A comparison between BNN and regression polynomial methods for the evaluation of the effect of soiling in large scale photovoltaic plants

ARTICLE *in* APPLIED ENERGY · AUGUST 2013

Impact Factor: 5.61 · DOI: 10.1016/j.apenergy.2013.03.023

4 AUTHORS, INCLUDING:

Alessandro Massi Pavan
Università degli Studi di Trieste

Diamando Pieri
City University of Seattle

Soteris A. Kalogirou
Cyprus University of Technology

Applied Energy 108 (2013) 392–401

Contents lists available at SciVerse ScienceDirect

Applied Energy

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

A comparison between BNN and regression polynomial methods for the evaluation of the effect of soiling in large scale photovoltaic plants

A. Massi Pavan [a], A. Mellit [b,c,*,1], D. De Pieri [d], S.A. Kalogirou [e]

[a] Department of Engineering and Architecture, University of Trieste, Via A. Valerio, 6/A, 34127 Trieste, Italy
[b] Faculty of Sciences and Technology, Renewable Energy Laboratory, University of Jijel, Jijel 18000, Algeria
[c] Unité de Développement des Équipements Solaires (UDES), Bousmail, Tipaza 42000, Algeria
[d] Suntrust S.r.l., Strada della Rosandra, 2, 34147 Trieste, Italy
[e] Department of Mechanical Engineering and Materials Science and Engineering, Cyprus University of Technology, P.O. Box 50329, Limassol 3603, Cyprus

HIGHLIGHTS

- The soiling effect can have a significant impact on the PV plant performance.
- Bayesian neural network performs better than polynomial regression model.
- Grid operators can benefit from the proposed technique.

ARTICLE INFO

Article history:
Received 18 September 2012
Received in revised form 8 March 2013
Accepted 10 March 2013
Available online 13 April 2013

Keywords:
Large scale photovoltaic plant
Soiling
Pollution
Maintenance
Bayesian NN
Polynomial regression

ABSTRACT

This paper presents a comparison between two different techniques for the determination of the effect of soiling on large scale photovoltaic plants. Four Bayesian Neural Network (BNN) models have been developed in order to calculate the performance at Standard Test Conditions (STCs) of two plants installed in Southern Italy before and after a complete clean-up of their modules. The differences between the STC power before and after the clean-up represent the losses due to the soiling effect. The results obtained with the BNN models are compared with the ones calculated with a well known regression model. Although the soiling effect can have a significant impact on the PV system performance and specific models developed are applicable only to the specific location in which the testing was conducted, this study is of great importance because it suggests a procedure to be used in order to give the necessary confidence to operation and maintenance personnel in applying the right schedule of clean-ups by making the right compromise between washing cost and losses in energy production.

1. Introduction

With reference to Figs. 1a and 1b [1,2], for the first time in 2011 the Italian photovoltaic market overcame the German one that has been the leader during the past 10 years. With more than 9 GWp installed in 2011, the Italian photovoltaic market is currently the biggest one worldwide. In 2011, more than 5% of the national electricity consumption has been satisfied by means of photovoltaic energy production [3]. The Italian experience confirms that photovoltaic plants can play an important role in the electricity market and that it will be a key energy production system in the near future.

As described by the IEA PV Power Systems (PVPS) Task 2 [4], module soiling can result from various mechanisms such as pollution, accumulation of dust or pollen, bird droppings or growth of lichen (particularly at the lower edge of framed modules).

Several studies have been carried out to estimate soiling losses in photovoltaic systems [5–11]. Garcia et al. [12] showed how modification in the incidence angle due to dust accumulation, impacts energy losses. This confirms the earlier findings of Hammond et al. [5] where the soiling losses were estimated to be in the range [2.3–7.7%] and are a function of the angle of incidence. Thevenard and Pelland [13] estimated the uncertainty in long term PV yield prediction due to soiling which was calculated to be 2%, while recently Riley and Littmann [14] have been shown that soiling losses can reach 11.5% in agricultural areas. Qasem et al. [15], demonstrated

* Corresponding author at: Faculty of Sciences and Technology, Renewable Energy Laboratory, University of Jijel, Jijel 18000, Algeria. Tel.: +213 (0)551 998 982.
E-mail address: a.mellit@yahoo.co.uk (A. Mellit).
[1] Associate member at the International Centre for Theoretical Physics, Strada Costiera, 11, 34151 Trieste, Italy.

0306-2619/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.apenergy.2013.03.023

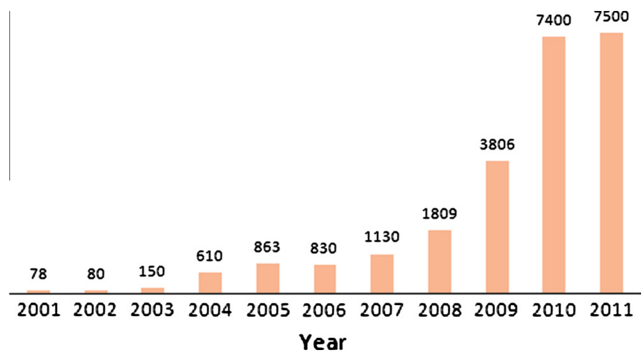

Fig. 1a. Germany – installed power in MWp.

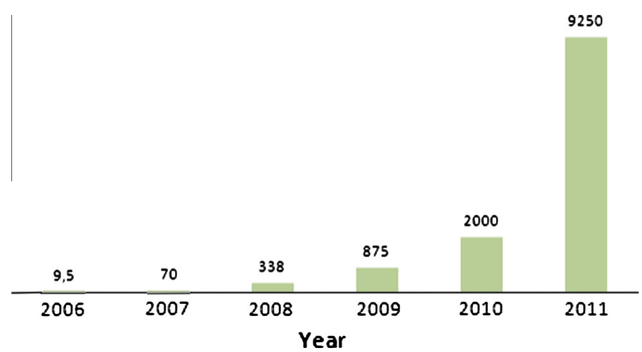

Fig. 1b. Italy – installed power in MWp.

that the effect of dust on photovoltaic modules depends on concentration and spectral transmittance. In a recent study in Cyprus [16,17] dust effects were investigated with attention to the sand carried over the sea from the Sahara desert and found that cleaning of the PV modules is necessary immediately after such a sand episode which usually ends with light rain which sticks the sand on the PV's surface.

It has been demonstrated that dirt and soiling strongly depend on the site location and climate [8], and the percentage reduction is difficult to be modeled or evaluated from case studies. Therefore, real PV plants performance monitoring is a better approach [18]. Moreover, the accumulation of dirt on solar panels (soiling) can have a significant impact on the PV system performance, and much of the information available is applicable only to the specific location in which the testing was conducted [6]. For these reasons a polynomial regression approach has been used in [8] and it is of great importance to investigate those results with the ones obtained with other methods.

The method given by Drews et al. [19] can be used when a monitoring system is not available and gives an indication of the so called "diffuse losses" that are a function of the soiling losses.

Marion et al. [20] described the different performance evaluation methods for grid-connected photovoltaic systems. These concern the performance ratio, the system yield and the PVUSA method. The performance ratio quantifies the overall effect of losses and thus cannot be used for determining the effect of soiling. The same for the system yield as this is introduced as a convenient way to compare the energy produced by systems with a different nominal power. Finally, the PVUSA method is certainly useful to calculate the soiling losses but cannot be used in this study as the wind speed is not available.

In order to calculate the effect of soiling on plants behavior, the idea in this study is to calculate the power produced at Standard Test Conditions (STCs – irradiance: 1000 W/m^2; cell temperature: 25 °C; solar spectrum: AM 1.5) by the system under study before and after a clean-up of their photovoltaic modules. The percentage difference between these two power values represents the losses due to the soiling effect.

Several methods are presented in literature to determine the power produced by a photovoltaic cell, module, or system. Physical models such as the one or two diodes models [21] are not easy to apply in the evaluation of the soiling effect as they require several parameters that are not always provided by the manufacturer of the system components, and they may not provide accurate results if a Large Scale Photovoltaic (LSPV) system is made of different PV module classes. Additionally, simple models such as the one discussed by Rahman and Yamashiro [22] use as an input the solar irradiance neglecting the effect of temperature and thus are not as accurate as needed.

Since, Artificial Neural Network (ANN) techniques represent a very good option where large databases of monitored data are available and with only a few information regarding the system are known [23], this type of model has been used for the comparison with the results obtained in [8].

This choice should also be intentional as polynomial models can have some limitations with respect to advanced algorithms used in ANNs such as the Bayesian regularization, genetic algorithm, particle swarm optimization [24], fuzzy logic [25] and other. Moreover, ANN-based models allow the overcome of deficiencies of conventional analytical techniques or numerical approaches [26]. Due to its generalization capability over other classical ANNs, Bayesian Neural Networks (BNNs) have been used for the development of the models utilized in the calculation of the soiling losses presented in this study [27,28].

Four BNN models have been developed in order to calculate the power at STC of two couples of photovoltaic strings forming two different photovoltaic fields before and after a complete clean-up of their modules.

The present paper is organized as follows: the following section is dedicated to the description of the two photovoltaic plants and of the collected datasets. Section 3 is on the multilayer perceptron and Bayesian neural network regularization; while in Section 4 the developed BNN models and results are presented. Finally, Section 5 gives conclusions on this work and perspectives for the future.

2. Photovoltaic plants and collected dataset

2.1. Photovoltaic plants

In order to show the effectiveness of the method employed at different conditions, the performance of two identical solar plants built at different locations in the countryside of Southern Italy have been studied. Both plants are located 100 km from Bari, longitude 41°7′1″, latitude 16°52′18″. The different conditions are related to the fact that a first plant is installed in a sandy site, while the second in a green one.

The nominal power of the two solar plants is 1 MWp, the nominal AC voltage is 20 kV, and the utilized conversion is of the centralized type (i.e., the photovoltaic fields are divided into two subfields connected to two inverters).

A schematic block diagram of the photovoltaic plants considered is depicted in Fig. 2a whereas a photograph of the second plant is shown in Fig. 2b. The strings are made of 20 series of connected PV modules, while groups of 16 strings are connected in parallel into 16 DC boards where fuses prevent over current into the strings [8]. The PV modules employed are produced by Q. Cells multi-crystalline silicon model QC-C04, whose electrical data are reported in Table 1a. The electrical parameters of a typical string made of 20 number 220 Wp power class modules are reported in

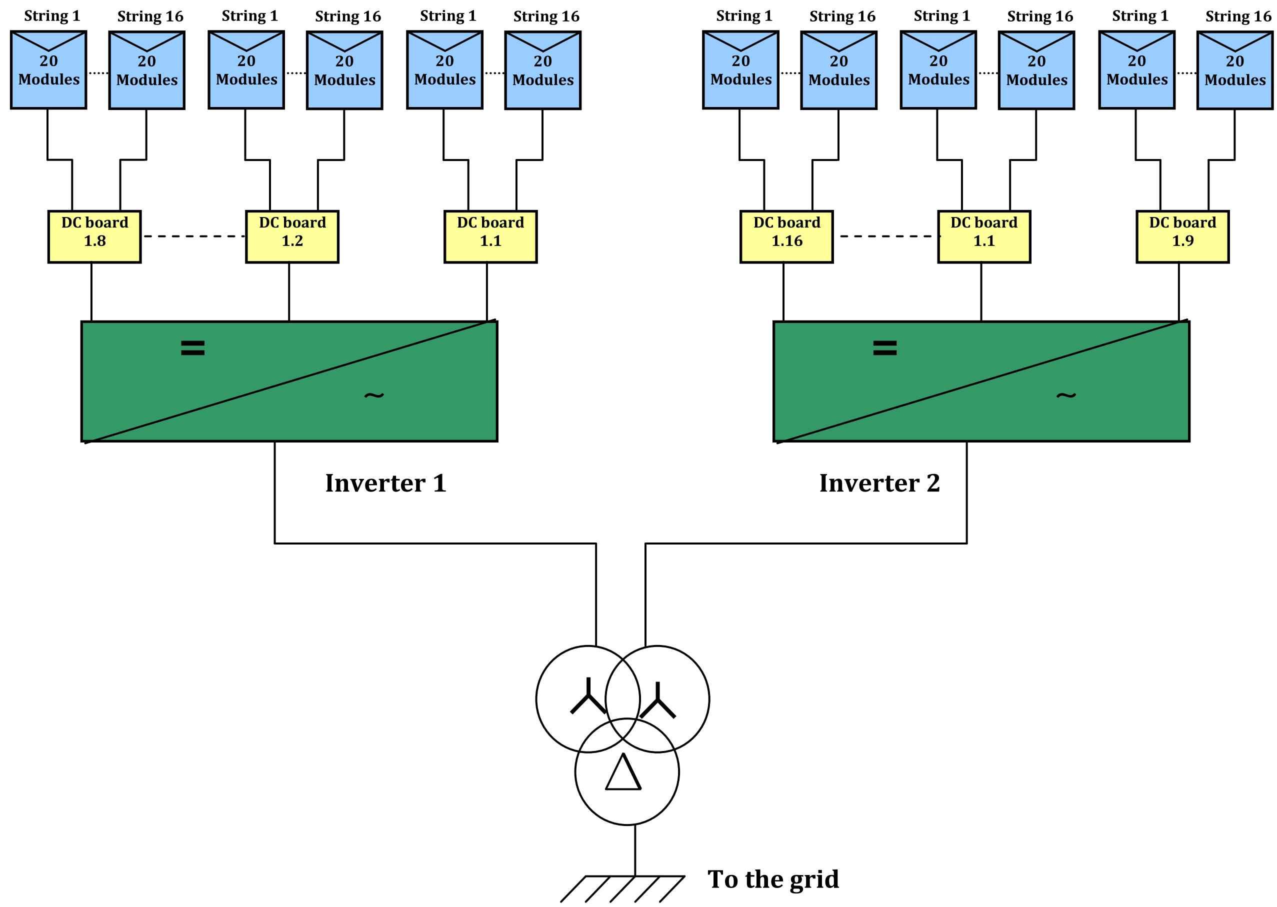

Fig. 2a. Schematic block diagram of the considered PV plants.

Fig. 2b. Aerial view of plant #2.

Table 1b. The two plants are equipped with two identical monitoring systems produced by Skytron® [29]. The monitoring systems consist of:

- A monocrystalline silicon solar cell Skytron® SOZ-3 for the measure of the on-plane solar irradiance (precision is ±5% in annual mean compared with a WMO class-1 pyranometer).
- Cell temperature contact probes, type Pt1000 (precision is ±0.8% at 100 °C).
- A number of channels for the measurement of the DC current (precision is ±2%).
- One channel for the measurement of the DC voltage (precision is ±5%).
- One data logger type SkyLog® that collects the electrical and climate data from the field.
- One server type SkyServ® connected to the web.

2.2. Collected dataset

In order to determine the PV plant behavior at STC, two datasets – one from soiled modules and the second from cleaned modules – of electrical and climate data (in-plane irradiance and cell temperature) have been collected for each plant. The spectral properties have not been measured and thus have not been considered in this study. However, as the geometrical conditions (latitude, tilt and azimuth angles) are the same both for dirty and cleaned modules and as the observation periods are rather long (7 weeks each), the spectral properties can be assumed to be constant. In any case this information is inherently included in the data measured as these concern actual data from an actual plant.

The PV modules have been operating for approximately 1 year before any cleaning was performed. The first acquisition period was June 21st–August 30th 2010, while the second one is September 2nd–October 24th 2010; the sample frequency is 15 min.

The following parameters have been measured for each plant:

- The current produced by the two couples of parallel connected PV strings, I_{DC}.
- The DC bus voltage, V_{DC}.
- The cell temperature, T_c.

Table 1a
STC electrical data for Q. Cells QC-C04 PV modules.

Nominal power (W)	210	215	220	225	230	235
Short circuit current (A)	8.10	8.20	8.30	8.40	8.45	8.55
Open circuit voltage (V)	35.90	36.10	36.25	36.35	36.40	36.50
Current at maximum power point (A)	7.45	7.55	7.65	7.75	7.85	7.95
Voltage at maximum power point (V)	28.30	28.60	28.80	29.00	29.20	29.40
Current/temperature coefficient (%/K)	+0.05					
Voltage/temperature coefficient (%/K)	−0.37					
Power/temperature coefficient (%/K)	−0.47					

Table 1b
Electrical parameter of a typical string made of 20 number 220 Wp QC04 photovoltaic modules.

Nominal power (kW)	4.4
Nominal voltage (V)	576
Maximum MPPT voltage (V) (at −10 °C)	651
Minimum MPPT voltage (V) (at 70 °C)	480
Maximum open circuit voltage (at −10 °C)	819

– The in-plane solar irradiance, G_i.

Fig. 3 shows an example of the monitored data (G_i, T_c) and the power output (P_{DC}) for a couple of PV strings calculated as the product of the measured current (I_{DC}) and voltage (V_{DC}).

3. Multi-layer perceptron and Bayesian neural network

The most commonly used neural network is the multilayer perceptron (feed-forward neural network). This type of ANN has been used successfully in several applications, including photovoltaic research studies, as reported in [25]. The ANN employed is a loop-free network which has its units arranged in layers, with only one input unit connected to the units in the next layer of the sequence. The first layer comprises a fixed input unit; there may be several layers of trainable 'hidden units' carrying an internal activation function, and finally, there is the output unit layer which is also trainable. Given the input and related output datasets it is possible to train the network by means of a supervised learning technique, generally known as back-propagation (BP) that allows adapting the unit weights. These changes are based on the differences between the known output (part of training dataset) and the one calculated by the ANN. In order to train the neural network fast, the following algorithms have been developed over the years: conjugate gradient (CG), Quasi-Newton (QN), Levenberg–Marquardt (LM) and Variable Learning Rate Back-Propagation (VLRBP). One of the problems that occur during neural network training is known as over-fitting in which the error on the given training data set is driven to a very small value, but when new training data are presented to the network, the output presents a non negligible error. This means that the network has memorized the training examples leading to a small error, but it has not learned to generalize to new situations. Bayesian regularization can greatly improve neural network's generalization ability [27]. The process of a Bayesian regularization consists in updating the weight and bias values according to Levenberg–Marquardt optimization technique. This minimizes a combination of squared errors and weights and determines the correct combination so as to produce a network that generalizes well.

Bayesian NNs are mainly BP networks with an additional ridge parameter added to the objective function. A brief description of the Bayesian inference method is as follows. In the Bayesian framework, the neural learning process is assigned to a probabilistic interpretation. The regularized objective function is given by:

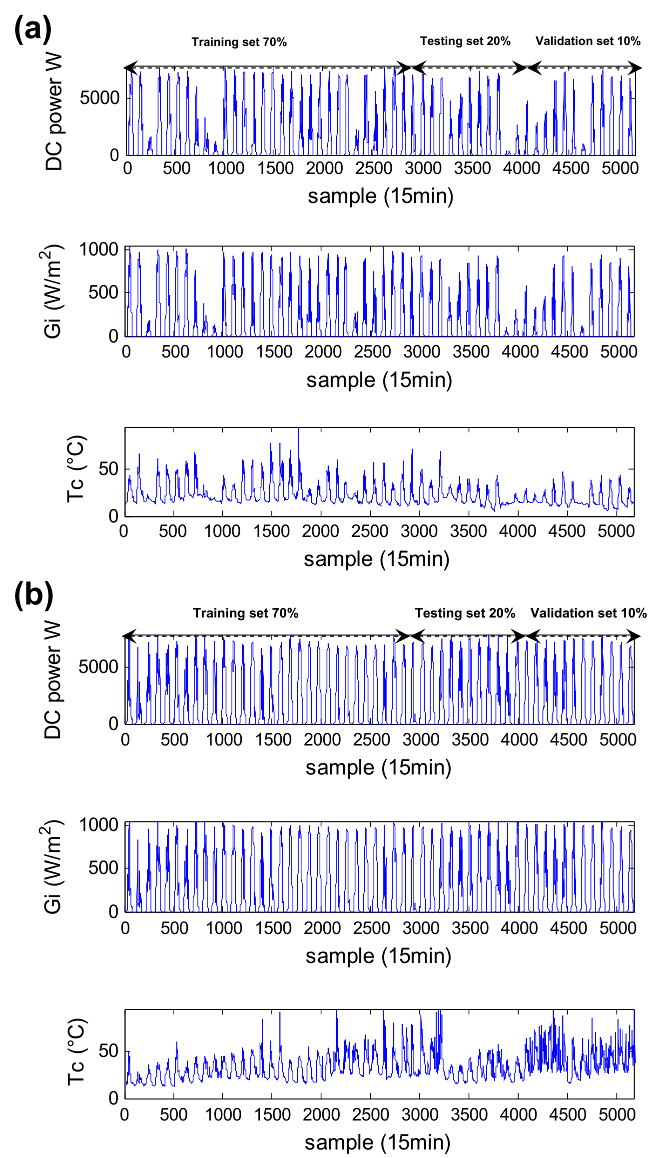

Fig. 3. Measured DC power for a couple of one string, in-plane solar irradiance and solar cell temperature of one PV plant: (a) cleaned PV modules and (b) dirty PV modules.

$$S = \beta E_e + \alpha E_w \tag{1}$$

where:

$$E_e(e) = \frac{1}{2}\sum_{i=1}^{n}(e_i)^2 = \frac{1}{2}\sum_{i=1}^{n}(y_i - \tilde{y}_i)^2 \tag{2}$$

$$E_w(w) = \frac{1}{2}\sum_{i=1}^{n} w_i^2 \quad (3)$$

while y and $\tilde{y}$ are the actual output and the calculated output respectively; n is the total number of parameters/samples in the network; and α and β are termed hyper-parameters (regularization parameter).

During training, the objective is to maximize the posterior distribution over the weights w, to obtain the most probable parameter values. The posterior distribution is then used to evaluate the predictions of the trained network for new values of the input variables. For a particular network A, trained to fit a dataset $D = \{x_i, t_i\}_{i=1}^{N}$ by minimizing an error function S given by Eq. (1), Bayes' theorem can be used in order to estimate the posterior probability distribution $p(w|D,\alpha,\beta,A)$ for the weights as follows [27]:

$$p(w|D,\alpha,\beta,A) = \frac{p(D|w,\beta,A)p(w|\alpha,A)}{p(D|\alpha,\beta,A)} \quad (4)$$

where $p(w|\alpha,A)$ is the prior density representative of the knowledge of the weights before any data is observed $p(D|w,\beta,A)$. This is the likelihood function that represents a model for the noise process on the target data. The factor $p(D|\alpha,\beta,A)$ is known as the evidence and is the normalization factor which guarantees that the total probability is 1, given by:

$$p(D|\alpha,\beta,A) = \int p(D|w,\beta,A)p(w,\alpha,A) \quad (5)$$

In order to evaluate the posterior distribution, the expressions for the prior distribution and the likelihood function must be defined [27]. The well-known Levenberg–Marquardt algorithm (LM) has been employed in order to minimize the objective function given in Eq. (1).

The BNNs have been chosen for following reasons [30]:

- They are difficult to over-train, as an evidence procedure provide an objective criterion for stopping, training and removing the need for a separate validation set to detect the onset of over-training.
- They are difficult to over-fit, because they calculate and train the effective number of parameters. This is considerably smaller than the number of weights and bias in a standard BP network.
- They are inherently insensitive to the architecture of the network as long as a minimal number of hidden neurons is employed.
- A large database is not necessary.

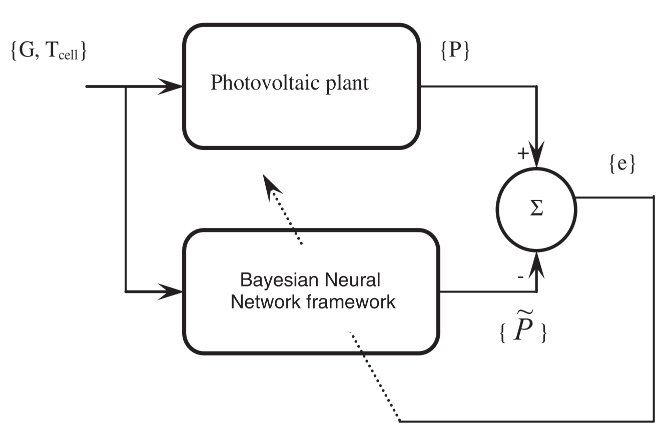

Fig. 4a. MLP-based schematic block diagram for training Bayesian regularization.

4. Models development and results

4.1. Prepocessing of the data

The following pre-processing, which makes the network more efficient, has been carried out on the four available datasets:

$$y = y_{\min} + (x - x_{\min})(x_{\max} - x_{\min})^{-1}(y_{\max} - y_{\min}) \quad (6)$$

where $x\varepsilon\,[x_{\min} - x_{\max}]$ is the original data value, and $y\varepsilon\,[y_{\min} - y_{\max}]$ is the corresponding normalized variable, while $y_{\min}$ and $y_{\max}$ have been set to -1 and 1 respectively.

Finally, for each PV plant and period a set of 4800 samples has been divided into three subsets: 70% of the samples of the total set have been used for the training of the BNNs, 20% for testing, and 10% for validation.

4.2. BNN-based model

A schematic block diagram of the BNNs used for the calculation of the STC power is depicted in Fig. 4a. The employed BNNs have three layers: an input layer, a single hidden layer, and an output layer. The first layer has two inputs (the solar irradiance and the cell temperature), while the output layer consists of a single output node that represents the power produced by the PV plant. The number of units in the hidden layer is estimated during the training process.

4.3. Polynomial Regression Model (PRM)

The model used by Massi Pavan et al. [8] for the calculation of the power produced at STC by the PV plant before and after the cleaning of their modules is the one proposed by Mayer et al. [31]:

$$P = a + b \cdot T_c \cdot G_i + c \cdot G_i + d \cdot G_i^2 \quad (7)$$

where T_c is the cell temperature, G_i the in-plane solar irradiance, a, b, c, and d are polynomial coefficients. The LM algorithm is used for the calibration of the models, and the polynomial coefficients found – reported in Appendix A – are then used for the calculation of the STC power by imposing a cell temperature of 25 °C and a solar irradiance of 1000 W/m^2.

4.4. Results and discussions

For each plant, a first dataset corresponding to 4 weeks of operation before the cleanup of their PV modules has been used to train the BNNs with the use of a soft computing program developed under MatLab© (Ver. 7.8, 2009). The same procedure has then been applied to a second dataset corresponding to 4 weeks of operations after the clean-up has been performed.

Different architectures of BNNs have been evaluated, and the best one has been chosen consisting of two units in the input layer,

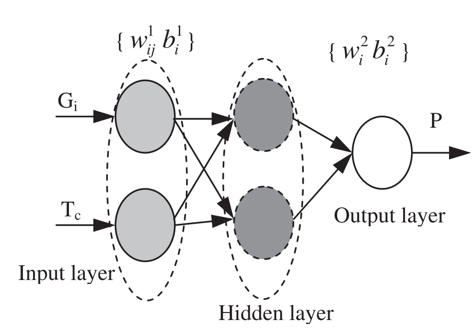

Fig. 4b. The obtained optimal neural network architecture.

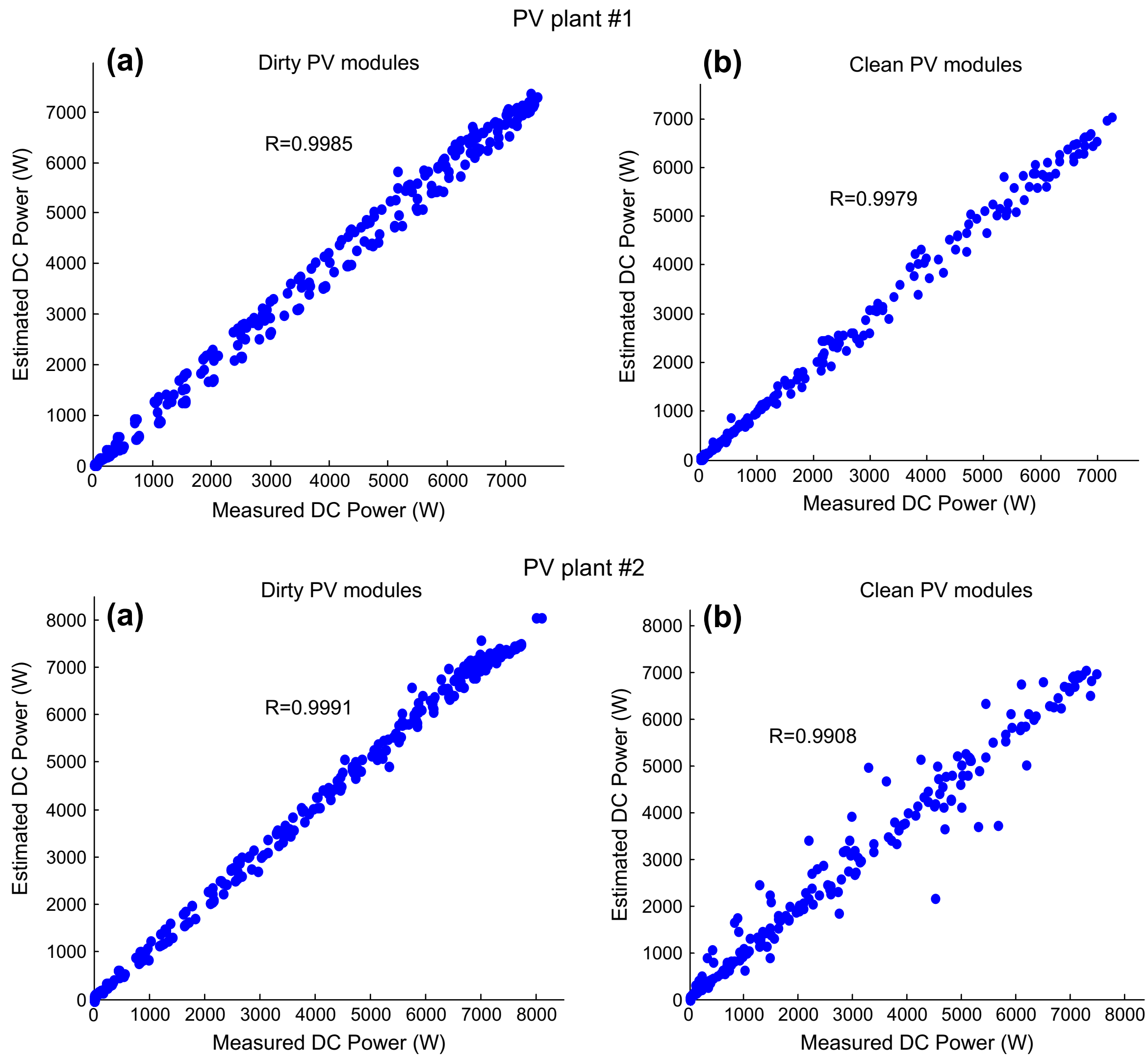

Fig. 5. Measured and estimated power production with BNN in both Photovoltaic pants: (a) dirty PV modules and (b) clean PV modules.

two in the hidden one, and one in the output layer as shown in Fig. 4b. Experimental investigations show that, even if the number of hidden units is increased the result does not change significantly, so this architecture is adopted due to its simplicity and accuracy.

The log-sigmoid function has been employed and thus the produced power can be written as:

$$\widetilde{P}(G_i, T_c) = w_1^2 c_1 + w_2^2 c_2 + b_1^2 \tag{8}$$

where:

$$c_1 = \frac{1}{1 + \exp\left(-\left(w_{1,1}^1 G_i + w_{1,2}^1 T_c + b_1^1\right)\right)} \tag{9a}$$

$$c_2 = \frac{1}{1 + \exp\left(-\left(w_{2,1}^1 G_i + w_{2,2}^1 T_c + b_2^1\right)\right)} \tag{9b}$$

The calculated weights (w) and bias (b) for the two PV plants under study before and after the clean-up are reported in Appendix A.

The comparisons between the produced and calculated power reported in Fig. 5 shows a very good performance of the designed BNNs; the mean correlation coefficient for the four cases is in fact greater than 0.99.

In order to evaluate the effectiveness of the designed BNN and polynomial models, three statistical errors have been calculated (the formulas are given is Appendix B):

- The Root Mean Square Error (RMSE).
- The Mean Absolute Error (MAE).
- The Mean Absolute Percentage Error (MAPE).

The results are shown in Table 2 and these confirm that the developed BNNs perform better with the data collected before the cleaning of the plants. This is due to the seasonal differences in the two observation periods.

In fact, the data relative to cleaned PV modules have been collected in autumn when the irradiance variability due to the presence of clouds has reduced the accuracy of the models.

Table 2
Error indexes between measured and estimated power by using BNN for 800 samples.

	RMSE (kW)	MAE (kW)	MAPE (%)
PV plant #1			
Dirty PV module	0.10	0.56	1.13
Clean PV module	0.14	0.12	1.37
PV plant #2			
Dirty PV module	0.12	0.09	1.21
Clean PV module	0.16	0.13	1.44

Table 3
Error indexes related to differences between measured and estimated produced power at STC using BNN and PRM models.

	R (%)	RMSE (kW)	MAE (kW)	MAPE (%)	Estimates power at STC (kW)	Losses (%)
PRM						
PV plant #1						
Dirty PV modules	99.80	0.27	0.13	2.0	7.45	5.52
Cleaned PV modules	99.81	0.22	0.10	8.4	7.66	
PV plant #2						
Dirty PV modules	99.96	0.22	0.08	2.3	7.66	1.01
Cleaned PV modules	99.61	0.36	0.16	6.1	7.73	
BNN						
PV plant #1						
Dirty PV modules	99.92	0.27	0.12	2.1	7.229	5.54
Cleaned PV modules	99.95	0.21	0.05	2.9	7.637	
PV plant #2						
Dirty PV modules	99.96	0.21	0.05	2.1	7.502	0.93
Cleaned PV modules	98.90	0.35	0.13	2.8	7.570	

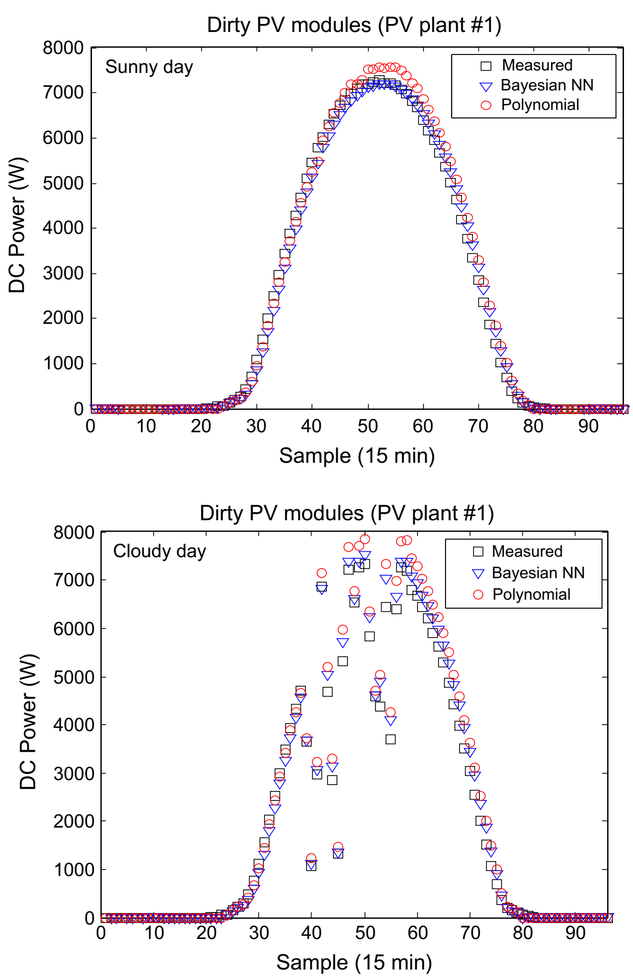

Fig. 6a. Comparison between BNN and PRM models when estimating the DC power for a couple of strings (PV plant #1) before cleaning the PV modules for a sunny and a cloudy day respectively.

Once the models have been developed and tested, the power losses (PL) due to the soiling effect have been calculated as follows:

$$PL = \frac{P_{STC}^{clean} - P_{STC}^{dirty}}{P_{STC}^{clean}} 100 \quad (10)$$

where P_{STC}^{clean} and P_{STC}^{dirty} are the power calculated at STC for clean and dirty PV modules respectively.

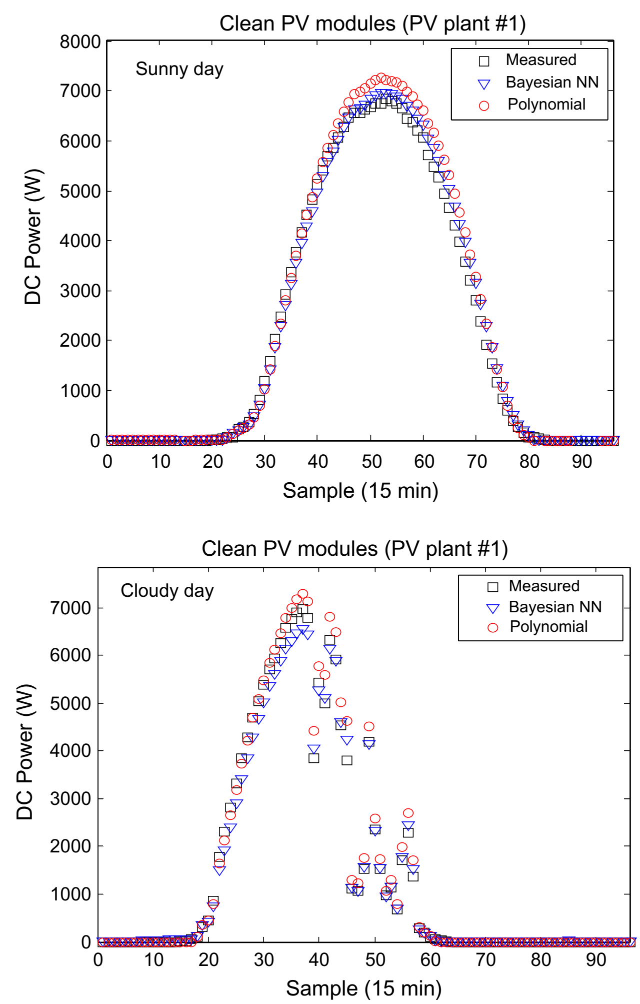

Fig. 6b. Comparison between BNN and PRM models when estimating the DC power for a couple of strings (PV plant #1) after cleaning the PV modules for a sunny and a cloudy day respectively.

As reported in Table 3, power losses calculated with the BNNs at STC are equal to 5.4% and 0.9% in plant 1 and plant 2 respectively.

Plant 1 and plant 2 are affected differently by the pollution phenomena because of the different type of the ground where they've been installed [8]. As was said before, plant number 1 is erected on a sandy ground while plant number 2 on a green one.

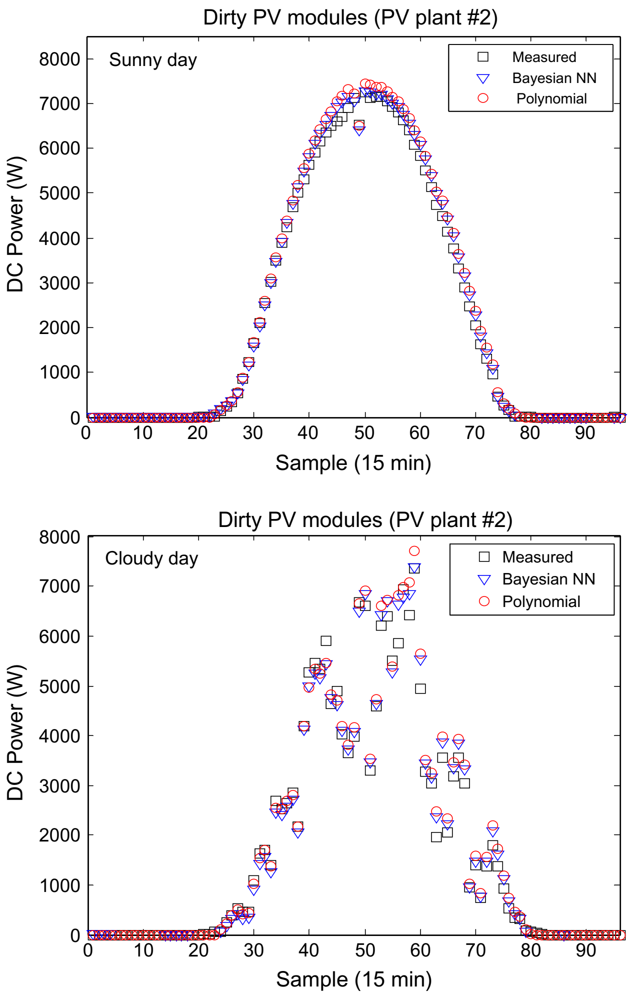

Fig. 6c. Comparison between BNN and PRM models when estimating the DC power for a couple of strings (PV plant #2) before cleaning the PV modules for a sunny and a cloudy day respectively.

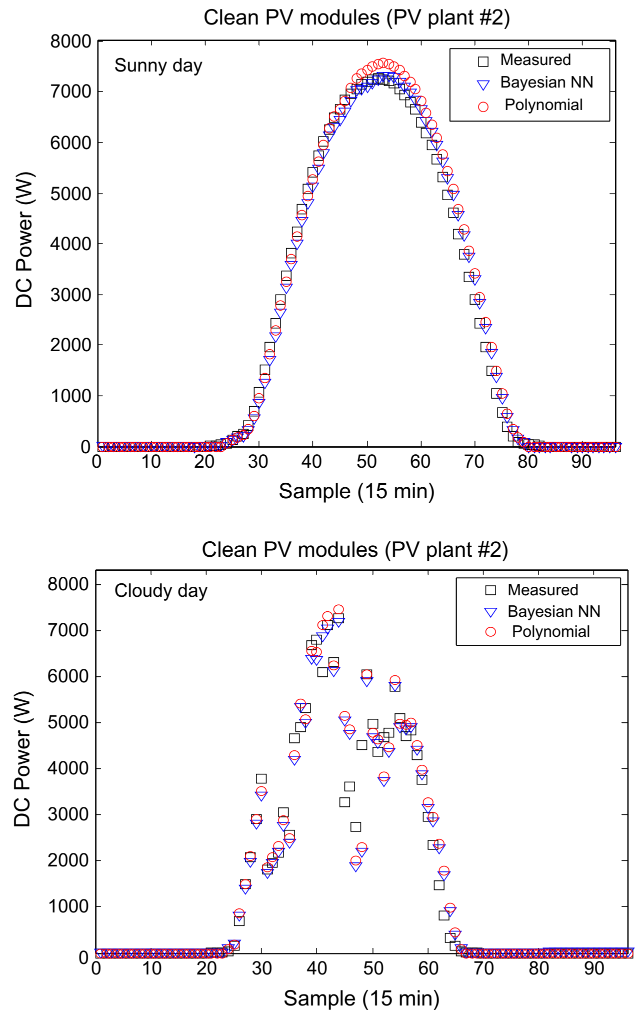

Fig. 6d. Comparison between BNN and PRM models when estimating the DC power for a couple of strings (PV plant #2) after cleaning the PV modules for a sunny and a cloudy day respectively.

Table 3 reports also the results given by the Polynomial Regression Model (PRM); these are very close to the ones given by the BNN models and the differences are not remarkable (5.5% versus 5.4% for plant 1, and 1.0% versus 0.9% for plant 2). The statistical errors reported in Table 3 shows that BNN models perform slightly better than PRM ones.

A comparison between the power calculated with the two different techniques is given in Figs. 6a, 6b, 6c and 6d for plants 1 and 2 respectively. For both plants, four different situation have been considered:

– Sunny day (before cleaning the PV modules).
– Sunny day (after cleaning the PV modules).
– Cloudy day (before cleaning the PV modules).
– Cloudy day (after cleaning the PV modules).

Figs. 6a, 6b, 6c and 6d show that the power calculated over a day with the two different techniques are very well correlated as confirmed by the correlation factor that is greater than 0.99. Moreover, it can be observed that the performance of the two models is not influenced by the weather conditions (sunny or cloudy day).

Given the fact that both BNN and PRM techniques provide reliable results when calculating percentage power losses, the differences in accuracy can be evaluated by examining the several error indexes. The following are the factors that influence the model development and its consequent behavior:

Table 4
MAPE between measured and estimated power by using BNN, RBFN and MLPN (PV plant #1, dirty PV module).

NN structures	MAPE (%)
MLPN ($2 \times 13 \times 1$)	1.1436
RBFN ($2 \times 21 \times 1$)	1.1253
BNN ($2 \times 2 \times 1$)	1.1312

– Intrinsic accuracy of the polynomial model $P = a + b \cdot T_c \cdot G_i + c \cdot G_i + d \cdot G_i^2$.
– Different objective functions: classic $E_e(e) = \frac{1}{2}\sum_{i=1}^{N}(y - \tilde{y})_i^2$ for the PRM model; modified objective function $S = \beta E_e + \alpha E_w$ for the BNN model.
– Choice of the initial values for the coefficients a, b, c and d.
– Choice of initial values (random) for weights and bias in the BNN model.
– Number of samples utilized for the determination of the polynomial coefficients and network training.
– The advantage of the BNN, is that the models are robust and the validation process, which scales as $O(N^2)$ in normal regression methods, is unnecessary [30].

– BNN models are robust since they are difficult to over-fit or over-train, they offer good accuracy and have the capability of generalization with respect to the classical NNs. This is proved by the data shown in Table 4 which reports the comparison between Multi-layer perceptron (MLPN), Radial basis function network (RBFN) and Bayesian neural network. As can be seen, for practically the same MAPE (1.1%), we have different structure, RBFN (21 neurons within the hidden layer), MLPN (13 neurons within the hidden layer).

5. Conclusion and perspectives

A simple and accurate model using BNN has been developed in order to evaluate the power losses of two photovoltaic plants. The results indicate that the losses due to dust accumulation on polycrystalline Si PV modules surface ranges from roughly 1% to 5% after 1 year of operation.

It has been demonstrated that estimating power at STC using the BNN model is more accurate than the one estimated by the polynomial model. Several factors affect the polynomial model accuracy which is also strongly influenced by the size of the database. In fact it has been noted that a large database is required to tune better the polynomial coefficients, while in the case of BNN, a 4000 samples dataset is largely sufficient to adapt the weights and bias of the network.

The developed methodology has a significant value for the operation and maintenance personnel that would schedule the clean-up intervention once the percentage loss of power is above an acceptable threshold. In countries like Italy, where investments on renewable energy are paid off proportionally to the produced energy, a PV plant underperformance directly affects the payback time and the return on investment indexes.

Grid operators can also benefit from the proposed technique. In fact, once their own PV plant model is created, they can predict its power output given the weather forecasts for the following day. Such information is useful for dispatch planning and it will have a fundamental role in the near future when considering the smart grid energy generation and distribution systems.

Future contribution concerns the evaluation of how the dust accumulation on PV modules surface affects the module behavior when considering different technologies (poly-crystalline Si, CIGS, CdTe, etc.). The test field for this investigation is already in place at the University of Trieste where data are being gathered.

Acknowledgments

Q. Cells Italia S.r.l. is kindly acknowledged for providing the information regarding the plants under study. The first author is grateful for the support provided by the Regione Friuli-Venezia Giulia under the GreenBoat Project (POR FESR 2007–2013, asse 1, attività 1.1b). Finally, the second author would like to thank the ICTP, Trieste, Italy.

Appendix A.

A.1. Polynomial coefficients

PV plant 1

Dirty PV modules:	$a = -3.1358$ $b = -0.0077$ $c = 8.5288$ $d = -0.0009$
Cleaned PV modules:	$a = -4.7530$ $b = -0.0150$ $c = 8.9873$ $d = -0.0007$

PV plant 2

Dirty PV modules:	$a = -11.3732$ $b = -0.0079$ $c = 9.2640$ $d = -0.0014$
Cleaned PV modules:	$a = -16.8737$ $b = -0.0065$ $c = 8.9998$ $d = -0.0011$

A.2. Weights and bias for BNNs

PV plant 1
Cleaned PV modules:

$w1 = [-0.0017 \quad 0.0004 \quad -52.7204 \quad -4.0518]$
$w2 = 10^4\ [-1.9024 \quad -0.0032]$
$b1 = [0.3864 \quad 0.3926]$
$b2 = 1.1377 \times 10^4$

Dirty PV modules:

$w1 = [-0.0032 \quad 0.0053 \quad -0.0050 \quad -0.0004]$
$w2 = 10^3[-8.1343 \quad -4.5679]$
$b1 = [2.1150 \quad 0.2607]$
$b2 = 9.8980\text{e}+003$

PV plant 2
Cleaned PV modules:

$w1 = [-0.2331 \quad -4.8670 \quad -0.0019 \quad 0.0014]$
$w2 = 10^4\ [0.0000 \quad -1.7939]$
$b1 = [-0.4681 \quad 0.4060]$
$b2 = 1.0854 \times 10^4$

Dirty PV modules:

$w1 = [-0.0051 \quad 0.0055 \quad -0.0044 \quad 0.0031]$
$w2 = 10^3\ [-6.3055 \quad -5.5721]$
$b1 = [0.1280 \quad 3.1044]$
$b2 = 8.8546$

Appendix B.

$$R = \frac{\sum_{i=1}^{N}(P_{m,i} - \overline{P}_{m,i})(P_{f,i} - \overline{P}_{f,i})}{\sqrt{\sum_{i=1}^{N}(P_{m,i} - \overline{P}_{m,i})}\sqrt{\sum_{i=1}^{N}(P_{f,i} - \overline{P}_{f,i})}}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N}(P_{m,i} - P_{f,i})^2}{N}}$$

$$\text{MAE} = \frac{\sum_{i=1}^{N}|(P_{m,i} - P_{f,i})|}{N}$$

$$\text{MAPE} = \left|\frac{\sum_{i=1}^{N}P_{m,i} - \sum_{i=1}^{N}P_{f,i}}{\sum_{i=1}^{N}P_{m,i}}\right| \times 100\%$$

References

[1] International Energy Agency. PVPS annual report; 2011. <http://www.iea.org/>.
[2] www.epia.org.
[3] www.gse.it.

[4] IEA PVPS Task 2. Cost and performance trends in grid-connected photovoltaic systems and case studies. Report IEAPVPS T2-06; 2007. <http://www.iea.org>.
[5] Hammond R, Srinivasan D, Harris A, Whitfield K, Wohlgemuth J. Effects of soiling on PV module and radiometer performance. In: 27th IEEE photovoltaic specialist conference, Hawaii, USA; 1997.
[6] Kimber A, Mitchell L, Nogradi S, Wenger H. The effect of soiling on large grid-connected photovoltaic systems in California and the Southwest region of the United States. In: 4th IEEE world conference on photovoltaic energy conversion, Hawaii, USA; 2006.
[7] Pang H, Close J, Lam K. Study on effect of urban pollution to performance of commercial copper indium diselenide module. In: IEEE 4th world conference on photovoltaic energy conversion, Waikoloa, HI; 2007. p. 2195–8.
[8] Massi Pavan A, Mellit A, De Pieri D. The effect of soiling on energy production for large-scale photovoltaic plants. Sol Energy 2011;85:1128–36.
[9] Qasem H, Betts TR, Gottschalg R. Effect of shading caused by dust on cadmium telluride photovoltaic modules. In: 36th IEEE photovoltaic specialists conference, Seattle, WS; 2011.
[10] Bouaouadja N, Bouzid S, Hamidouche M, Bousbaa C, Madjoubi M. Effects of sandblasting on the efficiencies of solar panels. Appl Energy 2000;65:99–105.
[11] Said SAM. Effects of dust accumulation on performances of thermal and photovoltaic flat-plate collectors. Appl Energy 1990;37:3–84.
[12] Garcia M, Marroyo L, Lorenzo E, Perez M. Soiling and other optical losses in solar-tracking PV plants in Navarra. Prog Photovoltaic Res Appl 2011;19:211–7.
[13] Thevenard D, Pelland S. Estimating the uncertainty in long-term photovoltaic yield predictions. Sol Energy 2011. http://dx.doi.org/10.1016/j.solener.2011.05.006.
[14] Riley CJ, Littmann B. Direct monitoring of energy lost due to soiling on first solar modules in California. IEEE J Photovoltaics 2013;3:336–40.
[15] Qasem H, Betts TR, Müllejans H, AlBusairi H, Gottschalg R. Dust-induced shading on photovoltaic modules. Prog Photovoltaic Res Appl 2012. http://dx.doi.org/10.1002/pip.2230.
[16] Agathokleous R, Kalogirou SA. On-site PV characterization and the effect of shading and soiling on their performance. In: Proceedings of SEEP conference, Dublin, Ireland; 2012. p. 262–7.
[17] Kalogirou SA, Agathokleous R, Panayiotou G, et al. On-site PV characterization and the effect of soiling on their performance. Energy 2013. http://dx.doi.org/10.1016/j.energy.2012.12.018.
[18] Kurokawa K, Komoto K, van der Vleuten P, Faiman D. Energy from the desert: VLSPV systems, socioeconomic, technical, financial and environmental aspects, London, Sterling, VA; 2010.
[19] Drews A, de Keizer AC, Beyer HG, Lorenz E, Betcke J, van Sark WGJHM, et al. Monitoring and remote failure detection of grid-connected PV systems based on satellite observations. Sol Energy 2007;81:548–64.
[20] Marion B, Adelstein J, Boyle K, Hayden H, Hammond B, Fletcher T et al. Performance parameters for grid-connected PV systems. In: 31st IEEE photovoltaics specialists conference and exhibition, 3–7 January, 2005. <www.nrel.gov/docs/fy05osti/37358.pdf>.
[21] Luque A, Hegedus S. Handbook of photovoltaic science and engineering. New York: John Wiley and Sons, Inc.; 2006. p. 167.
[22] Rahman MH, Yamashiro S. Novel distributed power generating system of PV-ECaSS using solar energy estimation. IEEE Trans Energy Convers 2007;2007(22):358–67.
[23] Hadi MA, Abdulrahman M, Al-Ibrahim M, Gurvinder S. Neuro-fuzzy based solar cell model. IEEE Trans Energy Convers 2004;19:619–24.
[24] Caputo D, Grimaccia F, Mussetta M, Zich RE. Photovoltaic plants predictive model by means of ANN trained by a hybrid evolutionary algorithm. In: IEEE international joint conference on neural networks, 18–23 July, 2010. p. 1–6.
[25] Mellit A, Kalogirou SA. Artificial intelligence techniques for photovoltaic applications: a review. Prog Energy Combust Sci 2008;34:574–632.
[26] Patra JC. Neural network-based model for dual junction solar cells. Prog Photovoltaic Res Appl 2011;19:33–44. http://dx.doi.org/10.1002/pip.985.
[27] MacKay DJC. A practical Bayesian framework for backpropagation networks. Neural Comput 1992;4:448–72.
[28] Yacef R, Benghanem M, Mellit A. Prediction of daily global solar irradiation data using Bayesian neural networks: a comparative study. Renew Energy 2012;48:146–54.
[29] http://www.skytron-energy.com/.
[30] Burden F, Winkler D. Bayesian regularization of neural networks. In: Livingstone DJ, editor. Artificial neural networks: methods and applications. Humana Press; 2008.
[31] Mayer D, Wald L, Poissant Y, Pelland S. Performance prediction of grid-connected photovoltaic systems using remote sensing. Report IEA-PVPS T2-07; 2008. p. 18.