Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 155 (2019) 463–470

Procedia Computer Science

www.elsevier.com/locate/procedia

The 9th International Conference on Sustainable Energy Information Technology (SEIT)
August 19-21, 2019, Halifax, Canada

Using Linear Regression and Back Propagation Neural Networks to Predict Performance of Soiled PV Modules

Salsabeel Shapsough *, Rached Dhaouadi, Imran Zualkernan

American University of Sharjah, University City, Sharjah, 26666, United Arab Emriates

Abstract

This paper presents a study on neural network-based modeling techniques and sensor data to estimate the power output of photovoltaic systems under soiling conditions. Predicting maximum power output under soiling conditions is considered an important and difficult problem and a variety of models using a host of factors including temperature and weather profiles have been proposed. This study used linear regression models and artificial neural networks and used only solar irradiation and ambient temperature, as well and the maximum power point (MPP) characteristic variables of photovoltaic (PV) modules obtained from online current-voltage (IV) tracers in the site of a PV installation. The two models were trained and validated using actual monitoring data of two 100-Watt PV modules installed in the UAE. One reference panel was cleaned on a weekly basis and the second panel was left to accumulate dust over the entire period between July 1, 2018 and 17 September, 2018. The results show that it is possible to predict maximum power output of soiled PV modules at about 97% accuracy. The proposed models perform at an accuracy comparable to more complex models in literature.

Keywords: PV; Neural Networks; Soiling;

1. Introduction

Solar energy has been recognized as one of the leading candidates to replace conventional, non-renewable energy sources. This alternative clean energy source offers solutions to some of the main challenges facing carbon-based

* Corresponding author. Tel.: +97-50-20-60-122
E-mail address: sshapsough@aus.edu

10.1016/j.procs.2019.08.065

fossil energy sources such as depletion of existing reserves, and damage to vital environmental resources caused by greenhouse gas emissions [1], [2]. Furthermore, as a consequence of recent advances made in the field of solar photovoltaic cells, solar energy may offer significantly-lower costs in the long run [3], [4]. However, wide-scale adoption of solar energy entails building solutions that provide efficient and stable energy. Efficiency in solar panels is defined as the amount of power produced as a function of solar irradiance and cell surface area. The amount of power that a solar facility can generate depends on its module fabrication, configuration, and the installation environment [5]. Past decades have witnessed major breakthroughs in PV module fabrication techniques that made it possible to reduce maintenance costs and to extract more power from single modules [6], [7]. There have been numerous studies that have proposed solutions to operational issues of solar panel management using concepts such as solar tracking and dynamic reconfiguration [8], [9]. Recent advances have made it so that solar facilities can be especially designed and manufactured to yield maximum output in a specific setting.

However, once in the field, the efficiency of a solar module is affected by its environment. Environmental variables such as shading [10], [11], air quality [12], and temperature [13] can influence how much power a solar system can generate, and hence causing intermittency [14], [15]. While it is difficult to control environmental factors, the ability to quantify the actual efficiency loss under various environmental conditions makes it possible to mitigate operational issues and to rectify system errors. Recent studies on maximizing the efficiency of solar modules incorporate the use of machine learning (ML) and artificial neural networks (ANN) to model and predict the performance of PV modules given a range of natural conditions [16]. Using in-situ measurements of modules' output, it is possible to forecast environmental factors such as solar irradiance and sun incidence [17], PV module performance metric such as output power [18] for grid integration and fault detection [19]. Such information is valuable in planning operation and maintenance (O&M) for PV microgrids [20] and for integrating of PV microgrids with large-scale grids [21].

Table 1. Literature review of perfromance prediction for soiled panels

Publication	Inputs	Model	Output	Best accuracy
[22] Pulipaka et al.	Particle size composition, incident horizontal irradiance	Neural network + multi linear regression	Power	98%
[23] Pulipaka and Kumar	Particle size composition, chemical composition, spectral behavior, title angle, irradiance	Neural network + multi linear regression	Power	98%
[24] Javed et al.	PM_{10}, wind speed, wind direction, ambient temperature, relative humidity, previous day PM_{10}, previous day wind speed, previous day relative humidity	Neural network	Cleanness index	54%
[25] Mani et al.	particle size composition, tilt angle, irradiance	Multi linear regression	power	97%
[26] Andrews and Pearce	Tilt angle, coating, short circuit current, module temperature, global irradiance, diffuse irradiance, snow depth	Linear model	Maximum power, short circuit current	
[27] You et al.	Relative humidity, temperature, wind speed, PM_{10}, $PM_{2.5}$, precipitation, rainy days	Linear model	Soiling loss	

One of the more complex environmental factors is soiling. Soiling is defined as the accumulation of dirt, snow, dust particles, and bird droppings on the surface of a solar panel. Soiling effectively reduces the amount of solar irradiance absorbed by the panel, and consequently the power generated. While periodic cleaning of the panels can reduce the impact of soiling, cleaning introduces new costs that may exceed the cost of decreasing power output at some point [28]. In such situations, maintaining solar power generation efficiency becomes an optimization problem [29]. Therefore, determining an optimal frequency of cleaning that maximizes power output while minimizing cleaning costs becomes the key to cost-effective, large-scale solar energy facilities. Most existing work on modeling and predicting the output of soiled panels depends largely on complex measurements to analyze the soiling particle's physical, chemical, and morphology characteristics [30]. As shown in Table 1, actors such as particle size, soil composition, and soiling layer depth are also commonly analyzed and used to model soil accumulation with high accuracy. This kind of analysis can pose a research overhead and add to the cost of instrumentation needed to run such projects. Alternatively, a neural network is capable of predicting the output of soiled networks using easily accessible data such as solar irradiance or weather forecasts would make it possible to study and predict the performance of large-scale solar facilities using historic data and climate forecasts.

This paper presents a study on prediction of PV output in soiling-susceptible regions using simple linear models and neural networks. The work studies the possibility of using easily-available weather information and temporal data exclusively to generate accurate predictions that are comparable to more complex models presented in literature.

Rest of the paper is organized as follows. Next section describes the PV system that was implemented to collect sensor data. This is followed by analysis of the collected data using linear regression and neural network models. The paper ends with a conclusion.

2. PV System and Data Collection

Fig. 1 shows a view of the PV system test facility. The PV system test facility included a standalone PV module of two panels of poly-crystalline silicon (Trina Solar TSM-PA05.08), each one of 100 Watt at peak power (Wp). Each panel supplied energy to a 12V battery and a resistive load through a maximum power point tracking (MPPT) charger. As Fig. 1 shows, one solar panel was used as a reference (i.e., the clean panel) and was cleaned on a weekly basis. The second identical panel, referred to as the soiled panel, was left to accumulate dust between 1st of July 2018 and 16th of September 2018. Maximum power and short current circuit were measured once every hour, every day.

Fig. 1. Solar Test Facility located in the parking area of the Engineering Building, American University of Sharjah, Sharjah, UAE.

The efficiency of a PV module is measured using current-voltage (I-V) and power-voltage (P-V) curves. The curves (shown in Fig. 2) demonstrates all possible combinations of current, voltage, and power that a module is capable of producing under specific conditions of temperature and solar irradiance. I-V tracing data are used to construct models that assist in tracking efficiency of solar energy generation, monitoring for preventative maintenance, as well as predicting output in the short-run and the long-run.

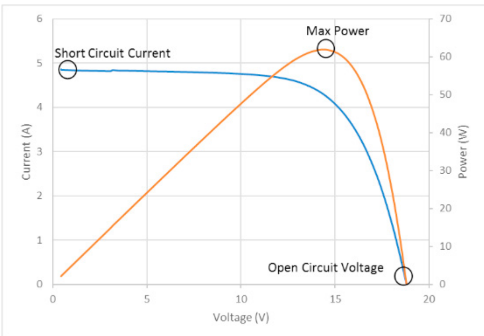


Fig. 2. I-V and P-V curves

From historical current and voltage data, it is possible to extract short circuit current (I_{SC}), open circuit voltage (V_{OC}), and maximum power point (P_{max}). The three values represent the maximum values that a PV module is capable of producing in a given setting, and can be used to compare performance of multiple modules, or the same module under different conditions. Environmental conditions that can influence IV curves include solar irradiance, ambient temperature, cell temperature, wind speed, wind direction. However, for the purpose of reducing instrumentation cost, only solar irradiance and ambient temperature were considered. Time-related data such as time of the day and month were also considered. This was to account for seasonal solar irradiance and sun incident angle as well. Day of the week was also considered as an input for the clean panel, in order to account for the cleaning schedule. Since the clean panel was cleaned once every Sunday, the weekday was included as an input variable to account for the fact that soil would accumulate between every two consecutive cleanings. Similarly, for the soiled panel, an index variable was introduced which described the number of days since last cleaning to account for the amount of soiling.

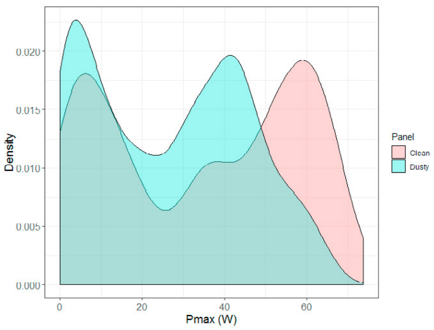
Data was collected using a remote monitoring system built using Internet of Things (IoT) technologies which was developed using virtual instrumentation and included an IV Tracer unit that communicates with a server through WiFi [21]. The IV characteristics of each panel were recorded every hour using a small microcontroller connected to a USB multimeter. The solar irradiance and temperature were measured every minute using a low-cost weather station placed on the same plane as the PV panels. The collected data was stored in a cloud server database. Power output of each panel and corresponding weather information were used to train software models which could then be used to predict the performance of modules in the future using only temporal and weather information. While modelling the clean panel aimed to detect any deviation from model "clean" behaviour, predictive power under soiling conditions makes it possible to distinguish power loss due to soiling from other, serious module fault symptoms.

3. Analysis and Results

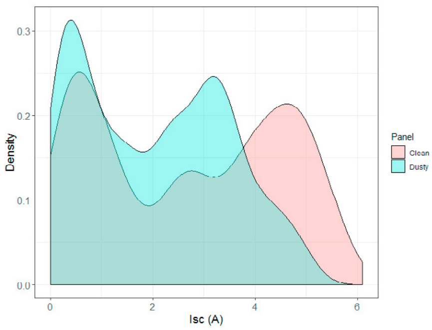
At the end of the testing period 943 data points were collected where each data point consisted of maximum power, short circuit current, solar irradiance, ambient temperature, and timestamp. The data was first cleaned of faulty readings. For example, the dataset was cleaned for missing values that were due to days when the system was under maintenance, as well as faulty zero value readings which were due to hardware malfunction either in the IV measurement unit or the weather station. The dataset was also cleaned for inconstant values were IV measurement did not correspond to the irradiance value. For each IV tracing cycle, the irradiance was first measured, then the two panels were traced one by one. Since tracing each panel took 60 seconds, there were incidents where cloud shading occurred while tracing one panel, but not the other. This was confirmed by viewing video surveillance of the setup. Since the current and voltage measured during sudden shading did not correspond to the value of irradiance measured at the beginning, these readings were discarded. The resulting, clean dataset consisted of 891 valid readings. These values were used to visualize and model the performance of the two panels. The algorithms used to model the data were linear regression, multiple linear regression, and a simple back propagation neural network. Those models were selected because they are the most popular in modeling solar power, as found in the literature. This is mainly to enable direct comparison with other models that use similar algorithms with different inputs.

3.1. Linear Regression

The distribution of the measured power is shown in Fig. 3. As shown in the figure, the distributions are not normal.



(a)



(b)

Fig. 3. Distribution of a) maximum power and b) short circuit current for each panel

The clean panel's output also displays higher variability and is skewed towards higher power and current values; which is assumed to be due to a higher range of values where the clean panel is able to generate more power at higher irradiance. The difference in variance is confirmed by performing a Levene test (P_{max}: Test Statistic = 83.756, p-value < 2.2e-16; Isc: Test Statistic = 119.42, p-value < 2.2e-16). Fig. 4 shows the maximum power against irradiance for each panel. While the output of the clean panel demonstrates a tight, linear relationship between solar irradiance and maximum power, the output of the dusty panel deviates and decreases under the same irradiance, as the soiling cycle progresses. As expected, the power varies mostly linearly with Irradiance. In addition, it can be seen that for the dusty panel, the P_{max} is reduced as the time progresses because the panel is getting dustier over time.

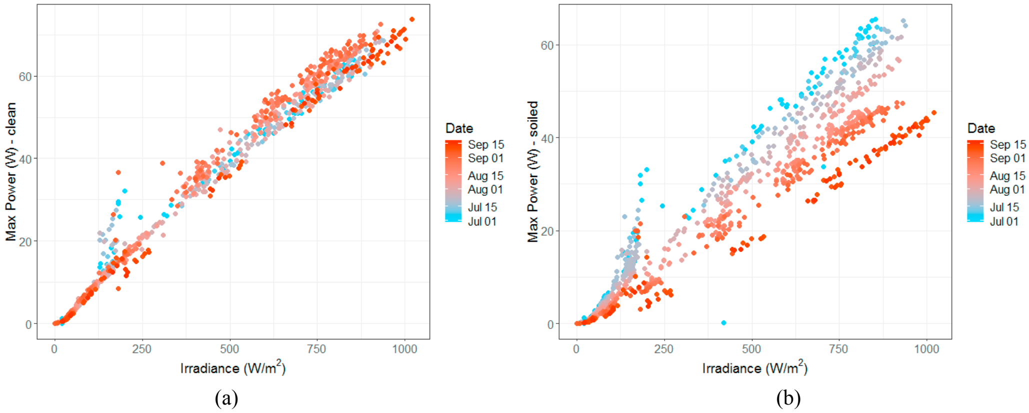


Fig. 4. Maximum power vs. irradiance for the a) clean panel and b) soiled panel, taking in date of reading as a factor

The simplest, most straight-forward approach predict P_{max} from is to use linear regression to model the output as a function of solar irradiance. The accuracy of the models was evaluated by calculate Root Mean Square Error (RMSE) and R square (R^2) values. As shown in Fig. 5(a), using linear regression, it is possible to predict the output of the clean panel to a high degree of accuracy (RMSE=0.044, R^2=0.981). However, once soiling is added to the equation, the linear model loses accuracy (RMSE=0.099, R^2=0.883). This is because the linear relationship between output power and irradiance is directly influence by the amount of soil deposited on the surface of the PV module.

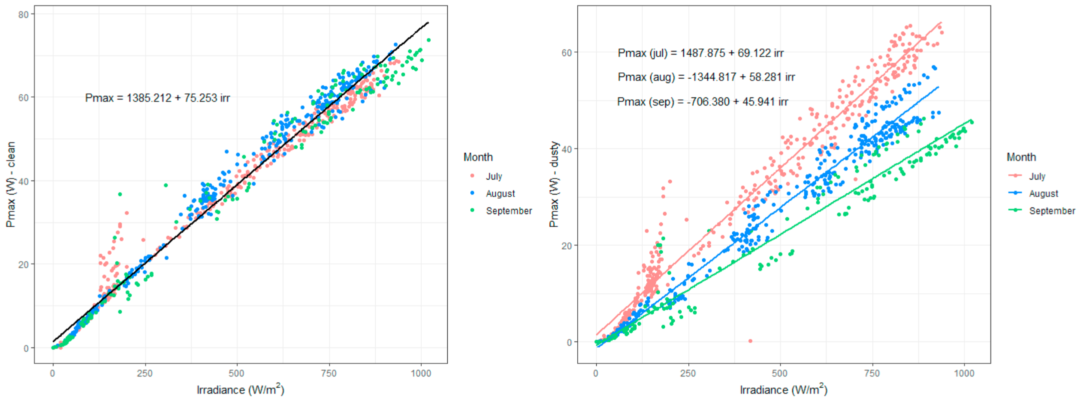


Fig. 5. (a) Modeling the output of the clean panel as a function of irradiance (b) Multiple linear regression to model the soiled panel for each month

This makes it more difficult to remotely know if the drop in power generation is due to soiling, or due to a more serious fault. As shown in Fig. 5(b), modeling the soiled panel can be approached by splitting the dataset into subsets based on the month. Building multiple linear models makes it possible to track the rate of change in the slope in the power-irradiance curve, where the output power for each month is represented by:

$$\text{Pmax} = \alpha * \text{irradiance} + \beta \qquad (1)$$

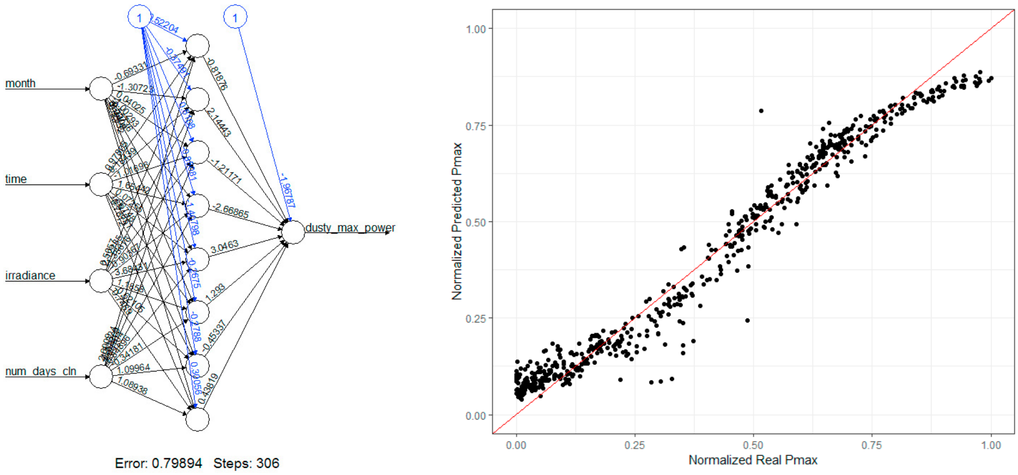
Where the slope of the curve (α) is a function of time (in months) since last cleaning. Dividing the model into temporal-based models results in higher accuracy in predicting the output of a soiled panel, based on how long it has been accumulating soil for. It is expected that monitoring for a long period of time (i.e. at least a year) would results in enough information to model the rate of change as a function of soiling period, therefore making it possible to predict power output using a linear model by predicting α and β.

Table 2. Statistical error indexes associated to differences between measured and predicted soiled panel's output for the linear models

		July	August	September
P_{max}	RMSE	0.059	0.043	0.077
	R^2	0.966	0.980	0.953
I_{SC}	RMSE	0.056	0.040	0.077
	R^2	0.967	0.981	0.949

3.2. Neural Networks

The problem of predicting the output of a soiled panel can also be approached from a neural network perspective. In addition to environmental factors such as solar irradiance and temperature, temporal data can be fed into a neural network to predict maximum power or short circuit current of the PV module in the future. To study the performance of neural network in predicting power using weather and temporal data only, a neural network was implemented using the neuralnet library in R [31]. The neural network for P_{max} and I_{SC} are shown in Fig. 6(a) and Fig. 7(a). The full dataset describing the performance of the dusty panel as well as the environmental information was used for training and testing. 10-fold cross validation was used in order to reduce bias as well as variance while training the network. Furthermore, in order to assess the goodness of fit, the cross validation was repeated in 100 iterations. The mean and standard deviation of the R^2 and RMSE values are shown in Table 3. Two neural networks were built for each P_{max} and I_{SC}, with one taking in the irradiance and temperature as well as temporal information as inputs, while the other took in irradiance only. Comparing the two uncovers that ambient temperature has no significant contribution to predicting PV output. Therefore, irradiance and temporal data are enough to build a simple neural network that can predict the output with great accuracy. Predicted vs. real P_{max} and I_{SC} values are shown in Fig. 6(b) and Fig. 7(b).



(a) (b)

Fig. 6. (a) Structure of the neural networks used to predict maximum power for the dusty panel (b) Predictions vs. real P_{max}

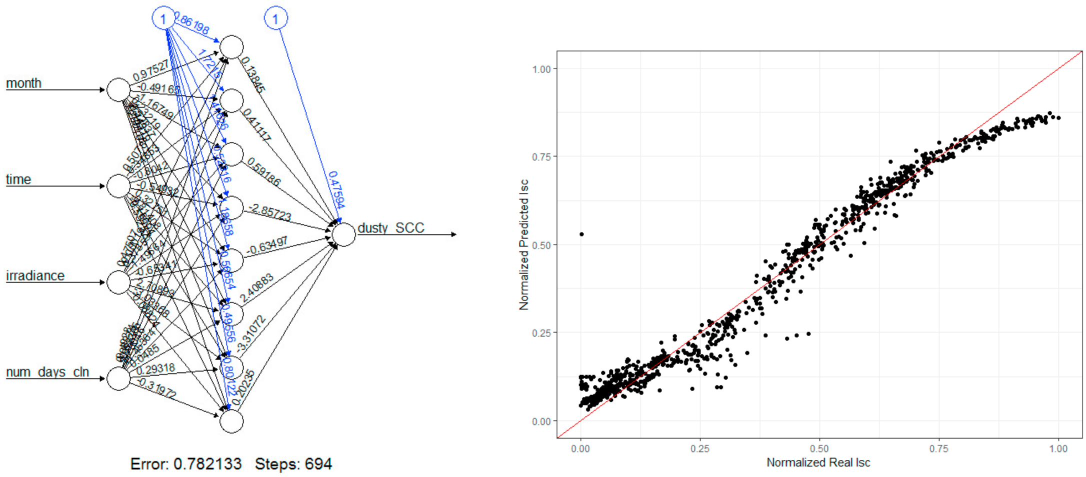


Fig. 7. (a) Structure of neural networks used to predict short circuit current (b) Predictions vs. real P_{max}

Table 3. Statistical error indexes associated to differences between measured and predicted soiled panel's output for the neural networks

Measurement	RMSE		R^2	
	Mean	Standard Deviation	Mean	Standard Deviation
P_{max} (irr+temp)	0.053	0.011	0.964	0.017
P_{max} (irr)	0.059	0.012	0.957	0.019
I_{SC} (irr+temp)	0.049	0.010	0.967	0.015
I_{sc} (irr)	0.050	0.009	0.967	0.013

Comparing the accuracy of the neural network against that of the multiple linear regression model no significant improvement. This suggests that either can be used to predict the performance of PV modules with great accuracy using only solar irradiance and temporal data. The accuracy is also comparable to that exhibited by complex models shown in works such as [22]. It is also possible to modify the time window for the linear model or the neural network structure to increase the accuracy depending on the soiling behavior in a specific environment.

4. Conclusion

Predicting the power yield of PV modules is key to providing stable solar energy in microgrids as well as integration with main power grids. There have been several works that studied the prediction of soiling patterns and solar power under soiling conditions. However, most existing work depend on the availability of a complex array of information incorporating sophisticated weather studies, as well as microanalysis of soiling particles in the area. This paper presents a study on predicting the output of soiled solar panels using openly available information including solar irradiance and ambient temperature. Two methods were used to model output of a soiled panel. The first method creates multiple linear models based on the time period since the last cleaning cycle, with each predicting output as a function of solar irradiance. The second method, on the other hand, models the soiled panel as a neural network, taking in date and time information, as well as irradiance and, optionally, temperature as an input. Both models were able to generate predictions with great accuracy.

Acknowledgements

The research reported here was supported in part by grant #SCRI-18-02, and Petrofac-Chair grant from the American University of Sharjah, UAE.

References

[1] M. T. Islam, N. Huda, A. B. Abdullah, and R. Saidur, “A comprehensive review of state-of-the-art concentrating solar power (CSP) technologies: Current status and research trends,” *Renewable and Sustainable Energy Reviews*, vol. 91, pp. 987–1018, Aug. 2018.
[2] K. Nuortimo, J. Härkönen, and E. Karvonen, “Exploring the global media image of solar power,” *Renewable and Sustainable Energy Reviews*, vol. 81, pp. 2806–2811, Jan. 2018.
[3] S. Comello, S. Reichelstein, and A. Sahoo, “The road ahead for solar PV power,” *Renewable and Sustainable Energy Reviews*, vol. 92, pp. 744–756, Sep. 2018.
[4] R. Ciriminna, F. Meneguzzo, M. Pecoraino, and M. Pagliaro, “Rethinking solar energy education on the dawn of the solar economy,” *Renewable and Sustainable Energy Reviews*, vol. 63, pp. 13–18, Sep. 2016.
[5] M. M. Fouad, L. A. Shihata, and E. I. Morgan, “An integrated review of factors influencing the perfomance of photovoltaic panels,” *Renewable and Sustainable Energy Reviews*, vol. 80, pp. 1499–1511, Dec. 2017.
[6] Z. Said, S. Arora, and E. Bellos, “A review on performance and environmental effects of conventional and nanofluid-based thermal photovoltaics,” *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 302–316, Oct. 2018.
[7] A. K. Pandey, V. V. Tyagi, J. A. Selvaraj, N. A. Rahim, and S. K. Tyagi, “Recent advances in solar photovoltaic systems for emerging trends and advanced applications,” *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 859–884, Jan. 2016.
[8] S. Malathy and R. Ramaprabha, “Comprehensive analysis on the role of array size and configuration on energy yield of photovoltaic systems under shaded conditions,” *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 672–679, Sep. 2015.
[9] Y. S. Khoo *et al.*, “Optimal Orientation and Tilt Angle for Maximizing in-Plane Solar Irradiation for PV Applications in Singapore,” *IEEE Journal of Photovoltaics*, vol. 4, no. 2, pp. 647–653, Mar. 2014.
[10] N. AL-Rousan, N. A. M. Isa, and M. K. M. Desa, “Advances in solar photovoltaic tracking systems: A review,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2548–2569, Feb. 2018.
[11] F. Belhachat and C. Larbes, “A review of global maximum power point tracking techniques of photovoltaic system under partial shading conditions,” *Renewable and Sustainable Energy Reviews*, vol. 92, pp. 513–553, Sep. 2018.
[12] H. Liu *et al.*, “The Impact of Haze on Performance Ratio and Short-Circuit Current of PV Systems in Singapore,” *IEEE Journal of Photovoltaics*, vol. 4, no. 6, pp. 1585–1592, Nov. 2014.
[13] A. Hasan, J. Sarwar, and A. H. Shah, “Concentrated photovoltaic: A review of thermal aspects, challenges and opportunities,” *Renewable and Sustainable Energy Reviews*, vol. 94, pp. 835–852, Oct. 2018.
[14] K. Engeland, M. Borga, J.-D. Creutin, B. François, M.-H. Ramos, and J.-P. Vidal, “Space-time variability of climate variables and intermittent renewable electricity production – A review,” *Renewable and Sustainable Energy Reviews*, vol. 79, pp. 600–617, Nov. 2017.
[15] P. Ravestein, G. van der Schrier, R. Haarsma, R. Scheele, and M. van den Broek, “Vulnerability of European intermittent renewable energy supply to climate change and climate variability,” *Renewable and Sustainable Energy Reviews*, vol. 97, pp. 497–508, Dec. 2018.
[16] S. Kalogirou and A. Sencan, “Artificial Intelligence Techniques in Solar Energy Applications,” *Solar Collectors and Panels, Theory and Applications*, Oct. 2010.
[17] A. Sharma and A. Kakkar, “Forecasting daily global solar irradiance generation using machine learning,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2254–2269, Feb. 2018.
[18] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, “Photovoltaic and solar power forecasting for smart grid energy management,” *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38–46, Dec. 2015.
[19] W. Chine, A. Mellit, V. Lughi, A. Malek, G. Sulligoi, and A. Massi Pavan, “A novel fault diagnosis technique for photovoltaic systems based on artificial neural networks,” *Renewable Energy*, vol. 90, pp. 501–512, May 2016.
[20] M. De Benedetti, F. Leonardi, F. Messina, C. Santoro, and A. Vasilakos, “Anomaly detection and predictive maintenance for photovoltaic systems,” *Neurocomputing*, vol. 310, pp. 59–68, Oct. 2018.
[21] R. J. Bessa, A. Trindade, C. S. P. Silva, and V. Miranda, “Probabilistic solar power forecasting in smart grids using distributed information,” *International Journal of Electrical Power & Energy Systems*, vol. 72, pp. 16–23, Nov. 2015.
[22] S. Pulipaka, F. Mani, and R. Kumar, “Modeling of soiled PV module with neural networks and regression using particle size composition,” *Solar Energy*, vol. 123, pp. 116–126, Jan. 2016.
[23] S. Pulipaka and R. Kumar, “Power prediction of soiled PV module with neural networks using hybrid data clustering and division techniques,” *Solar Energy*, vol. 133, pp. 485–500, Aug. 2016.
[24] W. Javed, B. Guo, and B. Figgis, “Modeling of photovoltaic soiling loss asa function of environmental variables,” *Solar Energy*, vol. 157, pp. 397–407, Nov. 2017.
[25] F. Mani, S. Pulipaka, and R. Kumar, “Characterization of power losses of a soiled PV panel in Shekhawati region of India,” *Solar Energy*, vol. 131, pp. 96–106, Jun. 2016.
[26] R. W. Andrews and J. M. Pearce, “Prediction of energy effects on photovoltaic systems due to snowfall events,” in *2012 38th IEEE Photovoltaic Specialists Conference*, 2012, pp. 003386–003391.
[27] S. You, Y. J. Lim, Y. Dai, and C.-H. Wang, “On the temporal modelling of solar photovoltaic soiling: Energy and economic impacts in seven cities,” *Applied Energy*, vol. 228, pp. 1136–1146, Oct. 2018.
[28] B. Stridh, “Evaluation of economical benefit of cleaning of soiling and snow in PV plants at three European locations,” in *2012 38th IEEE Photovoltaic Specialists Conference*, 2012, pp. 001448–001451.
[29] R. K. Jones *et al.*, “Optimized Cleaning Cost and Schedule Based on Observed Soiling Conditions for Photovoltaic Plants in Central Saudi Arabia,” *IEEE Journal of Photovoltaics*, vol. 6, no. 3, pp. 730–738, May 2016.
[30] S. C. S. Costa, A. S. A. C. Diniz, and L. L. Kazmerski, “Solar energy dust and soiling R&D progress: Literature review update for 2016,” *Renewable and Sustainable Energy Reviews*, vol. 82, pp. 2504–2536, Feb. 2018.
[31] Stefan Fritsch, Frauke Guenther, Marvin N. Wright, Marc Suling, and Sebastian M. Mueller, “Training of Neural Networks,” 07-Feb-2019.