Artificial Neural Network models for predicting the solar radiation as input of a concentrating photovoltaic system

C. Renno*, F. Petito, A. Gatto
Department of Industrial Engineering, University of Salerno,
Via Giovanni Paolo II, 132, 84084 Fisciano (Salerno), Italy.
*Corresponding author: crenno@unisa.it

Abstract

The energy production analysis of a system based on renewable technology depends on the inputs estimation accuracy. The solar energy is a free resource characterized by high variability; hence, its correct evaluation is a strategic factor for the feasibility of a solar system. In this paper a new methodological approach is presented in order to evaluate more accurately the electric and thermal energy production of a point-focus concentrating photovoltaic and thermal system (CPV/T). Two Artificial Neural Network (ANN) models for predicting solar global radiation and direct normal solar irradiance (DNI) are developed adopting different parameters such as climatic, astronomic and radiometric variables. In particular, a new combination of parameters is proposed in this paper and adopted first of all for the global radiation evaluation whose ANN model can be easily compared with the literature; the data are trained and tested by a multi layer perceptron (MLP). Hence, the results validation for the global solar radiation evaluation has encouraged to design an ANN model for the DNI by means of a similar variables set. The MLP network is trained, tested and validated for the hourly DNI estimation obtaining the MAPE, RMSE and $R^2$ statistical indexes values respectively equal to 5.72%, 3.15% and 0.992. Finally, the electric and thermal outputs of a point-focus CPV/T system are evaluated varying the concentration factor and cells number, and adopting as input the DNI evaluation results obtained by the ANN model presented in this paper. The CPV/T system outputs are estimated referring to the city of Salerno (Italy) under different meteorological conditions.

Key-words: artificial neural network, direct normal irradiance, global solar radiation, concentrating photovoltaic, thermal system.
1. Introduction

Renewable technologies play a relevant role in the energy production field. Their impact has become considerable in order to satisfy the energy demands of residential and industrial users [1]. Solar energy is widely recognized as the main renewable source; it constitutes a free resource largely available in the world [2]. The solar energy can be exploited by means of photovoltaic and solar systems in many applications as: demand balancing of electrical energy in national grids, reduction of environmental pollution, design and size of integrated energy systems. In particular, the concentrating photovoltaic and thermal systems (CPV/T) have been highly developed in the last years. Their main characteristic is to concentrate sunlight on a photovoltaic receiver by means of optical devices and then to decrease the solar cells area proportionally to the concentration factor (C) [3] equal to the ratio between the primary concentrator area and receiver area. High temperatures are also reached by means of the sunlight concentration [4]; so, it is necessary to cool the cells. The CPV/T systems usually adopt triple-junction cells, whose electric efficiency is less affected by the temperature increase [5]. Hence, the CPV/T systems allow the simultaneous production of electrical and thermal energy. These devices are more complex in comparison with the traditional photovoltaic systems and a standard configuration does not exist. In literature many CPV/T systems are present [6] and are different for optical [7], photovoltaic and thermal characteristics [8]. However, since the optics has to focus the sunlight on the cells, these systems can work only with the solar radiation direct component. For this reason it is basic to achieve an accurate evaluation of global and direct radiation. Many models have been developed in literature in order to evaluate the solar radiation. There are empirical [9], numerical and statistical models [10], physical models, etc., but the solar radiation prediction based on most of these models can't be accurate because of the intrinsic complexity of the problem.
The Artificial Neural Network (ANN) models are a very useful solution for problems which depend on many physical phenomena [11]. They adopt the long-term data series obtaining a higher level of reliability. Kalogirou has reported the ANN use in renewable energy systems applications [12]. Moreover, many ANNs have been developed in order to evaluate the global solar radiation. Azadeh et al. [11] estimated monthly the global solar radiation for six cities in Iran using climatic and meteorological data collected for six years. They have developed a multilayer feed-forward network which has back propagation (BP) with momentum, pruning and weight decay as training algorithm. Model inputs are: average maximum temperature, average minimum temperature, mean relative humidity, mean vapor pressure, total precipitation, mean wind speed and mean duration of sunshine. Wang et al. [13] developed two BP neural networks to evaluate the hourly global irradiance using the data of the National Renewable Energy Laboratory (NREL), collected in four years, normalized in [0.1, 0.9] and pre-processed. Levenberg-Marquardt (LM) is the training algorithm, while the network topology includes two hidden layers with 18 and 13 neurons. The transfer functions are respectively hyperbolic tangent and sigmoid. In [14] a generalized regression neural network (GRNN) has been employed to evaluate the solar radiation on tilted surface. In particular, radiometric and astronomical variables such as global solar radiation on horizontal surface, declination angles and hourly angles have been employed. Moreover, Celik et al. [14] collected data by an experimental grid-connected photovoltaic system with a tilt angle of the modules equal to the latitude of a location in Turkey; finally, the LM algorithm and cross-validation are used. In [15] a Gaussian model has been applied to evaluate the daily solar irradiance. A radial basic function network (RBF) is developed to calculate the Gaussian function amplitude. The model has been tested to use the minimum number of inputs such as the weather conditions and the duration of daylight. Finally, Amrouche and Le Pivert [16] have applied two feed-forward
ANNs with BP in order to estimate the daily global solar irradiance. The models exploit the local forecasting data; hence, the two ANNs can predict global radiation for locations where the measurements are not possible. The methodology is tested for two locations using US National Oceanic and Atmospheric Administration data.

Referring to systems that work only with the direct normal solar irradiance (DNI), the ANNs can also be used for predicting the DNI in order to estimate the amount of electrical and thermal energy produced by a CPV/T system. However, in [17] Yadav and Chandel have reviewed different ANN techniques for the solar radiation evaluation, but no techniques able to estimate the DNI have been considered. Mellit et al. [18] developed a feed-forward ANN to evaluate the hourly DNI and to compare it with an adaptive model. The network inputs are hourly temperature, humidity, sunshine duration and irradiance for the hour j, while the network output is the direct irradiance at the hour j+1; the neurons number in the hidden layer is 15. In [19] a feed-forward ANN has been applied for the clearness index evaluation of the DNI. The input data are chosen considering the functional dependence of the clearness index. The ANN has eight inputs: latitude, longitude, altitude, month of the year, local mean time, monthly mean hourly total rainfall, monthly mean hourly relative humidity, monthly mean duration of sunshine per hour. Data have been obtained by eleven stations in India.

In this paper two ANN models have been developed in order to evaluate the solar global radiation and DNI. First of all, different parameters such as climatic, astronomic and radiometric variables, have been introduced for the global radiation model. In particular, a new combination of parameters is proposed in this paper and adopted first of all for the global radiation evaluation whose ANN model can be easily compared with the literature; the data are trained and tested by multi layer perceptron (MLP). Hence, the results validation for the solar global radiation evaluation has encouraged to design the ANN
model for the DNI by means of a similar set of variables. Hence, a MLP network has been trained, tested and validated for the estimation of the hourly DNI. This is the first step to estimate the electric and thermal energy provided by a CPV/T system. Finally, in this paper a point-focus CPV/T system configuration has been also introduced and, using the DNI evaluation results, the electrical and thermal energy outputs of the system have been estimated.

2. The integrated artificial neural networks

An ANN is a mathematical tool used for a wide tasks variety: classification, data mining, pattern recognition, image compression, process modeling, etc. [11]. ANN are algorithms implemented in a computing program or electronic model based on the human brain functioning [20].

The ANN basic components are neurons which are simple processing elements interconnected and layered. Each connection among neurons is weighted and a training algorithm calculates these weights. Hence, each neuron computes a weighted sum based on the input variables values. A transfer function allows to determine the neurons output as a result of the input weighted sums. The ANN realization requires to define inputs, type of network, topology, training paradigm and transfer function. The ANN modeling allows to carry out the required output starting from corresponding input vectors without considering the assumption of any determinate relationship between the input and output.

There are many types of connection for the data transfer, the most used is the multi layer perceptron. MLP is a feed-forward ANN where data flow from the input layer to the output layer through a different number of hidden layers without any feedback loop. The hidden layers represent the network computation model core. MLP networks are able to learn complex relationships between input and output patterns, and they show a better approximation of the constants allowing a quick link between constant and non-constant
input values. The most selected learning rule for a MLP is the error Back-Propagation (BP) algorithm [11,13]. It calculates the gradient of the network error related to its modifiable weights. The BP learning approach can be implemented considering different topologies and training functions. This typical problem, when the ANN is developing, is solved by means of the cross-validation which is a validation technique to estimate how a model generalizes an independent data set. In particular, the k-fold cross-validation technique is applied to calculate the mean squared error (MSE) defining the fitting assessment. The MSE results can be used to select the best set of ANN parameters. Hence, the cross-validation is applied to choose the best model between different plausible ANN alternatives. The main choices are related to the training algorithm, the number of hidden layers and hidden neurons and the transfer functions. For the training algorithm, the BP paradigm allows to compare the gradient descent with the momentum and weight decay [11] to Levenberg-Marquardt algorithm [13, 15]. For the topology selection it is possible to consider one or two hidden layers [18], as assured by the universal approximation theorem. Finally for the transfer functions, the best solution for the output layer is a linear function [15], while, because of non-linearity problem, the best transfer function for the hidden layer is non linear. The most non linear selected functions are sigmoid or hyperbolic tangent, because their derivatives simplify the application of the BP algorithm.

3. Proposed methodology

In this section the elements of the ANNs used for the solar radiation modeling are described. The main aim of this paper is to develop an accurate model based on ANNs for predicting the hourly DNI in order to evaluate more accurately the energy performances of a point-focus CPV/T system. The DNI forecasting has been obtained using a heterogeneous variables set and adopting both experimental data and database. The radiometric climatic and astronomic parameters combination is not usually used in an
ANN to evaluate the solar radiation, even in the models which forecast the global radiation. Hence, this approach has been first adopted to estimate the global radiation whose ANN model can be easily compared with the literature. The multi layer perceptron (MLP) realized for the global radiation is validated by means of statistical parameters as, for example, the root mean squared error (RMSE) or the goodness of fit ($R^2$). So, the results validation for the global solar radiation evaluation has encouraged to design an ANN model for the DNI by means of a similar variables set. It is important to observe that this method can be used even for locations which don’t have a database previously measured. In Figure 1 the main steps of the proposed methodology are reported; this methodology can be adopted to estimate the CPV/T system potential by means of the direct solar irradiance ANN validation. In particular, the simulation values of the hourly DNI, calculated day by day, represent the CPV/T system proper input to evaluate its energy performances.

3.1. Training and validation process

The ANN designing process requires a large amount of data in order to train, to validate and to test the network. The data pre-process is needed to have a more efficient ANN training. Abnormal data have been initially excluded and then they should be normalized in order to avoid the variables measurement units and the range magnitude influence in the training phase. By means of the Matlab function “mapminmax” [21], a normalization interval [-1,1] has been defined. The data are divided into three subsets, the first represents the training set, used to compute the gradient and to update the network weights and biases by means of the training algorithm. The second is the validation set which calculates the error avoiding overfitting. The validation error decreases during the initial phase of training but, if the network data are in overfitting, it generally increases. Hence, in order to improve the network generalization capability, the early stopping technique is used. This
technique saves network weights and biases when the minimum validation error is reached.

The third data set is the test set which allows to evaluate the results in order to confirm the network predictive power and to compare different models. In this paper the ratio between the validation data amount and the whole data is 15%. Some papers [11,18] have fixed the validation set size at 10% and other at 30% [13] or 20% [20]. On the contrary, the test set size represents the 20% of the sample data. At the end of the test phase, the “mapminmax” function is also used in post-processing with the purpose to put the outputs into the original domain. The workflow for neural network designing presents several steps:

1. collecting and processing data;
2. neural network configuration;
3. initialization and training of the model using the training subset;
4. evaluation of ANN performances using the validation subset;
5. steps repetition from 2 using cross-validation;
6. network simulation and assessment obtained by cross-validation using the test subset;
7. comparison of the model results with the literature.

The models accuracy and the comparison with the literature is based on the some statistical indicators: mean absolute percentage error (MAPE), mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE) and goodness of fit ($R^2$). They are given by the following relationships:

\[
\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\hat{y}_i - y_i|}{y_i}
\]

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2
\]

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]

\[
\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|
\]

\[
R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}
\]
where \( n \) is the cardinality of dataset involved in the analysis, \( y_i \) is the variable to estimate, \( \bar{y} \) is the mean value of \( y_i \) and \( \hat{y}_i \) is the value calculated by the model. MAPE determines the accuracy and RMSE represents the standard deviation between predicted values and actual values; it is a good parameter to compare the forecasting error of different models related to the same variable. MAE measures as the predictions are close to observed values, while \( R^2 \) calculates the ratio between the variation evaluated by a regression model and the sample data variation; \( R^2 \) is an important parameter since it evaluates the general accuracy of a regression model.

3.2 ANN implementation for predicting the daily global radiation

The ANN for the global radiation estimation adopts experimental data and databases. The first have been obtained by means of a pyranometer, with a measuring and spectral range respectively equal to 0-2000 W/m\(^2\) and 335-2200 nm, and a platinum thermo-resistance. The second have been obtained by the Agro-meteorological Regional Center [22], which employs several automatic systems to measure the climatic and meteorological parameters, and by an experimental study of Italian Air Force [23]. The cross-validation allows to choose the main characteristics of the stable model: hidden layers number, hidden neurons number, transfer functions and training algorithm. The most important input parameters are air temperature and sunshine duration [18]. The sunshine duration gives an indication about the location cloudiness, that is defined as the sum of sub-period when the solar irradiance exceeds 120 W/m\(^2\) [24]; hence, it represents a radiometric variable. In order to define a model for the global radiation in different locations, latitude and longitude have to be included in the ANN inputs variables. The meteorological situation can be characterized by means of different inputs: relative humidity, vapor pressure, wind speed and precipitation [11]. The set of the input variables has been selected comparing two groups of heterogeneous parameters. At first a set of seven variables such as mean temperature
(1), sunshine duration (2), latitude (3), longitude (4), precipitation (5), vapor mean pressure (6) and wind mean speed (7), has been used. After training, validation and testing, wind speed and vapor pressure have been found to be the least incident. Hence, they have been replaced by two astronomic variables, unconventional in this kind of problem: daylight hours and declination angle. The first gives relevant information about the period of the year, the location considered and the cloudiness, especially if it is compared with the sunshine duration. The second defines the specific day considered. They have been evaluated by means of analytic expressions [25]. This new set of input allows an improvement of 33% in terms of MSE. Hence, the chosen input variables are: latitude, longitude, mean temperature, sunshine duration, total precipitation, daylight hours, declination angle. Consequently, the daily global solar radiation is a function of these parameters:

\[ G_{\text{global}} = f(L_t, L_g, T, SD, P, H, \delta) \quad (6) \]

The Equation 6 shows that the implemented ANN for predicting daily global solar radiation, takes into account of the different variables selected. These parameters constitute the input vectors for the ANN output estimation. Referring to the training algorithm, Levenberg-Marquardt and the gradient descent with momentum and weight decay have been compared. The cross-validation results show that the first algorithm is more precise and converges also quickly. On the contrary, the gradient descent uses much more time to reach the minimum point. Levenberg-Marquardt curve-fitting method is used to solve nonlinear least squared problems. These problems arise when a parameterized function has to fit in with a set of measured data points, minimizing the MSE function [26]. The transfer functions establish how the neurons carry information to the following layer. A sigmoid function has been selected as transfer function of the hidden layer comparing it with the
The number of hidden layers has been evaluated analyzing the solution with one or two hidden layers. The performance has been better with only one hidden layer and, then, the number of hidden layers certainly increases the computations but not the performances [18,19]. Finally, the cross-validation has also been employed in order to determine the ideal number of hidden neurons. The selection of the correct number of hidden neurons allows to balance the prediction performances and to avoid the overfitting.

The proposed neural network calculates the outputs as follow:

\[ y_k = f \left\{ \sum_{j=1}^{10} w_{ij} g \left( \sum_{i=1}^{n} p_{ji} x_{ik} + b_i \right) \right\} + a \]  \text{con } |K| = \text{No. patterns}  \tag{7}

where \( f \) is the linear function, \( g \) is the sigmoid function, \( p_{ji} \) is the weights matrix of hidden neurons \( j \) and input neurons \( i \), \( w_j \) is the vector of weights referred to the hidden neurons \( j \) and the output neuron, \( x_{ik} \) is the matrix of input, \( b_i \) is the vector of hidden neurons biases, \( a \) is the value of the output neurons bias, \( y_k \) is the value of the output for \( k \)-th day.

3.3 ANN implementation for predicting the hourly direct solar irradiance

The DNI forecast is essential in order to assess the potential of a CPV/T system. The DNI for a specific location is often calculated starting from the global irradiance data registered in the most radiometric stations. It is estimated by means of the decomposition model based on the regression between two indices: the clearness index \( k_t \) (horizontal global irradiance/horizontal extraterrestrial irradiance) and the direct solar transmittance \( k_b \) (direct normal irradiance/extraterrestrial irradiance) [27]. The relation between these indices is very complex and the traditional statistical methods are not efficient. Due to the problem of non-linearity, an ANN to evaluate the hourly DNI has been developed following the methodology introduced for the global radiation assessment. The data have been obtained between 2013 and 2014 by means of an experimental system located in Salerno (40.7730° N, 14.7901° W) with a sampling interval of one hour. The training data are referred to ten
months, while the validation and the test subsets are respectively of two and three months. There are not many studies about the DNI evaluation with the ANN; hence, a set of unconventional variables has been introduced. A MLP network has also been employed in order to have a quick link between constant and non-constant values, because the direct irradiance can suddenly change. The astronomic and radiometric input parameters chosen are: clearness index, declination angle, hour angle and global normal irradiance and

\[ DNI = f(k_v, \delta, HRA, G_{gn}) \]  

The Equation 8 reports the parameters that affect the hourly DNI estimation by means of the implemented ANN. These variables constitute the input neurons to compute the required output. The clearness index is the most relevant input; it can measure the filtering action of the atmosphere. The declination angle and hour angle allow to evaluate the output taking into account the specific day considered and its sunlight duration. Some climatic variables such as air temperature, dew-point temperature, precipitation and wind speed are two or three orders of magnitude less important than the relative air mass and clearness index [27]. Moreover, the hour angle can replace the relative air mass because the optical path length through the atmosphere is conditioned by the hour of the day. Finally, the global normal irradiance includes all the indications about meteorological and climatic parameters involved in the processes. The irradiance values considered are referred to a tilted surface; hence, they also trace the tilt and azimuth angle effects. As in the previous case, comparing Levenberg-Marquardt and the gradient descent with momentum and weight decay as training algorithm, the first method reaches convergence adopting less time. The cross-validation has allowed to chose one hidden layer with five neurons, while as transfer function for the hidden layer the sigmoid function has been selected comparing it with the hyperbolic tangent. The proposed neural network, focused on the DNI, calculates outputs with an expression similar to the one used for the global radiation:
In order to adopt the results of the DNI evaluation, in the following sections the point-focus CPV/T system considered is described together with the main equations.

### 4.1 CPV/T system description

A CPV/T system, differently from a CPV system that allows only the electric energy production, presents an active mechanism able to recover also thermal energy. Its main characteristic is to concentrate sunlight in order to increase the incident direct solar radiation and to decrease the photovoltaic area. Hence, the solar radiant power for photovoltaic area unit can be modified by means of optical devices varying the concentration factor:

\[
C = \frac{A_{\text{opt}}}{A_c} \cdot \eta_{\text{opt}}
\]  

(10)

where \(A_{\text{opt}}\) and \(A_c\) represent respectively the optics and cell area and \(\eta_{\text{opt}}\) the optical efficiency which depends on the optic device adopted. A CPV/T system presents three principal elements: receiver, optics and sun tracking system. The photovoltaic cells, usually triple-junction ones, and their active cooling mechanism are integrated into the receiver. The sunlight is selectively absorbed by different layers that constitute the triple-junction cells which present an open-circuit voltage that increases logarithmically with \(C\), [28]. Moreover, the concentration on triple-junction cells determines a temperature increase and a lower percentage decrease of the open circuit voltage, and then efficiencies of about 40% can be experimentally obtained [29]. For \(C\) values above 100, the cooling system allows both to preserve the cell electric efficiency and to obtain thermal energy. The optics allows to focus the sunlight on the receiver by means of a refractive (lens) or a reflective (mirrors) device. The refractive optics are cheaper than the reflective ones, but they have a lower performance mainly due to chromatic aberration problems [30]. Hence,
higher values of optical efficiency of about 90% can be reached adopting a reflective optics
as function of their transmission and reflectivity coefficients [31]. A CPV/T system, whose
size depends on the user energy demands, presents different configurations: point-focus,
line-focus and dense array.

In this paper a point-focus CPV/T system has been considered. In particular, a modular
configuration has been investigated in order to estimate the thermal and electrical energy
production as function of the DNI forecast by the ANN model. Hence, a basic scheme,
which can be increased or reduced as function of the user energy demands, is considered in
order to meet the electric, thermal and cooling needs. The point-focus basic configuration,
shown in Figure 2, consists of a small parabolic dish concentrator with InGaP/InGaAs/Ge
triple-junction solar cells placed in its focus. This structure can be replied in series building
a row and, then, in more rows in parallel that constitute the overall module. The cells are
arranged on a pipe where the cooling fluid (water–glycol solution) flows. The CPV/T
system is also equipped with a two-axis tracking system. In particular, the concentrated
sunlight is converted simultaneously into electricity and into thermal energy by means of
respectively the PV layer and the cooling fluid. The cells present an area of 1x1 cm², while
the mirror area, assuming a constant value of the optical efficiency, varies according to the
necessary C. Hence, in order to satisfy the user's energy demands, the proposed CPV/T
system can change its size according to the C value and the cells number. Once defined the
CPV/T system parameters, this system can be divided into different modules. Each module
presents a circulation pump for the refrigerant fluid and a thermostatic valve which sends
the fluid to a hot water storage once reached the temperature required. The tank allows to
meet different thermal energy user demands. The hot water can be employed for heating or
cooling needs by means of an absorber heat pump; the thermal storage is also equipped
with a boiler able to integrate the thermal energy (Figure 2). The CPV/T system is also
grid-connected and, then, it is possible to integrate the electricity from the network and to
give the surplus energy back to it. The proposed CPV/T system has a very flexible
configuration, whose parameters can be modulated in order to evaluate different solutions
from an energetic point of view [32]. The main aim of this paper is to estimate the CPV/T
system energy potential adopting the accurate DNI obtained by the ANN models.

4.2 CPV/T energy model

The CPV/T system energy outputs are strongly influenced by a set of heterogeneous
variables which depend on both external conditions and internal working parameters [33].
The first group represents the non controllable input conditions such as installation site,
direct solar radiation, atmospheric conditions. The ANN model for predicting the DNI
allows to use a more accurate input in order to estimate the CPV/T system energy
potential, once defined the installation site. Hence, the direct hourly DNI evaluated by
means of the ANN tool, can be integrated hourly and daily in order to assess the CPV/T
system daily, monthly and annual results. The second group of variables indicates the
working conditions values such as cells number, C, cell temperature and efficiency which
affect the CPV/T system working. As previously indicated, the cells number and C have
not been fixed in the model in order to investigate different CPV/T system sizes. The
CPV/T system model has been implemented in Matlab [21] since it can be easily interfaced
with the ANN toolbox employed. The model general assumptions adopted are: steady
state, radiation uniformly concentrated along multi-junction cells area and negligible
temperature gradients between cells and their substrate. Once defined the analysis temporal
level (hourly, daily, monthly or annual) the incident direct radiation for each triple junction
cell can be evaluated as:

\[ I_{dir} = (DNI_{int} \cdot sf) \cdot C \cdot A_c \cdot \eta_{opt} \] (11)
where DNI_{int} indicates the direct solar radiation integrated as function of the temporal level, C is the chosen concentration factor and A_c is the InGaP/InGaAs/Ge triple-junction cell area. The optical efficiency (\eta_{opt}) has been considered constant and equal to 0.865 considering the parabolic concentrator adopted [34]. The coefficient s_f represents a loss factor, whose value has been fixed at 0.9, which takes into account a non ideal tracking system. Hence, the electric energy production for each cell is:

\[ E_{el,c} = I_{dir} \cdot \eta_c \]  

The cell efficiency (\eta_c) constitutes a key parameter for the CPV/T electric energy evaluation; it is strongly interconnected to the cell operating temperature:

\[ \eta_c - \eta_{ref} = \sigma_t \cdot (T_c - T_{ref}) \]  

where T_{ref} is the reference temperature equal to 25°C and \eta_{ref} is the reference efficiency corresponding to the reference temperature value, obtained by some experimental curves [35-36]. The temperature coefficient \sigma_t indicates the efficiency percentage reduction as function of the temperature increase; according to the cell manufacturer instructions [37] its value has been set at -0.04 %/°C in a range 10°C/100°C.

The cell temperature (T_c) depends on many variables during the system operating, it can be approximately evaluated as [38]:

\[ T_c = T_{ref} + \frac{V_{oc}(T_c,C) - V_{oc}(T_{ref,C_0})}{\beta(C)} \]  

where V_{oc} (T_c, C) is the open circuit voltage function of the cell temperature and C, V_{oc} (T_{ref}, C_0) is the open circuit voltage function of the reference temperature and C equal to 1, \beta(C) is the tension thermal coefficient. It is possible to linearize [39], by means of some experimental results, the variables as a function only of C [40]:

\[ T_c = T_{ref} + \frac{V_{oc}(C) - V_{oc}(C_0)}{\beta(C)} \]  

Hence, considering a module composed of a variable cells number, the CPV/T system electric energy production can be estimated as:
where the module efficiency ($\eta_{\text{mod}}$) until 100 cells is equal to 0.9, $p_{\text{par}}$ is a loss factor depending on the radiation and is equal to 0.023 [34], $N_c$ represents the cells number which constitute the module and $\eta_{\text{inv}}$ is the inverter efficiency.

The system proposed in this paper allows to obtain electric energy but also to recover thermal energy from the triple-junction cells layer. Hence, the incident solar direct radiation $I_{\text{dir}}$, previously calculated, can be partially converted into thermal energy [36]:

\[
E_{\text{th,CPV/T}} = [(1 - \eta_{\text{el,CPV/T}}) \cdot I_{\text{dir}} \cdot N_c] - E_{\text{th,loss}}
\]

where the CPV/T system total electric efficiency takes into account the cells and module efficiency:

\[
\eta_{\text{el,CPV/T}} = \eta_c \cdot \eta_{\text{mod}} \cdot (1 - p_{\text{par}})
\]

The CPV/T system energy thermal losses $E_{\text{th,loss}}$ consider convective and radiative thermal losses related both to the front and to the back part of the system [41]. In particular, under the assumption of water circuit thermal insulation and considering the small active surface of multi-junction cells of about 1 cm$^2$, a low value of the thermal losses [34] has been evaluated by means of the following equation:

\[
E_{\text{th,loss}} = [\bar{h_c} \cdot (T_c - T_{\text{ref}}) + \varepsilon_c \cdot \sigma_{\text{ST}} \cdot (T_c^4 - T_{\text{ref}}^4)] \cdot A_c \cdot N_c
\]

where $\varepsilon_c$ is the cell emissivity equal to 0.85.

### 5. Results and discussion

The ANN modeling considers some specific steps: validation, implementation and simulation. In the validation step the network arrangement is evaluated considering its performances by means of data not treated during the training. In the implementation step a computing tool is adopted in order to obtain a stable network. In the simulation step the ANN model is applied to the test data set. In order to obtain an accurate model of a CPV/T system an input as the DNI provided by the ANN is required. Hence, the CPV/T system
model implementation allows to estimate the electric and thermal energy production when the system configuration is modified varying C and the cells number for each day or month.

5.1 ANNs for predicting daily global radiation and hourly direct irradiance

Different iterative steps have been developed in order to design the ANNs and to obtain the daily global radiation and the hourly DNI. The forecasting networks capabilities have been evaluated using dataset not considered in the training step of the networks. As shown in Figure 3, the ANN training process for predicting the global radiation has been implemented by means of data of four different stations near Salerno. The sampling data, collected in terms of latitude and longitude, have allowed to reach the best performances for Salerno. On the contrary, the training validation and simulation processes of the ANN to estimate the DNI have considered only experimental data, collected in one year for Salerno. The neural networks have been implemented using Matlab ANN-toolbox [21].

Many ANN configurations have been compared according to MSE in order to select the best networks features to evaluate the daily global radiation and the hourly DNI. Referring to the global radiation, the cross-validation results with alternative ANNs structures are presented in Table 1. The related MSE values allow to validate the network characteristics chosen in the designing process. The different solutions compared have been determined varying the network parameters one by one in order to isolate the specific feature effect. All the values are referred to normalized data and the best solution shows a MSE value equal to 0.001492. Similarly, in Table 2 alternative features for the ANN of the DNI are presented. A recurrent MSE value of 0.0009407 has been observed for the best ANN characteristics. In the Figures 4a and 4b the proposed ANNs structures for global radiation and DNI are shown. The ANN for predicting the global radiation presents an input layer of seven neurons; inputs are added using the weights matrix (p_{ij}) and biases vector (b_j) in
order to define the transfer function input of the hidden layer. The sigmoid function outputs are added using the output weights vector \( w_j \) and the output bias \( a \); finally, the model output is calculated by means of the output layer transfer function. The ANN representation for the DNI shows four input neurons and a hidden layer with five neurons.

The simulation stage for predicting the global radiation has been conducted in Salerno adopting experimental data referred to fifteen days of October and database values of March, July and November, because they present different climatic conditions. The DNI simulation step has only considered the experimental data of October, November and December. In the Figures 5a and 5b the global radiation and the DNI predicted trends, obtained by Matlab, have been compared with the measured data. According to these figures, the good agreements achieved between previsions and measured values are clearly proved. Hence, in Figure 5a the ANN model estimates with high correlation the global radiation referring to the different months considered. In Figure 5b, where a 180 values sample has been shown, the forecast capabilities of the hourly DNI seems even better than the daily global radiation forecast; this is due to the correspondence between sampled data trend and simulated values trend. Moreover, the better forecast capabilities of the hourly DNI networks are due to the relevance of the input variables \( G_{gn} \) in the predicting process. In fact, this represents a key radiometric parameter in the DNI evaluation. Therefore its use as input variable constitutes a strategic choice. Some statistical indicators have been calculated in order to validate these results and the scatter plots are shown in the Figures 6a and 6b. The scatter plots provide important indications related to the correlation between measured and predicted data. Figures 6a and 6b show both for the global radiation (Figure 6a) and the DNI (6b) points close to the linear function that represents the perfect correspondence between predicted and measured data. As reported in Figures 6a and 6b, daily global radiation and hourly DNI lower values are also well forecasted. This
represents a direct consequence of an important feature of the presented model: the input set selection. The balanced and heterogeneous set of radiometric, meteorological and astronomic variables have guaranteed a good prediction for lower values in changeable conditions. Hence, the ANN model can well predict situations of low irradiance as function of the input variable values.

The networks performances assessment has been quantified adopting some goodness test. The ANN statistical results for predicting the global radiation are shown in Table 3. The overall performances indicate a MAPE of about 4% and a MAE value of 117.2 Wh/m², which determine a low value of absolute error if compared with the mean daily solar radiation equal to 3732 Wh/m². The MAE and RMSE values are higher in July which represents the month with the highest daily global radiation. However, the model accuracy is good; hence, July shows the best MAPE results. Therefore, related to low-radiation months, such as November, or to months with high weather conditions variation, such as March, the performances are lower despite good RMSE and MAE values. Finally, the global solar radiation forecast based on these seven inputs show from a statistical point of view a high correlation coefficient represented by a $R^2$ value equal to 0.991. On the contrary, the ANN for predicting the DNI presents MAPE, RMSE, MAE and $R^2$ equal respectively to 5.72%, 3.15%, 11.6 W/m² and 0.992. These values guarantee correlation and good accuracy to predict the DNI, using astronomical and radiometric variables.

5.2 Literature comparison

The ANN models performances have been compared with some results presented in literature in terms of statistical indicators. This analysis allows an external validation of the approach used to develop the ANNs for predicting global radiation and DNI. In literature, the global radiation estimation in the energy applications is widely realized with ANNs, while the DNI prediction is not treated with the same spread. Hence, different models
present in the literature have been used in order to validate the proposed ANN models. In Table 4 the global radiation statistical parameters calculated in this paper are compared with some literature models. Azadeh et al. [11] show six different values for each statistical indicator because the prediction has been conducted for six different cities; in Table 4 the minimum and the maximum values have been considered. The MAPE of the proposed model is higher than the minimum value of Azadeh. With regard to RMSE, the 3.46 % value obtained by the model proposed in this paper is very close to results of Wang’s two models. The models differ because the first considers as input the solar irradiance from 6:00 am to 8:00 pm, while the second takes into account the 24 hours data [13]. It is interesting to underline that Wang et al. adopt a MLP with Levenberg-Marquardt algorithm, like the ANNs presented in this paper. Finally, the proposed model shows an excellent value of fit goodness ($R^2$) ensuring high correlation, especially if compared with Zervas et al. model [15], where the inputs number have been reduced expressing the weather conditions by means of a value referred to a scale from 1 to 6. According to Table 4, the comparison presented provides good results. In Table 4, the statistical indicators calculated for the hourly DNI modeling are also shown. The analysis has been developed comparing the RMSE achieved with the model of Kaushika et al. [19] and the $R^2$ value with Mellit et al. [18]. The proposed RMSE value is lower than the value presented in [19], although the model developed uses only four inputs while Kaushika et al. present eight inputs including geographic information. Moreover, in that network the samples data are divided into 70%, 15% and 15% respectively for training, validation and testing, and the radiometric and climatic variables have also been employed as in the ANN proposed. According to the $R^2$, Mellit et al. have implemented a feed-forward neural network less accurate than ANN proposed. Finally, the comparison between literature and proposed models has allowed to validate the results with a good agreement. Hence, both the ANNs
presented in this paper have shown high performances comparable with the outputs presented in literature.

5.3 CPV/T system energy evaluation

The presented model allows to estimate accurately the electric and thermal energy production of a CPV/T system that adopts the ANN outputs for predicting the hourly DNI. The CPV/T system energy production has been estimated referring to different temporal level (daily, monthly, annual) by means of the hourly results integration. While the daily energy production is affected by the input variability, the monthly or annual values give a relevant information referring to the CPV/T system feasibility studies, above all where measurements are not possible. The tested ANN model, with its heterogeneous set of inputs, can accurately estimate the DNI in each location. Hence, it is possible to identify the main technical characteristics of a CPV/T system efficient configuration. In particular, the principal simulation process variables are: installation site, cells number and size, optics system and C. A CPV/T system point-focus configuration with parabolic mirror and triple junction cells of 1 cm² has been considered in the simulation model. In the Figures 7a and 7b the annual electric and thermal energy production in Salerno are respectively reported for a 20 cells module and for different C values; the electric and thermal energy increase linearly and logarithmically respectively with cells number and C. A production increase of 59% and 69% respectively for electric and thermal energy has been observed when C increases between 300 and 500; the energy production percentage increase for C values higher than 500 decreases. In particular, the electric and thermal energy production increases are respectively 33% and 42% in the range 500 and 700 of C, and 23% and 30% when C varies between 700 and 900. Moreover, the thermal energy is about four times higher than the electric one and both follow the seasonal trend. The model presented is not only able to evaluate the energy production for each location, but even to distinguish the
electric and thermal hourly power variability. The Figures 8a and 8b show the electric and thermal hourly power for a cloudy, average irradiance and sunny day of January at Salerno considering 60 cells and a C value equal to 800. In the Figures 9a and 9b the same configuration has been employed in order to evaluate the hourly power in Salerno under the different climatic conditions of July. Each day considered in the analysis has been characterized by low climatic variations, because the purpose is to highlight the impact of different meteorological conditions on the energy production. The mean clearness index values of January days are respectively 0.0171, 0.110, 0.219 moving from the cloudy to the sunny day; for July the values are 0.0696, 0.297, 0.353. As expected, the high irradiance days are characterized by high clearness index values. In the worst case, a low irradiance day of January, the total electric and thermal daily energy productions are respectively equal to 0.11 kWh and 0.39 kWh. In the best case, a high irradiance day of July, 4.51 kWh and 20.07 kWh respectively of electric and thermal energy have been obtained. It is possible to note that the difference in terms of light hours between January and July is about 6 hours. The thermal and electric energy variation evaluation is fundamental in order to size a CPV/T system, because it also involves several financial and economic aspects. The electric and thermal energy variability for a CPV/T system in Salerno is shown in the Figures 10a and 10b. In particular, the daily average values have been calculated considering the electric and thermal hourly energy production for each day. The electric and thermal average energy production trend is included in the range of the deviation standard. Thus, the upper curve represents $\mu + \sigma$ while the lower is $\mu - \sigma$, where the parameters $\mu$ and $\sigma$ are the average values and the deviation standard of the hourly energy production in a day. As above said the simulation has considered 60 cells and a C value equal to 800. The standard deviation is included between 34% and 56% if compared to the
electric and thermal energy average production. March and September result the months with the highest variability, while December with the least.

6. Conclusions

In this paper two ANNs based models have been developed in order to forecast the global solar radiation and the hourly DNI. The proposed methodology has adopted a set of heterogeneous variables considering both experimental data and databases. The ANN validation related to the global radiation results has allowed to exploit the same approach for the DNI predicting that has been obtained using radiometric, climatic and astronomic parameters. The MLP realized has been able to predict the DNI for different locations thanks to the chosen input set. The validation tests have been performed using data measured for more than a year in Salerno (Italy) and have shown a good agreement between measured and predicted values for all the period considered. In particular, they have shown that the ANN models presented can estimate daily global radiation and DNI with satisfactory accuracy. Hence, adopting the ANN model results related to the DNI, the point-focus CPV/T system potential has been analyzed. The hourly DNI simulation values, day by day, have represented the CPV/T system model input which estimates the energy production according to the system configuration. The possibility of an accurate hourly solar input for the CPV/T system allowed to consider a more realistic production considering the sudden climatic changes. So, a RMSE of 3.15% for the ANN prediction allowed to realize an input profile for the CPV/T system according to different DNI distributions principally depending on the cloudiness. The CPV/T system configuration with sixty cell and C equal to 800, shows for Salerno a global electric and thermal daily production respectively equal to 0.11 kWh and 0.39 kWh in the worst case, and to 4.51 kWh and 20.07 kWh in the best situation. The main originality of this methodology is the ANNs models use with a new combination of input variables in order to forecast accurately
the CPV/T system configuration potential. This important feature leads to assess the system feasibility even where measured data are not available.

Nomenclature

A area (m²)
a output layer bias
ANN Artificial Neural Network
b_j vector of hidden layer biases
BP back propagation
C concentrating factor
CPV/T concentrating photovoltaic and thermal
DNI direct normal irradiance (W/m²)
E energy (kWh)
f output layer transfer function
G global radiation (Wh/m²)
g hidden layer transfer function
H daylight hours (h)
\bar{h}_c mean heat transfer coefficient (W/m² K)
HRA hour angle (°)
I incident direct radiation (kWh)
InGaP/InGaAs/Ge indium-gallium-phosphide/indium- gallium-arsenide/germanium
kb direct solar transmittance
kt clearness index
Lg longitude (°)
LM Levenberg-Marquardt
Lt latitude (°)
MAE  mean absolute error

MAPE mean absolute percentage error (%)

MLP multilayer perceptron

MSE mean squared error

N\text{cell} number of cells

NREL national renewable energy laboratory

p_{ij} array of hidden layer weights

p loss factor

P precipitation (mm)

PV photovoltaic

R2 goodness of fit

RMSE root mean squared error

RTD resistance temperature detector

SD sunshine duration (h)

sf safety factor

T temperature (°C)

V voltage (V)

w_j vector of output layer weights

x input array

n cardinality of dataset

y variable to estimate

y^{-} mean value of the variable to estimate

y^{\hat{}} estimated value of the variable to estimate

\mu stochastic variable mean value

\textbf{Greek symbol}
\[ \beta \text{ tension thermal coefficient (V/°C)} \]
\[ \delta \text{ solar declination angle (°)} \]
\[ \varepsilon \text{ emissivity coefficient} \]
\[ \eta \text{ efficiency} \]
\[ \sigma_{ST} \text{ Stefan–Boltzmann constant (W/m}^2 \text{ K}^4) \]
\[ \sigma \text{ stochastic variable standard deviation} \]
\[ \sigma_t \text{ temperature coefficient (%/°C)} \]

**Subscripts**

- \( oc \) open circuit
- \( c \) cell
- \( dir \) direct
- \( el \) electric
- \( gn \) global normal
- \( int \) integrated
- \( inv \) inverter
- \( o \) environment
- \( opt \) optic
- \( par \) parasitic losses
- \( ref \) reference
- \( th \) thermal

**References**


Triple-Junction Solar Cell for Terrestrial Applications, CTJ Photovoltaic Cell – 10 mm x 10 mm, Datasheets Emcore September 2012, Emcore Corporation.


Figure captions

Figure 1 Methodology flow-chart

Figure 2 CPV/T system scheme for different user energy demands

Figure 3 Geographic positions of the different stations for ANN global radiation training

Figure 4 Structure of the proposed neural network to estimate: (a) global radiation, (b) DNI.

Figure 5 Comparison between measured and predicted values: (a) daily global radiation, (b) hourly direct solar irradiance.

Figure 6 Scatterplot of measured and predicted values: (a) daily global irradiance, (b) hourly direct irradiance.

Figure 7 Daily production of the CPV/T system with 20 cells as function of C: (a) electric production, (b) thermal production.

Figure 8 Hourly power of a CPV/T system with 60 cells and C=800 in a day of January: (a) electric power, (b) thermal power.

Figure 9 Hourly power of a CPV/T system with 60 cells and C=800 in a day of July: (a) electric power, (b) thermal power.

Figure 10 Daily average variability of the hourly energy production of a module with 60 cells and C=800: (a) electric energy, (b) thermal energy.

Table captions

Table 1 ANN characteristics for global radiation predicting as function of the MSE values

Table 2 ANN characteristics for solar direct irradiance predicting as function of the MSE values

Table 3 Values of the statistical indicators for the proposed global radiation model

Table 4 Comparison between proposed and literature models for predicting global radiation and direct irradiance
<table>
<thead>
<tr>
<th>ANN Parameter</th>
<th>Alternatives</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training function</td>
<td>GD with momentum and weight decay</td>
<td>0.004578</td>
</tr>
<tr>
<td></td>
<td>Levenberg-Marquardt</td>
<td>0.001492</td>
</tr>
<tr>
<td>Transfer function</td>
<td>Tanh-Linear</td>
<td>0.002524</td>
</tr>
<tr>
<td></td>
<td>Sigmoid-Linear</td>
<td>0.001492</td>
</tr>
<tr>
<td>No. Hidden Layers</td>
<td>2</td>
<td>0.003164</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.001492</td>
</tr>
<tr>
<td>No. Hidden neurons</td>
<td>7</td>
<td>0.009610</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>0.001492</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>0.004313</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>0.024830</td>
</tr>
</tbody>
</table>

Table 1 ANN characteristics for global radiation predicting as function of the MSE values
<table>
<thead>
<tr>
<th>ANN parameter</th>
<th>Alternatives</th>
<th>Mean Squared Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training function</td>
<td>GD with momentum and weight decay</td>
<td>0.0796600</td>
</tr>
<tr>
<td></td>
<td>Levenberg-Marquardt</td>
<td>0.0014920</td>
</tr>
<tr>
<td>Transfer function</td>
<td>Tanh-Linear</td>
<td>0.0013540</td>
</tr>
<tr>
<td></td>
<td>Sigmoid-Linear</td>
<td>0.0009407</td>
</tr>
<tr>
<td>No. Hidden Layer</td>
<td>2</td>
<td>0.0009421</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.0009407</td>
</tr>
<tr>
<td>No. Hidden neurons</td>
<td>6</td>
<td>0.0031000</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>0.0009407</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>0.0010370</td>
</tr>
</tbody>
</table>

Table 2 ANN characteristics for solar direct irradiance predicting as function of the MSE values
<table>
<thead>
<tr>
<th></th>
<th>MAE [Wh/m²]</th>
<th>RMSE [Wh/m²]</th>
<th>MAPE [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall</td>
<td>117.2</td>
<td>135.2</td>
<td>4.17</td>
</tr>
<tr>
<td>March (database)</td>
<td>134.8</td>
<td>144.9</td>
<td>4.98</td>
</tr>
<tr>
<td>July (database)</td>
<td>128.2</td>
<td>152.6</td>
<td>2.06</td>
</tr>
<tr>
<td>November (database)</td>
<td>104.3</td>
<td>121.1</td>
<td>5.86</td>
</tr>
<tr>
<td>October (experimental)</td>
<td>85.3</td>
<td>98.4</td>
<td>3.48</td>
</tr>
</tbody>
</table>

Table 3 Values of the statistical indicators for the global radiation proposed model
### Literature Comparison

#### ANN for global radiation

<table>
<thead>
<tr>
<th>Model</th>
<th>MAPE [%]</th>
<th>RMSE [%]</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>4.17</td>
<td>3.46</td>
<td>0.991</td>
</tr>
<tr>
<td>Azadeh et al. [11]</td>
<td>3.00 - 11.0</td>
<td>2.60 - 5.20</td>
<td>0.980 - 0.860</td>
</tr>
<tr>
<td>Wang et al. [13]</td>
<td>-</td>
<td>3.31 - 4.50</td>
<td>0.991 - 0.964</td>
</tr>
<tr>
<td>Zervas et al. [15]</td>
<td>-</td>
<td>-</td>
<td>0.985</td>
</tr>
<tr>
<td>Celik et al. [14]</td>
<td>-</td>
<td>-</td>
<td>0.987</td>
</tr>
</tbody>
</table>

#### ANN for direct irradiance

<table>
<thead>
<tr>
<th>Model</th>
<th>MAPE [%]</th>
<th>RMSE [Wh/m²]</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>5.72</td>
<td>14.2</td>
<td>0.992</td>
</tr>
<tr>
<td>Kaushika et al. [19]</td>
<td>-</td>
<td>14.5</td>
<td>-</td>
</tr>
<tr>
<td>Mellit et al. [18]</td>
<td>-</td>
<td>-</td>
<td>0.967</td>
</tr>
</tbody>
</table>

Table 4 Comparison between proposed and literature models for predicting global radiation and direct irradiance
Figure 2

Diagram of a CPV/T system with a boiler, hot water storage, refrigeration demands, thermal demands, and electrical demands. The system includes a concentrator, cell, and hot/cold fluid paths.
Figure 4

INPUT LAYER

x0 = 1
x1: Latitude
x2: Longitude
x3: Temperature
x4: Sunshine duration
x5: Precipitation
x6: Declination angle
x7: Daylight hours

HIDDEN LAYER

INPUT LAYER

a

OUTPUT LAYER

Global Radiation

INPUT LAYER

x0 = 1
x1: Declination angle
x2: Hour angle
x3: Cleanliness index
x4: Global normal irradiance

HIDDEN LAYER

OUTPUT LAYER

DNI
Figure 5

(a) Daily global solar radiation [W/m²]

(b) Hourly direct solar irradiance [W/m²]
Figure 6

(a) Predicted global solar radiation vs. Measured global solar radiation [Wh/m²]

(b) Predicted direct solar irradiance vs. Measured direct solar irradiance [W/m²]
Figure 9

(a) CPV/T hourly electric power [kW]

(b) CPV/T hourly thermal power [kW]
Figure(s)

Click here to download high resolution image