Available online at www.sciencedirect.com

ScienceDirect

Solar Energy 123 (2016) 116–126

SOLAR ENERGY

www.elsevier.com/locate/solener

Modeling of soiled PV module with neural networks and regression using particle size composition

Subrahmanyam Pulipaka *, Fani Mani, Rajneesh Kumar

Department of Electrical and Electronics Engineering, Birla Institute of Technology and Science, Pilani, India

Received 1 September 2015; received in revised form 9 November 2015; accepted 10 November 2015
Available online 9 December 2015

Communicated by: Associate Editor Igor Tyukhov

Abstract

Particle size composition of the soil accumulated on a photovoltaic module influences its power output. It is therefore crucial to understand, quantify and model this soiling phenomenon with respect to particle size composition for predicting soiling losses. Five different soil samples from Shekhawati region in India are collected and relative percentage of standard particle sizes which are 2.36 mm, 1.18 mm, 600 μm, 300 μm, 150 μm, 75 μm and less than 75 μm are determined from sieve analysis. In order to understand and quantify the soiling effect, regression model is developed and to predict the power loss at various levels of irradiances, neural networks model is developed from the obtained experimental data. These models were compared and validated for the power output obtained at wide range of irradiance levels. It was concluded that regression can be used to analyze and quantify the particle size influence on the soiling losses of a PV module while neural networks are efficient in predicting the power output of a soiled panel. It was also observed that influence of 75 μm and lesser size particles is predominant on the power output at low irradiance levels (300–500 W/m^2) while it is the 150 μm particle size that impact the power output at higher levels of irradiance (1000–1200 W/m^2).

Keywords: Irradiance; Levenberg–Marquardt algorithm; Neural networks; Particle size composition; Regression; Soiling

1. Introduction

Desert environments are attractive locations for installing PV plants of small scale to large scale capabilities due to their typical environmental conditions, mainly high irradiance and low cloud cover. However, these areas predominantly experience no rainfall during which accumulation of soil or dust on the panel takes place. This accumulation known as soiling is said to reduce the power production of the modules by as high as 30% (Zorrilla-Casanova et al., 2013). It was the year 1944, when the studies regarding soiling phenomena surfaced and since then numerous contributions have been made in analyzing soiling effect (Sarver et al., 2013). However, it is the need of the hour to determine, quantify and model these losses using modern techniques or state of the art approaches. The soiling losses initially were extracted from the available performance data of a plant in California (Kimber et al., 2007). Due to advancement in technology and availability of various analytical tools, the developed systems (Caron and Littmann, 2013) have been able to measure soiling by tracking specific data. Gradually, these losses have also been quantified for large scale photovoltaic plants (Massi Pavan et al., 2011) Furthermore, to quantify the transmittance and spectral losses the minimum density of the soil accumulated (Burton et al., 2015) is analyzed. Studies have

* Corresponding author at: Department of Electrical and Electronics Engineering, BITS Pilani Vidya Vihar, Pilani, Rajasthan 333031, India.
E-mail address: pulipakasubbu@gmail.com (S. Pulipaka).

http://dx.doi.org/10.1016/j.solener.2015.11.012
0038-092X/

also proved that (Burton and King, 2014) the type of soil, in specific the particle size composition of the soil influences the spectral content of the incident light thereby leading to power losses.

A mathematical relationship between the dust on the module and the reduced electrical energy using equations of cumulative dust, solar radiation and energy was developed by Ketjoy and Konyu (2014) and the same relationship was studied experimentally by observing the *I–V* characteristics of soiled PV panel (Rao et al., 2014). Since soiling is a stochastic phenomena which is highly location oriented, there are various studies aimed at quantifying this phenomena at varied locations across the globe. For example, in Belgium (Appels et al., 2013) the soiling effect was studied by examining the physical properties of dust through scanning electron microscope and a solution to use a special coating was suggested. Similarly in the case of Colarodo, USA (Boyle et al., 2013) external glass plates similar to PV plates were used and the relation between the density of accumulation and light transmission was analyzed. Also, through an experiment conducted in Navarra, Spain (Garcia et al., 2011), a method was proposed to calculate the soiling losses based on the difference between the irradaince received by the cells and irradiance measured by the devices. It was an experiment in a solar park in Spain (Lorenzo et al., 2013) where the impact of the non homogenous deposits of dusts and their influence in power losses which poses threat to the life time of a panel was analyzed. A theoretical model for representing these soiling losses was proposed through experiments carried out in Southern Spain which is based on the percentage of the PV module surface exposed to soiling. The effect of soiling in CPV systems in Canberaa and Madrid is experimentally studied by Vivar et al. (2010) and the corresponding soiling losses were quantified. Also, through an experiment conducted in Southern Italy (Massi Pavan et al., 2013) models to evaluate the soiling losses were developed using Bayesian neural networks and regression polynomial and the best technique to predict the power losses was determined. In many of the aforementioned models authors have tried to estimate the power losses based on the density of the soil on the panel. However in (Fani et al., 2015) a neural network model of a soiled panel is developed to predict the soiling losses using particle size composition.

All the above mentioned works have used physical experiments to quantify soiling related effects and analytical tools to develop mathematical model that can ease the efforts to predict the power output of soiled panels. To carry forward the work related to soiling loss, this work develops a neural network model along with a regression based model to predict the power output and analyze the loss in a soiled panel. Neural network model can help in effective predictions while regression analysis is a statistical tool that can help in establishing the relationship between the independent input variables and the output variable. Therefore, this regression model can help in deciphering the effect or influence of particle size composition of the soil on the power losses. Both of these models were validated at wide range of irradiance levels from the experimentally obtained data and the results were analyzed. Five different soil samples were collected from Shekhawati Region in India which are used for experiments. This region is one of the most arid regions in India and analyzing the soiling losses in this region can help the upcoming large scale PV installations in this area. Although the specific models developed here 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 model the soiled panel through which the losses can be predicted.

2. Experimental methods

Five soil samples were collected from five different locations in Shekhawati region of Rajasthan in India. The soils collected belonged to the following regions namely, *Raghunathgarh* (Soil 1), *Neem Ka Thana* (Soil 2), *Khetri* (Soil 3), *Sikar* (Soil 4) and *Pilani* (Soil 5).

2.1. Sieve analysis

Sieve analysis is a technique adopted to assess the particle size distribution of a granular material. In this analysis, a box which contains seven sieves with standard sizes of 2.36 mm, 1.18 mm, 600 μm, 300 μm, 150 μm, 75 μm and a pan (to collect particle sizes less than 75 μm particle sizes) that are placed from top to bottom respectively. A 500 g sample of each soil is taken one at a time and is deposited in the top sieve of 2.36 mm size. Then this box is placed on an electrical shaking platform and power is switched on for 10 min. During this shaking, the particles of the respective sieve diameter remain in the sieve and other particles with diameter less than the sieve get transferred to the next sieve. The particles having size less than 75 μm get collected in the pan placed at the bottom most layer. After the power is switched off, the weight of soil particles deposited on each sieve is determined. Fig. 1 shows the output of the sieve analysis where the percentage of the standard particle sizes in the five soils is interpreted through a column chart representation. Soil 1 has higher composition (83.8%) of 150 μm particle size as well as the highest composition of this particle size among the five soils. Soil 2 has equal composition of 150 μm (32%) and 75 μm (35%) particle sizes and it has the highest composition of 75 μm particle size among these soils. Soil 3 and Soil 4 have 300 μm particle sizes in their composition in abundance (58% and 47% respectively). However soil 3 has highest 300 μm particle size composition among the soils. Soil 5 has 150 μm particle sizes mainly constituting its particle size composition (65.2%).

2.2. Artificial soiling experiment

In the artificial soiling experiment, all the 5 soil samples are spread over a PV panel one at a time and

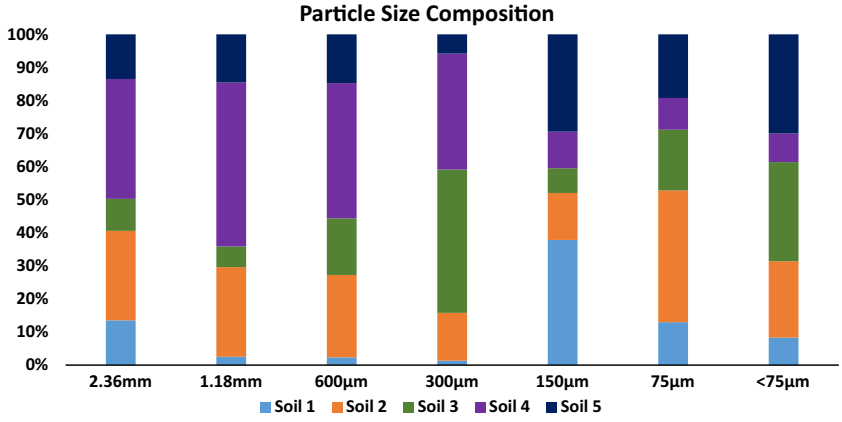

Fig. 1. Particle size composition of the soil samples.

the corresponding short circuit current (I_{SC}) and open circuit voltage (V_{oc}) readings were taken. These readings were obtained at varied levels of horizontal irradiance ranging from 200 to 1200 W/m^2. The pattern of the soil on the panel and its density are the factors of interest in an artificial soiling experiment. In order to nullify the dependency of soiling effect on these factors exhaustive data (*947 samples*) at different irradiances is taken and is verified to follow the same trend irrespective of the pattern of the soil on the panel. This data is taken for a set of 18 tilt angles (0°, 15°, 30°, 45°, 55°, 60°, 61°, 62°, 63°, 64°, 65°, 66°, 67°, 68°, 69°, 70°, 75°, and 90°). Power at the corresponding irradiance levels is calculated from V_{oc} and I_{sc} for the panels soiled with each of these soils. The experimental data collected for all the soils is represented as power vs irradiance plot in Fig. 2 at 18 tilt angles.

Out of all the 18 tilt angles for which experiment was conducted, data at the tilt angle 60° (30° from vertical) is chosen for developing the regression and neural network models. The reason behind choosing this tilt angle is its closeness to the latitude of Pilani (28.37°) where the experiment was conducted. The irradiance vs power output data of the 5 soiled panels at this tilt angle (60°) is graphically represented by a contour as shown in Fig. 3.This irradiance vs power output representation depicts the output power range at different levels of irradiances. For example when the panel is soiled with soil 1, the output power range at the irradiance level (230–500 W/m^2) is between 20 and 40 W. Similarly, between the irradiance levels of (550–900 W/m^2), the output power of this soiled panel is in the range of 40–60 W. Since the wireframe contour is a 2-d representation of a 3-d curve, the projection of the overlapping surface present in 3-d reoccur after the irradiance limit of every soil is reached. For example the presence of 40–60 W power range after 60–80 W range at 960 W/m^2 irradiance for soil 3 is due to the fact that there

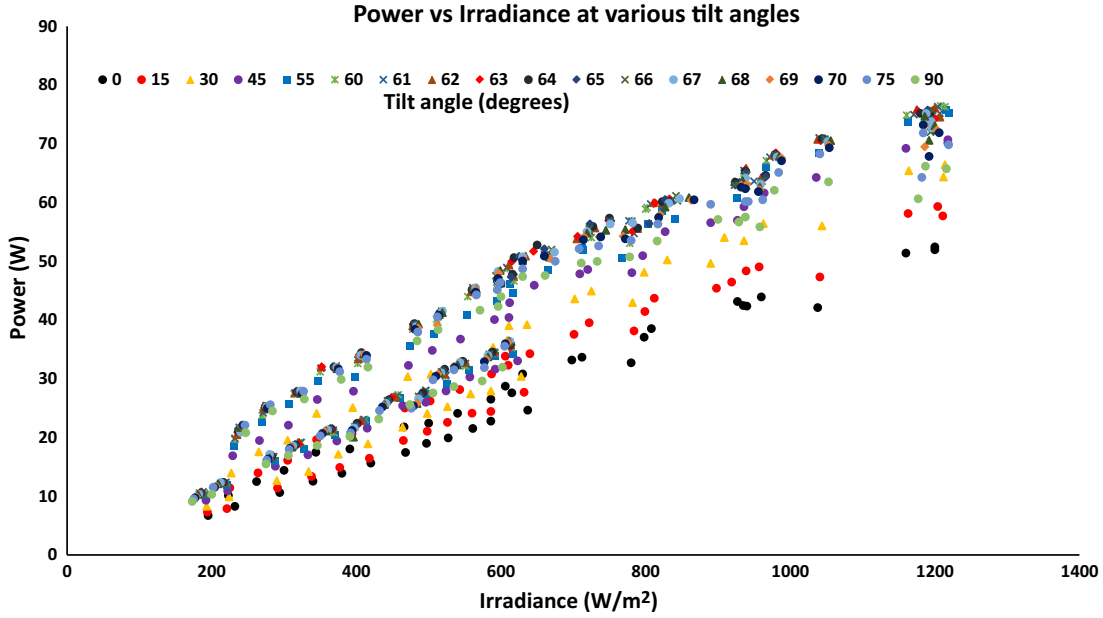

Fig. 2. Power vs irradiance representation of data collected for the soils at 18 different tilt angles.

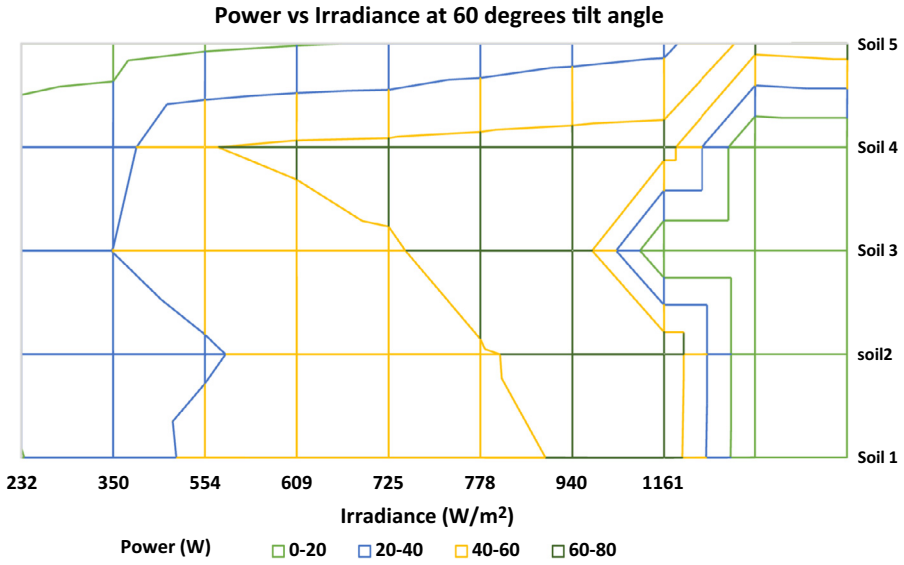

Fig. 3. Power vs irradiance graph of the data at 60° tilt angle.

are no values available for soil 3 after this irradiance level and the curves present after this level are the projections of overlapping surfaces in 3-d representation of irradiance vs power of each soil.

3. Artificial neural network (ANN) Model

Artificial Neural Networks have been significantly used to model and predict various solar photovoltaic phenomena ranging from predicting the solar radiation (Quesada-Ruiz et al,. 2015) to estimating the components of solar radiation (Elminir et al., 2005) or photovoltaic plant power prediction (Saint-Drenan et al., 2015). Furthermore these networks have also been able to predict the short term power output of large scale power plants (Mellit et al., 2014) as well as plants operating in arid environment (Mashaly et al., 2015). From aforementioned research works it can be clearly understood that neural networks are one of the best available mathematical techniques to model and predict photovoltaic power output under different environmental conditions. Hence, neural network model is developed in this research work to model and predict the soiling losses on a PV module.

An artificial neural network is a representation of system of interconnected neurons that can exchange messages among themselves. The specified connections are represented with weights which can be modified with respect to the desired model to obtain the target output. Multi-Layer Perceptron (MLP) model is a type of neural network that is made of an input layer, an output layer and one or more hidden layers as shown in Fig. 4. Each neuron of a layer is connected to the neuron of another layer and every connection is associated with a weight that determines the strength of the connection or the relation of that particular input variable to the output variable. This network approximates the nonlinear input–output relationships as defined by

$$B_j = \sum_{i=1}^{j} w_{ij} I_i + w_{0j} I_0 \tag{1}$$

where I_i ($i = 0,1,\ldots 8$) is the input node and $w_{i,j}$ is the weight between the nodes i and j. Each node value is carried through a transfer function to the next node. The most commonly used nonlinear transfer function known as sigmoid function in feed forward networks is given by

$$\sigma(u) = \frac{1}{(1+e^{-u})} \tag{2}$$

The MLP architecture also houses a bias node, b, in its input and hidden layers; the bias nodes are also connected to all the nodes in the subsequent layer. The N number of nodes in the input layer is equal to the number of process operating variables, whereas the number of output nodes K equals the number of process outputs. However, the number of hidden nodes L is an adjustable parameter whose magnitude is determined by the desired approximation and generalization performance of the network model.

At the offset, the neural network initializes the weights and biases of the network based on the range of inputs and targets provided to the network. With these initial weights and biases the neural network carries the training as shown in Fig. 4. The weights multiplied with the input, are added to the bias whose sum in turn goes as an input to the transfer function. This process continues till the end of the neural network. After obtaining the output, it is compared with the target. Consequently the gradient and other performance parameters like MSE, SSE or RMSE are calculated and based on the training algorithm the weights are updated. This updating of weights continues until the optimum value of performance function is reached. The commonly employed performance function is the RMSE defined as

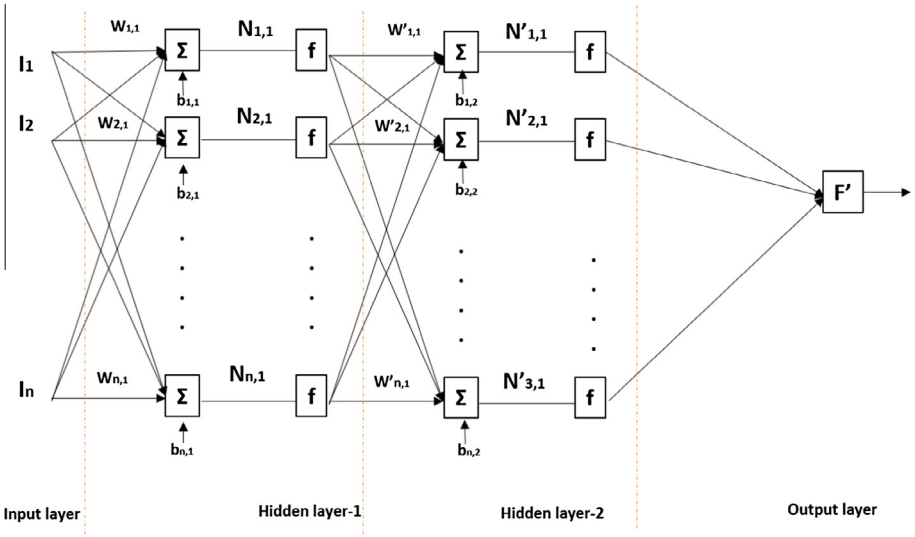

Fig. 4. Multilayer perceptron model of the neural network.

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}(t_{i,k} - o_{i,k})^2} \tag{3}$$

where $t_{i,k}$ is the target and $o_{i,k}$ is the output.

The ANN is then trained iteratively to obtain an optimal weight set $\{w_k\}$ and bias set $\{b\}$ that minimizes the performance function of root mean square error (RMSE) between the network outputs and the corresponding target value. The network thus obtained can be used to predict the target of any given input range.

For the case of this research work the data obtained through artificial soiling experiment is used to develop the neural network model. The input parameters for training this neural network are the percentage of different particle sizes present in each soil and the horizontal irradiance value represented by *I,* hidden layer 1 by *HL 1*, hidden layer 2 by *HL 2* and output *O* as shown below

$$I = [I_1 \quad I_2 \quad I_3 \quad I_4 \quad I_5 \quad I_6 \quad I_7 \quad I_8]^T \tag{4}$$

where $I_1 = \mathrm{S1}$, % of 2.36 mm size particle, $I_2 = \mathrm{S2}$, % of 1.18 mm size particle$I_3 = \mathrm{S3}$, % of 600 μm size particle, $I_4 = \mathrm{S4}$, % of 300 μm size particle$I_5 = \mathrm{S5}$, % of 150 μm size particle, $I_6 = \mathrm{S6}$, % of 75 μm size particle$I_7 = \mathrm{S7}$, % of <75 μm size particle, $I_8 = I$, incident horizontal irradiance

$$HL1 = \left[f(\textstyle\sum I_i * W_{i,1} + b_{i,1}) \ldots . f\left(\sum I_i * W_{i,11} + b_{i,11}\right)\right]^T \tag{5}$$

$$HL2 = \left[f\left(\sum_{j=1} f(\textstyle\sum I_i * W_{i,1} + b_i) * W_{j,2} + b_{j,2},\right) \ldots . f\left(\sum_{j=1} f(\textstyle\sum I_i * W_{i,1} + b_i) * W_{j,6} + b_{j,6},\right)\right]^T \tag{6}$$

$$O = [P] \tag{7}$$

Experimenting with different number of neurons in hidden layers, training the network and validating the network for minimum error the number of neurons in the hidden layers are determined. The ANN was created, trained and implemented by developing a MATLAB code (represented by Pseudo code) with feed forward network that uses Levenberg–Marquardt algorithm (LMA) for training.

3.1. Pseudo code

- Read The Inputs And Targets –
- ***I=xlsread('Jse.Xlsx', 'A1:H44')***
- ***O=xlsread('Jse.Xlsx', 'I1:I44')***
- ***Inputs=I'***
- ***Outputs=O'***
- Initialize The Neural Network – **Feed forward network**
- ***Net=feedforwardnet([11 5])***
- Configure The Neural Network – **Configure(Net)**
- ***Net=Configure(Net,Inputs,Outputs)***
- Train The Neural Network- **Trainlm(Net)**
- ***Net=Trainlm(Net,Inputs,Outputs)***
- Compare The Output With The Target By Calculating MSE, Gradient
- Continue The Training Until The Desired Value Of MSE Is Obtained
- Stop Training At The Epoch When The Optimum Performance Is Reached (MSE, Gradient)
- Plot The Training, Testing And Validation Regression Plots Along With The Performance

The developed code meets the custom requirements of developing the model from highly nonlinear data for better power prediction and it also offers high flexibility of altering the training algorithms and neurons in the hidden layer through iterations as compared to the MATLAB tool

box. LMA is the most widely used optimization algorithm which is mainly used to solve the problem of least squares curve fitting, namely nonlinear least squares minimization. Since the data obtained is highly nonlinear, LMA is used to give better prediction results. In this model, a RMSE of 10^{-3}, a minimum gradient of 10^{-7} and maximum iteration epoch of 52 were obtained. The training process would stop if any of these conditions were met. The initial weights and biases of the network were generated automatically by the program. The neural network model designed is operating with minimum errors with 11 neutrons in the first hidden layer and 5 neutrons in the second hidden layer as shown in Fig. 4.

After training this neural network the respective weights (w_{ij}) between two nodes of input layer to the first hidden layer $\{W_{\mathrm{I,HL1}}\}$, first hidden layer to the second hidden layer $\{W_{\mathrm{HL1,HL2}}\}$ and finally the second hidden layer to the output layer $\{W_{\mathrm{HL2,O}}\}$ are calculated and represented in the form of weight matrices as shown below.

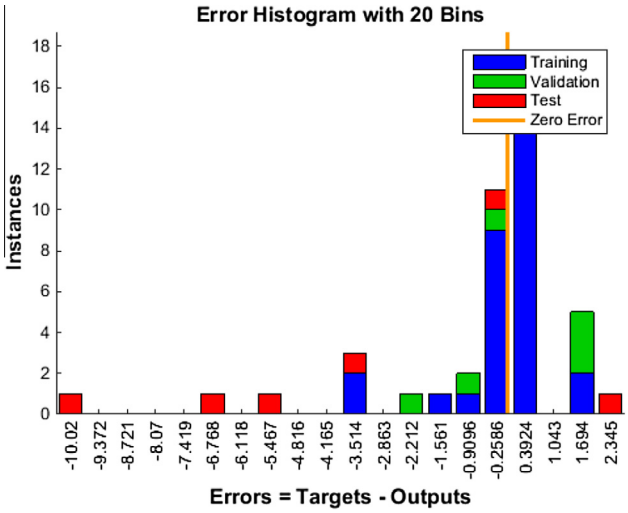

Fig. 5. Error histogram for the neural network model.

$$W_{\mathrm{I,HL1}} = \begin{Bmatrix} 1.79 & -1.18 & 1.33 & -0.99 & -0.1 & 0.08 & -0.56 & 0.95 & 0.73 & 1.46 & 1.86 \\ -0.2 & 0.97 & -0.5 & -0.03 & -0.9 & 0.61 & -0.72 & 0.06 & 0.15 & 1.54 & 0.88 \\ -0.1 & -0.63 & -0.9 & -0.73 & 0.7 & -0.6 & 1.05 & -0.53 & 0.81 & -0.42 & -0.55 \\ -0.9 & 0.34 & -0.1 & -0.55 & 0.56 & 0.22 & 0.16 & 0.82 & -0.6 & 0.94 & 0.48 \\ -0.5 & 0.11 & -1 & -0.85 & -0.2 & 0.71 & 1.23 & -0.85 & 0.41 & -0.13 & -0.93 \\ -0.3 & -0.83 & 0.54 & -0.16 & 0.65 & -0.88 & 0.49 & -0.39 & 0.48 & 0.6 & 0.49 \\ -0.03 & -0.75 & -0.3 & 0.399 & -1.1 & 0.81 & 0.08 & 0.8 & -1.3 & -0.01 & -0.74 \\ 1.44 & -0.83 & 0.88 & 0.289 & 1 & -0.97 & -0.32 & -0.8 & 0.77 & -0.47 & 0.75 \end{Bmatrix}$$

$$W_{\mathrm{HL1,HL2}} = \begin{Bmatrix} -0.7 & -1.1 & -0.9 & -0.8 & 1.64 \\ 1.64 & -0.3 & -0.4 & -0.5 & 1.22 \\ 0.46 & -0.3 & -1.5 & 0.9 & 0.16 \\ -0.97 & -1.6 & 0.25 & -0.2 & -0.5 \\ -0.38 & -0.6 & -0.7 & -0.6 & -0.2 \\ -1.05 & 0.65 & 0.86 & -0.4 & 0.2 \\ -0.15 & 0.44 & -0.8 & -0.7 & 0.46 \\ -0.13 & -0.7 & -0.6 & -0.3 & -1.2 \\ 0.84 & 0.97 & 0.56 & -0.3 & -0.4 \\ -0.45 & 0.58 & -0.9 & -0.8 & -0.7 \\ 0.08 & 0.46 & -0.3 & -0.1 & 0.24 \end{Bmatrix}$$

$$W_{\mathrm{HL2,O}} = \begin{Bmatrix} 1.04 \\ 0.84 \\ -0.3 \\ 0.28 \\ 0.88 \end{Bmatrix}$$

The error histogram of neural network training is shown in Fig. 5 and the data fitting plot is shown in Fig. 6. There are nearly 38 instances out of 45 data points with zero or negligible error. 70% of the data collected was used for the training, 15% of the data for validation and the rest 15% was used for testing.

A random validation was done at an irradiance level of $I = 232\ \mathrm{W/m^2}$ for soil 1. The power obtained through the model is observed to be 19.92 W while the experimental value at this irradiance is 19.64 W (see Fig. 3).

4. Multiple linear regression model

Apart from artificial neural networks, multiple linear regression is also a widely used mathematical tool in developing empirical formulae for photovoltaic phenomena ranging from solar collectors (Kicsiny, 2014) to solar radiation prediction (Ramedani et al., 2014), (Fu and Cheng, 2013) and maximum power point tracking (Massi Pavan et al., 2014). In recent times this mathematical tool is also being used to characterize the power losses of a PV module under practical conditions. For example (Makrides et al., 2014) have used linear regression to model the loss performance of grid connected PV modules under field conditions. (Mejia and Kleissl, 2013) have specifically analyzed and quantified soiling losses on PV modules in California using linear regression.

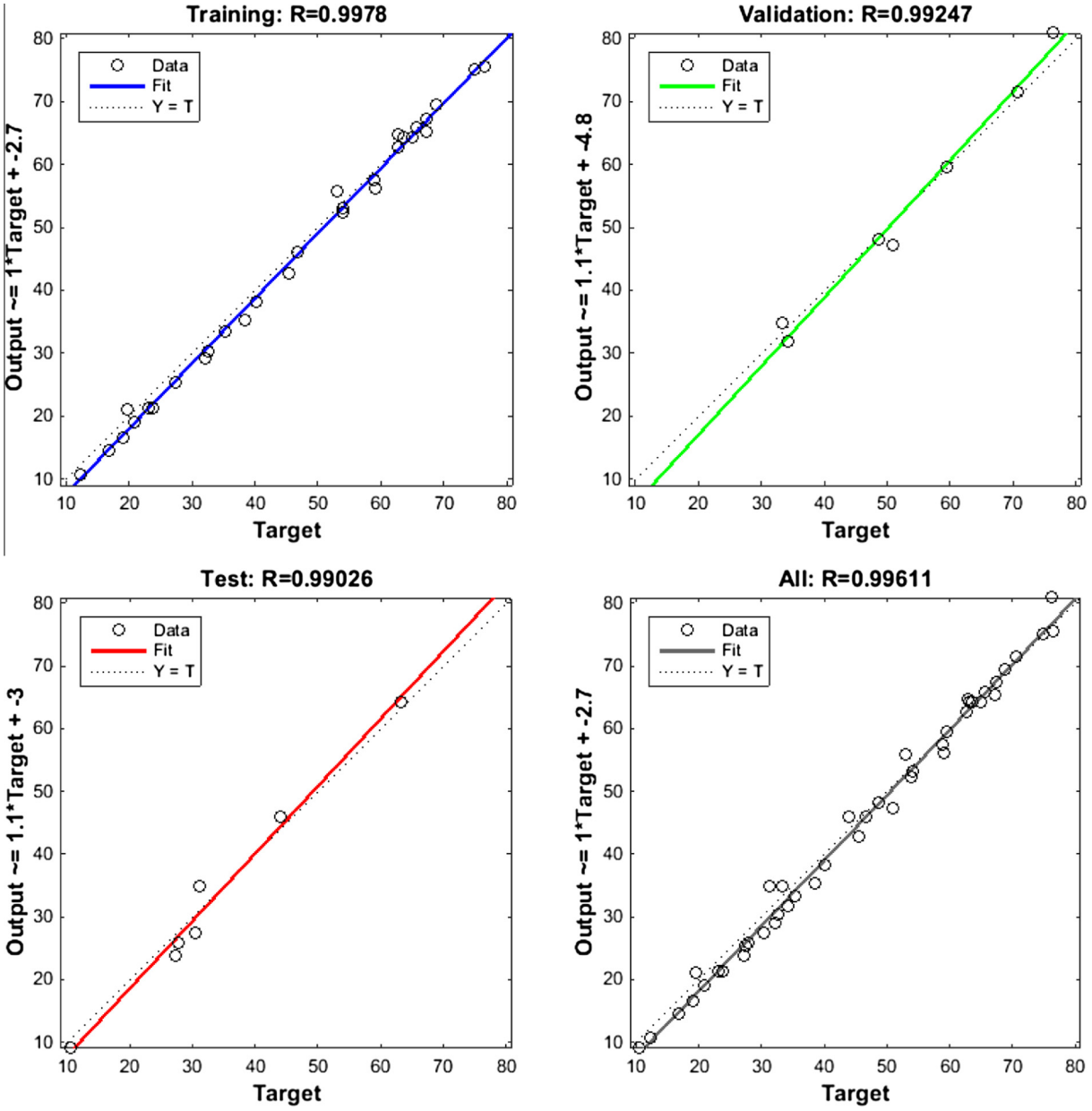

Fig. 6. Residual plot for the neural networks during training, testing, validating and overall performance.

Multiple linear regression analysis is a statistical process that helps to deduce the relationship between a set of independent variables on a dependent variable. This analysis helps in studying functional dependencies between input and output factors, implying that each input variables (x_1, x_2, x_3, ...) partially determines the level of output variable (y). Every value of the independent variable **x** is associated with a value of the dependent variable y.

For a data available with n observations of p independent variables in each sample, the regression line for variables x_1, x_2,..., x_p is defined as

$$\hat{y}_i - \bar{y} = \beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \cdots \beta pXp \quad (8)$$

This relation describes how the mean response of the model changes with the variables. The observed values for y vary about their means $\bar{y}$ and are assumed to have the same standard deviation σ. The fitted values b_0, b_1, ..., b_p estimate the parameters β_0, β_1, ..., β_p of the population regression line along with a residual. Since the observed values for y vary about their mean $\bar{y}$, the multiple regression model includes a term for this variation called residual. In simple terms the regression expresses the model as FIT + Residual, where $\beta 0 + \beta 1X1 + \beta 2X2 + \beta 3X3 + \cdots \beta pXp$ represents FIT and the residual term (denoted by ε) represents the deviations of the observed values y from their means $\bar{y}$, which are normally distributed with mean 0 and variance σ.

Hence, the representation of a multiple linear regression can be expressed as

$$y_i - \bar{y}_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon_i$$
$$\text{For } i = 1, 2, \ldots, n \tag{9}$$

The best fit for the model is obtained by minimizing the sum of the square of deviation from each data point to the fit line. The deviations are squared and then summed up and the least square estimates $b_0, b_1, b_2, \ldots, b_p$ are statistically computed.

The value fit by the equation in the form of $b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_p X_{ip}$ is represented as $\hat{y}_i$ and the observed value is denoted by y_i.

Hence the residual can be calculated as

$$e_i = y_i - \hat{y}_i \tag{10}$$

The ANOVA (Analysis of Variance) table is the measuring parameter of the regression which helps us in deducting the nature of the regression and curve fit. As specified earlier if we simplify the model of a regression it can be represented as

$$\text{DATA} = \text{FIT} + \text{Residual}$$

Table 1
ANOVA table.

	Degrees of freedom (df)	Sum of Squares(SS)	Mean square (MS)
Model	p	$\sum (\hat{y}_i - \bar{y})^2$	$\frac{\sum (\hat{y}_i - \bar{y})^2}{p}$
Error	$n-p-1$	$\sum (y_i - \hat{y}_i)^2$	$\frac{\sum (y_i - \hat{y}_i)^2}{n-p-1}$
Total	$n-1$	$\sum (y_i - \bar{y}_i)^2$	$\frac{\sum (y_i - \bar{y}_i)^2}{n-1}$

Which can be re- written from (8) (9) and (10) as

$$y_i - \bar{y}_i = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i) \tag{11}$$

where $(y_i - \bar{y}_i)$ the total variance in the response, $(\hat{y}_i - \bar{y})$ is variation in the mean response and $(y_i - \hat{y}_i)$ is the residual value. Squaring and summing these value results in

$$\sum (y_i - \bar{y}_i)^2 = \sum (\hat{y}_i - \bar{y})^2 + \sum (y_i - \hat{y}_i)^2 \tag{12}$$

Here, the R^2 which is square of sample correlation is expressed as

$$R^2 = \frac{\sum (y_i - \bar{y}_i)^2}{\sum (\hat{y}_i - \bar{y})^2} \tag{13}$$

The variance is estimated as

$$\sigma^2 = s^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - p - 1} = \text{MSE} \tag{14}$$

where MSE is the mean square error from which it can be seen that the estimate of the standard error (s) is the square root of MSE.

Hence the ANOVA table can be represented mathematically as given in Table 1.

The ANOVA table also has a column named F, which is the output of the F test performed on the model obtained. In an F-test the regression model is tested for the hypothesis of $\beta i \neq 0$ against the null hypothesis of $\beta i = 0$. The coefficients, standard error of the variables are calculated from these statistical expressions along with their t-stat and p-values are present in the summary output table of the regression. T-static is the estimated coefficient divided by its own standard error. This measures the standard deviations from zero of the coefficient and helps in deciding

Table 2
Regression analysis summary.

SUMMARY OUTPUT	
Regression Statistics	
Multiple R	0.98918
R Square	0.97847
Adjusted R Square	0.89879
Standard Error	3.02334
Observations	45

ANOVA

	df	SS	MS	F	Significance F
Regression	8	16207.112	2025.889	354.6169	7.445E-32
Residual	39	356.4846	9.14063		
Total	47	16563.597			

	Coefficients	Std. Error	*t* Stat	*P*-value	Lower 95%	Upper 95%
Intercept	−18.90408	10.179	−1.857	0.0708	−39.49	1.685
S1	0	0	65,535	0	0	0
S2	0	0	65,535	0	0	0
S3	0	0	65,535	0	0	0
S4	0.43968	0.1266	3.4716	0	0.183	0.695
S5	0.27400	0.1081	2.5341	0.0154	0.055	0.492
S6	0.52402	0.1402	3.7357	0.0005	0.240	0.807
S7	−4.4512	0.6477	−6.8719	3.21E−08	−5.761	−3.141
I	0.0612	0.00179	34.042	1.27E−30	0.057	0.064

whether the corresponding independent variable really belongs to the model. The p-value is the probability of observing a t-stat value of large magnitude, given the null hypothesis that the true coefficient value is zero. A low p value ($p < 0.05$) indicates one can reject the null hypothesis and include the variable in the regression equation.

The experimental data collected in terms of irradiance and power is used to develop the multi linear regression model, in specific an empirical equation of power in terms of particle size composition of soil ($S1$–$S7$) and the horizontal incident irradiance (I). The regression analysis was carried out using regression add-on in Microsoft excel. In this research work there are 8 independent variables, with 45 observations which determine the power P in each observation. The summary output of the regression analysis is represented in Table 2. Where the parameters ranging from R-square to ANOVA table and coefficients of variables along with their corresponding t-stat and p-values are represented.

Since the p value of all the coefficients is less than 0.05 at 95% confidence interval and their corresponding absolute value of t-stat variable is more than 2, all the variables can be included in the equation. The equation of the power output in terms of variables ($S1$–$S7$, I) is interpreted from the above model as given below.

$$P = 0.44 * S4 + 0.274 * S5 + 0.524 * S6 - 4.45 * S7 + 0.061 * I - 18.90 \quad (15)$$

It can be observed that this equation is independent of the variables $S1$, $S2$ and $S3$ thereby suggesting these particle sizes have no contribution in the power output of the soiled panel.

At an irradiance of $I = 232$ W/m^2 the power output for soil 1 with $S4 = 1.76$, $S5 = 83.8$, $S6 = 11.44$ and $S7 = 0.72$ using the regression model is obtained to be 21.32 W and the value from experiment at this irradiance is 19.64 W (Fig. 3). The residual plot of the regression model obtained is shown in Fig. 7. The difference between calculated value and experimental value is given by $R^2 = 0.9785$, thereby revealing that 97.85% of the output data obtained through regression is in lieu with the experimental data.

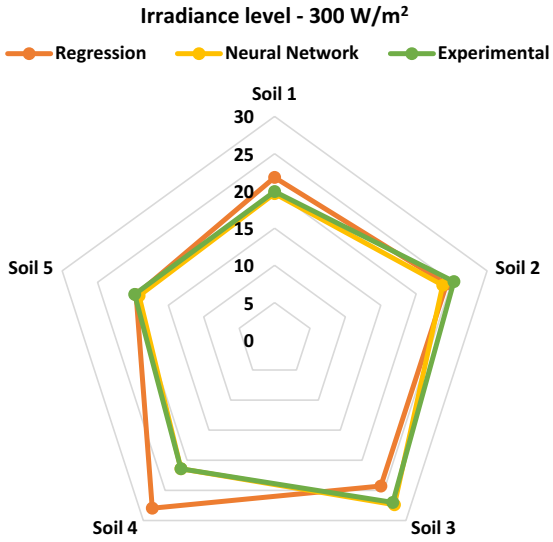

Fig. 8. Comparison of power output at 300 W/m^2.

5. Comparison of neural network and regression models

After developing a linear expression of power output through regression in terms of particle size composition and irradiance (15) and a neural network model to predict the power output of a soiled photovoltaic panel with the particle size composition and irradiance as inputs, a comparison between these models is done to determine the best model for a soiled panel to predict power output. For this purpose, these models are compared at different levels of irradiance extending from low (100 W/m^2) to high (1200 W/m^2) with respect to the corresponding experimental data. The samples of power output of these models at

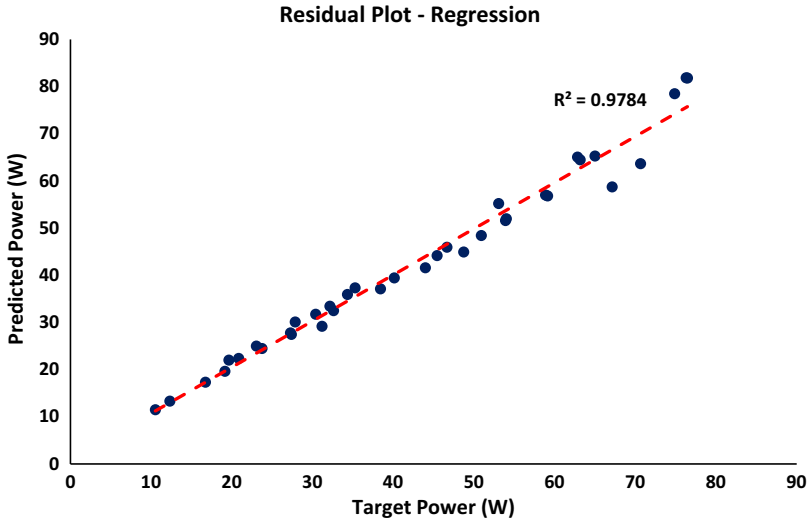

Fig. 7. Residual plot of regression model.

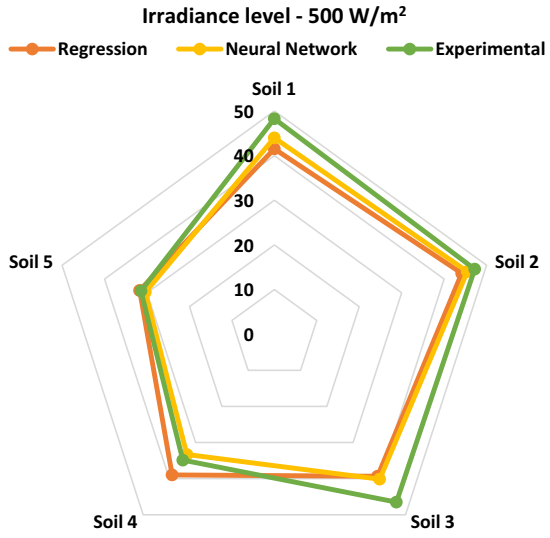

Fig. 9. Comparison of power output at 500 W/m^2.

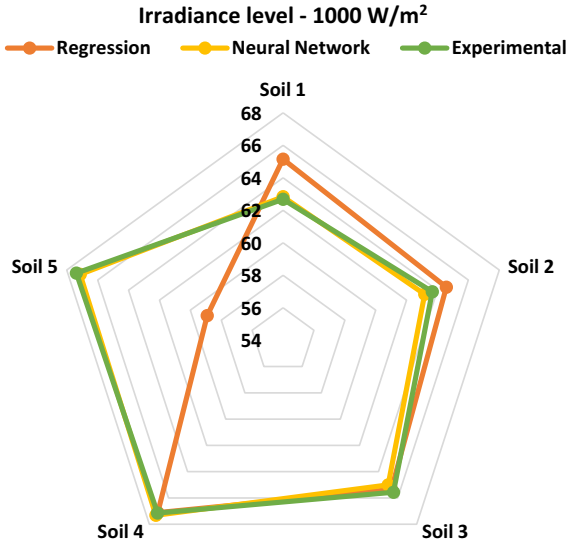

Fig. 10. Comparison of power output at 1000 W/m^2.

300 W/m^2, 500 W/m^2 and 1000 W/m^2 are compared with experimental results and are illustrated in Figs. 8–10. The nearest value to the experiments is highlighted in Table 3.

From Fig. 8, at an irradiance of 300 W/m^2 it can be seen that neural network model is predicting the power output accurately for all the soils, while regression model is unable to accurately estimate the output. The composition of particles of size 75 μm and below is least in Soil 1 (12%) and soil 4 (10%) as compared to other soils as shown in Fig. 1. In these soils regression is over estimating the power, while the rest of the soils whose least particle size composition is higher than 12% the regression model is under estimating the power.

From Fig. 9, at an irradiance of 500 W/m^2, we can see that regression underestimates the power output for all the soils except soil 4 which has least particle size composition of 75 μm among all the soils (8.5%, as shown in Fig. 1) particle size. While neural networks over estimate power for all soils with a maximum offset of 3 W.

From Fig. 10, at an irradiance of 1000 W/m^2 we observe that neural networks are in lieu with the experimental outputs while regression over estimates the power output for soil 1 and under estimates the power output of soil 5 at this irradiance level. It is also interesting to observe from Fig. 1, that the particle size composition of these two soils is similar in a way that they have 150 μm particle size as the main constituent in the particle size composition (83.8% for soil1 and 65.2% for soil 2) among the soils considered.

6. Conclusion

Overall, it appears that neural networks are somewhat, but not consistently better than multivariable regression models for determining the power output of soiled panel. On analyzing both models at different levels of irradiance, it can be inferred that their performance is being governed by the particle size composition of the soils. At lower as well as higher irradiances regression analysis fails to estimate the power output of soils with lower composition of 75 μm particles and higher concentration of 150 μm particle sizes respectively. This observation from the comparison can lead to insightful conclusions where it can be stated that the influence of 75 μm and lesser size particles is predominant on the power output at lower irradiance levels while it is the 150um particle size that may impact the power output at higher levels of irradiance.

Analyzing these models for each soil specifically interesting observations can be concluded.

- For soil 1 having high 150 μm particle composition regression can predict power accurately at 500 W/m^2 while for the other two ranges it is neural network that performs better.

Table 3
Comparison of Power output at 3 irradiance levels.

	Irradiance level −300 W/m^2			Irradiance level −500 W/m^2			Irradiance level −1000 W/m^2		
	Regression	Neural	Experiment	Regression	Neural	Experiment	Regression	Neural	Experiment
Soil 1	21.82	**19.64**	19.92	41.53	**43.99**	48.28	65.15	**62.84**	62.68
Soil 2	**24.29**	23.74	25.32	44.13	**45.44**	47.19	64.57	**63.17**	63.66
Soil 3	24.28	**27.38**	27	39.33	**40.13**	46.5	**65.35**	65.01	65.57
Soil 4	27.95	**21.4**	21.4	38.96	**33.32**	34.87	**67.12**	67.32	67.15
Soil 5	**19.55**	19.14	19.76	**31.73**	30.39	31.4	58.91	**67.13**	67.38

- For soil 2 having high composition of 75 μm particle composition regression can predict power accurately at 300 and 500 W/m^2 while for the 1000 W/m^2 it is neural network that performs better.
- For soil 3 having high composition of 300 μm particle composition regression can predict power accurately at 500 W/m^2 and 1000 W/m^2 while for the 300 W/m^2 it is neural network that performs better.
- For soil 4 having high composition of 1.18 mm particle composition only neural networks can predict power accurately at all the irradiance levels.
- For soil 5 having high composition of less than 75 μm particle composition regression can predict power accurately at only 300 W/m^2 while for the 500 W/m^2 and 1000 W/m^2 it is neural network that performs better.

Although regression model seemed to be failing to estimate the power losses very often, this model can however help us to decipher the relationship between the particle size of soil and the power output of the soiled panel. Therefore to conclude, regression based models are helpful in determining the relationship between the particle size compositions, irradiance with power but it is the neural networks that are efficient in power prediction given the particle size composition.

References

Appels, R., Lefevre, B., Herteleer, B., Goverde, H., Beerten, A., Paesen, R., Poortmans, J., 2013. Effect of soiling on photovoltaic modules. Sol. Energy 96, 283–291. http://dx.doi.org/10.1016/j.solener.2013.07.017.

Boyle, L., Flinchpaugh, H., Hannigan, M., 2013. Impact of natural soiling on the transmission of PV cover plates. Conference Record of the IEEE Photovoltaic Specialists Conference 3276–3278. http://dx.doi.org/10.1109/PVSC.2013.6745150.

Burton, P.D., King, B.H., 2014. Spectral sensitivity of simulated photovoltaic module soiling for a variety of synthesized soil types. IEEE J. Photovolt. 4 (3), 890–898. http://dx.doi.org/10.1109/JPHOTOV.2014.2301895.

Burton, P.D., Boyle, L., Griego, J.J.M., King, B.H., 2015. Quantification of a minimum detectable soiling level to affect photovoltaic devices by natural and simulated soils 5 (4), 1143–1149.

Caron, J.R., Littmann, B., 2013. Direct monitoring of energy lost due to soiling on first solar modules in California. IEEE J. Photovolt. 3 (1), 336–340. http://dx.doi.org/10.1109/JPHOTOV.2012.2216859.

Elminir, H., Areed, F., Elsayed, T., 2005. Estimation of solar radiation components incident on Helwan site using neural networks. Sol Energy 79 (3), 270–279. http://dx.doi.org/10.1016/j.solener.2004.11.006.

Fani, M., Subrahmanyam, P., Rajneesh, K., 2015. Modeling of soiled photovoltaic modules with neural networks using particle size composition of soil. In: presented at 42nd Photovoltaic Specialist Conference, New Orleans, USA.

Fu, C.L., Cheng, H.Y., 2013. Predicting solar irradiance with all-sky image features via regression. Sol. Energy 97, 537–550. http://dx.doi.org/10.1016/j.solener.2013.09.016.

García, M., Marroyo, L., Lorenzo, E., Pérez, M., 2011. Soiling and other optical losses in solar-tracking PV plants in navarra. Progr. Photovolt.: Res. Appl. 19 (2), 211–217. http://dx.doi.org/10.1002/pip.1004.

Ketjoy, N., Konyu, M., 2014. Study of dust effect on photovoltaic module for photovoltaic power plant. Energy Proc. 52, 431–437. http://dx.doi.org/10.1016/j.egypro.2014.07.095.

Kicsiny, R., 2014. Multiple linear regression based model for solar collectors. Sol. Energy 110, 496–506. http://dx.doi.org/10.1016/j.solener.2014.10.003.

Kimber, A., Mitchell, L., Nogradi, S., Wenger, H., 2007. The effect of soiling on large grid-connected photovoltaic systems in California and the Southwest Region of the United States. In Conference Record of the 2006 IEEE 4th World Conference on Photovoltaic Energy Conversion, WCPEC-4 (Vol. 2, pp. 2391–2395). http://dx.doi.org/10.1109/WCPEC.2006.279690.

Lorenzo, E., Moreton, R., Luque, I., 2013. Dust effects on PV array performance in-field observations with non-uniform patterns. Progr. Photovolt. Res. 22, 666–670.

Makrides, G., Zinsser, B., Schubert, M., Georghiou, G.E., 2014. Performance loss rate of twelve photovoltaic technologies under field conditions using statistical techniques. Sol. Energy 103, 28–42. http://dx.doi.org/10.1016/j.solener.2014.02.011.

Mashaly, A.F., Alazba, a.a., Al-Awaadh, a.M., Mattar, M.a., 2015. Predictive model for assessing and optimizing solar still performance using artificial neural network under hyper arid environment. Sol. Energy 118 (0), 41–58. http://dx.doi.org/10.1016/j.solener.2015.05.013.

Massi Pavan, a., Mellit, a., De Pieri, D., 2011. The effect of soiling on energy production for large-scale photovoltaic plants. Sol. Energy 85 (5), 1128–1136. http://dx.doi.org/10.1016/j.solener.2011.03.006.

MassiPavan, a., Mellit, a., Lughi, V., Pavan, a.M., Mellit, a., Lughi, V., 2014. Explicit empirical model for general photovoltaic devices: experimental validation at maximum power point. Sol. Energy 101 (July 2015), 105–116. http://dx.doi.org/10.1016/j.solener.2013.12.024.

MassiPavan, a., Mellit, a., De Pieri, D., Kalogirou, S.a., 2013. A comparison between BNN and regression polynomial methods for the evaluation of the effect of soiling in large scale photovoltaic plants. Appl. Energy 108, 392–401. http://dx.doi.org/10.1016/j.apenergy.2013.03.023.

Mejia, F.a., Kleissl, J., 2013. Soiling losses for solar photovoltaic systems in California. Sol. Energy 95, 357–363. http://dx.doi.org/10.1016/j.solener.2013.06.028.

Mellit, a., Massi Pavan, a., Lughi, V., 2014. Short-term forecasting of power production in a large-scale photovoltaic plant. Sol. Energy 105, 401–413. http://dx.doi.org/10.1016/j.solener.2014.03.018.

Quesada-Ruiz, S., Linares-Rodríguez, a., Ruiz-Arias, J.a., Pozo-Vázquez, D., Tovar-Pescador, J., 2015. An advanced ANN-based method to estimate hourly solar radiation from multi-spectral MSG imagery. Sol. Energy 115, 494–504. http://dx.doi.org/10.1016/j.solener.2015.03.014.

Ramedani, Z., Omid, M., Keyhani, A., Khoshnevisan, B., Saboohi, H., 2014. A comparative study between fuzzy linear regression and support vector regression for global solar radiation prediction in Iran. Sol. Energy 109, 135–143. http://dx.doi.org/10.1016/j.solener.2014.08.023.

Rao, A., Pillai, R., Mani, M., Ramamurthy, P., 2014. Influence of dust deposition on photovoltaic panel performance. Energy Proc. 54, 690–700. http://dx.doi.org/10.1016/j.egypro.2014.07.310.

Saint-Drenan, Y.M., Bofinger, S., Fritz, R., Vogt, S., Good, G.H., Dobschinski, J., 2015. An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation. Sol. Energy 120, 479–493. http://dx.doi.org/10.1016/j.solener.2015.07.024.

Sarver, T., Al-Qaraghuli, A., Kazmerski, L.L., 2013. A comprehensive review of the impact of dust on the use of solar energy: History, investigations, results, literature, and mitigation approaches. Renew. Sustain. Energy Rev. http://dx.doi.org/10.1016/j.rser.2012.12.065.

Vivar, M., Herrero, R., Antón, I., Martínez-Moreno, F., Moretón, R., Sala, G., Smeltink, J., 2010. Effect of soiling in CPV systems. Sol. Energy 84 (7), 1327–1335. http://dx.doi.org/10.1016/j.solener.2010.03.031.

Zorrilla-Casanova, J., Piliougine, M., Carretero, J., Bernaola-Galván, P., Carpena, P., Mora-López, L., Sidrach-De-Cardona, M., 2013. Losses produced by soiling in the incoming radiation to photovoltaic modules. Progr. Photovolt.: Res. Appl. 21 (4), 790–796. http://dx.doi.org/10.1002/pip.1258.