European Association for the Development of Renewable Energies, Environment and Power Quality

International Conference on Renewable Energies and Power Quality (ICREPQ'09)

Valencia (Spain), 15th to 17th April, 2009

Identification of Photovoltaic Array Model Parameters by Robust Linear Regression Methods

Maria Carmela Di Piazza, *Member IEEE*, Antonella Ragusa, *Member IEEE,* Gianpaolo Vitale, *Member IEEE*

Consiglio Nazionale delle Ricerche
Istituto di Studi sui Sistemi Intelligenti per l'Automazione
(ISSIA – CNR), sezione di Palermo, Via Dante, 12 90141 PALERMO, ITALY
TEL. +39 091 6113513 FAX +39 091 6113028
mariacarmela.dipiazza@ieee.org, ragusa@pa.issia.cnr.it, gianpaolo.vitale@ieee.org

Abstract. The aim of this paper is to propose an approach for photovoltaic (PV) sources modeling based on robust least squares linear regression (LSR) parameter identification method. On the basis of experimental data of solar irradiance, cell temperature and voltage and currents at maximum power points for a given PV array, correlation functions among the considered quantities are defined. By implementing these functions in a Matlab® Simulink model, accurate I-V characteristics for the considered array are obtained, managing only the solar irradiance. The method is validated comparing the computed and the experimental maximum power points (MPPs). Its effectiveness is proven to be better with respect of parameter identification methods based on discrete approaches and standard LSR method.

Key words
Models and simulation of renewable energy sources; Photovoltaic array; Statistics.

1. Introduction

Nowadays the more and more growing interest in applications of photovoltaic (PV) sources, with all the related problems of optimal exploitation, environment impact and grid stability, has determined a speed up of the research in this field. In particular several issues are, currently, under consideration such as the prediction of PV energy production, the optimal choice and design of the power converter interfacing the PV generator to the utility and the study of all the problems related to the power electronic control, including the maximum power point tracking (MPPT). In order to suitably face these problems an accurate modeling of the PV source is necessary. As a matter of fact, such modeling allows to assess the I-V curves of a PV array, for varying environmental conditions, and so to obtain an evaluation of the PV source electrical behavior.

Several method are presented in technical literature to model a PV generator. Some contributions [1]-[4] are based on a modeling of PV cells through equivalent circuits, others are based on a data base can be set up by measuring I-V curves in different environmental conditions [5]-[6]. If the first approach is used, a suitable technique for the model parameters extraction is needed. As for the parameter identification, in [7] a method that extract a PV panel model parameters on the basis of the datasheet values is presented. This method has the advantage to be rapid as no experimental measurement is required but it can not take into account the actual data spread due to panel characteristic tolerance. Another approach is to use experimental measurements obtained by a data acquisition system to identify the model [2], [8]-[9]. This method has the advantage to take into account the real behaviors of an experimental PV array. On the other hand a great amount of data has to be managed. The aim of this paper is to overcome the limitation of previously proposed method developing a PV array model according to its real experimental electrical behavior but without the need of managing many data. The adopted model is a four parameter model based on the single diode circuit representation of the PV cell. The parameter extraction is performed starting by the consideration of the temperature and MPPs voltage and current distributions versus solar irradiance. Such distributions show a data placement along a straight line that suggests the possibility to obtain such data by linear regressions. In such a way the model parameters can be identified by the knowledge of the only solar irradiance. Both a linear least squares regression (LSR) and a robust linear LSR have been used to identify the model parameters. The proposed approach is compared with an extraction method based on a discrete clustering of the experimental data. The presented methods are experimentally validated verifying the belonging of the experimental MPPs to the corresponding computed I-V curves. The paper is organized as follows: in section II a description of the PV experimental plant is given; in section III the PV source electrical model is presented. Section IV deals with the discrete approach to PV model parameter identification; section V gives the fundamentals on used linear regressions and explains their application to the case study. Finally in section VI a discussion and the experimental validation are presented.

2. Experimental Plant

The experimental data have been obtained by a 1.45kWp PV plant installed on a roof footbridge at the University of Palermo – Faculty of Engineering, set up by Italian National Agency for New Technologies, Energy and Environment (ENEA). The electrical features of the array, under standard test conditions, are the following: open circuit voltage, Voc(stc): 228.2V; short circuit current, Isc(stc): 9.2A; maximum power voltage, Vmp(stc): 185.5 V and maximum power current, Imp(stc): 8A.

In Fig. 1 a view of the PV array is reported. The PV plant is equipped with a data acquisition system that measures the following parameters: panels temperature, solar irradiance, DC voltage and current supplied by the solar array to the inverter, AC voltage, current and power supplied by inverter to the grid [10].

3. PV Source Electrical Model

A photovoltaic cell can be described by a mathematical model that gives the relation between voltage and current at its terminals.

The corresponding I-V characteristic equation is:

$$I = I_{ph} - I_{s1}\left[e^{\frac{q(V+IR_s)}{KT}} - 1\right] - I_{s2}\left[e^{\frac{q(V+IR_s)}{AKT}} - 1\right] - \frac{V+IR_s}{R_p} \quad (1)$$

where I_{ph} is the photo-generated current; K is the Boltzman constant = 1.38x10^{-23} J/K; q is the electronic charge = 1.6x10^{-19} C; T is the cell temperature [K]; A is the ideality factor of the diode; I_{s1} and I_{s2} are the saturation current due to diffusion and the saturation current due to recombination, respectively.

The I-V relation has a correspondent circuital model reported in Fig. 2. Because the presence of the recombination diode is relevant only at low voltage bias, a widely used simplification of the circuit implements only one diode, so the following I-V equation is obtained [3]:

$$I = I_{ph} - I_s\left(e^{\frac{q(V+IR_s)}{AKT}} - 1\right) - \frac{V+IR_s}{R_p} \quad (2)$$

If the shunt resistor R_p is neglected, the well known four parameter model is obtained [11]. This last simplification does not affect significantly the validity of the model. It should be noted that, in any case, equations (1) and (2), are non-linear and implicit so their analytical solution is not practicable.

In order to exploit the PV cells' characteristic to produce electrical energy, they are connected in series and in parallel and grouped in modules by manufacturer. Modules can be further combined into panels and the connection of the panels forms the entire photovoltaic array that gives the desired I-V characteristics. The whole assembly has a circuital model that is analogous to a four parameter model but its parameters must be scaled. In particular considering N_P cells in parallel and N_S cells in series :

$$\begin{cases} I_{ph,tot} = N_P\, I_{ph} \\ I_{s,tot} = N_P\, I_s \\ R_{s,tot} = R_s\, N_S / N_P \end{cases} \quad (3)$$

Hence, series connection increases the output voltage and parallel connection increases output current.

$$\begin{cases} I_{tot} = N_P I \\ V_{tot} = N_S V \end{cases} \quad (4)$$

This assumption is acceptable under the condition of uniform irradiance and temperature through the PV field. In case of partial shading, the electrical behavior of the single modules should be accounted for.

The four parameters model needs four information to be identified. In fact by rewriting for simplicity I-V equation as follows:

$$\begin{cases} I = I_o - e^{[(V+IR_s)K1+K2]} \\ where \\ I_{ph} = I_o;\ \frac{q}{AKT} = K1;\ I_s = e^{K2} \end{cases} \quad (5)$$

The four parameters to be identified are K_1, K_2, R_s and I_o. They can be obtained by the knowledge of the short-circuit current (I_{sc}), the open circuit voltage (V_{oc}) and the maximum power point current and voltage (I_{mp} and V_{mp}).

Fig. 1. View of the PV array.

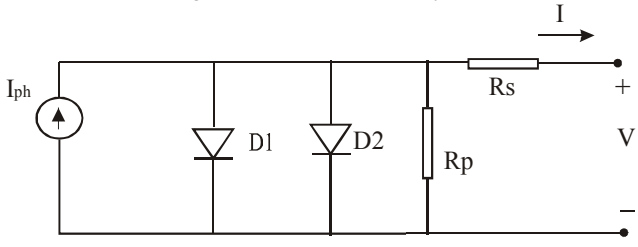

Fig. 2. Equivalent circuit of a PV cell.

By substituting these parameters into equation (5) and imposing that the voltage derivative of the power is zero at maximum power point, the following equations are obtained:

$$\begin{cases} I_{sc} = I_o - e^{[(I_{sc}R_s)K1+K2]} \\ 0 = I_o - e^{[(V_{oc})K1+K2]} \\ I_{mp} = I_o - e^{[(V_{mp}+I_{mp}R_s)K1+K2]} \\ \dfrac{\partial P}{\partial V} = I_o\left(1+K_1V_{mp}\right)e^{[(V_{mp}+I_{mp}R_s)K_1+K_2]} = 0 \end{cases} \quad (6)$$

From the above mentioned conditions the following relations for finding out the model parameters are obtained:

$$\begin{cases} K_1 = \dfrac{\dfrac{I_{mp}}{I_o - I_{mp}} + \log[1 - \dfrac{I_{mp}}{I_o}]}{2V_{mp} - V_{oc}} \\ R_s = \dfrac{V_{mp} - \dfrac{I_{mp}}{[I_o - I_{mp}]K_1}}{I_{mp}} \\ K_2 = \log[I_o] - V_{oc}K \\ I_{sc} = I_o - e^{(I_{sc}R_s)K_1+K_2} \end{cases} \quad (7)$$

The last equation has the form $I_o=f(I_o)$ and it could be solved by an iterative method; however the contribution of the exponential term is negligible because it contains a coefficient that corresponds to a reverse saturation current, as a consequence it can be assumed $I_{sc} = I_o$. It should be noticed that the obtained value of R_s takes into account the connecting cables resistance.

As for the knowledge of I_{sc} and V_{oc}, they are usually given by the manufacturer under standard test conditions, i.e. AM1.5, corresponding to a solar irradiance G_{stc}=1000 W/m^2 and a cell temperature T_{stc}=25 °C. Their values for each values of solar irradiance, G, and temperature, T, can be determined considering the following expressions.

$$\begin{cases} V_{oc} = V_{oc(stc)} + \Delta V_T(T - T_{stc}) \\ I_{sc} = \dfrac{G}{G_{stc}} I_{sc(stc)} \end{cases} \quad (8)$$

where $V_{oc(stc)}$ and $I_{sc(stc)}$ are the open circuit voltage and the short circuit current given by manufacturer under standard test conditions and ΔV_T= -0.08 V/°C. It should be noted that the V_{oc} variation depends (weakly) on the temperature, on the contrary irradiance has a large effect on short circuit current.

It is also possible to achieve the values of V_{oc} and I_{sc} through off line measurements on the PV plant. The V_{oc} can be obtained by a voltage measurement without the inverter; the measure of I_{sc} is more difficult because it requires the short circuit condition. However by performing some measurement also the coefficients of (8) can be deduced.

As for V_{mp}, I_{mp}, they could be deduced by data logger.

In the case under study the data logger acquires data each 10 minutes, the voltage and current when the inverter is running should correspond to V_{mp}, I_{mp}. Unfortunately this is not always true because there is no synchronization with the MPPT algorithm therefore some couples of data might have been acquired when the algorithm is searching for the maximum after a fast variation of solar irradiance. Moreover, with low solar radiation, the inverter remains disconnected and the data do not correspond to the maximum power.

On the other hand by measuring data it is possible to obtain informations about the real characterization of the plant, including the series resistance represented by the cables that connect the photovoltaic generator to the inverter. To this aim it is necessary to manage data in order to discriminate the "noise" represented by undesired measurements.

4. Discrete Parameter Identification

By performing measurements from June to October during all the day and sampling each 10 minutes, a set of more than 12000 couples of experimental values of current and voltage, corresponding to maximum power points (MPPs), for solar irradiance and temperature ranging between 7 W/m^2 and 1140 W/m^2 and 13°C and 58°C, respectively, have been acquired in all. Among these couples only 7150 are situated for solar irradiance and temperature ranging between 100 W/m^2 and 1100 W/m^2 and 20 °C and 50 °C, respectively.

An approach to find out V_{mp} and I_{mp}, from the data in order to identify the PV model parameters consists in a simple clustering of measured maximum power point data on the basis of defined intervals of solar irradiance and temperature. Then the estimation of MPP for each cluster is obtained using a normal distribution for the experimental data and performing a maximum likelihood estimation (MLE) of the distribution parameters.

In Fig. 3 an example of experimental measured MPPs, referred to irradiance range G=500÷600 W/m^2 and to temperature range T = 40÷45 °C, is reported. From Fig. 3 it is possible to observe a dispersion of experimental MPPs, which is present in all the intervals of measured data. However the problem consists on the correct choice of irradiance and temperature step intervals.

5. Linear Regressions

A. Least squares linear regressions

It is often necessary to obtain a mathematical relation between two (or more) variables. If ($X_1, X_2, ..., X_N$) and ($Y_1, Y_2, ..., Y_N$) are two set of data, by drawing the couples (X_i, Y_i) in a Cartesian diagram, a so-called dispersion diagram is obtained. In such a diagram if it is possible to identify a curve that approximates the drawn data, the curve is said interpolating curve.

The problem consists in the choice of the curve and determining its parameters. The simplest curve is the straight line. In order to find a good approximation, the straight line coefficients can be chosen so to minimize

the sum of the square of the distance between each point (X_i, Y_i) and the line obtaining the least square straight line.

The least square minimization may be also utilized with other kinds of curves. In formulas, the interpolating straight line may be written as:

$$Y = a_0 + a_1 X \tag{9}$$

and the coefficients that minimize the square of the distance between the line end the points are given by:

$$a_0 = \frac{\left(\sum_{i=1}^{N} Y_i\right)\left(\sum_{i=1}^{N} X_i^2\right) - \left(\sum_{i=1}^{N} X_i\right)\left(\sum_{i=1}^{N} X_i Y_i\right)}{N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2}$$

$$a_1 = \frac{N\sum_{i=1}^{N} X_i Y_i - \left(\sum_{i=1}^{N} X_i\right)\left(\sum_{i=1}^{N} Y_i\right)}{N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2} \tag{10}$$

The equation (9) estimates the variable Y on the basis of the value of the variable X, the line is said regression curve of Y on X.

The correlation coefficient evaluates the goodness of the fitting of data on the considered curve, the standard error measures the spread of data around the regression curve. The correlation coefficient value can vary in the range –1 and +1, if the value is zero there is not any linear correlation between the two variables X and Y. For linear case they are respectively given by:

$$r = \frac{N\sum_{i=1}^{N} X_i Y_i - \left(\sum_{i=1}^{N} X_i\right)\left(\sum_{i=1}^{N} Y_i\right)}{\sqrt{\left[N\sum_{i=1}^{N} X_i^2 - \left(\sum_{i=1}^{N} X_i\right)^2\right]\left[N\sum_{i=1}^{N} Y_i^2 - \left(\sum_{i=1}^{N} Y_i\right)^2\right]}} \tag{11}$$

$$s = \sqrt{\frac{\sum_{i=1}^{N} Y_i^2 - a_0\sum_{i=1}^{N} Y_i - a_1\sum_{i=1}^{N} X_i Y_i}{N}} \tag{12}$$

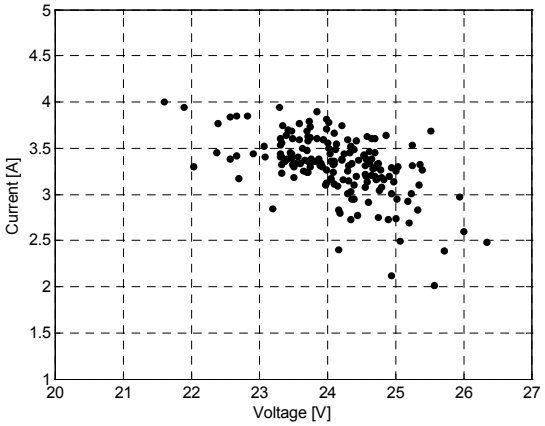

Fig 3. Experimental maximum power points of the PV array for G=500÷600 W/m2 and T = 40÷45 °C.

When more than two variables are involved a similar approach can be utilized; for example among three variables X, Y and Z a linear regression gives a plane in the form: $Z = b_0 + b_1 X + b_2 Y$ and the use of more complicated regression surfaces can be investigated.

B. Robust least squares linear regressions

Robust regression is a form of regression analysis devised to overcome some limitations of traditional regression methods.

Robust fitting is commonly used when the data contain outliers. In the presence of outliers, least squares estimation is inefficient and can be affected by inaccuracy. This is due to the shifting of the least squares estimates towards the outliers and to the corresponding altered increase of the estimates variance.

The application of the robust regression method for the PV model parameters identification is performed within Matlab® environment. In particular the embedded robustfit function is used to obtain the regression straight line coefficients estimates.

The *robustfit* function uses an iteratively re-weighted least squares algorithm, with the weights at each iteration calculated by applying the bisquare function to the residuals from the previous iteration. This algorithm gives lower weight to points that do not fit well. The results are less sensitive to outliers in the data as compared with ordinary LSR [13]-[15].

6. Data Analysis and Parameter Identification

The aim of this section is to obtain a simple relation that gives the V_{mp}, I_{mp} values versus solar irradiance and temperature. This is a typical three variables problem. The dispersion diagram of T versus G exhibits a linear trend. The linear correlation coefficient value in this case is 0.8183. It means that a linear relation between temperature and solar irradiance can be found by linear LSR.

By using eq. (10) the following relation is obtained

$$T = 26.377 + 0.023\,G \tag{13}$$

In figure 4 the dispersion diagram of T vs. G with superimposed the line defined by (13) is shown. The standard error value, s_{ls}, is 4.4636.

Thanks to this relation between T and G, the above mentioned problem is reduced to a two variables problem. However, it is well known that, in practice, different temperatures can correspond to the same radiation, according to other weather conditions as wind speed and environment temperature [12]. Using the linear regression, only one temperature is obtained for each irradiance value. This temperature is nearly the more probable, according to experimental data. This is demonstrated using an algorithm, developed by the authors, that gives the temperature value corresponding to the greater group of experimental data for given values of solar irradiance. Figure 5 shows that the temperature

trend, obtained by the described algorithm, fits the linear regression curve.

By observing the dispersion diagram of I_{mp} and V_{mp} versus solar irradiance G respectively, it can be noted that, in both diagrams, data are distributed in a straight way except for low irradiance where a high data dispersion is observed. The correlation coefficient value is 0.9612 for the I_{mp} and –0.2642 for the V_{mp}. In such cases the use of the linear LSR is not accurate enough. On the contrary with the use of linear robust LSR, low radiation data are considered as outliers and filtered.

In figure 6 the dispersion diagram of I_{mp} vs. G is reported with superimposed both the line obtained with linear LSR and with the robust linear LSR.

In figure 7 the dispersion diagram of V_{mp} vs. G is reported with superimposed both the line obtained with linear LSR and with the robust linear LSR.

The obtained equations for I_{mp} and V_{mp}, with LSR, are:

$$\begin{cases} I_{mp} = -0.155 + 0.0066\,G \\ V_{mp} = 169.170 - 0.0069\,G \end{cases} \tag{14}$$

and the standard errors, s_{ls}, are, respectively, 0.5206 and 7.0204

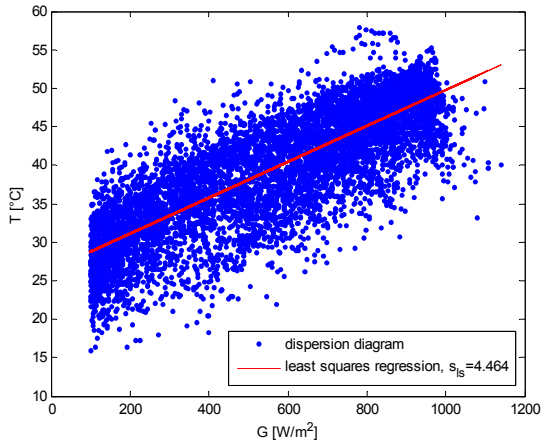

Fig. 4. Dispersion diagram of T vs. G and LSR line.

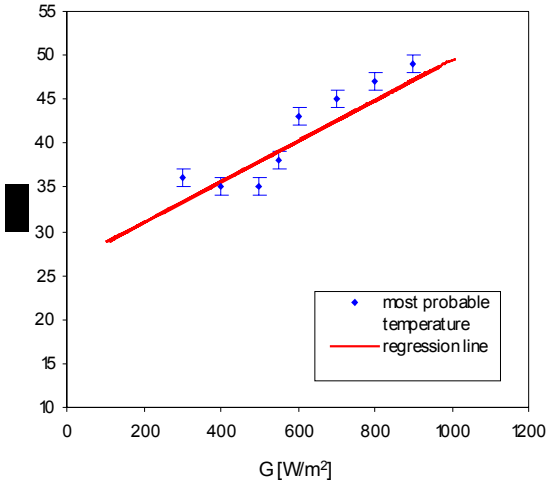

Fig. 5. T vs G regression line and temperature trend obtained by experimental data with ± 1 °C error bars.

With the robust LSR, the equations for I_{mp} and V_{mp} become:

$$\begin{cases} I_{mp} = -0.470 + 0.0071G \\ V_{mp} = 172.440 - 0.0115\,G \end{cases} \tag{15}$$

and the standard errors, s_r, decrease becoming, respectively, 0.3176 and 5.5450. Equations (13) and (15) are implemented in a Matlab® Simulink model that allows to obtain continuously the photovoltaic array model parameters, according to (7) and (8). As a result, an accurate modelling of the PV array is obtained, managing only the solar irradiance and considering the dependence of T, V_{mp} and I_{mp} on G.

7. Discussion and Experimental Validation

With the proposed approach, the PV array I-V characteristics are obtained, once the solar irradiance is given. In Fig. 8 the I-V characteristics, computed for G varying in the range 500÷570 W/m2, are shown, in comparison with the unique characteristic obtained when a discrete parameter identification is performed.

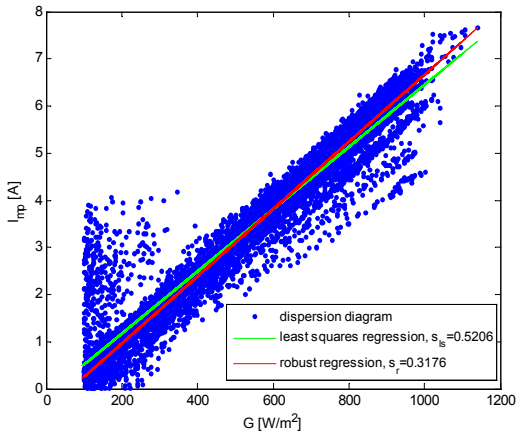

Fig. 6. Dispersion diagram of Imp vs. G with superimposed LSR and robust LSR lines.

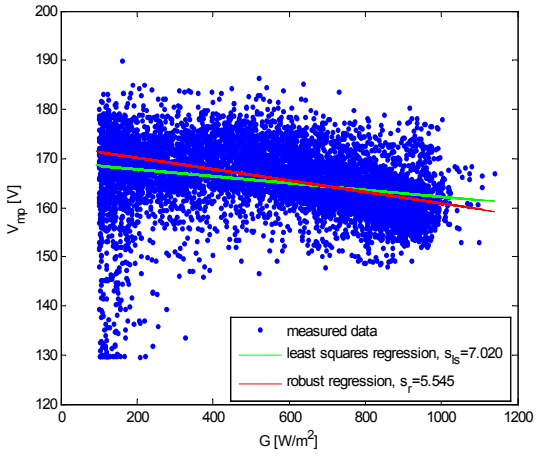

Fig. 7. Dispersion diagram of Vmp vs. G with superimposed LSR and robust LSR lines.

It is evident that the PV model parameters identification based on linear regressions overcomes the limitations of the step-by-step approach, giving a continuous representation of the PV source electrical behaviour.

In Fig. 9 the I-V characteristics, obtained when the PV model parameters identification is based on regressions (both LSR and robust LSR), are reported. It is possible to notice that, for low and high irradiance values, there is a deviation between the LSR-based and the robust LSR-based characteristics. This is due to the great spread of data at low irradiance and to the small amount of available data at high irradiance. On the other hand, a good matching between the LSR-based and the robust LSR-based characteristics is observed in the irradiance range where experimental data are better correlated to regression lines.

In Fig. 9 the MPPs are reported too. In particular the computed MPPs (both on LSR and robust LSR-based characteristics) and the experimental ones, for the different couples of values of irradiance and temperature, are shown. From Fig.9 it is evident that the experimental MPPs are very close to the computed MPPs on the robust LSR-based characteristics.

This result is confirmed by the values of the experimental powers at MPPs evaluated in comparison with the computed ones and reported in table I.

8. Conclusions

In this paper the parameters identification of a PV source model is performed by using a robust LSR method. Introducing correlation functions among the voltage and current at MPPs, the temperature and the solar irradiance, involved in the PV array electrical behaviour, the PV source model is accurately identified. Implementing the proposed method in a Matlab Simulink model, the I-V characteristics of the PV array are obtained, simply managing the solar irradiance.

The method is experimentally verified and it is demonstrated to be more effective with respect of parameter identification methods based on discrete approaches and standard LSR method.

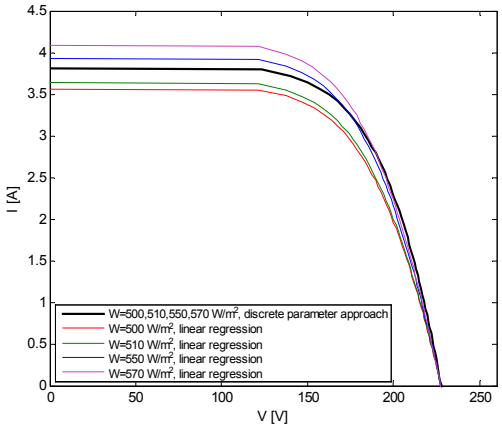

Fig. 8. Comparison of I-V characteristics obtained by linear regressions-based and discrete parameter identification approaches.

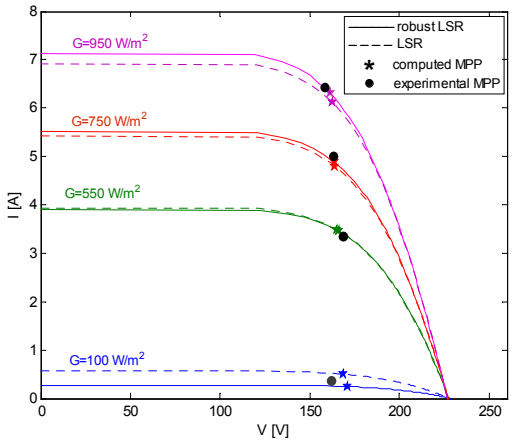

Fig. 9. Comparison of computed and experimental MMPs.

TABLE I – Comparison of Power at MPPs

G [W/m2]	LSR Pmp [W]	robust LSR Pmp [W]	Experimental Pmp [W]
100	85.07	41.63	58.46
550	574.75	573.54	564.44
750	786.36	799.34	817.85
950	994.34	1017.79	1018.50

References

[1] M. Veerachary, "PSIM Circuit-Oriented Simulator Model for the Nonlinear Photovoltaic Sources", IEEE Trans. on Aerospace and Electron.Sys., Vol.42, no.2, Apr. 2006, pp.735-740.

[2] J.A. Gow, C.D. Manning, "Development of a photovoltaic array model for use in power electronics simulation studies", IEE Proc. Electronic Power Applications, Vol.146, No. 2, Mar. 1999, pp.193 – 200.

[3] W. Xiao, W.G. Dunford, A. Capel, "A novel modeling method for Photovoltaic cells", in Proc. IEEE 35th Annual Power Electron. Spec. Conf., Aachen, Germany, 2004, pp.1950 – 1956.

[4] Yamashita, H.; Tamahashi, K.; Michihira, M.; Tsuyoshi, A.; Amako, K.; Park, M., "A novel simulation technique of the PV generation system using real weather conditions", in Proc. Power Conversion Conf., Apr.-2-5, 2002, pp.839-844.

[5] P.Sanchis, I. Echeverria, A. Ursua, O. Alonso, E. Gubia, L. Marroyo, "Electronic converter for the analysis of photovoltaic arrays and inverters", in *Proc. IEEE 34th Power Electron. Spec. Conf.*, vol. 4, 15-19 Jun. 2003, pp.1748-1753.

[6] Agilent application specific dc power supplies, Operating guide for solar array simulators Agilent models E4350B, E4351B, Sept. 2004, Agilent Part No. 5962-8206. [Online] Available: www.agilent.com/find/power.

[7] D. Sera, R. Teodorescu, P. Rodriguez, "PV panel model based on datasheet values", *IEEE International Symposium on Industrial Electronics ISIE 07*, June 4-7, 2007, pp.2392-2396.

[8] Di Piazza M.C., Serporta C., Vitale G., "A DC/DC Converter Based Circuit Model for a Solar Photovoltaic Array", *in Rec. 21th European Photov. Solar Energy Conf. and Exibition*, 4-8 Sept. 2006, Dresden, Germany, pp. 2726-2731.

[9] M. Cirrincione, M. C. Di Piazza, G. Marsala, M. Pucci, G. Vitale "Real Time Simulation of Renewable Sources by

Model-Based Control of DC/DC Converters**"**, *IEEE International symposium on Industrial Electronics, ISIE 2008*, 29 June 2 July 2008, Cambridge, UK.

[10] M.C Di Piazza, C Serporta.; G. Tine, G. Vitale, "Electromagnetic compatibility characterisation of the DC side in a low power photovoltaic plant", *in Proc. IEEE Intern. Conf. on Industr.Technology,* vol. 2, 8-10 Dec. 2004, pp.672 - 677.

[11] R. Chenni, M.Makhlouf, T.Kerbache, A.Bouzid, "A detailed Modeling for Photovoltaic Cells", *Elsevier, Energy 32, 2007*, pp.1724-1730.

[12] G. TamizhMani, L. Ji, Y. Tang, L. Petacci, C. Osterwald, "Photovoltaic module thermal/wind performance: Long-Term Monitoring and Model Development For Energy Rating", *NCPV and Solar Program Review Meeting*, 2003 pp.936-939.

[13] Holland, P.W., and R.E. Welsch (1977), "Robust Regression Using Iteratively Reweighted Least-Squares," Communications in Statistics: Theory and Methods, A6, pp.813-827.

[14] Huber, P.J. (1981), Robust Statistics, New York: Wiley.

[15] Street, J.O., R.J. Carroll, and D. Ruppert (1988), "A Note on Computing Robust Regression Estimates via Iteratively Reweighted Least Squares," The American Statistician, 42, pp.152-154.