Applied Energy 87 (2010) 442–451

Contents lists available at ScienceDirect

Applied Energy

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

Numerical method for the extraction of photovoltaic module double-diode model parameters through cluster analysis

L. Sandrolini *, M. Artioli, U. Reggiani

Department of Electrical Engineering, Alma Mater Studiorum – University of Bologna, Viale del Risorgimento 2, I-40136 Bologna, Italy

ARTICLE INFO

Article history:
Received 10 February 2009
Received in revised form 10 July 2009
Accepted 23 July 2009
Available online 8 September 2009

Keywords:
Renewable energy sources
Photovoltaic module
Equivalent circuit
Parameter extraction
Particle swarm optimization
Cluster analysis

ABSTRACT

A numerical procedure for the extraction of the double-diode model parameters of photovoltaic (PV) modules is described. A particle swarm optimization algorithm is used to fit the calculated current–voltage characteristic of a PV module to the experimental one. As no recurrent solution was found in the large number of simulations carried out, mainly because of the stochastic nature of the optimization algorithm, statistics in combination with cluster analysis was employed to give an insight into the PV module parameters. This approach allows one to obtain a set of parameters which is reasonable and representative of the physical system.

1. Introduction

The interest in photovoltaic energy is rapidly increasing worldwide because it is renewable, less polluting than conventional energy sources and readily available to all countries. The accurate modelling of PV modules is therefore of primary concern in order to integrate these energy sources in more complex systems (e.g., grid connected systems, remote facilities). In fact, the interfacing of photovoltaic sources to the grid requires one to use efficient power-electronic systems. Moreover, the design of converters with performing maximum power point tracking (MPPT) controllers may take advantage of accurate PV models. Another interesting aspect is that the insight provided by these models may be used as a quality control during the development stage to improve the devices. The simulation of the behaviour of PV modules can be conveniently done by means of equivalent electric circuits that were initially developed for single cells only and adapted to modules composed of several cells afterwards. These models make use of nonlinear lumped-parameter equivalent circuits whose parameters are determined from experimental current–voltage characteristics by means of analytical [1–5] or numerical [6–13] extraction techniques. The former techniques are attractive for their rapidity, as they require the knowledge of a few selected values only (e.g., short-circuit current, open-circuit voltage, current and voltage at the maximum power point, and slopes of the current–voltage characteristic at the axis intersections). However, they are based on simplified formulae and may be affected by the inaccuracy of measurements or calculations. Differently, the latter are based on algorithms that fit the calculated current–voltage characteristic to the entire experimental curve. An evident advantage is that a higher level of confidence of the extracted parameters may be obtained, as all points in the curve are used. A consequent drawback is that these techniques indeed require an extensive computation and their accuracy depends on the fitting algorithm, the fitting criterion, i.e., the objective function to minimize, and the starting values of the parameters to extract. These are particularly important as an inappropriate choice of initial values, often obtained by analytical techniques, may result in nonconvergence of the algorithm or unreliable results [6,8]. This paper discusses the use of a fitting strategy based on a particle swarm optimization (PSO) algorithm and points out how stochastic methods can lead to nonrepeatable results. In fact, random events show a trend when considered on a large scale basis, and therefore a large number of parameter extractions was carried out with two different fitting criteria. As no recurrent solution was found, statistics in combination with cluster analysis was employed to give an insight into the PV module parameters and to improve the parameter extraction.

The paper is organized as follows. Section 2 presents the equivalent circuit for PV modules and a description of the experimental tests that were carried out; besides, an outline of the parameter extraction procedure is given. In Section 3 the results obtained

* Corresponding author. Tel.: +39 0512093484; fax: +39 0512093588.
E-mail addresses: leonardo.sandrolini@mail.ing.unibo.it (L. Sandrolini), marcello.artioli@mail.ing.unibo.it (M. Artioli), ugo.reggiani@mail.ing.unibo.it (U. Reggiani).

0306-2619/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.apenergy.2009.07.022

with the proposed parameter extraction procedure are presented and commented.

2. Current–voltage characteristic of a photovoltaic module

2.1. Equivalent circuit

The well-known double-diode model depicted in Fig. 1 is chosen to represent the physics of a monocrystalline silicon PV module as it is more accurate than the single-diode model in describing its behaviour under low-irradiance conditions [1]. Further, this model is commonly accepted as reflecting the behaviour of polycrystalline silicon PV modules [14]. Both single- and double-diode models, developed and widely used in the past mainly for PV cells, have often been applied to PV modules and arrays, too [3,7,14,15]. One of the two diodes represents the diffusion current in the p–n junction, whereas the other is added to take the space–charge recombination effect into account. With reference to Fig. 1, the current versus voltage characteristic I–V is given by the equation

$$I - I_{ph} + I_{s1}\left[\exp\left(\frac{V+IR_S}{n_1V_T}\right) - 1\right] + I_{s2}\left[\exp\left(\frac{V+IR_S}{n_2V_T}\right) - 1\right] + \frac{V+IR_S}{R_P} = 0, \quad (1)$$

where

I_{ph} is the photocurrent
I_{s1}, I_{s2} are the diode saturation currents
n_1, n_2 are the diode ideality factors
R_S is the series resistance
R_P is the parallel or shunt resistance
$V_T = kT/q$ is the thermal voltage, where k is Boltzmann's constant
T is the PV module absolute temperature and q is the electron charge

Changeable meteorological conditions can be taken into account by introducing a dependence of the model parameters on environmental parameters such as irradiance and temperature [6,14].

2.2. Experimental

The extraction procedure was tested on the sets of data (voltage, current, temperature and irradiance) obtained in natural sunlight from six PV modules of the monocrystalline silicon type connected in series. The modules, mounted with the same tilt angle as shown in Fig. 2, are built by the same manufacturer and are composed of 72 solar cells and have a peak power of 150 W each. A data acquisition system based on National Instruments hardware (DAQPad-6015) and software (LabView) [16] was operated to retrieve for each PV module the experimental voltages and currents at different meteorological conditions when the working point was changed through a variable load. The experimental data sets were

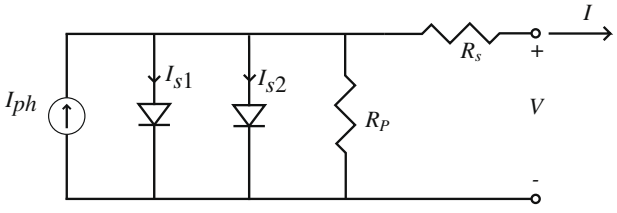


Fig. 1. Double-diode equivalent circuit for a photovoltaic module.

Fig. 2. PV modules connected in series used in the experiments; the installation is on the rooftop of the Department of Electrical Engineering, University of Bologna, Italy.

then classified according to different irradiance and temperature conditions in order to obtain homogeneous current–voltage characteristics. The equivalent circuit parameters for the PV modules were estimated with respect to the conditions of constant temperature of 25 °C and 1000 W/m^2 for the irradiance (measured with the pyranometer Deltaohm LP Pyra 03), in compliance with the EN 60904 standard [17].

2.3. Parameter extraction procedure

The procedure implemented to fit the PV module seven unknown parameters $(I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P)$ to each PV module experimental data set relies on a two-step procedure which

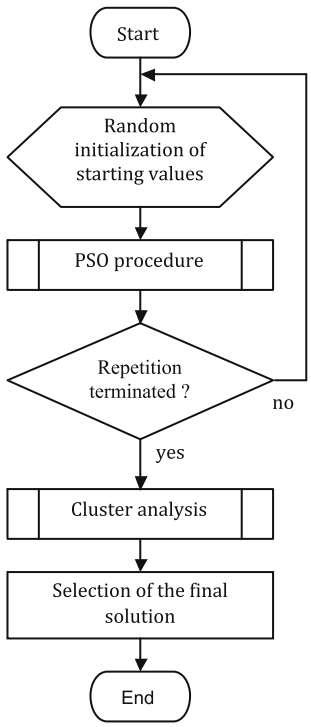


Fig. 3. Top-level flow-chart of the two-step procedure to extract the PV module parameters.

combines the PSO strategy [18,19] and cluster analysis [20–22] (see Fig. 3). The PSO optimization algorithm has been proved efficient in finding global minima for nonlinear problems. The PSO algorithm belongs to the so-called population based search methods (e.g., genetic algorithms, evolutionary programming, evolution strategies, genetic programming, etc.), that can solve a wide variety of optimization problems in a search space making use of the "swarm intelligence". The PSO concept is quite simple, and the few parameters to adjust make a general version of the algorithm applicable to several problems with slight changes only. The swarm strategy simulates the social behaviour of flocking organisms and places its effectiveness in information sharing and successful identification of a leader. Each particle in the swarm is a possible solution of an optimization problem and it moves with an adaptable velocity that depends on its own flying experience (i.e., cognitive term) and on the experience of the other particles in the swarm (i.e., social term). In practice, the former accelerates the particle towards its best position, whereas the latter towards the overall best position. The PSO algorithm is outlined in Fig. 4. The initialization of the swarm and the velocities is usually performed randomly in the search space, and the particles (i.e., the sets of seven unknown parameters) vary their position until either a relatively unchanging state is found, or a prefixed number of iterations is reached. The initial search space is a seven-dimensional space defined by fixed ranges of variability of the seven parameters of the double-diode equivalent circuit. The solution will be a seven element vector that minimizes a given objective function. The ith particle of a set of q individuals $(i = 1, 2, \ldots, q)$ in a seven-dimensional search space contains information about its current position and velocity, that can be represented as $X_i = \{x_{i1}, x_{i2}, \ldots, x_{i7}\}$ and $V_i = \{v_{i1}, v_{i2}, \ldots, v_{i7}\}$, respectively. The best particle position found in the considered iteration is stored as $P_i = \{p_{i1}, p_{i2}, \ldots, p_{i7}\}$. The equations that govern the current velocity and position of the ith particle in a seven-dimensional space are [23]

$$v_{ij} = w v_{ij} + c_1 U_1(0,1)(p_{ij} - x_{ij}) + c_2 U_2(0,1)(p_{gj} - x_{ij}), \quad (2)$$

$$x_{ij} = x_{ij} + v_{ij}, \quad (3)$$

where $j = 1, 2, \ldots, 7$, w is an inertia weight, g represents the index of the particle that gives the best position achieved by all particles of the population in the considered iteration, c_1 and c_2 are two positive constants, and $U_1(0,1)$ and $U_2(0,1)$ are two random functions that yield a real number in the range [0,1]. The term v_{ij} represents the rate of displacement of the ith particle along the jth direction of the seven-dimensional space in an iteration. All the q particles are supposed to move simultaneously during an iteration, that lasts a unit of time. The key concept for the particle swarm optimizer consists in changing the velocity of each particle towards its best result and the overall best solution. The inertia weight w in the first term of (2) is a parameter introduced to bias the search towards global rather than local minima, or vice versa. It can be a positive real constant or a linear or nonlinear function of time, i.e., of the number of iterations. Its introduction has improved performance of the method in a number of applications. A choice of equal values for the constants c_1 and c_2 seems to be the most effective. In [18,23], the recommended value is an integer number, e.g., 2, as larger values may cause some particles to wander in the search space, while smaller ones may address the swarm towards local minima. In (2), the second term represents a "cognition" model, whereas the third a "social" one [18,23]. The cognition model expresses the private thinking of a particle, that acts on its own experience, while the social one represents collaboration among particles. The model proposed by (2), (3) takes account of both aspects at the same time. In practice, in (2) the second term accelerates the particle towards its best position, whereas the third towards the overall best position.

In the proposed procedure, the parameters are initialized randomly within their search ranges and therefore are not conditioned by a set of initial values; rather, they are conditioned by the choice of proper starting search ranges. Another critical point in numerical techniques is the fitting criterion. For instance, it will be shown in the next section that minimizing the area between the experimental curve and that predicted with the extracted parameters [24] may not result in a good approximation of the experimental curve. Better results were obtained with a fitting criterion based on the distance between each point of the experimental characteristic and the corresponding one on the calculated curve. The main computational costs of this optimization procedure are proportional to the number of population particles, the maximum number of iterations needed to reach the prefixed convergence and the number of points that (1) has to fit to. The PSO algorithm (i.e., the first step of the extraction procedure) was implemented in Mathematica 6.0 [25]. The cluster analysis (i.e., the second step of the extraction procedure), also implemented in Mathematica 6.0, will be discussed later, once the partial results from the PSO step are presented. In fact, the cluster analysis acts as a correction filter on the data obtained with the PSO analysis.

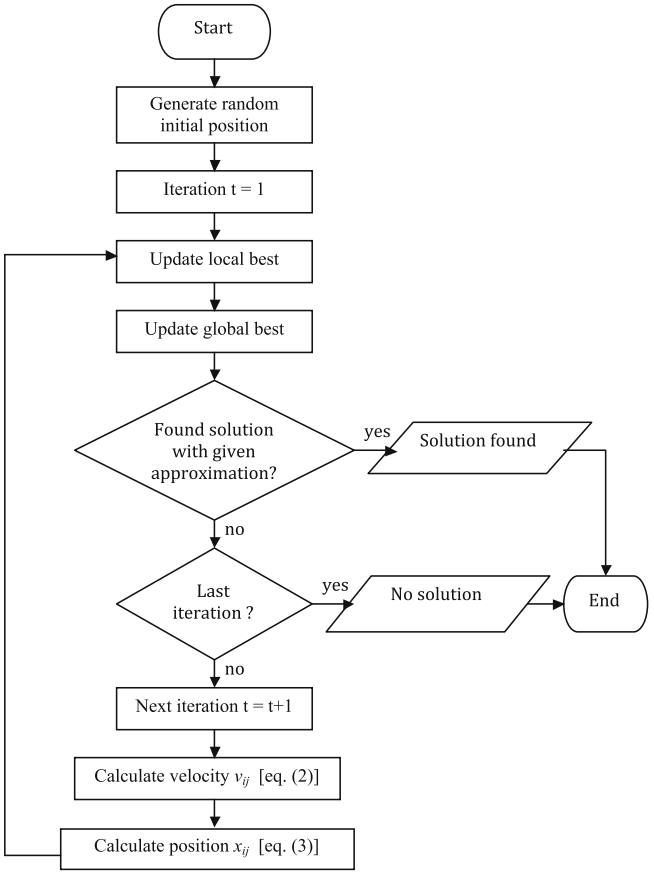


Fig. 4. PSO algorithm flow-chart.

3. Results and discussion

It has been observed that numerical extraction techniques which exploit data fitting algorithms (e.g., simplex algorithm, Marquardt–Levenberg algorithm, optimization algorithm) suffer from statistical instability and may result in parameters which are not realistic. Statistics has been used in [26–28] to take the experimental uncertainty into account and in [29] to process few experimental characteristics at various irradiance levels. In fact, it is acknowledged that outdoor testing of PV modules requires a vast

amount of data to ensure a reliable and statistically acceptable module characterization. Hybrid methods have been proposed, such as in [30], where a neural network was used to guide the evolutionary direction of a genetic or PSO algorithm. In [29], a number of possible solutions (i.e., PV cell parameters) were obtained by the repeated application of an iterative least-square algorithm to the set of experimental curves. The parameter frequency distributions clearly show that even when a specific irradiance level and a given temperature are considered, it is unlikely to obtain recurrent values for the PV module parameters. This statistical analysis approach was initially used also in the case of the proposed PSO procedure of parameter extraction. In order to investigate and assess its validity, the extraction procedure was repeatedly applied to the measured current–voltage characteristic of each PV module a large number of times. Several sets of the seven parameters that fit (1) to each experimental curve with different fitting qualities (and for two different fitting criteria, the area criterion and the distance criterion) were found. As the six PV modules are of the same model and built by the same manufacturer, the extracted parameters were expected to be only slightly different; in this way, some variability of the measurement can be taken into account. The data post-processing was implemented in Mathematica 6.0.

3.1. Fitting and convergence

Table 1 shows a summary of the simulations made using the area criterion with a requested quality of 0.0001 for each PV module, that is, the area of the calculated characteristic differs from the corresponding measured one less than 0.01%. Thus, a zero quality means that the areas of the calculated and experimental characteristics are equal. The table contains the information related to the number of executed/convergent runs (where a convergent run is a run which reaches the requested quality and represents a solution), the mean and best quality obtained plus the number of clusters where the solutions can be grouped (this will be discussed in Section 3.3). Repeated fits of (1) to each PV module experimental characteristic do not show recurrent results, mainly because of the stochastic nature of the PSO algorithm. The over 4000 runs completed (about 700 for each module) showed a very high convergence ratio: more than 90% of the runs reached the requested quality. According to this criterion, the lower the quality index the better should be the fitting because of the smaller difference between the areas of the calculated and measured curves. The mean quality index is good even if evaluated over the entire set of the simulations; this means that the nonconvergent runs also yield a good quality index. Of course, this index is smaller (better) if evaluated over the set of the convergent runs only.

The same kind of information was extracted from the simulations made using the distance criterion with a requested quality of 0.01 (see Table 2). This criterion requires that each point of the calculated characteristic must have a normalized distance from the corresponding point of the measured curve not greater than 1%, where the normalization value is the maximum current value of the experimental characteristic. Although the requested quality index is much higher than that used with the area criterion, the convergence ratio is null over a smaller but significant number of runs (2400), i.e., no simulation reached the requested quality. This fact immediately suggests that the distance criterion is stricter than the area one, and thus it is unlikely to reach convergence within the same number of iterations. It can be noticed that the PV modules show different mean fitting qualities; the best and worst one are obtained by the fourth and sixth PV modules, respectively.

Table 3 shows the results of more than 5000 simulations made using the same distance criterion and a requested quality of 0.02, which is a weaker convergence request than the previous ones. In this case, the convergence ratio of all PV modules but one is not null and slightly lower than the ratio obtained with the area criterion, confirming the major severity of the distance criterion. The nonconvergence of all simulations for PV module #6 may be due to the behaviour of this module, which is different than the others as it is shown in Fig. 5; this was not evident when the area fitting criterion was used.

3.2. Distribution of parameters

The analysis of the spatial distribution of the solutions reveals some interesting facts. For each parameter the solution space is split into ten equal-sized bins (classes) and the occurrences of

Table 1
Summary of simulations carried out with the area criterion (quality 0.01%).

	Executed runs	Convergent runs	Convergence ratio	Overall mean quality	Convergent mean quality	Convergent best quality	Convergent run clusters	Convergent run cluster dimension
All modules	4140	3796	0.917	1.72×10^{-4}	1.66×10^{-5}	0.	1	2.73
Module #1	690	639	0.926	1.10×10^{-4}	1.71×10^{-5}	1.60×10^{-16}	1	2.18
Module #2	690	634	0.919	1.81×10^{-4}	1.79×10^{-5}	0.	1	2.13
Module #3	690	628	0.910	9.68×10^{-5}	1.70×10^{-5}	0.	1	1.85
Module #4	690	636	0.922	2.16×10^{-4}	1.62×10^{-5}	1.62×10^{-16}	1	2.32
Module #5	690	635	0.920	2.87×10^{-4}	1.52×10^{-5}	2.78×10^{-15}	1	2.29
Module #6	690	624	0.904	1.39×10^{-4}	1.62×10^{-5}	0.	1	1.88

Table 2
Summary of simulations carried out with the distance criterion (quality 1%).

	Executed runs	Convergent runs	Convergence ratio	Overall mean quality	Convergent mean quality	Convergent best quality	Overall run 2clusters
All modules	2400	0	0	1.90×10^{-2}	–	–	1
Module #1	400	0	0	1.92×10^{-2}	–	–	1
Module #2	400	0	0	1.66×10^{-2}	–	–	3
Module #3	400	0	0	1.78×10^{-2}	–	–	3
Module #4	400	0	0	1.63×10^{-2}	–	–	1
Module #5	400	0	0	1.95×10^{-2}	–	–	1
Module #6	400	0	0	2.45×10^{-2}	–	–	1

Table 3
Summary of simulations carried out with the distance criterion (quality 2%).

	Executed runs	Convergent runs	Convergence ratio	Overall mean quality	Convergent mean quality	Convergent best quality	Convergent run clusters	Convergent run cluster dimension
All modules	5400	3922	0.726	2.08×10^{-2}	–	1.50×10^{-2}	1	0.064
Module #1	900	802	0.891	2.04×10^{-2}	1.94×10^{-2}	1.71×10^{-2}	1	0.008
Module #2	900	798	0.887	1.97×10^{-2}	1.86×10^{-2}	1.51×10^{-2}	3	0.005;0.001;0.002
Module #3	900	775	0.861	1.97×10^{-2}	1.89×10^{-2}	1.64×10^{-2}	1	0.036
Module #4	900	778	0.864	1.98×10^{-2}	1.85×10^{-2}	1.50×10^{-2}	1	0.021
Module #5	900	769	0.854	2.06×10^{-2}	1.95×10^{-2}	1.80×10^{-2}	1	0.008
Module #6	900	0	0	2.47×10^{-2}	–	–	–	

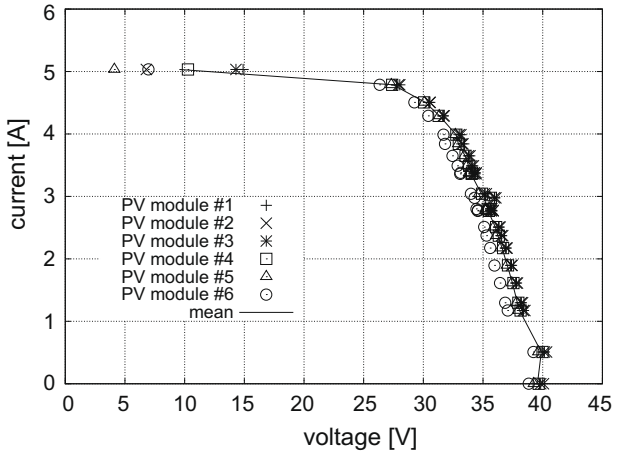


Fig. 5. Measured current–voltage characteristics of the six PV modules at 1000 W/m^2 and 25 °C.

the parameters that fall into each bin are counted. The size of each bin is one tenth of the solution space width (see Table 4). The overall frequency distributions of all seven parameters for all six PV modules obtained with the area fitting criterion are presented in Fig. 6, where the content of Table 5 is plotted as histograms. More bins could be used but this would worse the readability of the tables and figures which follow without altering the sense of the analysis. The peaks of the overall distribution of each parameter belong to different classes (in this and in the following tables the peaks are boxed, together with their percentage). This occurs also for the parameters extracted from each PV module alone; for instance, Table 6 presents the results obtained for the PV module #1. The same considerations can be replicated for the parameters extracted with the distance fitting criterion, as shown in Table 7 for all PV modules and in Table 8 for the PV module #1.

The graphical evaluation of the fit can be done with respect to any PV module experimental curve (with the exception of the PV module #6) as the characteristics are very close to each other, as shown in Fig. 5. Referring to the solutions obtained with the area fitting criterion, it has to be noticed that the values which fall in the peak classes of the overall parameter distributions are uncorrelated with the quality of the solutions, as Figs. 7 and 8 show. In the former all parameters fall outside the peak classes (grey boxes) but the fitting to the experimental characteristic of the PV module #4 is good, whereas in the latter a poor fit to the PV module #4 is depicted notwithstanding that almost all parameters belong to the peak classes. Moreover, it is not feasible to choose the mean values of the overall distributions as representative of each parameter. In fact, the mean values do not fall into the peak classes and may give a bad fitting, too, as Fig. 9 shows. With the distance fitting criterion the fitting to the experimental curves is rather good using the mean values of the overall distributions that, as in the previous case, do not belong to the peak classes (see Fig. 10 where the PV module #4 is used to assess the fit).

Another way to select a set of parameters which possibly represent the PV modules may be to choose the mean value of each peak class. Again, using the area fitting criterion these mean values fit the experimental curves poorly (see Fig. 11); instead, as shown in Fig. 12, the distance criterion gives a fair fitting, although worse than that showed in Fig. 10.

With respect to published results, all the parameters but n_2 show feasible values; indeed, the second diode ideality factor seems large in comparison with n_1. It is noticed that the normal approach for single PV cells is to assume the diode ideality factors N_1 and N_2 equal to 1 and 2, respectively. In the case of a PV module where p cells are connected in series, it is reasonable to think that the diode ideality factors are of the order of $n_1 = p \cdot N_1$ and $n_2 = p \cdot N_2$ [3,7,15]. The modules considered in the paper are composed of 72 PV cells in series, so one could expect $n_1 = 72$ and $n_2 = 144$. The data in Figs. 7–13 show that the values found for n_1 are consistent with the above hypothesis, whereas the values for n_2 are slightly larger. Further investigations are therefore required on this point.

In light of these facts, the area fitting criterion seems to be less satisfactory than the distance one when the PV module parameters are calculated as the mean of a set of solutions. Thus, the importance of choosing an appropriate fitting criterion is evident. Anyhow, a question still remains, whether the extracted parameters calculated as the mean values either of the overall distributions

Table 4
Solution space of $I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P$ (convergent solutions only) for both criteria.

	Initial search space	Area criterion all modules	Area criterion module #1	Distance criterion all modules	Distance criterion module #1
Iph	[4.90, 5.20]	[4.90, 5.20]	[4.90, 5.20]	[5.00, 5.18]	[5.00, 5.18]
Is1	$[0.00, 1.00 \times 10^{-6}]$	$[6.35 \times 10^{-10}, 1.00 \times 10^{-6}]$	$[7.31 \times 10^{-9}, 9.95 \times 10^{-7}]$	$[6.37 \times 10^{-10}, 1.00 \times 10^{-6}]$	$[7.41 \times 10^{-9}, 9.98 \times 10^{-7}]$
Is2	$[0.00, 1.00 \times 10^{-3}]$	$[3.05 \times 10^{-7}, 9.98 \times 10^{-4}]$	$[1.97 \times 10^{-6}, 9.92 \times 10^{-4}]$	$[3.52 \times 10^{-8}, 1.00 \times 10^{-3}]$	$[2.39 \times 10^{-6}, 9.97 \times 10^{-4}]$
n1	$[0.00, 2.50 \times 10^{2}]$	$[6.46 \times 10^{1}, 2.45 \times 10^{2}]$	$[8.53 \times 10^{1}, 2.38 \times 10^{2}]$	$[6.89 \times 10^{1}, 1.24 \times 10^{2}]$	$[7.71 \times 10^{1}, 1.01 \times 10^{2}]$
n2	$[0.00, 1.00 \times 10^{3}]$	$[1.38 \times 10^{2}, 9.92 \times 10^{2}]$	$[1.38 \times 10^{2}, 9.74 \times 10^{2}]$	$[1.17 \times 10^{2}, 8.78 \times 10^{2}]$	$[1.31 \times 10^{2}, 5.02 \times 10^{2}]$
Rs	[0.00, 4.00]	$[1.78 \times 10^{-2}, 3.99]$	$[3.38 \times 10^{-2}, 3.99]$	$[5.86 \times 10^{-1}, 1.15]$	$[6.41 \times 10^{-1}, 9.78 \times 10^{-1}]$
Rp	$[0.00, 1.50 \times 10^{4}]$	$[3.18 \times 10^{1}, 1.50 \times 10^{4}]$	$[3.45 \times 10^{1}, 1.48 \times 10^{4}]$	$[2.27 \times 10^{2}, 1.50 \times 10^{4}]$	$[4.27 \times 10^{2}, 1.50 \times 10^{4}]$

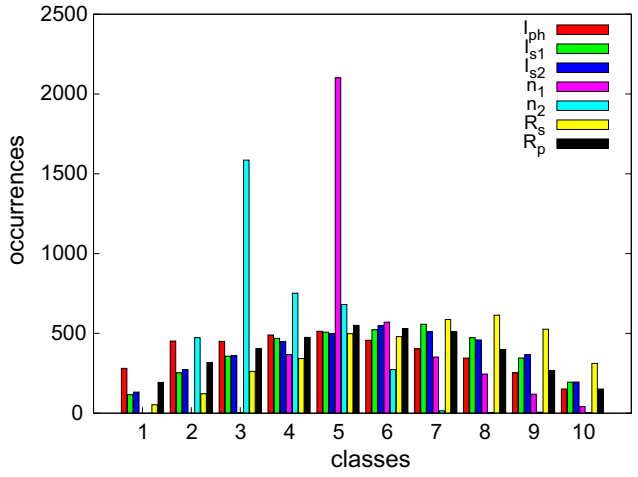


Fig. 6. Overall frequency distributions of $I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P$ (area fitting criterion, see Table 5).

or of the peak classes form a solution which has a physical meaning or not. Cluster analysis for the solutions was then adopted to address this issue.

3.3. Clusters of solutions

As mentioned before, after a number of possible solutions are obtained the main problem consists in recognizing which solution is the most representative of each PV module. Each solution is a vector of seven parameters (seven dimensions) and does not allow either a direct visual inspection or any plotting. Thus, an automated and intelligent (in other words, unsupervised) technique has to be used. The proposed technique is depicted in the flowchart of Fig. 14.

Clustering a multidimensional data set provides a systematic way to partition such data into groups based on patterns having similar features. Although there exist hybrid evolutionary algorithms based on clusters [20–22], at this stage the search for clusters was implemented solely as a post-processing procedure on the database of the extracted parameters. This is mainly due to

Table 5
Overall frequency distributions of $I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P$ (area fitting criterion, 0.01% quality, all modules).

	Class #1	Class #2	Class #3	Class #4	Class #5	Class #6	Class #7	Class #8	Class #9	Class #10
Iph	281	452	450	490	513 (0.14)	457	403	345	253	152
Is1	116	253	357	469	508	523	557 (0.15)	473	346	194
Is2	132	274	361	449	499	548 (0.14)	512	459	367	195
n1	0	0	1	367	2101 (0.55)	570	352	245	119	41
n2	0	473	1586 (0.42)	752	681	274	15	5	6	4
Rs	53	122	262	342	498	480	587	614 (0.16)	526	312
Rp	193	317	404	475	550 (0.14)	529	510	399	268	151

Table 6
Frequency distributions of $I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P$ (area fitting criterion, 0.01% quality, PV module #1 only).

	Class #1	Class #2	Class #3	Class #4	Class #5	Class #6	Class #7	Class #8	Class #9	Class #10
Iph	51	77	77	81	90(0.14)	73	61	50	45	34
Is1	18	43	67	75	84	85	100(0.16)	85	49	33
Is2	24	46	72	71	78	91(0.14)	81	77	63	36
n1	0	0	0	38	350(0.55)	116	62	43	20	10
n2	0	74	282(0.44)	127	97	49	5	2	1	2
Rs	8	14	40	59	81	88	113(0.18)	95	84	57
Rp	44	39	75	75	86	87(0.14)	79	72	54	28

Table 7
Overall frequency distributions of $I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P$ (distance fitting criterion, 2% quality, all modules).

	Class #1	Class #2	Class #3	Class #4	Class #5	Class #6	Class #7	Class #8	Class #9	Class #10
Iph	0	0	0	568	709	872	973(0.25)	655	142	3
Is1	147	283	403	445	516	529	563(0.14)	498	339	199
Is2	261	388	461	476	523	588(0.15)	479	356	245	145
n1	0	0	3	3517(0.90)	402	0	0	0	0	0
n2	0	226	1468(0.37)	1203	1017	7	0	0	1	0
Rs	0	852	3070(0.78)	0	0	0	0	0	0	0
Rp	162	315	409	489	544	606(0.15)	492	424	312	169

Table 8
Frequency distributions of $I_{ph}, I_{s1}, I_{s2}, n_1, n_2, R_S, R_P$ (distance fitting criterion, 2% quality, PV module #1 only).

	Class #1	Class #2	Class #3	Class #4	Class #5	Class #6	Class #7	Class #8	Class #9	Class #10
Iph	1	2	3	115	159	222 (0.25)	215	81	69	33
Is1	42	77	103	110	96	107	118	127 (0.14)	74	46
Is2	62	79	102	124	134	141 (0.16)	100	72	54	32
n1	0	0	0	760 (0.84)	96	25	17	2	0	0
n2	0	116	320 (0.36)	232	223	4	0	3	2	0
Rs	56	458 (0.51)	386	0	0	0	0	0	0	0
Rp	35	73	93	119	129	143 (0.16)	104	84	76	44

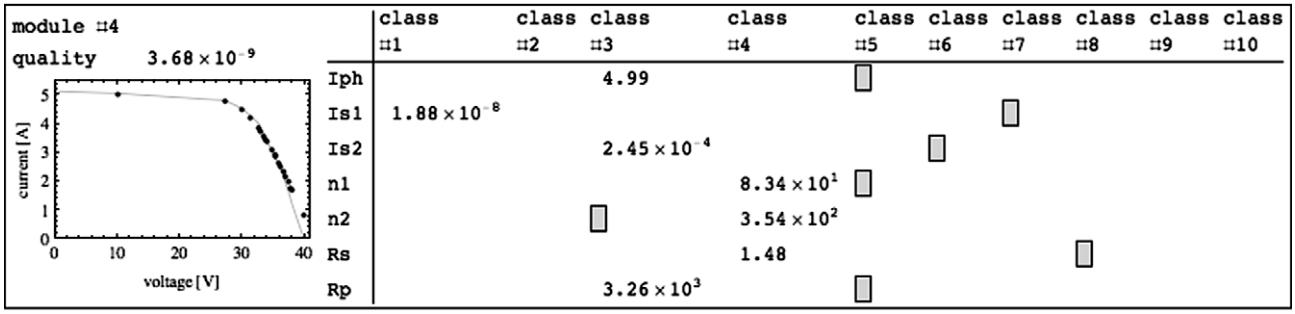


Fig. 7. Good fitting solution (area fitting criterion) whose parameters do not fall into the peak classes for the PV module #4.

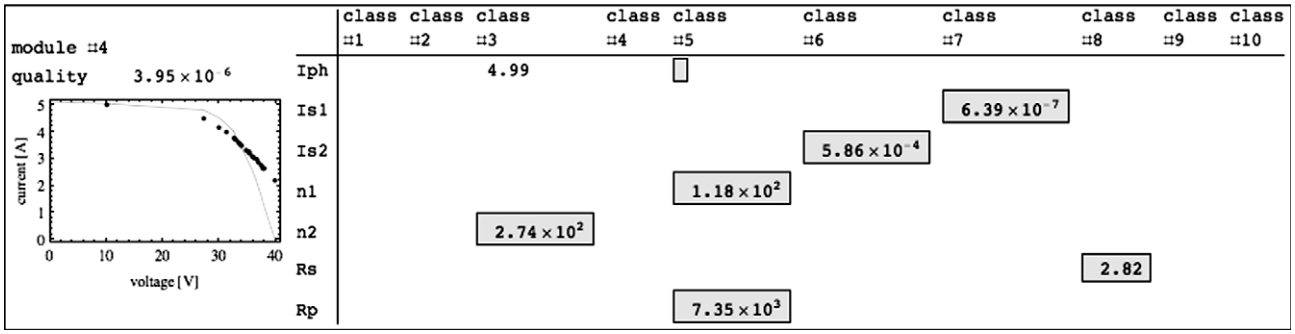


Fig. 8. Poor fitting solution (area fitting criterion) with six of the seven parameters which fall into the peak classes for the PV module #4.

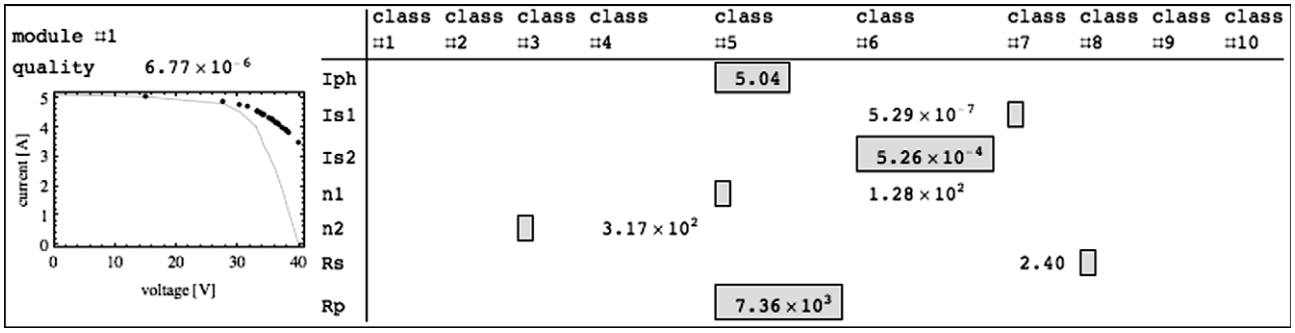


Fig. 9. Mean values of overall distributions (area fitting criterion) and their fit to the characteristic of the PV module #1.

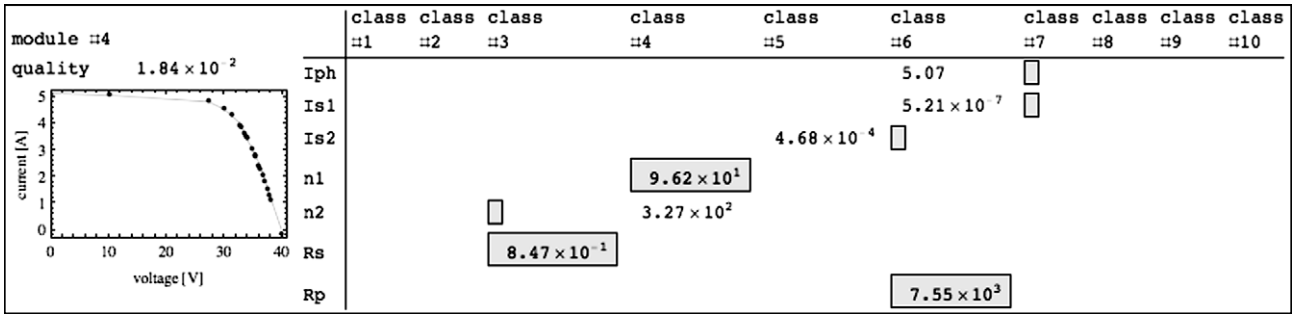


Fig. 10. Mean values of overall distributions (distance fitting criterion) and their fit to the characteristic of the PV module #4.

the fact that, although these hybrid methods are intended to improve the evolutionary algorithm performance, they do not alter the stochastic nature of the algorithm itself and result in an increased computational burden. It is worthwhile to mention that, in general, different clustering techniques may result in a different layout of clusters because each technique has its own dissimilarity criteria (i.e., distance function), initial guessing methods, etc. Mathematica 6.0 offers indeed a great variety of criteria and methods; it

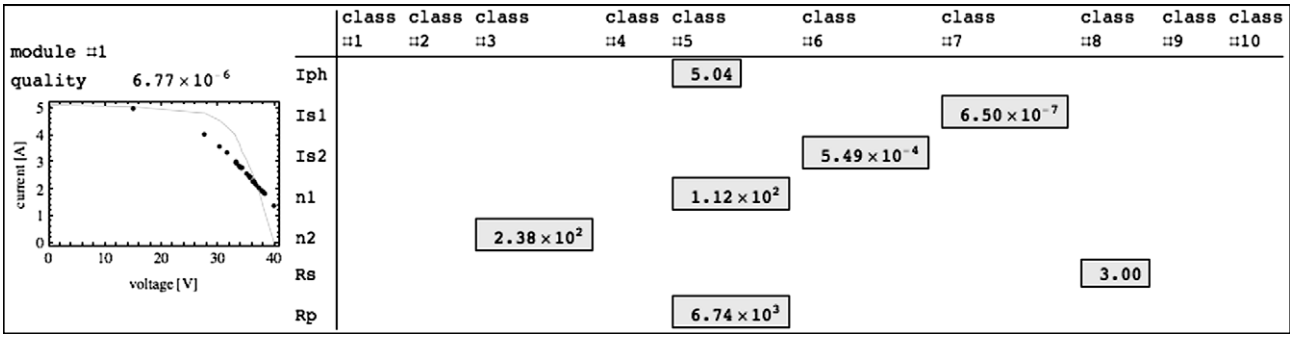


Fig. 11. Mean values of the peak classes (area fitting criterion) and their fit to the characteristic of the PV module #1.

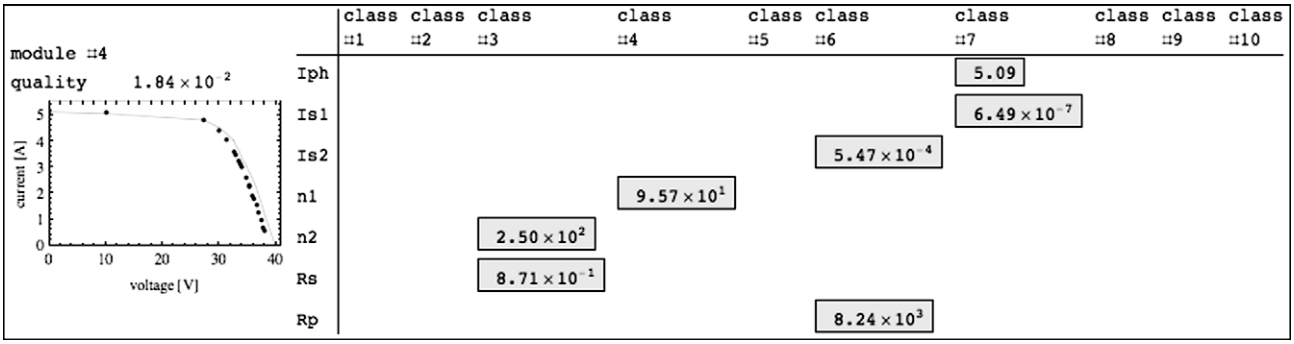


Fig. 12. Mean values of the peak classes (distance fitting criterion) and their fit to the characteristic of the PV module #4.

module #2

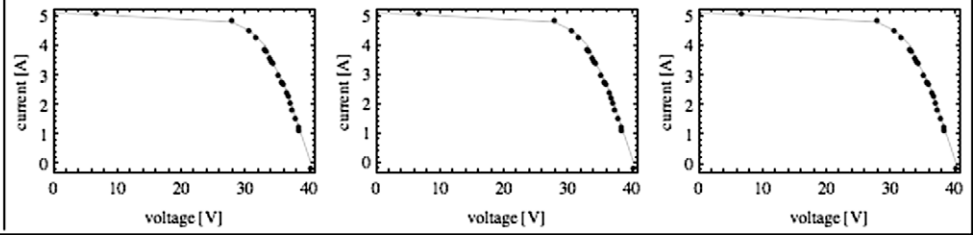
	cluster 1	cluster 2	cluster 3
elements	201	323	274
min distance	6.87899	3.46384	4.61578
max distance	4768.39	4346.23	5581.75
volume	0.00549236	0.00136736	0.00179349
density	36596.3	236221.	152775.
mean Iph	5.08525	5.07559	5.07128
mean Is1	5.06967×10^{-7}	5.12576×10^{-7}	5.16589×10^{-7}
mean Is2	0.000431847	0.000481793	0.000452849
mean n1	96.863	96.8177	97.0049
mean n2	328.623	336.361	322.392
mean Rs	0.863222	0.865502	0.857966
mean Rp	2812.72	7250.33	11530.1
mean fitting			

Fig. 13. Details of the clusters found for the PV module #2.

was chosen to use the Euclidean norm as the dissimilarity function and silhouette statistics to test how well the data are clustered [25].

With reference to Table 1, the last two columns of the table report the number of clusters found in the sets of the convergent runs and the cluster dimension in order to point out groups of similar solutions: the solutions with the area fitting criterion form a single cluster for each PV module and for all PV modules together. Naturally, when solutions gather in a single cluster, the parameter mean values of the cluster solutions coincide with the mean values of the parameter distributions. The dimension of the clusters, calculated by multiplying all the parameter solution ranges, is

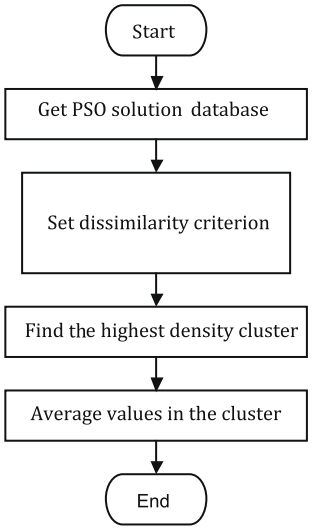


Fig. 14. Clustering analysis flow-chart.

comparable to that of the initial search space (4.5 for all simulations); thus, whichever mean is calculated, it would be obtained from parameters spread over ranges rather large with respect to the initial search ones. This may explain the poor fit obtained when this fitting criterion was employed (see Figs. 9 and 11); the set of parameter mean values may not correspond to a solution of adequate quality. Table 3 shows that, with the exception of the PV module #2, single clusters are found when the distance criterion is used, and that all solutions together gather in a single cluster. However, the cluster dimensions are much smaller than in the previous case (so are the ranges of the extracted parameters, that is, the solutions are less spread and closer to a value that can be therefore representative for the whole group) and that may explain the good fit obtained when a solution is formed as the mean of the extracted parameters, as Figs. 10 and 12 show. The cluster dimension seems then to indicate whether the mean values of the parameter distributions can adequately represent the PV modules. In addition, when more than one cluster is found, as for the PV module #2, the mean values of the distributions can not be indicative of a reasonable solution. In fact, a different set of parameters can be obtained from each cluster, as Fig. 13 shows. As it can be noticed, each set may still fit well (1) to the experimental characteristic. One then may resort to statistics and choose the set of parameters corresponding to the cluster which has the highest density (that is the ratio between the number of solutions and the cluster dimension). Cluster analysis applied to the solutions (given by the PV modules from #1 to #5) yields the results already reported in Table 3 and plotted in Fig. 15, where for each PV module the mean values of the parameters of the solutions belonging to a cluster are normalized to the parameter mean values of all solutions. Fig. 15 shows that the parameters extracted from the PV modules #1, #3, #4 and #5 are very similar and close to the mean values; among the three clusters found for the PV module #2, cluster 2 (the highest density one) seems to give the best results as it is composed of solutions whose parameters are closer to the parameters of the other modules and to the mean values.

4. Conclusions

A large amount of data is necessary when numerical algorithms are employed to have a reliable and statistically acceptable electrical characterization of PV modules. A large number of parameter extractions were carried out using a PSO algorithm and two different fitting criteria in order to obtain the parameter distributions. It is shown that the fitting criterion based on distance of the experimental characteristics is more appropriate than that on the area. In general, it is not feasible to choose the mean values of the overall distributions as representative of each parameter; cluster analysis of the solutions is then implemented and shows that the cluster dimension seems to indicate whether the parameter mean values represent adequately the PV modules. When the solutions gather in different clusters, the analysis suggests to choose the set of parameters in correspondence with the highest density cluster. Further steps in the analysis may imply clustering techniques to be implemented directly in the PSO algorithm. The results obtained are deemed satisfactory and the procedure could find an useful application to the development of PV cells.

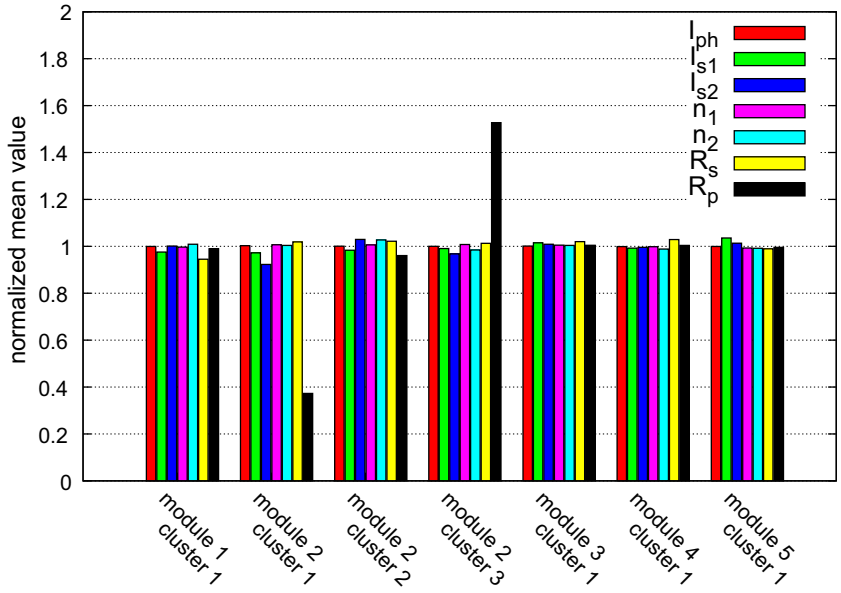


Fig. 15. Normalized mean values of the extracted parameters for the PV modules under exam.

Acknowledgment

The experiments performed in this work were financially supported by the Italian University Ministry (MUR) under a Program for the Development of Research of National Interest (PRIN Grant No. 2002098425_002).

References

[1] Chan DSH, Phang JCH. Analytical methods for the extraction of solar-cell single- and double-diode model parameters from I–V characteristics. IEEE Trans Electron Dev 1987;ED-34(2 pt. 1):286–93.
[2] Charles JP, Bordure G, Khoury A, Mialhe P. Consistency of the double exponential model with the physical mechanisms of conduction for a solar cell under illumination. J Phys D – Appl Phys 1985;18:2261–8.
[3] De Blas MA, Torres JL, Prieto E, Garcia A. Selecting a suitable model for characterizing photovoltaic devices. Renew Energy 2002;25(3):371–80.
[4] Ortiz-Conde A, Garcia Sanchez FJ, Muci J. New method to extract the model parameters of solar cells from the explicit analytic solutions of their illuminated i–v characteristics. Solar Energy Mater Solar Cells 2006;90(3):352–61.
[5] De Soto W, Klein S, Beckman W. Improvement and validation of a model for photovoltaic array performance. Solar Energy 2006;80(1):78–88.
[6] Gottschalg R, Rommel M, Infield DG, Kearney MJ. The influence of the measurement environment on the accuracy of the extraction of the physical parameters of solar cells. Meas Sci Technol 1999;10(9):796–804.
[7] Nakanishi F, Ikegami T, Ebihara K, Kuriyama S, Shiota Y. Modeling and operation of a 10 kW photovoltaic power generator using equivalent electric circuit method. In: Proceedings of 28th photovoltaic specialists conference; 2000. p. 1703–6.
[8] Jervase JA, Bourdoucen H, Al-Lawati A. Solar cell parameter extraction using genetic algorithms. Meas Sci Technol 2001;12(11):1922–5.
[9] Chegaar M, Ouennoughi Z, Guechi F. Extracting dc parameters of solar cells under illumination. Vacuum 2004;75:367–72.
[10] Haouari-Merbah M, Belhamel M, Tobias I, Ruiz J. Extraction and analysis of solar cell parameters from the illuminated current–voltage curve. Solar Energy Mater Solar Cells 2005;87(1–4):225–33.
[11] Chegaar M, Azzouzi G, Mialhe P. Simple parameter extraction method for illuminated solar cells. Solid State Electron 2006;50:1234–7.
[12] Duffie JA, Beckman WA. Solar engineering of thermal processes. third ed. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2006.
[13] Liu CC, Chen CY, Weng CY, Wang CC, Jenq FL, Cheng PJ, et al. Physical parameters extraction from current–voltage characteristic for diodes using multiple nonlinear regression analysis. Solid State Electron 2008;52(6):839–43.
[14] Gow JA, Manning CD. Development of a photovoltaic array model for use in power–electronics simulation studies. IEE Proc – Electric Power Appl 1999;146(2):193–200.
[15] Celik A, Acikgoz N. Modelling and experimental verification of the operating current of mono-crystalline photovoltaic modules using four- and five-parameter models. Appl Energy 2007;84(1):1–15.
[16] National Instruments Corporation. LabView v. 8.2, National Instruments Corporation; 2007.
[17] EN 60904-1. Photovoltaic devices – part 1: measurement of photovoltaic current–voltage characteristics. Technical Report, CENELEC; 1993.
[18] Kennedy J, Eberhart RC. Particle swarm optimization. In: IEEE international conference on neural networks, Perth, Australia; 1995. p. 1942–8.
[19] Eberhart RC, Shi Y. Particle swarm optimization: developments, applications and resources. In: 2001 congress on evolutionary computation, Seoul, Korea; 2001. p. 81–6.
[20] Sarkar M, Yegnanarayana B, Khemani D. Clustering algorithm using an evolutionary programming-based approach. Pattern Recogn Lett 1997;18(10):975–86.
[21] Kennedy J. Stereotyping: improving particle swarm performance with cluster analysis. In: Proceedings of the IEEE conference on evolutionary computation, ICEC, vol. 2. IEEE, Piscataway, NJ, USA, California, CA, USA; 2000. p. 1507–12.
[22] Omran M, Salman A, Engelbrecht A. Dynamic clustering using particle swarm optimization with application in image segmentation. Pattern Anal Appl 2006;8(4):332–44.
[23] Shi Y, Eberhart RC. A modified particle swarm optimizer. In: IEEE international conference on evolutionary computation, Anchorage, USA; 1998. p. 69–73.
[24] Chan DSH, Phillips JR, Phang JCH. A comparative study of extraction methods for solar cell model parameters. Solid State Electron 1986;29(3):329–37.
[25] Wolfram Research, Inc. Mathematica v. 6.0, Wolfram Research, Inc., Champaign, IL, USA; 2007.
[26] Veissid N, Bonnet D, Richter H. Experimental investigation of the double exponential model of a solar cell under illuminated conditions: considering the instrumental uncertainties in the current, voltage and temperature values. Solid State Electron: Int J 1995;38(11):1937–43.
[27] Bouzidi K, Chegaar M, Bouhemadou A. Solar cells parameters evaluation considering the series and shunt resistance. Solar Energy Mater Solar Cells 2007;91(18):1647–51.
[28] Kim HS, Morris BG, Han SS, May GS. A comparison of genetic and particle swarm optimization for contact formation in high-performance silicon solar cells. In: Proceedings of the international joint conference on neural networks. Hong Kong, China: Institute of Electrical and Electronics Engineers Inc.; 2008. p. 1531–5.
[29] Lasnier F, Ang TG. Photovoltaic engineering handbook. Bristol, Great Britain: Adam Hilger; 1990.
[30] Li Y. Hybrid intelligent approach for modeling and optimization of semiconductor devices and nanostructures. Comput Mater Sci 2009;45(1):41–51.