Theoretical and Applied Climatology (2022) 150:715–729
https://doi.org/10.1007/s00704-022-04166-6

ORIGINAL PAPER

Evaluation of energy extraction of PV systems affected by environmental factors under real outdoor conditions

Muhammed A. Hassan[1] · Nadjem Bailek[2,3] · Kada Bouchouicha[4] · Abdelhameed Ibrahim[5] · Basharat Jamil[6] · Alban Kuriqi[7,8] · Samuel Chukwujindu Nwokolo[9] · El-Sayed M. El-kenawy[10,11]

Received: 3 January 2022 / Accepted: 3 August 2022 / Published online: 30 August 2022

Abstract

The global agenda to increase the renewable energy share has driven many countries and entities to harness solar energy from solar photovoltaic (PV) systems. However, the power generation of PV systems is strongly affected by climate conditions. Therefore, the main objective of this study is to analyze and predict the power generation of different PV technologies under arid desert climate conditions on an hourly basis. Two areas have been considered as case studies: Adrar in Algeria and Alice Springs in Australia. A total of nine physical models and input parameter combinations from six different power plants have been used and tested for the suitability of the proposed models for predicting the power yield of PV power plants depending on solar irradiance and other meteorological variables. Then, an ensemble learning technique is applied to improve the performance capabilities of the best-fit input combinations. The results reveal that the global irradiance, ambient air temperature, and relative humidity combination are the most related to the PV power output of all technologies under all-sky conditions and provide effective and efficient performance with the proposed ensemble learning, with an estimated accuracy of over 99%.

Keywords Clean energy · Desert · Renewable energy · Photovoltaic · Power generation

1 Introduction

The increasing global awareness and research interest in reducing the end-use of conventional energy resources (i.e., fossil fuels) are mainly due to their excessive emission rates of greenhouse gases, which are among the key factors behind the current climate change issues. Renewable energy resources conveniently serve as a direct substitute, especially when their costs are competitive and mature technologies (Maji et al. 2019; Bouchouicha et al. 2020a). These resources are abundant worldwide and, most importantly,

We confirm that this manuscript is original, has not been published before, and is not currently being considered for publication elsewhere.

✉ Muhammed A. Hassan
mhd.zidan17@cu.edu.eg

✉ Nadjem Bailek
bailek.nadjem@univ-tam.dz

1 Mechanical Power Engineering Department, Faculty of Engineering, Cairo University, Giza 12613, Egypt

2 Sustainable Development and Computer Science Laboratory, Faculty of Sciences and Technology, Ahmed Draia University of Adrar, Adrar, Algeria

3 Faculty of Engineering, Architectures Nisantasi University, Istanbul, Turkey

4 Unité de Recherche en Energies Renouvelables en Milieu Saharien (URERMS), Centre de Développement Des Energies Renouvelables (CDER), 01000 Adrar, Algeria

5 Computer Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt

6 Department of Computer Sciences, Universidad Rey Juan Carlos, Tulipán s/n, 28933 Móstoles (Madrid), Spain

7 CERIS, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal

8 Civil Engineering Department, University for Business and Technology, Pristina, Kosovo

9 Department of Physics, Faculty of Physical Sciences, University of Calabar, Calabar, Nigeria

10 Department of Communications and Electronics, Delta Higher Institute of Engineering and Technology, Mansoura, Egypt

11 Faculty of Artificial Intelligence, Delta University for Science and Technology, Mansoura 35712, Egypt

have a lower carbon footprint (El-Shimy 2017; Tseng 2017; Li et al. 2019; Hassan et al. 2021a).

As a well-established clean power production approach, solar power generation is gaining renewal research momentum for its offered benefits (Slimani et al. 2020). Solar photovoltaics (PV) is considered a mainstream option in the power sector. An increasing number of countries generate more than 20% of their electricity using PV systems. Nations that have shown a great interest in a significant shift to solar energy over the past 3 years include Egypt, Brazil, Mexico, Algeria, Pakistan, Turkey, and the Netherlands (IRENA 2020; Bailek et al. 2017a, 2017b).

Despite the benefits of clean electricity generation, it also has some limitations. The main drawback of those technologies, especially PV, is the stochastic and intermittent nature of the resource. The generated PV power varies mainly due to continuous changes in solar irradiance (Bouchouicha et al. 2019). This, as described by Ehsan et al. (Ehsan et al. 2017), results in several issues when connecting the PV system to the grid. Li et al. (Li et al. 2019) found that the intermittent solar irradiance affects the general dispatch ability function of the generated power. This critical issue can be reduced by accurately predicting the PV power production. The grid operator can directly intervene to increase system efficiency and energy balance in stand-alone or grid-connected modes.

Accurate estimates of the energy performance of PV systems are valuable for the planning and operational security of the power systems (Wu et al. 2021; Razmjoo et al. 2019). Many highly accurate models have been proposed for different locations around the globe. Mazumdar et al. (Mazumdar et al. 2014) derived a statistical method to empirically model the ramping behavior of utility-scale solar PV power output for short time scales. The analysis was carried out in terms of ramp rate (i.e., ramp up or ramp down), which is the change in the generated power of the PV system. It was reported that the proposed model could be used to estimate the frequencies of PV ramp events. Hassan et al. (Hassan et al. 2021b) developed genetically optimized models based on an autoregressive exogenous neural network to predict PV power production in different sub-hourly time steps (i.e., 5 to 60 min) at different desert locations in Algeria and Australia. The proposed models are sufficiently accurate, with relative random error components varying between 10 and 20%. The performances of their proposed models were also higher when having smaller prediction horizons. Bouchouicha et al. (2020a) used linear and non-linear approaches to estimate the electric power production of a 20 MWp plant installed in the Adrar region, South of Algeria, based on instantaneous radiometric and meteorological data in 15-min time intervals. According to their results, all artificial neural network-based models are superior in prediction accuracy and performance stability. Cascade-forward neural network-based models provided the most reliable predictions. Trigo-González et al. (Trigo-González et al. 2019) estimated the hourly energy production in energy yield. A multiple linear regression model was presented to determine the hourly PV production using the performance ratio factor based on selected technologies (cadmium telluride and multi-crystalline silicon). Their linear regression model was validated against an independent dataset and showed a root mean square error of around 8% for San Pedro de Atacama plants and 7 to 16% for Antofagasta plants. Wang et al. (Wang et al. 2021) compared the performances of different prediction models for forecasting PV power production. A simple efficiency, temperature correction, and one-diode model were proposed for each PV configuration. The simple efficiency model overestimated the power output of PV modules by approximately 10%, except for cadmium telluride (CDTE) PV modules.

Furthermore, it was reported that the one-diode model has the best accuracy for predicting monocrystalline (Mono-Si) power output and polycrystalline (Poly-Si) PV modules. Regressive/linear methods can be classified as linear, multiple-linear stationery, and non-linear stationary methods. These methods estimate the correlation between a dependent variable (i.e., produced power) and independent variables (i.e., predictors) and require high-quality historical data on PV output and weather conditions to enhance the PV power estimates.

Machine learning models could be too complicated for some end users. However, it is standard today to use advanced machine learning methods in data analysis and forecasting of many solar energy systems. For instance, Dolara et al. (Dolara et al. 2015) compared three physical models to predict the power output of a PV cell. The models were based on equivalent circuits of three, four, and five input parameters: photo-generated current, reverse saturation current, diode ideality factor, series resistance, and shunt resistance. Experimental data of 29 PV modules were collected for Milano, Italy, and the data was used to calibrate the models. With fast computations, the adopted calibration method showed good accuracy for the considered PV modules. Alam et al. (Alam et al. 2015) proposed a scheme for modeling and identifying the maximum power output of a stand-alone PV generator. To establish the performance of PV, a static model (i.e., with non-varying inputs) was developed in a MATLAB/Simulink environment, considering ambient air temperature and solar irradiance as real-time variables. Then, two cases of constant solar radiation with the static model and constant solar radiation with the dynamic PV model were analyzed and compared. The PV model was validated with the experimental data for 30 days, and results were analyzed for each 30-min interval. It was concluded that the developed model has a high accuracy of 99.72%. Liu et al. (Liu et al. 2018) developed a two-stage model to estimate the percentage of prediction intervals for

PV power output. The generalized regression neural network, extreme machine learning, and Elman neural network were integrated using the optimized backpropagation genetic algorithm to develop a weight-varying combination forecast model. The non-parametric kernel density estimation method was adopted to estimate the prediction intervals concerning the statistical distribution of the errors of the earlier deterministic point predictions. It was reported that the proposed algorithm produced much higher accuracy results than the conventional approaches.

This study conducts a baseline study for desert areas in Australia and Algeria to provide needed information for future developments. Many studies have been conducted to monitor challenges for solar energy in desert areas, e.g., radionuclides (Aba et al. 2018; El-Kenawy et al. 2022), PV plants' performance (Aoun et al. 2019), improvement of sustainable energy systems (Bailek et al. 2018), and passive air pollution (Tang and Al-Dousari 2006). However, to the authors' knowledge, the evaluation of PV power production of various PV technologies under the arid desert climate is unavailable. The intermittent nature of PV production poses a significant obstacle in integrating PV systems into the electric grid. As a result, accurate predictions are needed. Within the scope of this paper, regression models based on multiple environmental factors affecting PV power generation of various types of photovoltaic panels in six typical arid desert areas in Australia and Algeria are established and tested, taking into consideration the technological features of photovoltaics, along with the actual characteristics of the operation settings and climatic conditions for considered sites in hourly time scales. Then, an effective ensemble-learning approach is used to improve the performance capabilities of the optimal (best-fit) input combinations for a more accurate estimate.

2 Materials and methods

2.1 Data collection

Desert climate is experienced in arid regions. It is characterized by excessive evaporation and very low precipitation, ranging between 25 and 200 mm annually (Vaughn 2005; Sikka 1997). Dry climate regions cover 26.22% of the global land (Kottek et al. 2006). Adrar, located in Algeria, is the second-largest town in the Algerian desert in the southern region of Algeria. It is characterized by energy-rich solar resources (Bouchouicha et al. 2017, 2015; Bailek et al. 2020a) and relatively flat terrain, where the highest point reaches 421 m. The region receives annual global solar irradiation higher than 2200 kWh/m^2, with around 3500 sunshine hours (mostly clear-sky days). Alice Springs is located in Australia's interior desert region. It is part of the northern territory of Australia, with marginal rainfall. The area receives an average global horizontal irradiation of ~ 6.17 kWh/m^2/day with a daily average sunshine duration of more than 9 h (Darula et al. 2010).

Desert areas tend to have clear skies for most of the year, making it easier to forecast the PV output power compared to, e.g., tropical and temperate climates. However, other factors should be considered. For instance,

- It is widely established that the performance of solar PV systems is degraded with increasing temperatures (Rezk and Fathy 2017). Therefore, the actual outdoor performance of the solar PV cells needs to be quantified before exploring the large-scale deployment of PV plants.
- Sandstorms are frequent, and wind speeds are higher in typical desert areas, which results in a considerable deviation from the expected performance of PV panels based on standard test conditions (Mostefaoui et al. 2019).
- The dust accumulation rate is typically higher, and the typical frequency of PV cleaning is smaller since many such plants are installed remotely (Mostefaoui et al. 2019; Huang et al. 2016).
- The climate in the studied regions is mostly clear throughout the year, hence the arid desert climate classification. However, this does not mean all days of the year have clear skies. The examined areas have their considerable shares of cloudy and rainy days (Weatherspark. 2022).

Long-term measurements of the PV parameters and the relevant meteo-solar parameters are used in this study. For Adrar, these datasets are obtained from the Renewable Energy Research Unit in the Saharan Region (URERMS) site, which is located at 27°53′N latitude and 00°16′W longitude and has an elevation of 269 m. For Alice Springs, similar datasets are obtained from the Desert Knowledge Australia Solar Centre (DKASC), located at 23°46′S latitude and 133°52′E longitude, at 558 m (Desert Knowledge Australia Solar Centre - Download Data. 2020). As shown in Table 1, various photovoltaic technologies in six plants have been considered in this study, namely monocrystalline, hybrid silicon (heterojunction "HIT" cells), amorphous silicon, cadmium telluride, and polycrystalline.

All data that passed the simple quality tests were used for the study in this work. The data quality check considers all monitored parameters' physical and statistically possible, and extremely rare limits, as detailed in Hassan et al. 2021c. For instance, the tests ensured that the ratio between the ground-level global horizontal solar irradiance and the extraterrestrial horizontal irradiance (i.e., the global clearness index) is above 0.0 (considering daylight hours) and below 1.2. The tests also ensured that the minimum wind speed and humidity are $\geq$ 0.0 m/s and 0.0%, respectively, and the maximum relative humidity is $\leq$ 100%. Besides, the automated check highlighted any short-term data gaps

Table 1 General specifications of the six considered PV modules

Technology (ID)	Monocrystalline silicon (MOS1)	HIT silicon (HIS)	Amorphous silicon (AMS)	Monocrystalline silicon (MOS2)	Cadmium telluride (CDT)	Polycrystalline silicon (POS)
Location	Alice Springs	Alice Springs	Alice Springs	Adrar	Alice Springs	Alice Springs
Data source	DKASC	DKASC	DKASC	URERMS	DKASC	DKASC
Array rating	215 W	6.3 kW	6.0 kW	7.0 kW	7.0 kW	5.0 kW
Data period	2013–2015	2014–2016	2012–2014	2013–2017	2012–2013	2012–2015
Manufacturer	Trina Solar	SANYO	Kaneka	BJ POWER	First Solar	Yingli Solar
Module	TSM-175DC01	HIT-210 NKHE5	G-EA060	BJP-250SA	FS-272	YL245P-29b
Maximum power (P_{max}) [W]	175	210	60	250	72	245
Maximum power voltage (V_{pm}) [V]	36.2	41.3	67	36.99	66.7	30.2
Maximum power current (I_{pm}) [A]	4.85	5.09	0.9	8.768	1.09	8.11
Open circuit voltage (V_{oc}) [V]	43.9	50.9	91.8	30.75	88.7	37.8
Short circuit current (I_{sc}) [A]	5.3	5.57	1.19	8.131	1.23	8.79
Maximum system voltage (V_{dc}) [V]	1000	1000	530	1000	600	600
Temperature coefficient of P_{max} [%/°C]	−0.45	−0.3	−0.23	−0.469	−0.25	−0.45
Temperature coefficient of V_{oc} [V/°C]	−0.35	−0.127	−0.305	−0.334	−0.222	−0.33
Temperature coefficient of I_{sc} [mA/°C]	0.05	1.67	0.0752	0.052	0.04%	0.06
Panel efficiency (%)	13.7	17.1	6.3	15.22	10.07	15.3
Module area [mm×mm]	1581×809	1580×798	990×960	1652×994	1200×600	1650×990

in the recorded datasets. Finally, the few missing or omitted data points (usually near sunset and sunrise) have been re-filled using the two-directional exponential smoothing (Hassan et al. 2021c). The recorded datasets consist of solar PV power production and other meteo-solar parameters, namely global horizontal solar irradiance, ambient air temperature, relative humidity, and wind speed. All parameters are averaged from the original 5-min resolution to obtain the hourly average values. The period of measurements varies from two to five years, depending on the station (Table 1). Features with higher bounds will dominate and affect the calculation process. Therefore, it is essential to scale and normalize data to guarantee that all features lay in the same bounds and are treated similarly by the physical and machine learning models. One of the simple ways to scale data is the min–max scaler, in which data features are scaled and bounded between the range of 0 and 1 using the min–max scaler. Figure 1 depicts the frequency distributions of normalized power production from each plant throughout the periods of data collection shown in Table 1. The produced power from these plants is normalized based on the peak capacity of each plant, hence the Wh/Whp unit (the "p" subscript is for "peak").

The original data sets can be divided into three subsets based on the corresponding sky conditions, represented by the global clearness index, as shown in Table 2. The global clearness index is the ratio between the ground-level glob irradiance and the corresponding extraterrestrial horizontal irradiance (Hassan et al. 2021d) . It can be seen that 64.78–65.69% of the input datasets correspond to clear sky conditions, 22.53–23.47% are registered for partly cloudy sky conditions, while 11.75–12.31% correspond to cloudy sky conditions, as presented in Table 2. This indicates that the study areas are predominately clear sky weather conditions (desert environment). This facilitates the prediction of PV power generation using relatively simple models. Cloud formations are frequent and unpredictable, unlike cloudy sky conditions, leading to the relatively poor prediction accuracy of simple models. On the other hand, Table A1 in the supplementary material provides quantitative statistical summaries of the different measured meteo-solar parameters during measurement periods.

In general, the performance of the PV power system is influenced by electrical and solid-state material characteristics. However, meteo-solar parameters, such as global horizontal irradiance (*GHI*), average air temperature (*TEM*), wind speed (*WSP*), and relative humidity (*REH*), have been frequently reported as the most influential variable in determining the instantaneous PV power output, with different degrees of influence. Table A2 in the supplementary material shows that the Pearson correlation coefficients (*R*) between PV power output and GHI are the strongest, ranging between 0.870 and 0.970 for the six studied plants. The other meteorological parameters are less correlated to the produced power, but the correlation coefficients are still considerable.

2.2 Ensemble learning models

A weighted sum ensemble is an ensemble learning approach that combines the predictions from multiple models, where the contribution of each ensemble

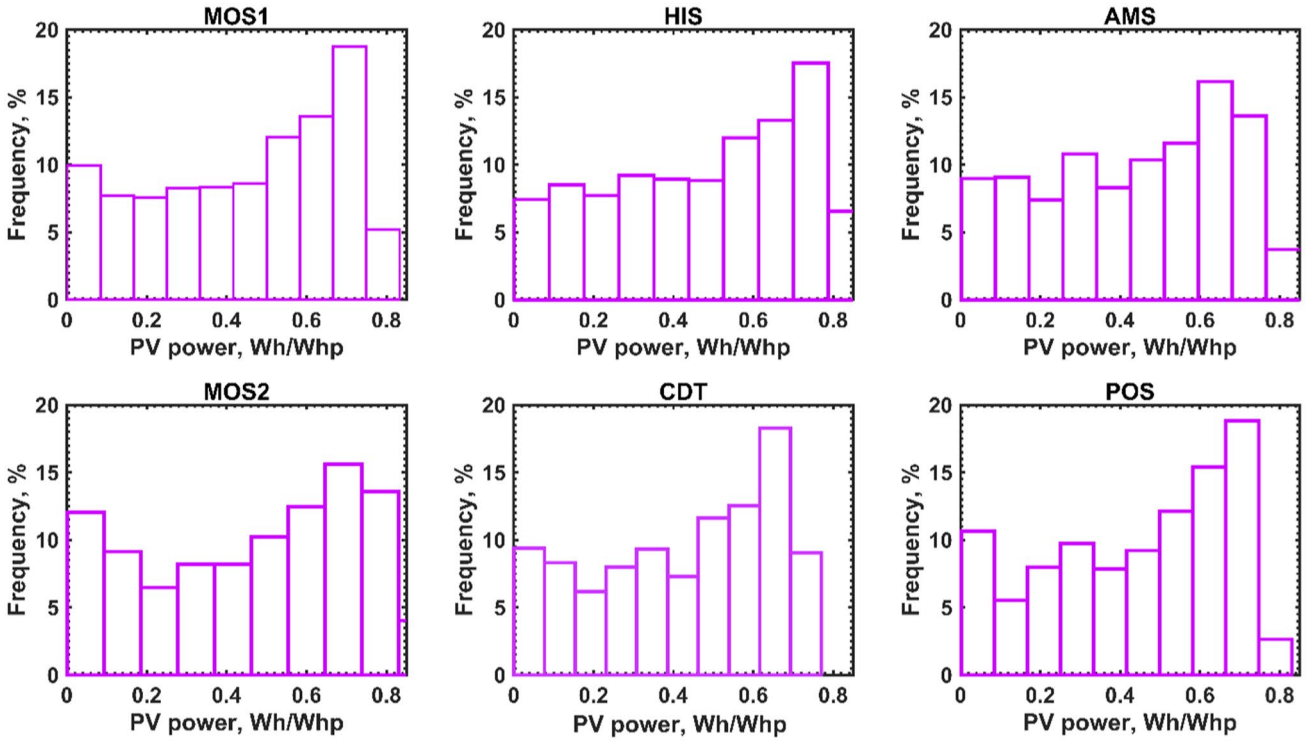

Fig. 1 Frequency histograms of normalized PV power of each technology

Table 2 Frequency of data corresponding to the four categories of sky conditions for all selected technologies

Sky conditions	Clearness index range	Frequency [%]					
		MOS1	HIS	AMS	MOS2	CDT	POS
Overcast sky	0–0.35	10.62	10.02	10.65	12.77	10.86	10.42
Partially cloudy sky	0.35–0.55	10.10	10.87	9.81	14.25	9.42	9.82
Partially clear sky	0.55–0.65	10.35	11.02	10.91	13.33	10.21	10.81
Clear sky	> 0.65	68.93	68.09	68.62	59.64	69.51	68.94

member to the final prediction is to be weighted proportionally to its capability or skill. The ensemble learning approach is adopted to achieve better performance than the performance obtained from single methods. Five machine learning techniques were selected as follows: Decision Tree Regressor (DTR), Random Forest Regressor (RFR), MLP Regressor, Support Vector Regression, and K-Neighbors Regressor. Due to their better classification performances, these algorithms were employed to implement the ensemble using K-Neighbors Regressor (KNR-ensemble).

The Random Forest (RF) algorithm is a successful and widely employed ensemble technique (Mao and Wang 2012). RF algorithm can be used for regression and classification. The considerable interest in RF is gained due to its immunity to noise and accuracy compared to other single classifiers. No reasonable change is expected in the RF tree due to small changes in training data because of the hierarchical architecture of tree classifiers. The main drawback of the RF algorithm is the high variance. However, RF usually performs better than the decision tree (DT) algorithm.

Multilayer perceptron (MLP) with two or more hidden layers is considered an artificial neural network (ANN). MLP is one of the excellent algorithms for classification and regression (Keshtegar et al. 2022; Bouchouicha et al. 2020b). This is due to MLP's ability to learn with a non-linear decision boundary. It is very flexible to give a reasonable solution to real-world tasks. MLP has many artificial neurons and connections named processing elements (PEs). These PEs emulates the human nervous system operations based on a particular training algorithm.

Support Vector Regression (SVR) is also a robust algorithm (VanDeventer et al. 2019; Liu et al. 2021). SVR has the flexibility to define how much error is acceptable in our model and will find an appropriate line (or hyperplane in

higher dimensions) to fit the data. This can be achieved by tuning the tolerance of falling outside the acceptable error rate and an acceptable error margin.

The K-Neighbors Regressor (KNR) algorithm could use the similarity measure technique to classify cases or samples after storing the variable samples (Qun'ou et al. 2021). The KNR algorithm classifies data using the nearest samples or points. An adjustable parameter k, nearest neighbors, can be updated to force the model to be flexible. The default value of the k parameter is one.

Polar Rose Guided Whale Optimization algorithm based on the Dynamic Adaptive technique (AD-PSO-Guided WOA) has been used for feature selection in the present investigation. The main target of the individual Adaptive Dynamic Polar Rose Algorithm combined with the Guided Whale Optimization Algorithm in the exploitation group is to move toward the optimal or best solution. The main target of the individuals in the exploration group is to search the area around the leaders. The change (update) between the agents of the population groups is working dynamically (Ghoneim et al. 2021). Algorithm 1 shows the complete steps of computations in the AD-PSO-Guided WOA algorithm. The algorithm starts with initializing the population parameters, fitness or objective function, number of required iterations, and the parameters needed to start the AD-PSO-Guided WOA algorithm. The fitness function is then evaluated for all populations for the best solution. The algorithm converts all the available solutions to binary ones by the following equation.

$$X_d^{(t+1)} = \begin{cases} 0 & \text{if Sigmoid } (X^*) < 0.5 \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

$$\text{Sigmoid } (X^*) = \frac{1}{1 + e^{-10(X^* - 0.5)}} \tag{2}$$

where X^* is the best individual. The algorithm searches for and updates the best solution at the end of iterations. If the algorithm is stacked, it starts to select three random search solutions X_{rand1}, X_{rand2}, and X_{rand3}, to be used in updating the current search agents (solutions) position based on the following equation.

$$X(t+1) = w_1 X_{rand1} + z w_2 \left(X_{rand2} - X_{rand3}\right) + (1 - z) w_3 \left(X - X_{rand1}\right) \tag{3}$$

where $z = 1 - \left(\frac{t}{Max_{iter}}\right)^2$ at iteration t and maximum iterations Max_{iter}. The fitness function F_n is then calculated for each X_i from this form called Guided WOA. Otherwise, the fitness function F_n will be calculated using the PSO algorithm for each X_i. The algorithm ends by the end of iterations, selecting the best solution.

2.3 Regression models

Recent literature has established that an empirical model for estimating PV power output can be developed by employing linear and multi-linear regression models (Trigo-González et al. 2019; Azevedo Dias et al. 2017). In addition, the diurnal fluctuation of PV power production equally follows linear and non-linear trends (Dolara et al. 2015). This can be attributed to the linear and non-linear fluctuations of the influential meteo-solar parameters, e.g., the global horizontal solar irradiance. The empirical models developed in this study are based on the correlation between meteo-solar parameters and PV power (Mostefaoui et al. 2019). This compares the two approaches (regression and ensemble learning) for predicting the PV power output in the desert environment. The nine regression models are expressed as

$$P_{pv} = b_1\, GHI + a \tag{4}$$

$$P_{pv} = b_1\, TEM + a \tag{5}$$

$$P_{pv} = b_1\, WSP + a \tag{6}$$

$$P_{pv} = b_1\, GHI + b_2\, TEM + a \tag{7}$$

$$P_{pv} = b_1\, GHI + b_2\, WSP + a \tag{8}$$

$$P_{pv} = b_1\, GHI + b_2\, REH + a \tag{9}$$

$$P_{pv} = b_1\, TEM + b_2\, REH + b_3\, WSP + a \tag{10}$$

$$P_{pv} = b_1\, GHI + b_2\, TEM + b_3\, WSP + a \tag{11}$$

$$P_{pv} = b_1\, GHI + b_2\, TEM + b_3\, REH + a \tag{12}$$

where P_{pv} is the produced power, and a, b_1, b_2, and b_3 are the fitted regression coefficients.

2.4 Model evaluation

About 80% of the dataset collected at each location was used for fitting/training the regression and ensemble models. The remaining 20% was employed to test the reliability of the developed models. The sampling and assignment of observations to the two subsets were performed randomly based on a uniform distribution instead of chronological partitioning. This is to reduce the dependency of the developed models on the specific data used in the fitting process and to ensure an

Algorithm 1 The binary AD-PSO-Guided WOA algorithm.

Initialize population $X_i(i = 1,2,\ldots,n)$ with size n, the fitness function F_n, maximum iterations t_{max}
Initialize parameters $a, A, C, l, r_1, r_2, r_3, w_1, w_2, w_3, t = 1$
Evaluate fitness function F_n for each X_i
Find the best individual X^*
Convert solution to binary [0 or 1]
while $t < t_{max}$ **do**
  if $(t\%2 == 0)$ **then**
    for $(i = 1; i < n + 1)$ **do**
      if $(r_3 < 0.5)$ **then**
        if $(|A| < 1)$ **then**
          Update current search agent position as $X(t+1) = X^*(t) - A.D$
        else
          Select three random search agents X_{rand1}, X_{rand2}, and X_{rand3}
          Update (z) by the exponential form of $z = 1 - \left(\frac{t}{Max_{iter}}\right)^2$
          Update current search agent position
          $X(t+1) = w_1 * X_{rand1} + z * w_2 * (X_{rand2} - X_{rand3}) + (1 - z) * w_3 * (X - X_{rand1})$
        end if
      else
        Update current search agent position as $X(t+1) = D'.e^{bl}.cos(2\pi l) + X^*(t)$
      end if
    end for
    Calculate fitness function F_n for each X_i from Guided WOA
  else
    Calculate fitness function F_n for each X_i from PSO
  end if
    Find the best individual X^*
  Convert updated solution to binary by
  $X_d^{(t+1)} = \begin{cases} 0 & \text{if} Sigmoid(X^*) < 0.5 \\ 1 & otherwise \end{cases}, Sigmoid(X^*) = \frac{1}{1+e^{-10(X^*-0.5)}}$
  Update a, A, C, l, r_3
  Set $t = t + 1$
end while
Return X^*

equivalent performance of the models when handling new datasets (Bouchouicha et al. 2019).

The metrics used for performance evaluation include the mean bias error (MBE), the root mean square error (*RMSE*), the relative root mean square error (*RRMSE*, %), and Pearson's correlation coefficient (*R*), all calculated to present non-dimensional error estimates. The main calculation formulas for the six metrics are as follows (Muzathik et al. 2011; Jamil et al. 2018; Almorox et al. 2020, 2021; Bailek et al. 2017)

$$\text{MBE} = \frac{1}{M}\sum_{m=1}^{M}\left(\widehat{Y_m} - Y_m\right) \tag{13}$$

$$RMSE = \sqrt{\frac{1}{M}\sum_{m=1}^{M}\left[\widehat{Y_m} - Y_m\right]^2} \tag{14}$$

$$RRMSE = \frac{RMSE}{\overline{Y_m}} \times 100 \tag{15}$$

$$R = \frac{\sum_{m=1}^{M}\left(\widehat{Y_m} - \overline{\widehat{Y_m}}\right)\left(Y_m - \overline{Y_m}\right)}{\sqrt{\left[\sum_{m=1}^{M}\left(\widehat{Y_m} - \overline{\widehat{Y_m}}\right)^2\right]\left[\sum_{m=1}^{M}\left(Y_m - \overline{Y_m}\right)^2\right]}} \tag{16}$$

where M is the number of observations in the subset, $\widehat{Y_m}$ and Y_m are the m^{th} estimated and observed PV power values, respectively, and $\overline{\widehat{Y_m}}$ and $\overline{Y_m}$ are the arithmetic means of the estimated and observed values, respectively.

3 Results and discussion

Firstly, linear and multiple linear regression techniques were used to study the relationship between meteo-solar parameters and power production of various PV technologies in desert climate conditions. As a result, for each PV technology in this study, performance analyses employed six multi-linear (MLR) and three linear regression (LIR) models for MOS1, HIS, AMS, MOS2, CDT, and POS technologies (Eqs. (4) to (12)). The fitted coefficients of all models are shown in Table 3.

In the first step and to obtain an objective overview of models' performances, only three technologies are selected for the preliminary analysis, namely MOS1, HIS, and AMS technologies. Figure 2 compares the different categories of models based on the test data. To enhance the readability of this figure, only R and $RMSE$ values are displayed. The values of R and $RMSE$ of PV power output reported best fits using multi-linear regression models, compared to the corresponding performances of linear regression models in the test stage.

Various comparisons are also made to assess the performance of the selected and tested correlations under arid desert climates. The estimation results of the power output of MOS and AMS technologies using the different models depicted that for the MOS technology, the developed models registered higher values of $RRMSE$ compared to AMS technology. In addition, the results show that the performance results of the different linear regression models are close to each other, except for the first LIR model (Eq. (4)), where it is noticed that the magnitudes of R and $RRMSE$ values are more significant, especially for AMS technology. As such, the best-performing model from the linear regression category (Eq. (4)) is selected based on the error metrics of the testing subsets.

It follows from Table 4 that the different multiple-linear regression models produce a wide range of R values (ranging from 0.4030 for model #7 to 0.9762 for model #9), as well as a wider range of estimated $RRMSE$s (ranging from 11.3389% for model #9 to 48.46% for model #7). Model #9, categorized as a multiple-linear model with the inputs of global horizontal irradiance, ambient temperature, and relative humidity, shows superior performance in terms of testing error measures. It is also depicted that model #8 (another multiple-linear model) performs close to model #8. Regarding Rs, model #9 emerges as the best-fitted model for HIS and AMS technologies. In contrast, model #8 yielded the best performance for MOS1 technology.

The relative ability of the models to predict the PV Power output is, a priori, a function of sky conditions. So far, the models have been analyzed under all-sky conditions. However, this section analyzes them under specific sky conditions (clear, partially cloudy, and cloudy skies), as mentioned in Table 2. The so-called Taylor diagrams are used to obtain an analytical description of the two best-performing regression models under different sky conditions for all technologies. The Taylor diagram, shown in Fig. 3, indicates that considering the sky conditions, each model behaves differently in predicting the PV power output of all technology. It is observed that model #9 shows better estimates under overcast and partially cloudy sky conditions and produces equal error estimates to those of model #8 under relatively clearer skies. This implies that the global irradiance, air temperature, and relative humidity are more related to the PV power output of all technology, followed by global irradiance with temperature and wind speed.

Next, the proposed ensemble learning techniques with the top-performing input (global horizontal irradiance, ambient temperature, and relative humidity) were used to predict the power production of various PV technologies in desert climate conditions. Figure 4 compares the relative performance of the individual and ensemble learning models, compared to the best-performing regression model, in terms of percentage drops in $RMSE$s. It can be seen from Fig. 4 that the $RMSE$s drops do not exceed the value of 50% for the results of DTR, MLP, SVR, and RFR models for the various PV technologies; except for MOS technology, the percentage exceeds this value. In contrast, the ensemble learning techniques reported a percentage exceeding 75%. Generally, all the individual selected models have different $RMSE$ drops varied from technology to technology. Also, as expected, the proposed ensemble model registered higher $RMSE$ drop values than individual models. Therefore, it can be inferred that the proposed ensemble model outperformed individual models.

Table 3 Regression models and their fitted coefficients in this study

Technology	MOS1				HIS				AMS			
Correlation	a	b_1	b_2	b_3	a	b_1	b_2	b_3	a	b_1	b_2	b_3
$P_{pv} = b_1 GHI + a$	0.02451	0.00079	0	0	0.04869	0.00079	0	0	0.01809	0.00079	0	0
$P_{pv} = b_1 TEM + a$	0.34210	0.00406	0	0	0.37139	0.00407	0	0	0.21846	0.00790	0	0
$P_{pv} = b_1 WSP + a$	0.23293	0.06370	0	0	0.27015	0.06413	0	0	0.31230	0.03187	0	0
$P_{pv} = b_1 GHI + b_2 TEM + a$	0.17557	0.00086	−0.00709	0	0.19951	0.00086	−0.00700	0	0.09837	0.00082	−0.00359	0
$P_{pv} = b_1 GHI + b_2 WSP + a$	0.03474	0.00079	−0.00368	0	0.04903	0.00079	−0.00012	0	0.03503	0.00079	−0.00532	0
$P_{pv} = b_1 GHI + b_2 REH + a$	−0.04773	0.00084	0.00166	0	0.00425	0.00082	0.00105	0	−0.02809	0.00082	0.00111	0
$P_{pv} = b_1 TEM + b_2 REH + b_3 WSP + a$	0.51254	−0.00417	−0.00469	0.05079	0.53293	−0.00356	−0.00482	0.05295	0.48890	−0.00040	−0.00436	0.01861
$P_{pv} = b_1 GHI + b_2 TEM + b_3 WSP + a$	0.17108	0.00085	−0.00712	0.00182	0.19056	0.00086	−0.00703	0.00357	0.10017	0.00082	−0.00354	−0.00094
$P_{pv} = b_1 GHI + b_2 TEM + b_3 REH + a$	0.17189	0.00086	−0.00703	0.00005	0.24794	0.00085	−0.00782	−0.00073	0.08138	0.00083	−0.00327	0.00023
Technology	MOS2				CDT				POS			
Correlation	a	b_1	b_2	b_3	a	b_1	b_2	b_3	a	b_1	b_2	b_3
$P_{pv} = b_1 GHI + a$	0.06157	0.00075	0	0	0.01719	0.00075	0	0	0.02645	0.00076	0	0
$P_{pv} = b_1 TEM + a$	0.46875	0.00019	0	0	0.26666	0.00574	0	0	0.30187	0.00491	0	0
$P_{pv} = b_1 WSP + a$	0.41323	0.01753	0	0	0.34214	0.01912	0	0	0.33284	0.02672	0	0
$P_{pv} = b_1 GHI + b_2 TEM + a$	0.26945	0.00087	−0.00859	0	0.12303	0.00081	−0.00487	0	0.17319	0.00083	−0.00662	0
$P_{pv} = b_1 GHI + b_2 WSP + a$	0.06755	0.00075	−0.00195	0	0.05621	0.00077	−0.01136	0	0.05586	0.00077	−0.00930	0
$P_{pv} = b_1 GHI + b_2 REH + a$	−0.09342	0.00082	0.00738	0	−0.02163	0.00078	0.00099	0	−0.03866	0.00081	0.00155	0
$P_{pv} = b_1 TEM + b_2 REH + b_3 WSP + a$	0.55429	−0.0026	−0.00409	0.01842	0.58845	−0.00303	−0.00514	0.01045	0.61539	−0.00429	−0.00496	0.01797
$P_{pv} = b_1 GHI + b_2 TEM + b_3 WSP + a$	0.28148	0.00087	−0.00861	−0.00376	0.13213	0.00081	−0.00446	−0.00521	0.17569	0.00083	−0.00654	−0.00130
$P_{pv} = b_1 GHI + b_2 TEM + b_3 REH + a$	0.29633	0.00086	−0.00907	−0.00074	0.17909	0.00079	−0.00597	−0.00081	0.20103	0.00082	−0.00715	−0.00038

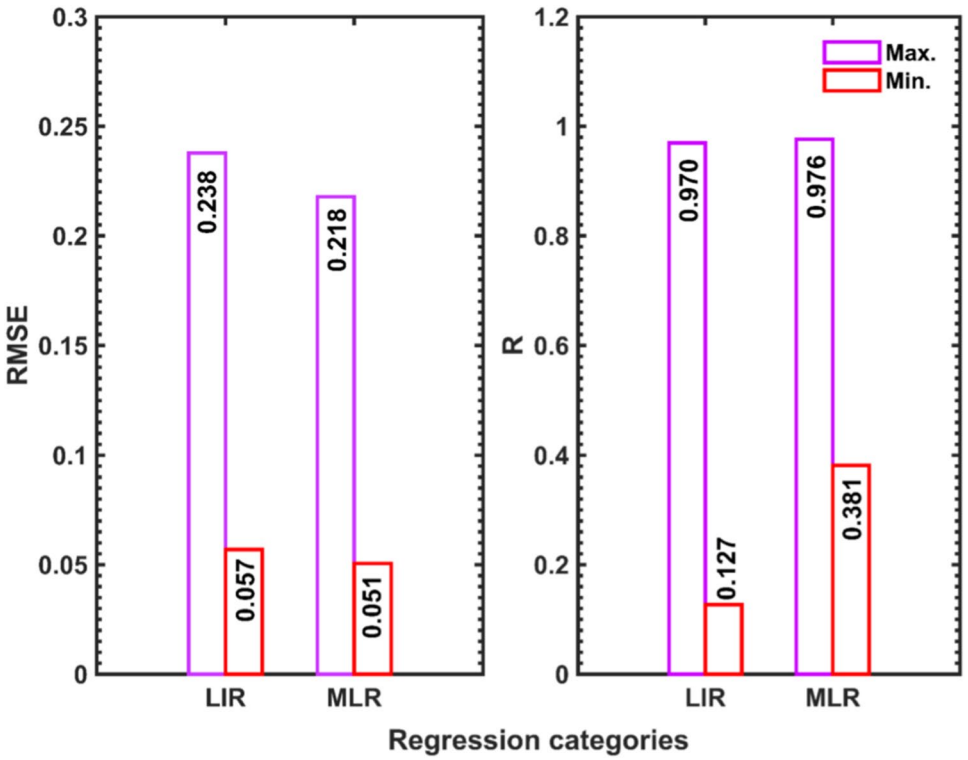

Fig. 2 Comparison between the best- and worst-performing multi-linear and-linear regression models

Table 4 Testing performance indices of PV power models (based on normalized P_{pv} values) for MOS1, HIS, and AMS technologies

Technology		MOS1		HIS		AMS	
Regression types	Model #	*RRMSE* [%]	*R* [−]	*RRMSE* [%]	*R* [−]	*RRMSE* [%]	*R* [−]
Linear (LIR)	1	16.8182	0.9462	16.9015	0.9444	12.7577	0.9696
	2	51.4559	0.1271	49.7051	0.1543	51.1892	0.2426
	3	50.5470	0.2701	47.9829	0.2784	51.9474	0.2150
Multiple-linear (MLR)	4	12.6411	0.9707	13.1090	0.9660	11.3755	0.9760
	5	16.8443	0.9461	16.9019	0.9444	12.7639	0.9696
	6	15.7902	0.9531	16.4949	0.9468	12.0656	0.9729
	7	47.7975	0.4030	45.5822	0.4246	48.4609	0.3815
	8	12.6247	0.9708	13.0868	0.9661	11.3921	0.9760
	9	12.6449	0.9707	12.9035	0.9671	11.3389	0.9762

The error estimates values for the proposed ensemble model using each PV module technology data are calculated and summarized in Table 5. Table 5 shows that the proposed ensemble model generated a higher numerical range of error values corresponding to maximum *RMSE* and *RRMSE* values of 0.0323 and 6.3820% for MOS2 technology. This is followed by CDT technology, with an *RMSE* of 0.0315 and an *RRMSE* of 5.9625%. It is worth mentioning that the MOS2 plant is the only one located at Adrar, where seasonal passing clouds take place. This demonstrates the importance of incorporating cloud parameters into PV power prediction models, which will be further examined in future works.

On the other hand, it was observed that the model provided the best performance for MOS1 technology, corresponding to maximum MBE and *RRMSE* values of − 0.0004 and 1.5109%. Moreover, the correlation coefficient of the results obtained using the proposed model for almost all the PV technologies used in the present work exceeded 0.99, with a maximum of 0.9994 for MOS1 technology.

Table A3 in the supplementary material compares the results of the proposed ensemble-learning technique with alternative models developed in previous studies in terms of *RRMSE*. It is clear from the table that the proposed approach has substantially lower error values (down to *RRMSE* of 1.5109%) compared to conventional machine-learning

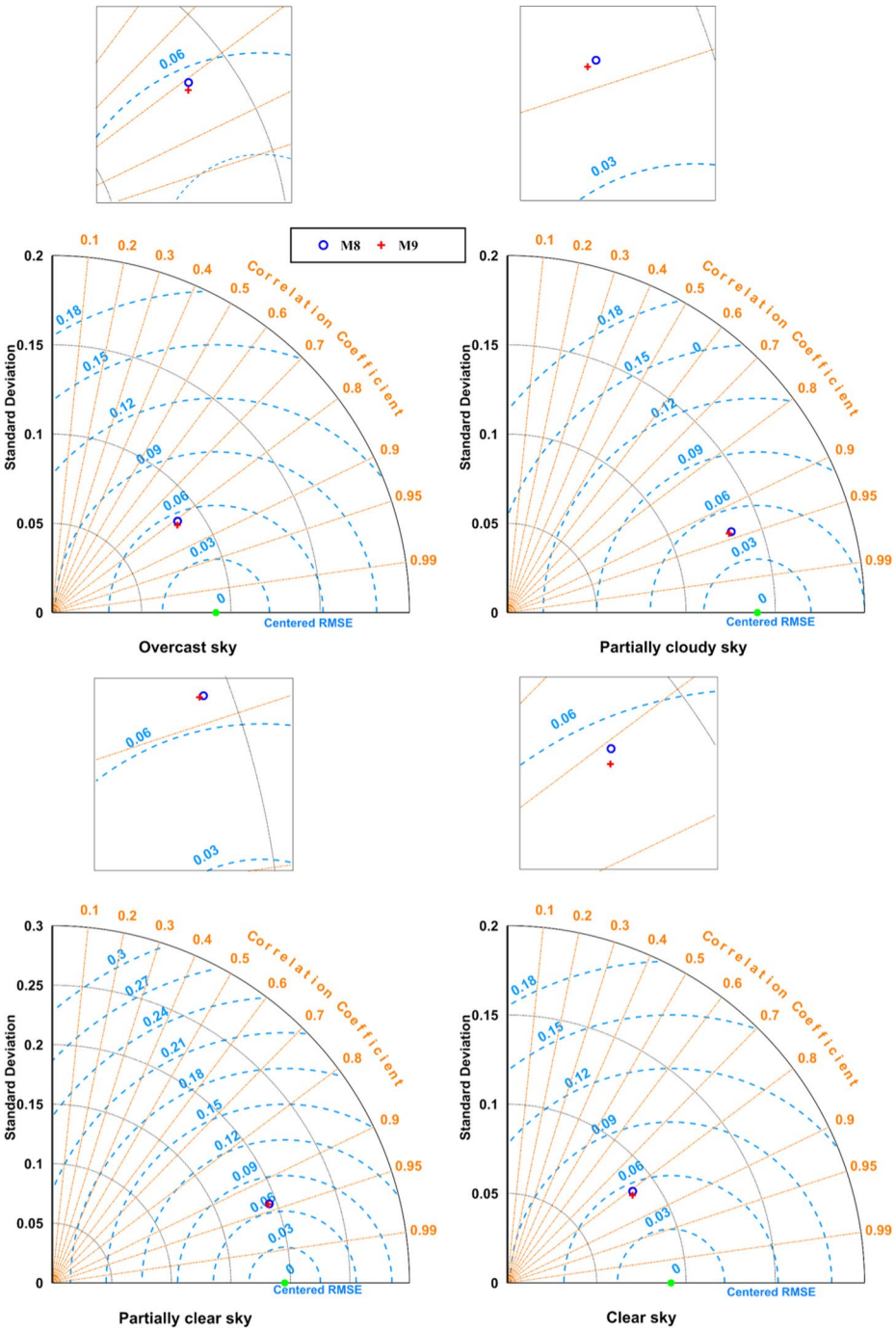

Fig. 3 Evaluation of the two best-performing regression models under different sky conditions for all technologies

algorithms, such as ANN, SVR, and MLR, with *RRMSE*s ranging between 3.6 and 9.546%.

In general, multi-linear regression models offer competitive performance under different sky conditions. Their performance decreases as the sky cloudiness increases. Besides, when integrating global solar radiation, ambient air temperature, and relative humidity measurements as typical inputs, the regression models' performance improves generally and achieves efficient performance with the proposed ensemble learning, with an estimated accuracy of over 99%.

Finally, the solar and climate data variables contain enough information to predict the power generation of different PV technologies accurately. Also, it should be stated that PV modules of the different technologies operate in an outdoor environment with numerous fluctuations in other operating conditions such as the cloud and dust pollutants parameters, leading to deterioration in the output power of

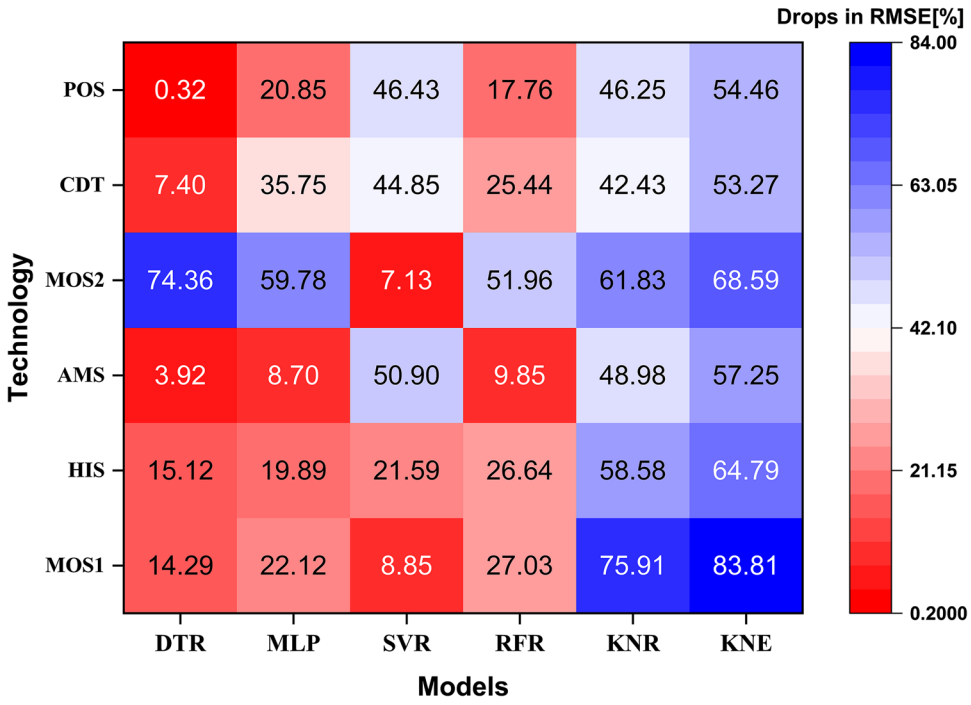

Fig. 4 Percentage drops in the *RMSE* when replacing model #9 (best-performing regression model) with machine learning models

Table 5 PV power prediction errors (normalized values) when applying the proposed ensemble model for each technology

Technology	MBE [−]	*RMSE* [−]	*RRMSE* [%]	*R* [−]
MOS1	−0.0004	0.0093	1.5109	0.9994
HIS	−0.005	0.0216	3.7252	0.9972
AMS	−0.0060	0.0216	3.7587	0.9973
MOS2	−0.0115	0.0323	6.3820	0.9948
CDT	−0.0135	0.0315	5.9625	0.9953
POS	−0.0093	0.0265	4.6954	0.9962

the PV modules. However, the effect of these factors varies slightly among the different PV technologies used for power generation. The dust pollution effect depends on the local area where the PV system is mounted and the site's local environmental conditions (Abd El-Wahab et al. 2018; Darwish et al. 2018). This will be covered in future studies.

4 Conclusions

This study proposes new regression and ensemble-learning models by studying six different desert sites in the Algerian Big South desert and the Australian Northern Territory to achieve accurate estimations of the performance of PV power plants running in desert areas. Six PV module technologies were selected for the analysis. A feature selection method was developed to enhance the ensemble-learning models using the AD-PSO-Guided WOA algorithm. The proposed approach considers the photovoltaic system's technological features and the actual characteristics of the operation settings and climatic conditions for experiment sites (global irradiance, ambient temperature, and relative humidity). The following points summarize the findings of the study:

- Multi-linear regression models offer competitive performances under the different sky conditions, with their performances declining as the sky becomes heavier.
- By incorporating global solar irradiance, ambient air temperature, and relative humidity measurements as model inputs, the performances of the regression models generally improve.
- With these inputs, the ninth developed regression model showed *RRMSE* values of up to 11.33%, based on the normalized values of the PV power output.
- The proposed K-Neighbors Regressor ensemble model showed a reduction of 83.8% in the *RMSE* of the top-performing regression model, with an estimated accuracy of over 99%.
- The drops in *RMSE* do not exceed 50% for DTR, MLP, SVR, and RFR-based models for the various PV technologies, except for MOS technology. In contrast, the ensemble learning techniques reported a percentage exceeding 75%.
- Generally, all selected individual models have different percentage reductions in *RMSE* that vary from one technology to another. However, the proposed ensemble model registered higher percentage reductions in *RMSE* values.

- The proposed ensemble model generated a higher numerical range of error values corresponding to maximum *RRMSE* values of 6.382% for the MOS2 technology.
- The ensemble model also provided the best performance for MOS1 technology, corresponding to a maximum *RRMSE* of 1.511%.
- The correlation coefficients of the proposed model for almost all PV technologies adopted in the present work exceeded 0.99, with a maximum of 0.9994 for the MOS1 technology.
- It is concluded that the proposed model best fits all examined PV technologies, which are the most suitable for desert locations. It also outperforms conventional models in the literature by reducing the *RRMSE* by up to 6.32 folds.

However, it should be noted that this study primarily focused on meteorological parameters' overall influences without considering the ground and atmospheric parameters, such as dust accumulation and cloud formations, which are not regularly measured in meteorological networks. In future works, these parameters will be considered for more precise predictors. It is also recommended to re-evaluate the proposed models under other climate zones, including tropical and temperature climates.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00704-022-04166-6.

Acknowledgements The authors would like to thank URER-MS for supporting this research. The authors also acknowledge the contribution of the DKASC in providing most of the used data.

Author contribution MH: investigation, conceptualization, data curation, visualization, formal analysis, writing—original draft, writing—review and editing. NB: conceptualization, data curation, visualization, formal analysis, writing—original draft, writing—review and editing, validation. KB: conceptualization, data curation, writing—original draft, writing—review and editing. AI: methodology, software, formal analysis, writing—original draft, writing—review and editing, validation. BJ: conceptualization, writing—original draft, writing—review and editing. AK: resources, writing—original draft, writing—review and editing. SN: writing—original draft, writing—review and editing. EE: resources, software, writing—original draft, validation.

Funding Open access funding provided by The Science, Technology & Innovation Funding Authority (STDF) in cooperation with The Egyptian Knowledge Bank (EKB). This work was supported by the Portuguese Foundation (FCT) for Science and Technology through the project PTDC-CTA-OHR-30561/2017 (WinTherface).

Data availability The authors confirm that the data supporting the findings of this study are available within the article.

Declarations

Consent to participate As the research team in this current contribution, we have voluntarily agreed to participate in this research study.

Consent for publication We want to consent to publish identifiable details, including text, material and methods, figures, and tables published in the Journal.

Conflict of interest The authors declare no competing interests.

References

Aba A, Al-Dousari AM, Ismaeel A (2018) Atmospheric deposition fluxes of 137Cs associated with dust fallout in the northeastern Arabian Gulf. J Environ Radioact 192:565–572

Abd El-Wahab RH, Al-Rashed AR, Al-Dousari A (2018) Influences of physiographic factors, vegetation patterns and human impacts on aeolian landforms in arid environment. Arid Ecosyst 8:97–110. https://doi.org/10.1134/S2079096118020026

Alam MS, Alouani AT, Azeem MF (2015) Efficient prediction of maximum PV module output power through dynamic modeling. Sustain Energy Technol Assess 11:27–35. https://doi.org/10.1016/j.seta.2015.06.001

Almorox J, Arnaldo JA, Bailek N, Martí P (2020) Adjustment of the Angstrom-Prescott equation from Campbell-Stokes and Kipp-Zonen sunshine measures at different timescales in Spain. Renew Energy 154:337–350. https://doi.org/10.1016/j.renene.2020.03.023

Almorox J, Voyant C, Bailek N, Kuriqi A, Arnaldo JA (2021) Total solar irradiance's effect on the performance of empirical models for estimating global solar radiation: an empirical-based review. Energy 236:121486. https://doi.org/10.1016/j.energy.2021.121486

Aoun N, Bouchouicha K, Bailek N (2019) Seasonal performance comparison of four electrical models of monocrystalline PV module operating in a harsh environment. IEEE J Photovoltaics 9(4):1057–1063. https://doi.org/10.1109/jphotov.2019.2917272

Bailek N, Bouchouicha K, El-Shimy M, Slimani A, Chang K-C, Djaafari A (2017b) Improved mathematical modeling of the hourly solar diffuse fraction (HSDF)-Adrar, Algeria case study. Sol Energy 4(2):8–12.

Bailek N, Bouchouicha K, Aoun N, EL-Shimy M, Jamil B, Mostafaeipour A (2018) Optimized fixed tilt for incident solar energy maximization on flat surfaces located in the Algerian Big South. Sustain Energy Technol Assess 28:96–102. https://doi.org/10.1016/j.seta.2018.06.002

Bailek N, Bouchouicha K, Abdel-Hadi YA, El-Shimy M, Slimani A, Jamil B, Djaafari A (2020a) Developing a new model for predicting global solar radiation on a horizontal surface located in Southwest Region of Algeria. NRIAG J Astron Geophys 9:341–349. https://doi.org/10.1080/20909977.2020.1746892

Bailek N, Bouchouicha K, Hassan MA, Slimani A, Jamil B (2020b) Implicit regression-based correlations to predict the back

temperature of PV modules in the arid region of south Algeria. Renew Energy. https://doi.org/10.1016/j.renene.2020.04.073
Bailek N, Bouchouicha K, El-Shimy M, Slimani A (2017a) Updated status of renewable and sustainable energy projects in Algeria. Econ Variable Renew Sources Electric Power Prod. 519–528.
Bouchouicha K, Razagui A, Bachari NI, Aoun N (2015) Mapping and geospatial analysis of solar resource in Algeria. Int J Energy Environ Econ 23:735–751
Bouchouicha K, Aoun N, Bailek N, Razagui A (2017) Solar resource potentials in Algeria. Economics of Variable Renewable Sources for Electric Power Production, El-Shimy, M., Ed., Germany: Lap Lambert Academic, Omniscriptum
Bouchouicha K, Bailek N, Mahmoud ME-S, Alonso JA, Slimani A, Djaafari A (2018) Estimation of monthly average daily global solar radiation using meteorological-based models in Adrar, Algeria. Appl Solar Energy. 54:448–455. https://doi.org/10.3103/S0003701X1806004X
Bouchouicha K, Hassan MA, Bailek N, Aoun N (2019) Estimating the global solar irradiation and optimizing the error estimates under Algerian desert climate. Renewable Energy 139:844–858. https://doi.org/10.1016/j.renene.2019.02.071
Bouchouicha K, Bailek N, Razagui A, Mohamed E-S, Bellaoui M, Bachari NEI (2020a) Comparison of artificial intelligence and empirical models for energy production estimation of 20 MWp solar photovoltaic plant at the Saharan Medium of Algeria. Int J Energy Sector Manag 15:119–138
Bouchouicha K, Bailek N, Bellaoui M, Oulimar B (2020b) Estimation of solar power output using ANN model: a case study of a 20-MW Solar PV Plan at Adrar, Algeria. In: Hatti M (ed) Smart Energy Empowerment in Smart and Resilient Cities. Springer International Publishing, Cham, pp 195–203
Darula S, Kittler R, Kocifaj M (2010) Luminous effectiveness of tubular light-guides in tropics. Appl Energy 87:3460–3466
Darwish ZA, Kazem HA, Sopian K, Alghoul MA, Alawadhi H (2018) Experimental investigation of dust pollutants and the impact of environmental parameters on PV performance: an experimental study. Environ Dev Sustain 20:155–174. https://doi.org/10.1007/s10668-016-9875-7
de Azevedo Dias CL, Castelo Branco DA, Arouca MC, Loureiro Legey LF (2017) Performance estimation of photovoltaic technologies in Brazil. Renew Energy 114:367–375. https://doi.org/10.1016/j.renene.2017.07.033
Desert Knowledge Australia Solar Centre - Download Data. (n.d.). https://dkasolarcentre.com.au/download/notes-on-the-data. (Accessed 18 Apr 2020)
Dolara A, Leva S, Manzolini G (2015) Comparison of different physical models for PV power output prediction. Sol Energy 119:83–99
Ehsan RM, Simon SP, Venkateswaran PR (2017) Day-ahead forecasting of solar photovoltaic output power using multilayer perceptron. Neural Comput Appl 28:3981–3992
El-Kenawy E-SM, Ibrahim A, Bailek N, Bouchouicha K, Hassan MA, Jamil B, Al-Ansari N (2022) Hybrid ensemble-learning approach for renewable energy resources evaluation in Algeria. Comput Mater Continua 71:5837–5854
Mohamed ES (ed) (2017) Economics of variable renewable sources for electric power production. LAP LAMBERT Academic Publishing
Ghoneim SSM, Farrag TA, Rashed AA, El-Kenawy E-SM, Ibrahim A (2021) Adaptive dynamic meta-heuristics for feature selection and classification in diagnostic accuracy of transformer faults. IEEE Access 9:78324–78340. https://doi.org/10.1109/ACCESS.2021.3083593
Hassan MA, Akoush BM, Abubakr M, Campana PE, Khalil A (2021a) High-resolution estimates of diffuse fraction based on dynamic definitions of sky conditions. Renewable Energy 169:641–659. https://doi.org/10.1016/j.renene.2021.01.066
Hassan MA, Bailek N, Bouchouicha K, Nwokolo SC (2021b) Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks. Renewable Energy 171:191–209. https://doi.org/10.1016/j.renene.2021.02.103
Hassan MA, Khalil A, Abubakr M (2021c) Selection methodology of representative meteorological days for assessment of renewable energy systems. Renewable Energy 177:34–51. https://doi.org/10.1016/j.renene.2021.05.124
Hassan MA, Abubakr M, Khalil A (2021d) A profile-free non-parametric approach towards generation of synthetic hourly global solar irradiation data from daily totals. Renewable Energy. https://doi.org/10.1016/j.renene.2020.11.125
Huang C, Bensoussan A, Edesess M, Tsui KL (2016) Improvement in artificial neural network-based estimation of grid connected photovoltaic power output. Renewable Energy 97:838–848
IRENA (2020) Renewable energy statistics 2020
Jamil B, Siddiqui AT (2018) Estimation of monthly mean diffuse solar radiation over India: performance of two variable models under different climatic zones. Sustain Energy Technol Assess 25:161–180
Keshtegar B, Bouchouicha K, Bailek N, Hassan MA, Kolahchi R, Despotovic M (2022) Solar irradiance short-term prediction under meteorological uncertainties: survey hybrid artificial intelligent basis music-inspired optimization models. Eur Phys J Plus 137:362. https://doi.org/10.1140/epjp/s13360-022-02371-w
Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. Meteorol Z 15:259–263
Li L-L, Wen S-Y, Tseng M-L, Wang C-S (2019) Renewable energy prediction: a novel short-term prediction model of photovoltaic output power. J Clean Prod 228:359–375
Liu L, Zhao Y, Chang D, Xie J, Ma Z, Sun Q, Yin H, Wennersten R (2018) Prediction of short-term PV power output and uncertainty analysis. Appl Energy 228:700–711. https://doi.org/10.1016/j.apenergy.2018.06.112
Liu Y, Wang L, Gu K (2021) A support vector regression (SVR)-based method for dynamic load identification using heterogeneous responses under interval uncertainties. Appl Soft Comput 110:107599. https://doi.org/10.1016/j.asoc.2021.107599
Maji IK, Sulaiman C, Abdul-Rahim AS (2019) Renewable energy consumption and economic growth nexus: a fresh evidence from West Africa. Energy Rep 5:384–392
Mao W, Wang F-Y (2012) Chapter 8 - Cultural modeling for behavior analysis and prediction, In: W. Mao, F.-Y. Wang (eds), New advances in intelligence and security informatics, Academic Press, Boston pp. 91–102. https://doi.org/10.1016/B978-0-12-397200-2.00008-7.
Mazumdar BM, Saquib M, Das AK (2014) An empirical model for ramp analysis of utility-scale solar PV power. Solar Energy 107:44–49. https://doi.org/10.1016/j.solener.2014.05.027
Mostefaoui M, Ziane A, Bouraiou A, Khelifi S (2019) Effect of sand dust accumulation on photovoltaic performance in the Saharan environment: southern Algeria (Adrar), Environmental Science and Pollution. Research 26:259–268. https://doi.org/10.1007/s11356-018-3496-7
Muzathik AM, Nik WBW, Ibrahim MZ, Samo KB, Sopian K, Alghoul MA (2011) Daily global solar radiation estimate based on sunshine hours., International Journal of Mechanical and Materials. Engineering 6:75–80
Qun'ou J, Lidan X, Siyang S, Meilin W, Huijie X (2021) Retrieval model for total nitrogen concentration based on UAV hyper spectral remote sensing data and machine learning algorithms – a case study in the Miyun Reservoir, China. Ecol Indicators 124:107356. https://doi.org/10.1016/j.ecolind.2021.107356

Razmjoo A, Shirmohammadi R, Davarpanah A, Pourfayaz F, Aslani A (2019) Stand-alone hybrid energy systems for remote area power generation. Energy Rep 5:231–241

Rezk H, Fathy A (2017) A novel optimal parameters identification of triple-junction solar cell based on a recently meta-heuristic water cycle algorithm. Sol Energy 157:778–791. https://doi.org/10.1016/j.solener.2017.08.084

Sikka DR (1997) Desert climate and its dynamics. Curr Sci 72:35–46

Slimani A, Tanjaoui MN, Boutadara A, Saihi L, Bailek N, Adel MS, Koussa K (2020) A PV-active power filter interface scheme for three phase balanced system. Int J Eng Res Afr 46:125–145. https://doi.org/10.4028/www.scientific.net/JERA.46.125

Tang H, Al-Dousari A (2006) Air pollution background study in Bouby an Island of Kuwait. Int J Sustain Dev Plan 1:326–341

Trigo-González M, Batlles FJ, Alonso-Montesinos J, Ferrada P, del Sagrado J, Martínez-Durbán M, Cortés M, Portillo C, Marzo A (2019) Hourly PV production estimation by means of an exportable multiple linear regression model. Renewable Energy 135:303–312. https://doi.org/10.1016/j.renene.2018.12.014

Tseng M-L (2017) Using social media and qualitative and quantitative information scales to benchmark corporate sustainability. J Clean Prod 142:727–738

VanDeventer W, Jamei E, Thirunavukkarasu GS, Seyedmahmoudian M, Soon TK, Horan B, Mekhilef S, Stojcevski A (2019) Short-term PV power forecasting using hybrid GASVM technique, Renew. Energy 140:367–379

Vaughn DM (2005) Arid Climates BT- Encyclopedia of World Climatology, in: Springer Netherlands, Dordrecht, pp 85–89. https://doi.org/10.1007/1-4020-3266-8_16.

Wang M, Peng J, Luo Y, Shen Z, Yang H (2021) Comparison of different simplistic prediction models for forecasting PV power output: assessment with experimental measurements. Energy 224:120162. https://doi.org/10.1016/j.energy.2021.120162

https://weatherspark.com/y/42374/Average-Weather-in-Adrar-Algeria-Year-Round, Weatherspark. (2022).

Wu W, Li M-F, Xu X, Tang X-P, Yang C, Liu H-B (2021) The transferability of random forest and support vector machine for estimating daily global solar radiation using sunshine duration over different climate zones. Theoret Appl Climatol 146:45–55. https://doi.org/10.1007/s00704-021-03726-6