Accurate thermal prediction model for building-integrated photovoltaics systems using guided artificial intelligence algorithms

L. Serrano-Luján[a], C. Toledo[b,c,*], J. M. Colmenar[a], J. Abad[c], A. Urbina[c]

[a] *Department of Computer Science and Statistics, Rey Juan Carlos University (URJC), Calle Tulipán S/N, 28933 Móstoles (Madrid), Spain*
[b] *Italian National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA). Centro Ricerche Portici, largo Enrico Fermi 1 80055, Portici (NA), Italy*
[c] *Departments of Electronics and Applied Physics, Technical University of Cartagena (UPCT). Plaza del Hospital 1, 30202 Cartagena (MU), Spain.*

Abstract

Progress in development of building-integrated photovoltaic systems is still hindered by the complexity of the physics and materials properties of the photovoltaic (PV) modules and its effect on the thermal behavior of the building. This affects not only the energy generation, as its active function and linked to economic feasibility, but also the thermal insulation of the building as part of the structure's skin. Traditional modeling methods currently presents limitations, including the fact that they do not account for material thermal inertia and that the proposed semi-empirical coefficients do not define all types of technologies, mounting configuration, or climatic conditions. This article presents an artificial intelligence-based approach for predicting the temperature of a PV module based on local outdoor weather conditions (ambient temperature, solar irradiation, relative outdoor humidity and wind speed) and indoor comfort parameters (indoor temperature and indoor relative humidity) as inputs. A combination of two algorithms (Grammatical Evolution and Differential Evolution) guides to the creation of a customized expression based on the Sandia model. Different data-sets for a fully integrated PV system were tested to demonstrate its per-

*L. Serrano-Luján: lucia.serrano@urjc.es; C. Toledo: agrivoltaics.project@enea.it

formance on three different types of days: sunny, cloudy and diffuse showing relative errors of less than 4% in all cases and including night time. In comparison to Sandia model, this method reduces the error by up to 11% in conditions of variability of sky over short time intervals (cloudy days).

Keywords:
BIPV, PV module temperature, Module temperature estimation, Grammatical evolution, Differential evolution, Machine learning

---

1. Introduction

Photovoltaic (PV) technology is considered a mature technology, with installed capacity worldwide growing exponentially in the past decades and reaching 760.4 GW_p at the end of 2020 [1]. Although 2020 could see a slight decline in the growth rate due to COVID19 negative impact, it had nevertheless added an extra 140 GW_p in 2020, and it is expected that the TeraWatt milestone could be reached in 2024 according to the International Energy Agency forecast [2]. Similarly, the global 9% (2020) share of installed capacity for the generation of electricity by photovoltaic systems continues to increase, as well as the electricity production yield of this installed capacity, which is improving due to advances in the technological optimization of the PV systems. Those advances have led to a very good capacity factor, in the range of more than 25% for many countries. One of the main impacts on the performance of operating systems are the temperature losses, which arise from the high temperatures that the modules may reach when irradiance and ambient temperature are high; any module operating at temperature above T = 25°C (the standard test conditions at which nominal power is measured) will have a loss with respect its nominal peak power. The correct determination of the operating module temperature is necessary to calculate the losses which will affect the output power of the PV module and the electricity generated in a certain period of time. The environmental conditions at which the PV module operates change daily and seasonally and make difficult to obtain good prediction of thermal losses in a given geo-

graphical location although good databases of daily or monthly average values of environmental parameters are available [3, 4, 5]. In particular, the error of thermal model predictions depends strongly on the ambient conditions: sunny, cloudy or variable days create a different thermal pattern which makes precise calculations very challenging.

Along with the PV conversion process (partly converted into heat by recombination and other internal mechanisms), the heat transfer mechanisms (defined by the system configuration) and the characteristics of module encapsulating material determine the PV module temperature. The coexistence of complex interacting processes is the main difficulty to achieve a simple, general and accurate method to predict the temperature of PV modules integrated in built structures, this task has been elusive to date. Prior research efforts have focused on the development of models based on empirical correlations of different parameters by different approaches, normally taking a steady state approach and expressing the PV module temperature as a function of ambient conditions (ambient temperature, wind speed and incident solar irradiance) and including some adjustment by encapsulation material or system properties. In this sense, Skoplaki and Palyvos [6] reviewed a wide set of correlations that have been proposed in the literature for specific mounting conditions, and, in another study, the authors proposed a semi-empirical correlation independent of the system configuration by introducing a dimensionless mounting parameter that define the integration level of the installation [7]. This highlights the need for predictive tools for non-conventional applications, as is the case of building integrated photovoltaics (BIPV), because of the difficulty to obtain accurate predictions of the thermal behavior at outdoor operating conditions.

Commonly, BIPV system configurations suffer from overheating due to the lack of natural ventilation from the rear surface of the PV module which seriously compromises the energy yield, the thermal comfort of the building and the degradation of the panel. Some widely-used operating cell temperature methods have been analyzed in the literature for different BIPV configurations. Davis et al. [8] analyzed the NOCT (nominal operating cell temperature) procedure for

crystalline BIPV panels mounted vertically, and showed that the model underpredicts the module temperature by approximately 20°C. Alonso-García and Balenzategui [9] studied the response of NOCT to different types of model encapsulations (including glass-glass and glass-tedlar), technologies, orientation and tilt angles. They highlighted that the method is not the most adequate for this kind of application as the working conditions substantially differs to those defined by the procedure. D'Orazio et al. [10] investigated the thermal performance for different roof configurations in Italy for one year using the two most common predictions models: NOCT and SNL (Sandia National Laboratory). The experimental results demonstrate a difference between predicted and calculated temperature around 10°C and suggest an optimization of the empirical coefficient for the use of SNL model which allows to reduce the error under 2.5% for these type of PV modules under these climatic conditions. Chatzipanagi et al. [11] also investigated the thermal behavior in real working conditions for different BIPV system configuration and technologies using as reference NOCT and the equivalent cell temperature (ECT) models. In this case, the indoor temperature of the experimental house is kept nearly constant along the year and the results show differences around 5°C. Assoa et al. [12] presented a comparison of nine thermal models in a rooftop partially integrated configuration in France (with two air layers between PV module and roof). The results show that although good prediction have achieved, the accuracy of the considered model strongly depends on the season, and most especially the irradiance level. The authors also presented a dynamic prediction model for different integration levels, based in the correct choice of the convective heat transfer coefficient, with an accuracy in a range of 2.6°C and 4.1°C for clear sky conditions. The error increases (around 9°C) at high irradiation levels which correspond to partial cloud periods where diffuse component is predominant [13].

As can be noted, the complexity on the estimation of the thermal behavior for non-conventional applications of PV derives not only from the irradiance levels, the sky conditions or the convective heat transfer coefficient. The absorbed radiation as heat inside the layers of the PV module induces a thermal inertia

which increases the estimation error. This behavior has been observed by some researchers [14, 9] and studied in detail for different technologies and orientations by our research group [15]. A short-term variability of the ambient conditions, such as passing clouds, also represents a key challenge for an accurate estimation of thermal performance. All this indicates the need for validated predictive tools based on data-driven approaches which allows to build unbiased statistical models which help to break down the barriers to widespread application of BIPV.

The storage of information dealing with historical photovoltaic performance and environmental parameters, has led to the application of novel computational techniques in the field. Thousands of studies in the literature show results on the application of Artificial Intelligence (AI) to photovoltaic generators. They hunt models with better prediction accuracy than the classic models, which are based on the physical, chemical, mechanical and/or thermal basis. Most of the application of AI to PV generators deals with the prediction of the maximum power point tracking, fault detection and output power/efficiency. The most applied AI techniques are: artificial neural networks, fuzzy logic, genetic algorithms and their hybrid models [16]. Meanwhile, other studies are focused on forecasting and modelling of meteorological data, basic modelling of solar cells and sizing of photovoltaic systems [17].

Previous works focused on the study of the thermal behaviour of conventional PV systems, consisting on module temperature prediction by applying AI techniques such as Artificial Neural Network (ANN), Back-propagation Neural Network (BPNN), Support Vector Machines (SVM), Genetic Programming (GP), and a wide range of variation and combination of them. Table 1 shows the studies identified in the literature as those proposing a new method to estimate the module temperature by applying AI techniques. None of them considers BIPV systems.

The proposed work aims to study the thermal behaviour by obtaining a model to accurately predict the module temperature (poly-crystalline Silicon PV) from environmental parameters. In particular, a combination of Grammat-

Ref	Year	Input parameters	AI technique	Accuracy assessement	PV Technology
[18]	2013	T_a, G, HR, W_s, P, V_{oc}, I_{sc}, T_c	MLP-ANN	ME_poly = -0.21 °C, MAE_poly = 0.23 °C ME_mono = -0.06 °C, MAE_mono = 0.11 °C	Mono-Si, Poly-Si
[19]	2014	T_a, G	ANN	E_{REL} = 1.16016e-3, R= 9.90215e-1	Mono-Si
[20]	2014	T_a, G	ANN	MAPE = 2.5659%	Poly-Si
[21]	2015	T_a, G, HR, W_s	SVM, MLP	RMSE_SVM = 1.5 °C RMSE_MLP = 1.4 °C	CdS/CdTe
[22]	2015	T_a, G, HR, W_s	ANN	R = 95.9%, MBE = 0.41 °C, RMSE = 0.1 °C, MPE = 4.5%	Poly-Si
[23]	2015	T_a, G, W_s	BP-ANN	MAPE = 6.398% - 11.645%	–
[24]	2016	V, I, d	ANN, ANFIS	RMSE_ANFIS = 2.5235 °C, MAPE_ANFIS = 0.6566, R_ANFIS = 0.7996 RMSE_ANN = 2.4368 °C, MAPE_ANN = 0.6413, R_ANN = 0.8167	Poly-Si
[25]	2016	T_a, G, W_s	MLP-ANN	RMSE = 2.67 °C, MAPE = 8.18%, R = 0.99	CPV-MJ
[26]	2017	T_a, G, W_s, P	BP-ANN	MaxError = 4.25 °C, SD = 1.03 °C, RMSE = 2.26 °C, MAE = 1.78 °C	Poly-Si
[27]	2017	T_a, G, W_s, W_d, HR, P_a, P	ANFIS	MAE=0.06729, RMSE = 3.4553, R = 95.56%	Mono-Si
[28]	2018	T_a, G, HR, W_s, P	SVM, MLP, RBF, TBM	RMSE_SVM = 3.2230 °C, MAPE_SVM = 5.2130%, R_SVM = 0.9348 RMSE_MLP = 2.8046 °C, MAPE_MLP = 5.4955%, R_MLP = 0.9510 RMSE_RBF = 2.7362 °C, MAPE_RBF = 5.4056%, R_RBF = 0.9534 RMSE_TBM = 2.3985 °C, MAPE_TBM = 5.3241%, R_TBM = 0.9552	Mono-Si
[29]	2018	T_a, G, W_s, W_d, HR	RBF-ANN	Cloudy day: E_{REL}_min = 7.53%, E_{REL}_max = 0.04%, MAPE = 2.16% Sunny day: E_{REL}_min = 4.47%, E_{REL}_max = 0.01%, MAPE = 1.18%	PV/T
[30]	2019	T_a, G	ANN	E_{REL} = 4%	a-Si
[31]	2020	T_a, G, W_s, HR	GP	MBE = 3.03% - 3.96%	Poly-Si
[32]	2020	T_a, G, W_s, HR, Cloud cover	ANN	cvRMSE = 19.81%	Poly-Si
[33]	2020	T_a, G, W_s,W_d, HR, P_a	DFPIO-SVM	MAPE = 0.2883%	–

Table 1: Literature review of application of Artificial Intelligence to obtain a model to predict the module PV temperature.

ical Evolution and Differential Evolution, very well-known Artificial Intelligence techniques, is applied to datasets obtained from historical data. With this aim, datasets are classified into three day-type groups: datasets belonging to sunny days, those belonging to cloudy days, and finally datasets belonging to diffuse light-type days. This approach has produced a new set of models with predictive capacities that have been tested on experimental results and have delivered errors that are lower than other existing models.

The rest of the article is organized as follows. Next, the previous experimental methods and thermal models found in the literature, as well as the classification of days by sky condition, are shown in section 2. Section 3 describes the algorithm approach, the applied techniques, the design of the experiments and the obtained results. Finally, section 4 presents the main conclusions of the study.

2. Experimental methods and thermal models

In this section the experimental set-up that was used to measure the indoor and outdoor, ambient and module temperatures included in the database is described; a summary of the main thermal models for PV modules and the description of standard thermal parameters used to calculate operational temperatures of the PV modules are also described.

2.1. Simplified building integrated PV systems

A BIPV system has been designed and built using commercial PV panels of crystalline silicon and a few additional materials. The final "building" is a cube in which each face is comprised of a photovoltaic panel, an aluminium frame and a few patches of polycarbonate used to complete the structure. In detail: the PV modules for the structure are Atersa Mod. SHS 100 (100 W_p, power conversion efficiency 14%) and occupy 78% of East (E) and West (W) faces, 83% of North (N) and South (S) faces and 86% of the Horizontal (H) face; the structure includes polycarbonate in a small patch 9% for East and West faces and the bottom (100%) of the structure. Aluminium profile frames (40 mm x 40 mm, Bosh Rexroth®) which comprises respectively 18%, 17% and 14% for E/W, N/S and H faces are used to fix the PV modules. Those percentages are important to describe the thermal conditions of the closed building structure, since internal temperature is very different compared to ambient temperature. Usually, thermal models of PV cubes only consider one ambient temperature on both sides of the module but in this case the experimental system presented in this article provides thermal conditions in which the PV module is in contact with an external ambient temperature and an internal ambient temperature which are different. Several temperatures on each side (inner or outer) of each face (E, W, N, S, H) of the structure have been measured and monitored every five minutes during two years. The collected database is used to train the AI tool and, once the models and parameters are determined, the database is also used to test them, simulating a predictive behaviour (which is tested using

previously measured temperatures). A detailed description of the structures and the monitoring system is provided in reference [34]. Additionally, an on-site meteorological station provides ambient data: temperature, humidity, wind speed and direction; irradiance data are measured with a shadow-band Delta-T pyranometer (Model Suhshine BF5) which provides global and diffuse irradiance in the horizontal plane and a Kipp-Zonen pyranometer (Model CMP3) which provides global irradiance at different orientations and which is used to test the accuracy of the transposition irradiation models used for some calculations [35]. For the results presented in this article, focused on the horizontal surface, direct data from the pyranometers have been used. Data stored in the database is available on request for collaborations (although this article is focused on crystalline silicon technology, the database contains data for a-Si:H, CdTe and organic PV modules).

2.2. PV thermal models and their limitations

As indicated in the introduction, two main approaches are used in the photovoltaic industry to calculate the temperature of operational PV modules. The most common and simple is based on the Ross model, which is also known as the NOCT model [36] because this parameter is included by all manufacturers in its PV modules technical sheets. The Ross model is a linear model based on an empirical temperature parameter (NOCT) which is measured in the laboratory under controlled illumination at irradiance G = 800 W/m^2 and T = 20°C ambient temperature; in these ambient conditions the PV module reaches an operational temperature when kept at maximum power point which is provided by the manufacturer (the NOCT). With this parameter, the module temperature at any other ambient conditions of irradiance and temperature is provided by a simple linear calculation:

$$T_m = T_a + \frac{NOCT - 20}{800} G, \tag{1}$$

where G is given in W/m^2, T_a is the ambient temperature and T_m is the PV module operating temperature in the given conditions. The accuracy of this

model is good, but any small deviation from the real temperature has strong implications, since the main loss of power generation in the PV module is due to an unwanted increment of temperature which reduces power output relative to its standard test conditions (STC) at which the industry measures and provides all electrical parameters of the PV modules. Therefore, a small mismatch in real time of the Ross model will generate extra losses in the electrical power generation (in addition to the losses already predicted by the Ross model by using the NOCT parameter). A better model is the so-called Sandia model (after Sandia National Laboratory, USA), which proposed the model [37]. The advantage provided when compared to Ross model is the inclusion of wind speed in order to increase the accuracy of the model. It is less used because it requires monitorization of wind speed in a given location. The functional dependence of the temperature of the PV module is exponential, with two parameters included in the exponent, the first one is equivalent to NOCT and the second one accounts for the new dependence on wind speed (W_s):

$$T_m = T_a + e^{(a+b \cdot W_s)} G, \tag{2}$$

where a and b are empirical parameters that are measured for specific PV modules, being the main difference between them the materials used for cover (glass of different kinds) and backsheet (plastics, often EVA) and frames (aluminium). Although Sandia model outperforms Ross model due to the inclusion of the effect of wind, the difference of accuracy when a large set of days are monitored is small, and strongly dependant on atmospheric conditions (variability of irradiance between clear, cloudy or diffuse days, wind speed). Furthermore, an important ambient parameter is not used in these models: the relative humidity, which has a strong effect on the thermal behaviour of the modules. Additionally, only one ambient temperature is considered in both models, assuming that the PV module is mounted on an open rack with both sides in contact with same ambient. For many PV applications this is no longer the case, since PV modules are increasingly being used in BIPV systems, where the module is part

of the building or in "agrivoltaic" systems when modules are incorporated in greenhouses [38, 39, 40]. In these applications, the module is affected by two ambient temperatures: one outer temperature and one inner temperature in both respective PV module sides; also the relative humidity is very different in both cases and has not been considered in the thermal models so far. The purpose of this study is to improve the prediction of thermal module temperature in operating conditions by including ambient temperature and humidity in both sides of the module, wind speed and direction.

Humidity is directly related to the thermal conductivity between the PV module and the air [41], and given its multi-functional character, it is one of the main factors which determines the thermal comfort, energy consumption and human health in buildings [42]. Surprisingly humidity is not included in the most used models to calculate module temperatures. Given the difficulty to develop an easy-to-implement, reliably and accurately function to describe the relationship between both sides taking into account a wide range of technologies, environment conditions and designs, there is a growing need to use AI tools to help predict thermal characteristics of the PV module, overcoming key technical obstacles, and thus accelerating the penetration of BIPV in the market.

2.3. Classification of days by sky condition using clearness index as indicator

One of the key proposals of this work is the classification of days with the aim of providing different models for each selected type. In order to identify the day type, the clearness index K_t is used as descriptor. Clearness index is defined as the ratio of the global horizontal irradiance (GHI) and the horizontal extraterrestrial irradiance (G_e), as shown in (3).

$$K_t = \frac{GHI}{G_e} \tag{3}$$

However, this parameter is not independent of solar elevation angle, so the

clearness index modified by Perez et al. [43] is used:

$$K'_t = \frac{K_t}{1.031\ e^{\left(\frac{-1.4}{0.9+\frac{9.4}{AM}}\right)}} \tag{4}$$

where AM is the air mass that has been calculated from empirical equation given by Kasten and Young [44].

Based on the modified clearness index and following Ineichens sky condition classification [45], three categories are defined:

$$Sky\ condition = \begin{cases} Clear, & : \ K_t^{'} > 0.65 \\ Intermediate, & : \ 0.3 < K_t^{'} \leq 0.65 \\ Cloudy, & : \ K_t^{'} < 0.3 \end{cases} \tag{5}$$

Once the sky condition is defined, days are classified based on the predominant category on the records of each day. Note that Cloudy label defined for equation (5) for definition of Sky condition differs from "Cloudy" type, used to classify the days. The database contained days belonging to the three defined types: (a) *Sunny* days denoting clear sky conditions for almost all the records. Solar irradiance profile for sunny days has shown as example in figure 1. (b) *Diffuse* days that are associated to constant cloudy periods along the day, so diffuse irradiance are the major contributor to the global component of the sunlight. Figure 2 shows the irradiance profile for days classified under this category showing that the diffuse horizontal irradiance (DHI) is the main component of the total irradiance. Finally, (c) *Cloudy* days which indicates variability of sky conditions over short time intervals. Figure 3 shows typical irradiance profiles for this category.

Figure 4 shows the frequency of the records normalized for each day which allows to classify the day according to each category.

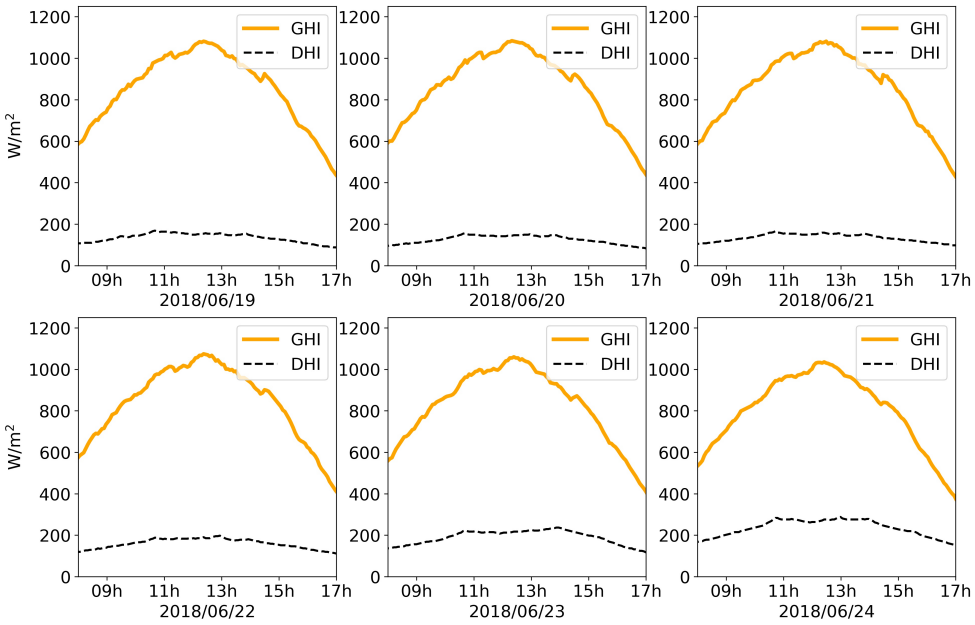


Figure 1: Profile of solar irradiance for *Sunny* days. GHI Global horizontal irradiance and DHI diffuse horizontal irradiance.

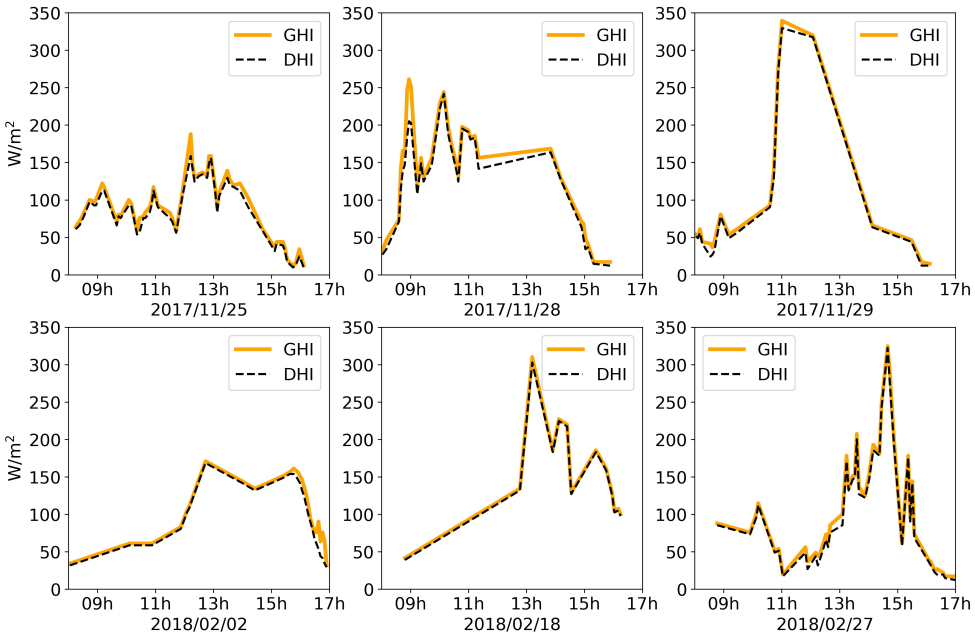


Figure 2: Profile of solar irradiance for *Diffuse* days. GHI Global horizontal irradiance and DHI diffuse horizontal irradiance.

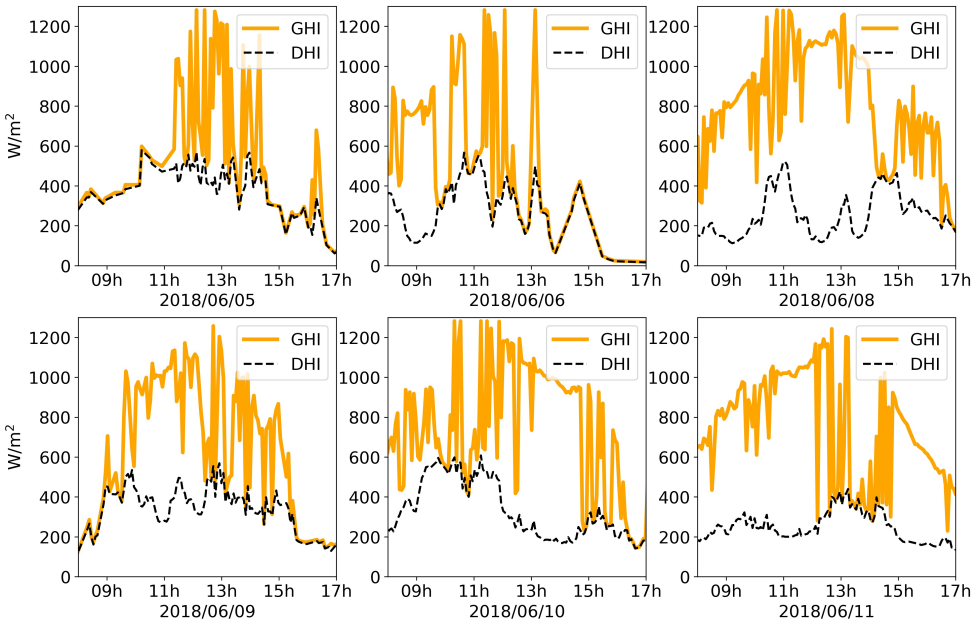
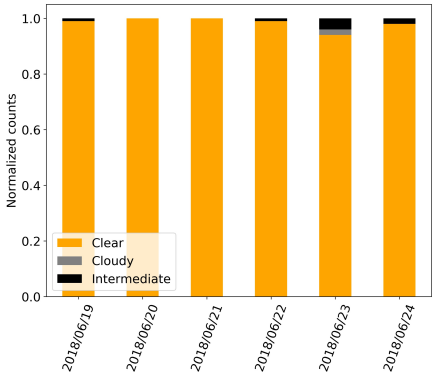
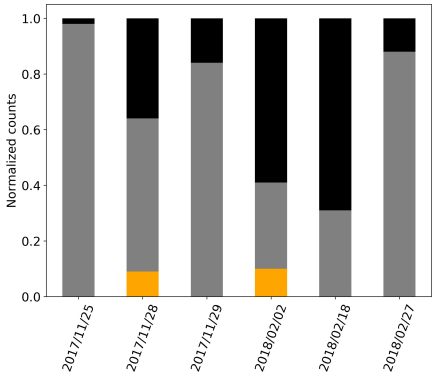


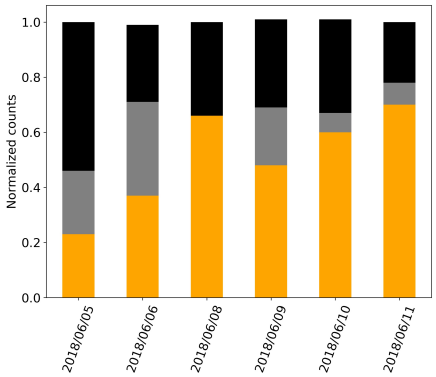
Figure 3: Profile of solar irradiance for *Cloudy* days. GHI Global horizontal irradiance and DHI diffuse horizontal irradiance.



(a) Sunny



(b) Diffuse



(c) Cloudy

Figure 4: Frequency of records classified by sky conditions.

3. Algorithmic approach and results

This section is devoted to show how AI is applied to the collected database to get a highly accurate model to predict the module temperature.

3.1. Optimization algorithm

In order to obtain a customized model for the temperature, we have addressed the question as an optimization problem. To this aim, we have used a combination of two different algorithms: Grammatical Evolution (GE) [46] and Differential Evolution (DE) [47]. GE and DE are AI methods which belong to the family of the metaheuristic algorithms. Their combined performance has been successfully applied before on other target problems [48] [49].

The optimization process is described in Figure 5. As it can be seen, GE generates a set of parameterized models which are produced after the rules stated in a grammar. The parameter values are set after the DE process, which uses the training data to find the best parameter values for each model. Then, the optimized models are returned to GE with an associated error value. After that, GE ranks the optimized models considering their quality and, using this information, generates a new set of models repeating the process. The number of iterations of this loop is stated in the configuration of the algorithms. Finally, the best model obtained is returned.

One of the main advantages of this approach is the ability to include particular features of the problem in the optimization process by means of the grammar. More precisely, stating an initial structure for the prediction models in the grammar allows the algorithm to reduce the search space and, therefore, obtain better solutions.

In this case, an initial structure of the Sandia model is included in the grammar, taking into account the following parameters:

- T_m: Module temperature (°C)
- T_a: Ambient temperature (°C)

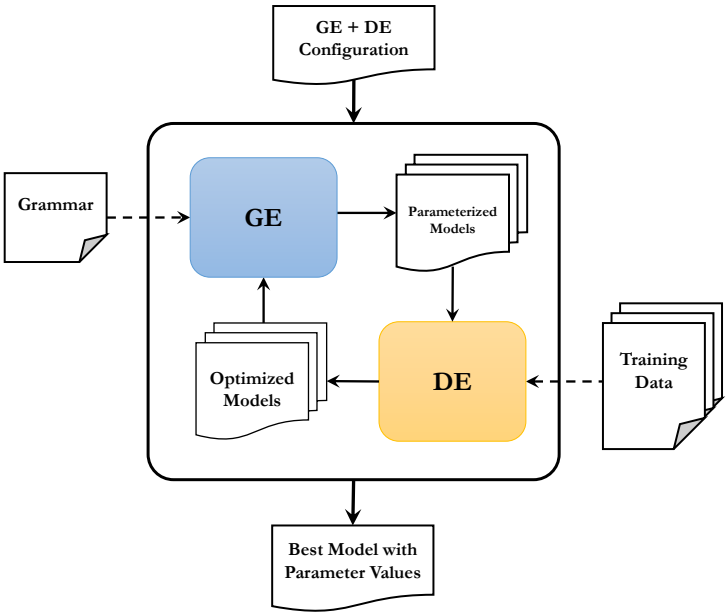


Figure 5: GE + DE optimization process.

- W_s: Wind speed (m/s)
- G: Irradiance (W/m^2)
- T_i: Interior temperature (°C)
- W_d: Wind direction (°)
- HR_{out}: Relative humidity outside (%)
- HR_{in} Relative humidity inside (%)

Notice that both humidity parameters (HR_{out} and HR_{in}) were included in the grammar despite they are not considered in the Sandia model. Finally, the target variable for the obtained models is the module temperature (T_m).

Figure S1 in the Supplementary Material represents an extract of the grammar in Backus-Naur Form (BNF) format designed for finding a predictive model for temperature including the previously described information.

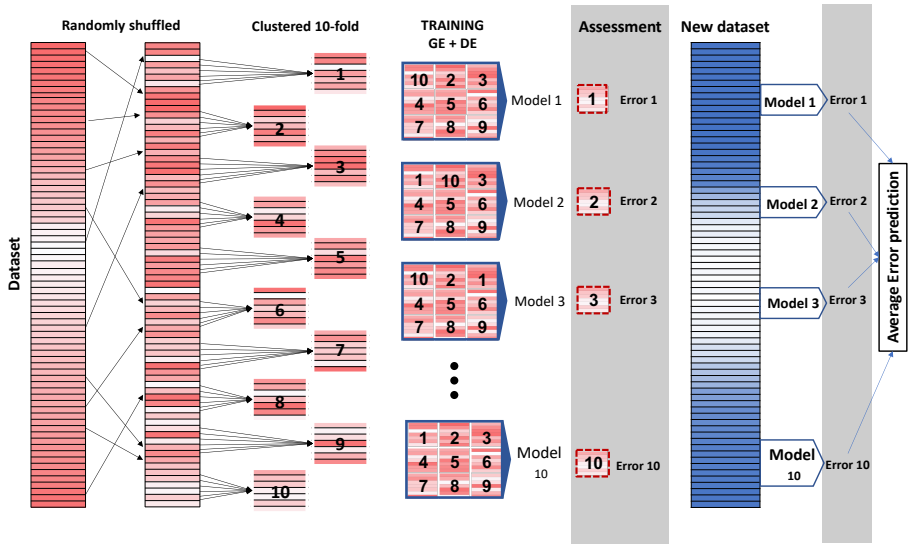


Figure 6: Design of the experiments for GE + DE applied to photovoltaic module datasets.

3.2. Design of experiments

The collected dataset was divided into three parts classified per type of day, i.e. *Sunny*, *Cloudy* or *Diffuse*, as defined in Section 2.3.

GE+DE algorithm requires two types of dataset groups: training and test. The two groups are built following a 10-fold process. Figure 6 shows how the dataset is divided in this process producing a set of models. As seen in the figure, a set of records from the same day-type is randomly shuffled as the first step (left-hand side of the figure, in red). Secondly, the data is cut into ten same-size groups named folds. Nine of them are used as training group, using for the testing the outcast one. This process is repeated 10 times, selecting a different test fold for each iteration, therefore obtaining 10 models from the execution of the algorithm. These models are assessed obtaining different error metrics (see column "Assessment" in the figure) corresponding to the end of the training phase.

In order to get a prediction for a new dataset, the test phase is applied. Figure 6 shows the steps of this phase in blue colors in the right-hand side of the figure. In this case, the 10 models are evaluated for a new dataset from the

same day-type producing 10 different predictions whose average is obtained. Finally, the error of the ensemble of models is obtained averaging the individual errors.

3.3. Verification of the obtained models

To verify the prediction ability of the models, which are obtained by the GE+DE algorithm, the relative error is calculated as follows:

$$E_{REL} = \frac{1}{n} \sum_{i=1}^{n} \frac{(v[i] - p[i]) \cdot 100}{v[i]} \tag{6}$$

where $v[i]$ represents the measured module temperature (°C), $p[i]$ the predicted module temperature (°C) and n corresponds to the total number of values, i.e. number of records in the test dataset.

3.4. Analysis of the results

Once the general process has been described, the results of calculations carried out with the experimental datasets are presented and discussed in this section.

A number of 2000 rows of day-type data were randomly chosen as training input, and 220 rows fit the test. Ten experiments were executed per day-type, combining training and test datasets, from which 10 models were obtained per day-type. Since there are three types of day, a total of 30 models were produced, which are shown in tables S1, S3 and S2 in the Supplementary Material.

All obtained models predict the same parameter, i.e. the module temperature (T_m) and, although their shape was guided by the same grammar, their final layout varies from one experiment to another, which is a typical result when applying AI techniques.

The relative error of the obtained models is in the intervals [2.46 - 2.97], [3.33 - 4.19] and [1.55 - 2.25] for *Sunny*, *Cloudy* and *Diffuse* days respectively.

Since the grammar directs the exploration of the algorithm, the 30 obtained models begin with exactly the same expression (see tables S1, S3) and S2,

inspired the in Sandia model: $T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)}$. Nevertheless, the shape of the last part of each expression changes in each one of the models, as well as the values of the parameters.

Table 2 shows the best models obtained in the training phase for *Sunny*, *Cloudy* and *Diffuse* types of day. Notice that the first part of the models, corresponding to the fixed Sandia structure is removed for the sake of space. The table also shows the relative error as well as the values of the parameters obtained by GE+DE.

	Day-type		
	Sunny	Cloudy	Diffuse
GE+DE	$-w_7 \cdot HR_{out}^{w_8} \cdot \log \mid w_5 - HR_{in} \mid$	$\cdot w_8 \cdot HR_{in}^{w_6}$	$-\log \mid w_7 - HR_{out} \mid \cdot w_5 \cdot HR_{in}$
E_{REL}	2.07	3.34	1.55
w_1	0.378	0.033	-0.151
w_2	0.508	0.917	1.133
w_3	-3.605	-3.26	-4.694
w_4	-0.057	-0.091	-0.025
w_5	18.248	0	-0.002
w_6	0	-0.319	0
w_7	3.522	0	74.289
w_8	-0.529	0.726	0

Table 2: Best obtained models, and their parameter values, for sunny, cloudy and diffuse day-type datasets. Note that all models start with the expression $T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)}$ which is removed from the table.

Table 3 shows a comparison between the Sandia model and the GE+DE models prediction for both training and test datasets. In particular, the table shows the average relative error obtained by the Sandia model and the ensemble of models generated by GE+DE. Notice that the improvement in the training phase is noticeable (reaching up to 15.35% in the Cloudy day-type). Besides and, more importantly, the improvement over a new dataset (test phase) reaches a 10.74% in the Cloudy days).

Figure 7 shows the average GE+DE prediction in comparison with the actual temperature and the Sandia model. As seen, the inclusion of relative humidity by the GE+DE algorithm produces a higher accuracy of the prediction of the module temperature, specially for the cloudy days, since they present less stable conditions.

	Test			Training		
Day-type	GE+DE	Sandia	Impr.	GE+DE	Sandia	Impr.
Sunny	2.07	13.10	11.03	3.75	11.62	7.87
Cloudy	3.34	18.69	15.35	3.99	14.73	10.74
Diffuse	1.55	15.01	13.46	3.42	12.73	9.31

Table 3: Comparison between the GE+DE models and the Sandia one for each day-type. All values are in %.

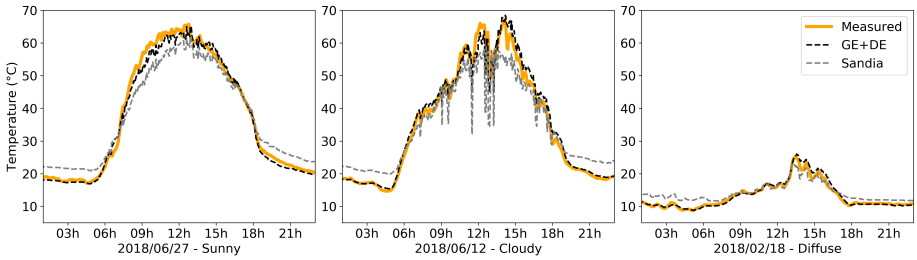


Figure 7: Measured module temperature during three type-days, compared to Sandia's prediction and obtained day-type models' predictions.

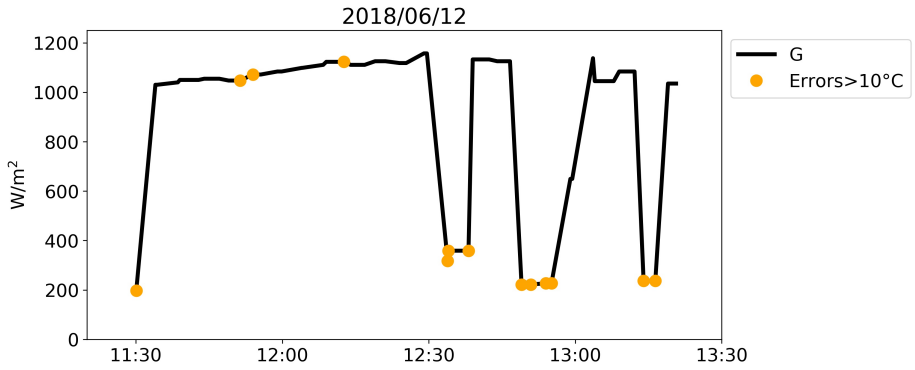


Figure 8: Solar irradiance profile in a cloudy day (G). The yellow dots indicate the measurements where the Sandia model shows differences greater than 10°C in the estimation of the module temperature.

Figure 8 provides a clear picture of the current limitations of using traditional approaches in which the model can reach errors of more than 10°C in sky conditions with high variability of solar irradiation in short time intervals, since they do not account for the module's thermal inertia.

Despite the fact that studies in the literature show that differences in predicted and calculated temperatures around 10°C cause percentage variations on the annual energy predicted that are generally less than 5% [13, 10, 50, 9], which is acceptable for most applications, uncertainties in module temperature calculation would have a significant impact on its passive function. Although the most important parameter to calculate electricity production by a PV module is the irradiance, the module temperature is the second most important, with an high impact on output losses (with respect standard conditions), and on the other hand, it affects the comfort conditions inside the building where the modules are attached and indirectly affects the energy consumption required for cooling (or heating) by air condictioning systems.

Thus, from a multifunctional perspective that includes both energy generation and energy demand requirements for buildings, the proposed methodology aids in obtaining a more precise estimation of the thermal behavior of a module integrated in the building envelope, covering different climatic and mounting conditions as well as the variation of the environmental parameters during the night. As seen in Figure 7, there is a strong influence of irradiance in conventional models such as the Sandia one, which leads to significant errors in temperature estimation during the nighttime.

4. Conclusions

The photovoltaic module temperature varies during the day, influenced by the PV technology, the mounting configuration and climatic conditions. BIPV systems temperature behaves differently from ground-mounted PV, since there is no air circulation at the rear side of the first configuration and so the thermal transmission behaves differently from the not integrated approach. The

prediction models for PV module temperature in the literature are focused on ground-mounted PV. So, when applying those models to BIPV configuration, even though some adaptations were proposed, i.e., Sandia and Ross models, the prediction errors are quite significant. In this study, Sandia-based BIPV-adapted models are obtained by the application of AI algorithms on the collected data.

An available database from a monitored BIPV device, where inside and outside ambient conditions were recorder each 5 minutes during 2 years, was used to produce models with a combination of Grammatical Evolution (GE) and Differential Evolution (DE) algorithms. A grammar inspired on the Sandia equation shape was designed for the application of GE, adding the relative humidity and inside temperature as input parameters. Three day-types were identified from the dataset, i.e. *Sunny*, *Cloudy* and *Diffuse*, and 10 accurate models were obtained per each day-type. The average relative errors of the obtained models were 2.07%, 3.34% and 1.55% for *Sunny*, *Cloudy* and *Diffuse* day-types, respectively. The Sandia model was applied to the same dataset, obtaining relative errors of 13.01%, 18.69% and 15.01% for the three day-types respectively. Thereby, the application of GE+DE is underlined here as a reliable method to obtain novel models to accurately predict the BIPV module temperature. The historic record and monitorization of indoor and outdoor ambient conditions of BIPV systems is key in order to apply AI methods. The obtained models are specific for BIPV configurations and are easily applicable to similar PV generators, since the above mentioned parameters could be obtained by measuring the indoor building conditions (or the target temperature of an air conditioning system) and by loading public open-source datasets dealing with the outside climatic conditions, such as those provided by NASA, PVGIS and State Meteorological Agencies.

Acknowledgements

This work was supported by Project PID2019-104272RB-C55, funded by Agencia Estatal de Investigacin (AEI-MICINN, Spain, including FEDER funds), Spanish MCIU/AEI/FEDER, UE, under grant ref. PGC2018-095322-B-C22; and Comunidad de Madrid y Fondos Estructurales de la Unión Europea with grant ref. P2018/TCS-4566. Carlos Toledo is grateful for postdoctoral fellowship 21227/PD/19. Fundación Séneca. Región de Murcia (Spain).

References

[1] IRENA, Solar Energy Data. Installed Capacity Trends.
URL https://www.irena.org/solar

[2] IEA, World energy outlook 2020.
URL https://www.iea.org/reports/world-energy-outlook-2020

[3] European Commission - Joint Research Centre, Photovoltaic Geographical Information System (PVGIS).
URL http://re.jrc.ec.europa.eu/pvg_tools/en/tools.html

[4] National Renewable Energy Laboratory (NREL), NSRDB: National Solar Radiation Database.
URL https://nsrdb.nrel.gov

[5] NASA, EarthData.
URL https://earthdata.nasa.gov

[6] E. Skoplaki, J. Palyvos, Operating temperature of photovoltaic modules: A survey of pertinent correlations, Renewable Energy 34 (1) (2009) 23–29. doi:10.1016/j.renene.2008.04.009.
URL http://www.sciencedirect.com/science/article/pii/S0960148108001353

[7] E. Skoplaki, A. Boudouvis, J. Palyvos, A simple correlation for the operating temperature of photovoltaic modules of arbitrary mounting, Solar Energy Materials and Solar Cells 92 (11) (2008) 1393–1402. doi:https://doi.org/10.1016/j.solmat.2008.05.016.
URL https://www.sciencedirect.com/science/article/pii/S0927024808001918

[8] M. W. Davis, A. H. Fanney, B. P. Dougherty, Prediction of Building Integrated Photovoltaic Cell Temperatures*, Journal of Solar Energy Engineering 123 (3) (2001) 200–210. doi:10.1115/1.1385825.
URL http://dx.doi.org/10.1115/1.1385825

[9] M. Alonso Garca, J. Balenzategui, Estimation of photovoltaic module yearly temperature and performance based on Nominal Operation Cell Temperature calculations, Renewable Energy 29 (12) (2004) 1997–2010. doi:10.1016/j.renene.2004.03.010.
URL http://www.sciencedirect.com/science/article/pii/S0960148104001260

[10] M. D'Orazio, C. Di Perna, E. Di Giuseppe, Experimental operating cell temperature assessment of BIPV with different installation configurations on roofs under Mediterranean climate, Renewable Energy 68 (2014) 378–396. doi:10.1016/j.renene.2014.02.009.
URL http://www.sciencedirect.com/science/article/pii/S0960148114000895

[11] A. Chatzipanagi, F. Frontini, A. Virtuani, BIPV-temp: A demonstrative Building Integrated Photovoltaic installation, Applied Energy 173 (2016) 1–12. doi:10.1016/j.apenergy.2016.03.097.
URL http://www.sciencedirect.com/science/article/pii/S0306261916304299

[12] Y. B. Assoa, L. Mongibello, A. Carr, B. Kubicek, M. Machado, J. Merten, S. Misara, F. Roca, W. Sprenger, M. Wagner, S. Zamini,

T. Baenas, P. Malbranche, Thermal analysis of a BIPV system by various modelling approaches, Solar Energy 155 (2017) 1289–1299. doi:10.1016/j.solener.2017.07.066.
URL http://www.sciencedirect.com/science/article/pii/S0038092X17306515

[13] Y. B. Assoa, L. Gaillard, C. Mnzo, N. Negri, F. Sauzedde, Dynamic prediction of a building integrated photovoltaic system thermal behaviour, Applied Energy 214 (2018) 73–82. doi:10.1016/j.apenergy.2018.01.078.
URL https://www.sciencedirect.com/science/article/pii/S0306261918300928

[14] L. Maturi, G. Belluardo, D. Moser, M. Del Buono, BiPV System Performance and Efficiency Drops: Overview on PV Module Temperature Conditions of Different Module Types, Proceedings of the 2nd International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2013) 48 (2014) 1311–1319. doi:10.1016/j.egypro.2014.02.148.
URL http://www.sciencedirect.com/science/article/pii/S187661021400410X

[15] C. Toledo, R. Lpez-Vicente, J. Abad, A. Urbina, Thermal performance of PV modules as building elements: Analysis under real operating conditions of different technologies., Energy and Buildings 223 (2020) 110087. doi:10.1016/j.enbuild.2020.110087.
URL http://www.sciencedirect.com/science/article/pii/S0378778820300955

[16] K. S. Garud, S. Jayaraj, M.-Y. Lee, A review on modeling of solar photovoltaic systems using artificial neural networks, fuzzy logic, genetic algorithm and hybrid models, International Journal of Energy Research 45 (1) (2021) 6–35, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/er.5608.

doi:https://doi.org/10.1002/er.5608.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/er.5608

[17] R. Ghannam, P. V. Klaine, M. Imran, Artificial Intelligence for Photovoltaic Systems, in: R.-E. Precup, T. Kamal, S. Zulqadar Hassan (Eds.), Solar Photovoltaic Power Plants: Advanced Control and Optimization Techniques, Power Systems, Springer, Singapore, 2019, pp. 121–142. doi:10.1007/978-981-13-6151-7_6.
URL https://doi.org/10.1007/978-981-13-6151-7_6

[18] G. Ciulla, V. Lo Brano, E. Moreci, Forecasting the Cell Temperature of PV Modules with an Adaptive System, International Journal of Photoenergy 2013 (2013) e192854, publisher: Hindawi. doi:10.1155/2013/192854.
URL https://www.hindawi.com/journals/ijp/2013/192854/

[19] I. Ceylan, O. Erkaymaz, E. Gedik, A. E. Grel, The prediction of photovoltaic module temperature with artificial neural networks, Case Studies in Thermal Engineering 3 (2014) 11–20. doi:10.1016/j.csite.2014.02.001.
URL https://www.sciencedirect.com/science/article/pii/S2214157X14000069

[20] S. I. Sulaiman, N. Z. Zainol, Z. Othman, H. Zainuddin, Cuckoo search for determining Artificial Neural Network training parameters in modeling operating photovoltaic module temperature, in: Proceedings of 2014 International Conference on Modelling, Identification Control, 2014, pp. 306–309. doi:10.1109/ICMIC.2014.7020770.

[21] M. Piliougine, L. Mora-López, J. Carretero, M. Sidrach-de Cardona, Characterisation of hourly temperature of a thin-film module from weather conditions by artificial intelligence techniques (2015).
URL http://hdl.handle.net/10630/10280

[22] M. Almaktar, H. A. Rahman, M. Y. Hassan, I. Saeh, Artificial neural network-based photovoltaic module temperature estimation for tropical

climate of Malaysia and its impact on photovoltaic system energy yield, Progress in Photovoltaics: Research and Applications 23 (3) (2015) 302–318, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/pip.2424. doi:https://doi.org/10.1002/pip.2424.
URL https://onlinelibrary.wiley.com/doi/abs/10.1002/pip.2424

[23] Yujing Sun, Fei Wang, Zhao Zhen, Zengqiang Mi, Chun Liu, Bo Wang, Jing Lu, Research on short-term module temperature prediction model based on BP neural network for photovoltaic power forecasting, in: 2015 IEEE Power Energy Society General Meeting, 2015, pp. 1–5, iSSN: 1932-5517. doi:10.1109/PESGM.2015.7286350.

[24] J. T. Dzib, E. J. A. Moo, A. Bassam, M. Flota-Bauelos, M. A. E. Soberanis, L. J. Ricalde, M. J. Lpez-Snchez, Photovoltaic Module Temperature Estimation: A Comparison Between Artificial Neural Networks and Adaptive Neuro Fuzzy Inference Systems Models, in: A. Martin-Gonzalez, V. Uc-Cetina (Eds.), Intelligent Computing Systems, Communications in Computer and Information Science, Springer International Publishing, Cham, 2016, pp. 46–60. doi:10.1007/978-3-319-30447-2_4.

[25] C. Cancro, S. Ferlito, G. Graditi, Forecasting the working temperature of a concentrator photovoltaic module by using artificial neural network-based model, AIP Conference Proceedings 1766 (1) (2016) 090004, publisher: American Institute of Physics. doi:10.1063/1.4962110.
URL https://aip.scitation.org/doi/abs/10.1063/1.4962110

[26] H. Zhu, W. Lian, L. Lu, P. Kamunyu, C. Yu, S. Dai, Y. Hu, Online Modelling and Calculation for Operating Temperature of Silicon-Based PV Modules Based on BP-ANN, International Journal of Photoenergy 2017 (2017) e6759295, publisher: Hindawi. doi:10.1155/2017/6759295.
URL https://www.hindawi.com/journals/ijp/2017/6759295/

[27] A. Bassam, O. May Tzuc, M. Escalante Soberanis, L. J. Ricalde, B. Cruz, Temperature Estimation for Photovoltaic Array Using an Adaptive Neuro

Fuzzy Inference System, Sustainability 9 (8) (2017) 1399, number: 8 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/su9081399.
URL https://www.mdpi.com/2071-1050/9/8/1399

[28] O. May Tzuc, A. Bassam, P. E. Mendez-Monroy, I. S. Dominguez, Estimation of the operating temperature of photovoltaic modules using artificial intelligence techniques and global sensitivity analysis: A comparative approach, Journal of Renewable and Sustainable Energy 10 (3) (2018) 033503, publisher: American Institute of Physics. doi:10.1063/1.5017520.
URL https://aip.scitation.org/doi/abs/10.1063/1.5017520

[29] J. Li, Z. Zhao, Y. Li, J. Xiao, Y. Tang, Short-term PV/T module temperature prediction based on PCA-RBF neural network, IOP Conference Series: Earth and Environmental Science 121 (2018) 052045, publisher: IOP Publishing. doi:10.1088/1755-1315/121/5/052045.
URL https://doi.org/10.1088/1755-1315/121/5/052045

[30] A. Hegazy, E. Shenawy, M. Ibrahim, Determination of the PV Module Surface Temperature Based on Neural Network using Solar Radiation and Surface Temperature, ARPN J. Eng. Appl. Sci 14 (2019) 494–503.

[31] A. Sohani, H. Sayyaadi, Employing genetic programming to find the best correlation to predict temperature of solar photovoltaic panels, Energy Conversion and Management 224 (2020) 113291. doi:10.1016/j.enconman.2020.113291.
URL https://www.sciencedirect.com/science/article/pii/S019689042030830X

[32] D. E. Jung, C. Lee, K. H. Kim, S. L. Do, Development of a Predictive Model for a Photovoltaic Modules Surface Temperature, Energies 13 (15) (2020) 4005, number: 15 Publisher: Multidisciplinary Digital Publishing Institute. doi:10.3390/en13154005.
URL https://www.mdpi.com/1996-1073/13/15/4005

[33] S. Fan, S. Cao, Y. Zhang, Temperature Prediction of Photovoltaic Panels Based on Support Vector Machine with Pigeon-Inspired Optimization, Complexity 2020 (2020) e9278162, publisher: Hindawi. doi:10.1155/2020/9278162.
URL https://www.hindawi.com/journals/complexity/2020/9278162/

[34] C. Toledo, L. Serrano-Lujan, J. Abad, A. Lampitelli, A. Urbina, Measurement of thermal and electrical parameters in photovoltaic systems for predictive and cross-correlated monitorization, Energies 12 (4) (2019). doi:10.3390/en12040668.
URL http://www.mdpi.com/1996-1073/12/4/668

[35] C. Toledo, A. M. Gracia Amillo, G. Bardizza, J. Abad, A. Urbina, Evaluation of solar radiation transposition models for passive energy management and building integrated photovoltaics, Energies 13 (3) (2020). doi:10.3390/en13030702.
URL https://www.mdpi.com/1996-1073/13/3/702

[36] Richard G. Ross, Interface design considerations for terrestrial solar cell modules, in: 12th IEEE Photovolt. Spec. Conference, 1976.

[37] D. L. King, W. E. Boyson, J. A. Kratochvill, Photovoltaic Array Performance Model, Tech. Rep. SAND2004-3535, Sandia National Laboratories (2004).

[38] R. H. E. Hassanien, M. Li, W. Dong Lin, Advanced applications of solar energy in agricultural greenhouses, Renewable and Sustainable Energy Reviews 54 (2016) 989–1001. doi:10.1016/j.rser.2015.10.095.
URL https://doi.org/10.1016/j.rser.2015.10.095

[39] C. S. Allardyce, C. Fankhauser, S. M. Zakeeruddin, M. Grtzel, P. J. Dyson, The influence of greenhouse-integrated photovoltaics on crop production, Solar Energy 155 (2017) 517–522. doi:10.1016/j.solener.2017.06.044.
URL http://www.sciencedirect.com/science/article/pii/S0038092X17305510

[40] A. Yano, M. Cossu, Energy sustainable greenhouse crop cultivation using photovoltaic technologies, Renewable and Sustainable Energy Reviews 109 (2019) 116–137. doi:10.1016/j.rser.2019.04.026.
URL http://www.sciencedirect.com/science/article/pii/S1364032119302394

[41] J. ho Choi, J. Hyun, W. Lee, B.-G. Bhang, Y. K. Min, H.-K. Ahn, Power performance of high density photovoltaic module using energy balance model under high humidity environment, Solar Energy 219 (2021) 50–57, special Issue on Floating Solar: beyond the state of the art technology. doi:https://doi.org/10.1016/j.solener.2020.10.022.
URL https://www.sciencedirect.com/science/article/pii/S0038092X20310823

[42] D. Enescu, A review of thermal comfort models and indicators for indoor environments, Renewable and Sustainable Energy Reviews 79 (2017) 1353–1379. doi:https://doi.org/10.1016/j.rser.2017.05.175.
URL https://www.sciencedirect.com/science/article/pii/S1364032117308109

[43] R. Perez, P. Ineichen, R. Seals, J. Michalsky, R. Stewart, Modeling daylight availability and irradiance components from direct and global irradiance, Solar Energy 44 (5) (1990) 271–289. doi:10.1016/0038-092X(90)90055-H.
URL http://www.sciencedirect.com/science/article/pii/0038092X9090055H

[44] F. Kasten, A. T. Young, Revised optical air mass tables and approximation formula, Applied Optics 28 (22) (1989) 4735–4738. doi:10.1364/AO.28.004735.
URL http://ao.osa.org/abstract.cfm?URI=ao-28-22-4735

[45] P. Ineichen, Five satellite products deriving beam and global irradiance val-

idation on data from 23 ground stations, Tech. rep., Universite de Geneve. International Energy Agency (Feb. 2011).

[46] M. O'Neill, C. Ryan, Grammatical evolution, IEEE Trans. Evolutionary Computation 5 (4) (2001) 349–358.

[47] R. Storn, K. Price, Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces, Journal of Global Optimization 11 (4) (1997) 341–359.

[48] J. Colmenar, J. Hidalgo, S. Salcedo-Sanz, Automatic generation of models for energy demand estimation using grammatical evolution, Energy 164 (2018) 183–193. doi:https://doi.org/10.1016/j.energy.2018.08.199.
URL https://www.sciencedirect.com/science/article/pii/S0360544218317353

[49] N. Lourenço, J. M. Colmenar, J. I. Hidalgo, S. Salcedo-Sanz, Evolving energy demand estimation models over macroeconomic indicators, in: Proceedings of the 2020 Genetic and Evolutionary Computation Conference, GECCO '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 11431149. doi:10.1145/3377930.3390153.
URL https://doi.org/10.1145/3377930.3390153

[50] P. Trinuruk, C. Sorapipatana, D. Chenvidhya, Estimating operating cell temperature of BIPV modules in Thailand, Renewable Energy 34 (11) (2009) 2515–2523. doi:10.1016/j.renene.2009.02.027.
URL http://www.sciencedirect.com/science/article/pii/S0960148109000962

Nomenclature

Artificial Intelligence

ANFIS Adaptive Neuro Fuzzy Inference Systems

ANN Artificial neural network

BP-ANN Back Propagation Neural Network

DE Differential Evolution

DFPIO Pigeon-inspired Optimization Improvement

GE Grammatical Evolution

GP Genetic Programming

MLP Multilayer perceptron

RBF Radial Basis Function neural network

SVM Support Vector Machine

TBM Tree Boost Method

Photovoltaics

a-Si Amorphous Silicon

BIPV Building Integrated Photovoltaics

CdS/CdTe Capped Cadmium Culphide/Cadmium Telluride solar cells

CPV Concentrator Photovoltaic module

MJ Multi-junction solar cells

Mono-Si Mono-crystalline silicon

NOCT Nominal Operating Cell Temperature

Poly-Si Poly-crystalline silicon

PV Photovoltaic

PV/T Photovoltaic-thermal system

Statistical Indicators

cvRMSE Coefficient of variation of the RMSE

E_{REL} Relative error

MAE Mean absolute error

MAPE Mean absolute percentage error

MaxError Maximum error

MBE Mean bias error

ME Mean error

MPE Mean percentage error

R Correlation coefficient

RMSE Root mean square error

SD Standard deviation

Variables

d Duty cycle

G Solar irradiation

HR Relative humidity

I Current

I_{sc} Short circuit current

P Power output

P_a Atmospheric pressure

T_a	Ambient temperature
T_c	Cell temperature
T_m	Module temperature
V	Voltage
V_{oc}	Open circuit voltage
W_d	Wind direction
W_s	Wind speed

Supplementary Material

Table S1: Models obtained for *Sunny*-type days.

Fold	Model	Parameters								Error
		w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	
1	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} + w_6 \cdot HR_{out}^{w_5}$	0.546	0.514	-3.600	-0.094	-0.469	-25.884	0.000	0.000	2.33
2	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - HR_{in}^{w_8} + HR_{in}^{w_6} + \log \mid w_5 - HR_{in} \mid - w_5 + HR_{out}^{w_8}$	0.394	0.634	-3.709	-0.091	5.455	-0.551	0.000	-1.042	2.25
3	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - w_7 \cdot HR_{out}^{w_8} \cdot \log \mid w_5 - HR_{in} \mid$	0.377	0.508	-3.605	-0.057	18.248	0.000	3.522	-0.529	2.07
4	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - w_7 \cdot HR_{out}^{w_6} \cdot \log \mid w_7 - HR_{in} \mid$	0.393	0.499	-3.568	-0.066	0.000	-0.984	17.215	0.000	2.13
5	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - w_5 + HR_{in}^{w_8} + \log \mid w_7 - HR_{out} \mid$	0.561	0.452	-3.589	-0.063	6.812	0.000	17.004	-2.106	2.32
6	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - \log \mid w_6 + W \mid + \log \mid w_8 \cdot HR_{out} \mid$	0.416	0.479	-3.604	-0.068	0.000	304.172	0.000	5.057	2.46
7	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - w_6 \cdot HR_{out}^{w_8} \cdot \log \mid w_5 - HR_{in} \mid$	0.390	0.498	-3.613	-0.066	17.800	9.560	0.000	-0.982	2.20
8	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - w_8 - W \cdot w_8 + W + \log \mid w_6 + W \mid$	0.253	0.494	-3.615	-0.061	0.000	73.047	0.000	1.008	2.42
9	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - w_6 \cdot HR_{out}^{w_8} \cdot \log \mid w_7 - HR_{in} \mid$	0.428	0.462	-3.534	-0.061	0.000	8.432	17.352	-0.719	2.31
10	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3 + w_4 \cdot W_s)} - \log \mid w_8 - HR_{out} \mid \cdot w_6 \cdot HR_{out}^{w_7}$	0.429	0.484	-3.603	-0.071	0.000	6.050	-0.739	50.344	2.19

```
# Model structure: fixed initial part plus additional recursive expression.
<func> ::= w1*Ta+w2*Ti+(exp(w3+w4*Ws))*G <op> <recExpr>
# Expressions
<recExpr> ::= <expr> | <expr> <op> <recExpr>

<expr> ::= <param> <op> <var> | <param> <op> (<var>)^(<param>) |
        exp(abs(<param> <op> <var>)) | log(abs(<param> <op> <var>))

# Parameters for DE
<param> ::= w5|w6|w7|w8https://www.overleaf.com/project/6017340
    b0d36a098212fd46f

# Additional input variables
<var> ::= Wd|HRout|HRin

# Operands
<op> ::= +|-|*
```

Figure S1: Grammar for temperature modeling.

Table S2: Models obtained for *Diffuse*-type days.

		Parameters								
Fold	**Model**	w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	Error
1	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} + \log \mid w_5 - HR_{out} \mid \cdot w_6 \cdot HR_{in}$	-0.157	1.137	-4.547	-0.079	0.000	-0.002	74.225	0.000	1.811
2	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - \log \mid w_7 - HR_{out} \mid \cdot w_6 \cdot HR_{in}$	-0.157	1.137	-4.547	-0.079	0.000	-0.002	74.225	0.000	2.248
3	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - w_8 - HR_{in}^{w_8} \cdot \log \mid w_6 + HR_{out} \mid + e^{\mid w_8 \cdot HR_{out} \mid} - w_7 - HR_{out}^{w_7}$	-0.083	1.075	-4.212	-0.157	0.000	-11.383	-1.417	-0.011	1.987 3
4	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} + w_6 \cdot HR_{in}^{w_5} \cdot w_5 \cdot HR_{out}^{w_8}$	-0.098	1.104	-4.303	-0.113	-2.510	14.483	0.000	0.774	2.223
5	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - \log \mid w_6 - HR_{out} \mid \cdot w_5 \cdot HR_{in}$	-0.178	1.159	-4.539	-0.141	-0.003	73.409	0.000	0.000	2.110
6	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - w_6 \cdot HR_{in}^{w_8} - w_6 \cdot HR_{out}^{w_6} \cdot \log \mid w_5 + HR_{in} \mid$	-0.110	1.089	-4.355	-0.126	-22.598	-0.206	0.000	-2.360	1.895
7	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} + w_5 \cdot HR_{in}^{w_6} + \log \mid w_8 - HR_{in} \mid \cdot w_5 \cdot HR_{in}$	-0.130	1.114	-4.555	-0.085	0.001	-0.938	0.000	22.425	2.014
8	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} + w_7 \cdot HR_{out}^{w_7} \cdot w_7 \cdot HRin \cdot \log \mid w_8 - HR_{in} \mid$	-0.100	1.076	-4.371	-0.108	0.000	0.000	-1.692	22.846	1.855
9	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - w_5 + \log \mid w_8 - HR_{in} \mid \cdot w_7$	-0.037	1.052	-4.344	-0.063	0.550	0.000	0.092	22.234	1.877
10	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - \log \mid w_7 - HR_{out} \mid \cdot w_5 \cdot HR_{in}$	-0.151	1.133	-4.694	-0.025	-0.002	0.000	74.289	0.000	1.550

Table S3: Models obtained for *Cloudy*-type days.

Fold	Model	Parameters								Error
		w_1	w_2	w_3	w_4	w_5	w_6	w_7	w_8	
1	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_8 \cdot HR_{in}^{w_7} \cdot w_6 - HR_{in}^{w_5}$	0.02	0.931	-5.834	-0.071	-2.041	4.976	-0.318	1.724	3.495
2	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_6 \cdot HR_{in}^{w_5} \cdot \log \mid w_6 + HR_{in} \mid$	0.008	0.946	-2.674	-0.080	-0.890	-0.551	0.000	0.000	3.525
3	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_7 \cdot HR_{in}^{w_5}$	0.036	0.913	-3.229	-0.091	-0.320	0.000	0.708	0.000	4.193
4	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} - w_5 \cdot HR_{in}^{w_7} \cdot w_7 + HR_{in}^{w_5}$	0.019	0.917	-1.519	-0.092	-0.388	0.000	-0.322	0.000	3.378
5	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot \log \mid w_7 + HR_{in} \mid \cdot w_8 \cdot HR_{in}^{w_6}$	0.021	0.930	-3.207	-0.091	0.000	-0.901	-0.522	1.022	4.039
6	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_7 \cdot HR_{in}^{w_8} + e^{\mid w_5 + HR_{out} \mid}$	-0.083	0.993	-5.261	-0.077	0.000	0.000	4.615	-0.334	4.106
7	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_8 \cdot HR_{in}^{w_7}$	0.034	0.915	-3.375	-0.093	0.000	0.000	-0.316	0.816	3.993
8	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_8 \cdot HR_{in}^{w_6}$	0.033	0.917	-3.260	-0.091	0.000	-0.319	0.000	0.726	3.33
9	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_8 \cdot HR_{in}^{w_7} \cdot \log \mid w_8 - HR_{in} \mid + \log \mid w_7 - HR_{out} \mid - \log \mid w_8 \cdot HR_{out} \mid$	-0.062	0.968	-2.073	-0.081	0.000	0.000	-0.986	0.361	3.739
10	$T_m = w_1 \cdot T_a + w_2 \cdot T_i + G \cdot e^{(w_3+w_4 \cdot W_s)} \cdot w_8 \cdot HR_{in}^{w_7} \cdot \log \mid w_7 - HR_{in} \mid$	0.026	0.925	-2.076	-0.085	0.000	0.000	-0.825	0.273	3.799