estimate the performance and energy productivity of photovoltaic modules in real operating conditions. Sol. Energy (ISSN: 0038-092X) 110, 543–560. http://dx.doi.org/10.1016/j.solener.2014.09.043.

Tossa, A.K., Soro, Y.M., Thiaw, L., Azoumah, Y., Sicot, L., Yamegueu, D., Lishou, C., Coulibaly, Y., Razongles, G., 2016. Energy performance of different silicon photovoltaic technologies under hot and harsh climate. Energy (ISSN: 0360-5442) 103, 261–270. http://dx.doi.org/10.1016/j.energy.2016.02.

Trina Solar Co., 2023. Global. https://www.trinasolar.com/en-glb. (Accessed 21 April 2023).

University of Wisconsin, 2012. Solar energy laboratory, TRNSYS 17:A transient system simulation program. http://www.trnsys.com.

Vikram Solar Limited, 2023. Kolkata, west bengal, india. https://www.vikramsolar.com. (Accessed 21 April 2023).

Wang, ven., Huang, T., 2021. Applications of Deep Learning in Biomedicine. Systems Medicine, Academic Press, ISBN: 9780128160787, pp. 29–39. http://dx.doi.org/10.1016/B978-0-12-801238-3.11507-7,

Wilberforce, T., Biswas, M., 2022. A study into proton exchange membrane fuel cell power and voltage prediction using artificial neural network. Energy Rep. (ISSN: 2352-4847) 8, 12843–12852. http://dx.doi.org/10.1016/j.egyr.2022.09.104.

Wu, L., Chen, Z., Long, C., Cheng, S., Lin, P., Chen, Y., Chen, H., 2018. Parameter extraction of photovoltaic models from measured I-V characteristics curves using a hybrid trust-region reflective algorithm. Appl. Energy (ISSN: 0306-2619) 232, 36–53. http://dx.doi.org/10.1016/j.apenergy.2018.09.

Wuxi Suntech Power Co., 2023. Wuxi, China. http://suntech-power.com. (Accessed 21 April 2023).

Xu, Y., Kong, X., Zeng, Y., Tao, S., Xiao, X., 2014. A modeling method for photovoltaic cells using explicit equations and optimization algorithm. Int. J. Electr. Power Energy Syst. (ISSN: 0142-0615) 59, 23–28. http://dx.doi.org/10.1016/j.ijepes.2014.01.017.

Xu, S., Wang, Y., 2017. Parameter estimation of photovoltaic modules using a hybrid flower pollination algorithm. Energy Convers. Manage. (ISSN: 0196-8904) 144, 53–68. http://dx.doi.org/10.1016/j.enconman.2017.04.042.

Ye, W., Herdem, M.S., Li, J.Z., Nathwani, J., Wen, J.Z., 2022. Formulation and data-driven optimization for maximizing the photovoltaic power with tilt angle adjustment. Energies 15 (22), 8578. http://dx.doi.org/10.3390/en15228578.

Yu, K., Liang, J.J., Qu, B.Y., Cheng, Z., Wang, H., 2018. Multiple learning backtracking search algorithm for estimating parameters of photovoltaic models. Appl. Energy (ISSN: 0306-2619) 226, 408–422. http://dx.doi.org/10.1016/j.apenergy.2018.06.010.

Yu, K., Qu, B., Yue, C., Ge, S., Chen, X., Liang, J., 2019. A performance-guided JAYA algorithm for parameters identification of photovoltaic cell and module. Appl. Energy (ISSN: 0306-2619) 237, 241–257. http://dx.doi.org/10.1016/j.apenergy.2019.01.008.

Zhejiang Chint New Energy Development Co. - Astronergy Solar Inc., 2023. Hangzhou shi, Zhejiang Sheng, China - Pomona CA, USA. http://energy.chint.com/en. (Accessed 21 April 2023).