

# Solar Energy Radiation Forecasting Method



Anuj Gupta, Kapil Gupta, and Sumit Saroha

## 1 Introduction

Energy demands around the world are expected to have more than double by 2050 and to more than triple by the end of century. Growing improvements in the existing energy grid would not be enough to meet the world's energy demand. Global security, economic growth and the quality of life are closely related to sufficient supplies of clean energy. The most daunting challenges for the world is to find energy resources to meet the rising demands of the planet. Solar prediction is a milestone to these challenges. Solar prediction depend on several factors such as characteristics of solar power plants which convert sun's energy to electric power, scattering process, knowledge of the Sun's path, nature of atmosphere etc. [1]. Solar forecasting information is necessary for the operation and planning for the future. Forecasting information provide grid operators with means to forecast and align electricity production and consumption and set up bilateral contract negotiations between suppliers and customers. Precise prediction methods increase the quality of the energy supplied to the grid and reduce the extra costs associated with ancillary equipment [2]. Based on the input data types and forecasting time horizons requirement various prediction approaches are introduced. For a very short time scale, on-site measurements are sufficient for the time series model. Intra-hour forecast obtained from a ground based sky imager with a high spatial and temporal resolution. Cloud motion vector forecast based on intra-day satellite images shows good

---

A. Gupta (✉) · K. Gupta

Department of Electronics and Communication Engineering, Maharishi Markandeshwar (Deemed to be University), Mullana-Ambala, India

S. Saroha

Department of Electrical Engineering, Guru Jambheshwar University of Science and Technology, Hisar, India

results. These forecast based on NWP models. Photovoltaic system integrated with grid required information up to minimum 2 days ahead or even beyond. Different type of solar power systems exist in the solar forecasting such as solar concentrating system and solar non-concentrating system [3]. DNI is highly correlated to a concentrated photovoltaic system. Measurement of DNI is very important for the operation and management of concentrated solar thermal power plants. DNI is highly affected by the number of factors like as dust storms, air pollution, cirrus clouds which degrade the DNI up to 30%. For non-concentrated solar PV systems, the primary element to measure is GHI which is less sensitive to error in DNI.

## 2 Research Motivation

India is facing Energy crises along with the world. There is a substantial difference in energy demand and supply. As our country progresses towards development day by day, this gap is rising and addressing this situation is very important in order to continue the ascending direction of our country. In order to address this situation, a range of options with a strong emphasis on renewable energy are considered. A lot of researcher and academics are engaged in developing tools, models and algorithms in today solar system. In today's dynamic world forecasting is a critical part of business planning with greater penetration of renewable energy resources and implementation of power deregulation in industry. Forecasting of solar power has become a major issue in power systems. Following needs of the markets, various techniques are used to forecast the solar radiation. As a result, it is anticipated that the thorough analysis would assist potential researchers as well as utility operators in gaining useful insight into the need for solar power output and forecasting models. The knowledge gained can also assist the government and energy market participants in making more efficient and beneficial decisions regarding solar power system implementation.

## 3 Solar Radiation Component

Three essential and fundamentals component are assessed for measurement of the solar irradiance [4].

### 3.1 *Direct Normal Irradiance*

Direct Normal Irradiance can be defined as the total amount of solar radiation obtained in a direct path from the sun at the horizontal earth surface with no atmospheric losses. Such amount of radiation is very important for the concentration of solar thermal system like concentrated solar power and concentrated photo voltaic.

### 3.2 *Diffuse Horizontal Irradiance*

Diffuse Horizontal Irradiance can be characterized by as the amount of solar radiation obtained from the sun on an indirect path on the horizontal surface, when it has been spread by air molecules, aerosol particles, cloud particles or other particles [5].

### 3.3 *Global Horizontal Irradiance*

Global Horizontal Irradiance is the cumulative amount of radiation received by the surface horizontal to the ground [6]. It consists of both DNI and DHI.

The three fundamental component of solar irradiation can be related to each other using the following equation:

$$GHI = DNI + DHI \times \cos(Z) \quad (1)$$

where Z represent the solar zenith angle.

## 4 **Need of Solar Forecasting**

The necessity of forecasting is for the operation and planning for the future. However the need of forecasting is given below

- Solar generation is variable in nature
- Necessity for successful bilateral contract negotiations between suppliers and customers
- Operational planning decision which are used to describe the economic location, type and scale of solar power plants
- Solar forecast provide grid operators with means to forecast and align electricity production and consumption
- Decision on expansion and enhancement of transmission, augmentation of generation, planning of distribution and exchange of regional electricity

## 5 **Solar Forecasting Methodologies**

The determinant factor for classifying methodologies for solar forecasting is different-2 forecast horizons. Precise forecasting can enable grid operators to create balance between consumption and production [7]. Table 1 shows three forms of horizons: intra-hour, intra-day, and day ahead.

1. For very short time scale various time series models such as an Artificial Neural Network, Autoregressive Integrated Moving Average, and Persistence model used for forecast solar irradiance [8, 9]
2. For short time irradiance forecasting, solar irradiance largely depend on the observation based on the temporal developments of clouds, may be used as a basis
  - For the sub-hour range, cloud data is collected from sky images ground based with high spatial resolution may be used to predict solar irradiance.
  - For 30 min up to 6 h solar irradiance depends on cloud motion vector from satellite photos.
3. For long term horizon, from 4–6 h ahead numerical weather prediction model perform better than the satellite based forecasts [9, 10]
4. There are also integrated techniques to derive an optimized forecast for the different-2 time horizon

## 6 State of Art for Solar Irradiance Forecast

As Per literature, forecasting methods are categorized into three types: statistical method, physical method and ensemble method.

### 6.1 Physical Methods

The physical methods are depending on the Total Sky Imagers (TSI), Numerical Weather Prediction (NWP) and physical parameters include temperature, cloudcover and humidity etc [11].

**Table 1** Relationship between time horizon, prediction model and related operations

Horizon type	Forecast horizon	Granularity time step	Events	Forecasting models
Intra hour	15 min to 2 h	30 s to minute	Ramping events, variability related operations	Total sky imagers and /or time series modles and NWP models and /or satellite imagery
Intraday	1–6 h	Hourly	Load following forecasting	
Day ahead	1–3 day	Hourly	Unit commitment, transmission scheduling, day ahead markets	

## Numerical Weather Prediction

The numerical weather prediction depends on atmospheric physics. For forecasting the future weather state, current observations of the weather are forecast using the assimilation process. NWP model performance is good for the horizons of 1 day to multi days ahead [12]. NWP process as follows:

- Step 1: In the initial stages, satellite and sky images ground based used to collect the current weather condition of an atmosphere. Assimilation process is used for processing the current weather state which is a very critical and complex process.
- Step 2: In the second phase, the most dominant atmospheric equation such as thermodynamic equation, Newton second law for fluids are integrated and solved [3].

Well, a known example of NWP models is worldwide model, regional model and weather research & forecasting model (WRF) model (Table 2). We can differentiate them in term of input parameters and spatial resolution [1].

## Cloud Imagery and Satellite Models

The situations of clouds are analyzed by cloud imagery with high spatial resolution. They detect the variability of clouds and predict global irradiance up to 6 h ahead. Solar irradiance is highly affected by cloud cover and cloud optical depth. Information about the clouds helps to predict the solar irradiance using total sky imagers for very short term forecasting. Some researchers develop their own TSIs while other researchers use commercially available TSIs such as TSI-800 [4].

## 6.2 Statistical Methods

Prediction method depends on the previous time series data of solar irradiation as input and does not depend on the internal phase of the model. Persistence model, ARIMA, ANN, Fuzzy logic etc. include in the statistical method.

**Table 2** Comparison of various NWP model

	Model	Time step	No. of layers	Resolution (km)	Agency
Mesoscale model	HRRR	15 min	50	3	NOAA
	NAM	6 h	60	12	NCEP
	RAP	1 h	50	13	NOAA
	WRF	Depends on user	Depends on user	1	NCAR
Worldwide model	ECMWF	3 h	91	25	–
	GFS	6 h	64	28	NOAA

### Time Series Model

Time series model predicts future value by consider previously observed value. Observation measured over time it may be hourly, daily and weekly. The sequence of data could be random and mainly focus on the pattern of the data. The pattern of the data should be recognizable and predictable for forecasting techniques. Auto-correlation Function (ACF) & Partial Correlation Function (PCF) used to identifying the pattern [5].

Time series is expressed as:

$$y(t) = s(t) + R(t) + T(t) \quad \text{Where } t = -1, 0, 1, 2, 3, \dots \dots \dots \quad (2)$$

$S(t)$  = Seasonal term,  $R(t)$  = Random term,  $T(t)$  = Trend term Stock market, revenue forecasting, economic forecasting, budgetary analysis, sales forecasting also utilize the application of time series method.

One of the benchmark model in the solar irradiance forecasting is an ARIMA model. Moving Average (MA), Auto Regressive (AR), & Autoregressive Moving Average (ARMA) is the variants of the ARIMA model. First one is moving average and the second one autoregressive. ARIMA is the most commonly used model for evaluating the relationship between real and expected performance. ARIMA model is the statistical tool to analyze the relationship between actual and forecasted output. ARIMA use three main steps for the forecasting: model identification, estimation of parameters and diagnostic checking [6]. There are seasonal and non-seasonal time series models that can be used for forecasting. An ARIMA is describe by three elements: p, d, q. where “p” is for autoregressive term, “q” is for moving average term and d is the number of differencing required to make the time series stationary. Mathematically, Autoregressive AR can be expressed as:

$$y_t = a + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t$$

$$= a + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots \dots \dots \phi_p y_{t-p} + \varepsilon_t \quad (3)$$

$y_t$  represents the actual value,  $\phi_i$  is model parameter and  $\varepsilon_t$  represent the random error,  $a$  and  $p$  are the constant term. This equation represents the linear relationship between the predicted value and the past value with some random error and constant term. Whereas the Moving Average equation represent the past value as a dependent variable

$$y_t = \eta + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t = \eta + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots \dots \dots \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4)$$

$\theta_j$  represent the model parameter,  $\eta$  represent the mean of the time series and  $q$  is the order of the model.

Combining Eqs. (1) and (2) become ARMA and mathematically can be expressed as

$$y_t = a + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (5)$$

Here  $p$  is for autoregressive and  $q$  is for moving average.

ARIMA is very popular for users due to the advantage of statistical expertise, the latest version of MATLAB makes it easier with the “Econometric Modeler app” available in MATLAB 2018 and 2019 [7]. In comparison to ARIMA, this approach requires an additional coefficient differencing operator, i.e. ( $p, d, q$ ). The ARIMA’s mathematical expression is

$$\phi(L)(1-L)^d y_t = \theta(L)\varepsilon_t \quad (6)$$

$\phi$  and  $\theta$  is model parameter,  $\varepsilon_t$  is random parameter and  $L$  denote lag operator and  $d$  represent differencing operator.

### Persistence Model

It is also known as naïve predictor. Persistence model is very simple as comparison to other forecasting model. It forecast the future value based on previous value [8].

$$x_{t+1} = x_t \quad (7)$$

The performance of persistence model is better when changes in weather pattern are little.

### Artificial Neural Network

The working of the neural network is similar to the human brain which takes the decision based on biological neurons. Neurons in the human brain perform the different-2 types of parallel processing, pattern recognition etc. The same phenomenon can be used to solve non-linear math problems in modeling, image analysis, and in other fields [10]. The ANN use different-2 algorithm to predict solar irradiation such as: scaled conjugate gradient, levenberg marquardt algorithm, pola-riberie conjugate gradient etc. This techniques trained model to map the input and output to obtain the best value. Support vector machine, radial basis network, multilayer perceptron and Hopfield network include under the artificial neural network. ANN process is carried out in three stages: (1) Design phase (2) Training phase

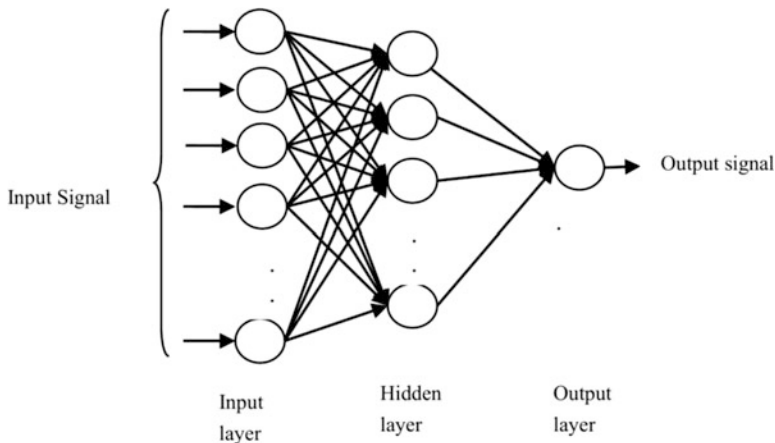


Fig. 1 Basic Architecture of ANN

(3) Validation phase. First stage consists of input parameters, neural network type, hidden neurons, In the training phase weight of the neuron are modified and in validation stage forecasting of solar irradiance based on trained weight [11]. Basic Architecture of artificial neural network is shown in Fig. 1.

The MLP structure is one of the important forms of neural network. This MLP structure consist an input layer, hidden layer and output layer. Hidden layer was characterized as a number of hidden neuron, input and output layer denoted by vector  $p$  &  $q$  respectively.

$$q = q(p : w) = \sum_{i=0}^h \left[ wif \left( \sum_{j=0}^d w_{ij}x_j \right) \right] \tag{8}$$

$j$  &  $i$  represent the weights and biases while vector  $w$  supervises the non-linear mapping. Babak Jahani et al. compared the empirical, artificial neural network and artificial neural network with a genetic algorithm optimization technique to predict the global solar radiation. The Genetic algorithm was used in the model to reduce the error in predictive results [13] Premalatha Neelamegam et al. proposed two artificial neural network model with different combinations of inputs, the accuracy of the model was measured based on MAE, RMSE and  $R^2$  [14]. Voyant et al. presented a review of solar radiation forecasting using machine learning techniques. According to the author standalone models such as: artificial neural network, linear regression, random forest, support vector machine performed well in the forecasting field while hybrid model are viable way to improve the accuracy of prediction model [15].



### Support Vector Machine

It is a form of machine learning introduced in 1995 by Cortes and Vapnik with statistical learning. Firstly, this particular approach is developed for pattern recognition and is now enthusiastically used for various technologies such as image retrieval, fault diagnosis, regression computation and forecasting etc. [16]. The time series is used to train a model that is as simple as neural network model and there is no question of over fitting curve, struck to local minima in SVM [17]. Essentially, it uses the mapping function to map the input vector  $(x_1 + x_2 + x_3 + \dots \dots .x_n)$  to the output  $(y_1 + y_2 + y_3 + \dots \dots .y_n)$ . The equation with SVM can be represented as

$$y = \sum_{i=1}^n \phi_i k(x, x_i) + b \tag{9}$$

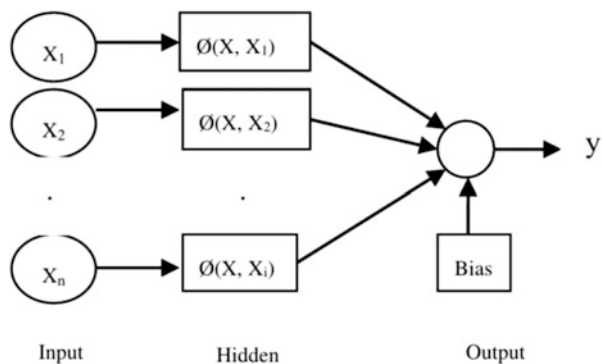
where  $y$  is output function and  $b$  is bias and the basic architecture of SVM shown in Fig. 2.

Jie shi et al. used SVM model to forecast the solar power. The entire data is dividing into four groups based on all seasons. The categorized data are feed into four SVM developed models. The performances of developed models are evaluated using RMSE and MAE and performance of all developed models outperform bench mark model [18].

### Markov Chain

It represents a deterministic cycle that used to forecast wind and solar irradiance. The procedure of deterministic cycle is essential reliant on the neighboring states i.e. the current state variables are dependent on the former one. Similarly, the next state variables are reliant on the current one [19] as shown in Fig. 3.

Fig. 2 Architecture of SVM



This procedure is described by a sequence of finite random digit. Let  $\{y_n, n=0,1,2,\dots\}$ . The sequence for the current state  $i$  at  $n$ th time can be shown as

$$y_n = i$$

Whereas the likelihood of next condition in  $j$  is  $P_{ij}$ . i.e.

$$P\{y_{n+1} = j | y_n = i, y_{n-1} = i_{n-1}, \dots, y_1 = i_1, y_0 = i_0\} = P_{ij} \quad (10)$$

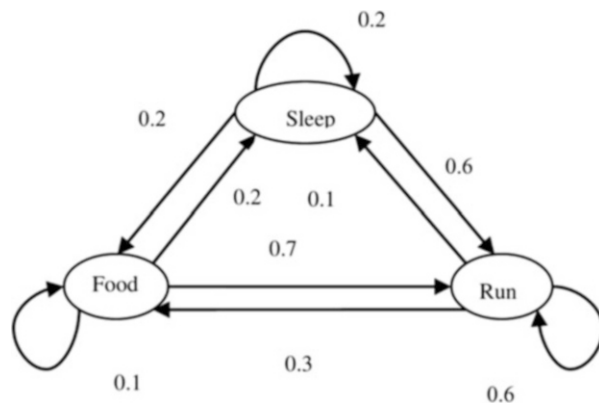
This equation shows the next state having the dependency on the present state.

To estimate the power of a photovoltaic device, Sanjari et al. developed a markov chain model. The input parameters were radiant energy and relative humidity. The proposed model outperforms other approaches in terms of MAPE results [20].

### 7 Empirical Model

Empirical modeling is a genetic term for activities that create model by observation and experiments. Samani and Hargreaves present first empirical model in 1982. Now number of model have evolved by changing the various factor such as altitude, latitude, angular position, tilt angle, air particle dispersion, water vapor content, hours of sunlight, max temp, lowest temperature, cloud cover index etc. Empirical model is a mathematical technique used to forecast solar irradiance by creating a linear or nonlinear connection between climatologically and solar variables [21]. Nadjem Bailek et al. addressed mathematical models for obtaining a accurate diffuse solar radiation. The developed models were dividing into three categories based on sunshine period and clearness index. The performance of all three models were evaluated using MAPE, RMSE and U95 (Uncertainty Factor) and compared with the eight models discussed in the literature [22].

Fig. 3 Markov chain process

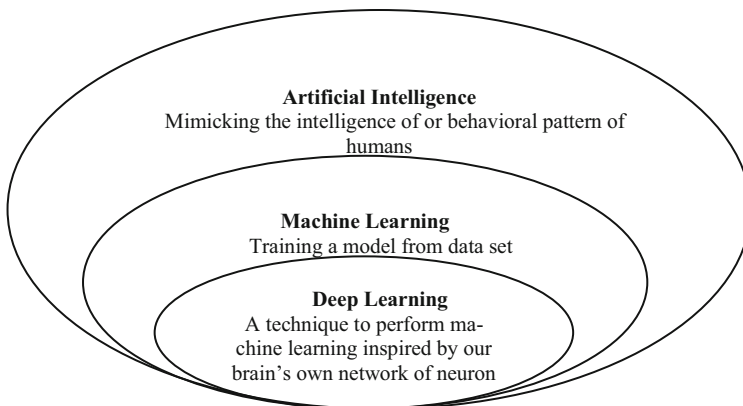


## 8 Deep Learning

This term deep learning introduced in 1986 by Rina Dechter & over the past years, deep learning has become very prevalent. It is also named as deep structured learning, is a branch of machine learning which intern is the subset of artificial intelligence; Machine learning is a technique for achieving artificial intelligence through algorithms trained with data, whereas artificial intelligence is a technique for enabling a machine to act like a person as shown in Fig. 4. On the other hand, Deep learning is a set of statistical machine learning techniques used to learn feature hierarchies which is often based on ANN. Here, learning can be supervised, unsupervised or semi-supervised. The application of deep learning algorithm such as: CNN, RNN and DBN are used in computer vision, image processing, audio recognition, speech recognition etc. [23].

Deep learning is a modern substitute for machine learning; we can have a variety of structured and unstructured data in various forms and aspects from every region of the world. Structural data can be easily drowned out while unstructured data could take decades to provide relevant information. Deep learning is used to deal with a huge amount of data simply known as big data which is taken from various medium such as social medium, online platform i.e. e-commerce, internet engine search so on. This abundance amount of data is smoothly accessible. It can be shared through fintech application such as mobile payment applications etc.

Wang et al. present a forecasted model using deep learning techniques. The author applies pre-processing technique to improve the performance of forecasted model [1]. Melit et al. conducted a review on machine learning techniques. According to the author, deep learning techniques and numerical weather forecast with extracting features use to generate long term photovoltaic power generation and for determine the time dependence information in forecasting the performance of convolution neural network and recurrent neural network were better [24].



**Fig. 4** Deep learning is a subset of machine learning

## 9 Hybrid Method

These models are used to enhancing the precision of forecasted models. There are many factors that are not considered in the individual model by a model needed to perform more accurately. The hybrid approach is about integrating two or more methods for determining the forecast. Various data decomposition techniques used with forecasted models to increase the accuracy of forecasted models [25] (Table 3).

## 10 Factors Influencing Solar Radiation Forecasting

There are some other factors/parameters that affect the accuracy of model forecasting directly or indirectly. The solar forecasting depends on forecast horizons, geographical condition, day/night value and normalization, testing period, climatic variability and pre-processing technique.

### 10.1 Input Parameter Selection

Solar energy is an important aspect of solar radiation forecasting but it is unavailable for many places due to measuring device cost, upkeep and calibration. So, we need input parameters for estimating the solar radiation. The input parameter may be temperature, pressure, humidity, solar zenith angle, precipitation, latitude, longitude, wind direction, wind speed, sunshine duration [48–50].

### 10.2 Forecast Horizon

The time horizon concept is concerned with the span of time duration which the model is used for prediction. Time duration can range from a few seconds to many hours. As per literature, four type of time horizon exist such as: very short term, short term, mid-term and long term forecasting [48–51].

### 10.3 Climatic Variability

The variables in the input data may be systemic, endogenous and exogenous. On various combinations of input parameter different model behave differently. The model's efficiency suffers as the number of insignificant meteorological variables used. As a result, the necessary parameters must be chosen to improve a model's efficiency. To predict solar radiation, M.A. Behrang et al. proposed two models using neural network based on various combinations of a climatologically variable [52].

**Table 3** Study of solar forecasting techniques

Author (year)	Location	Input value	Prediction method	Time horizon	Evaluation metrics	Forecast variables	Results
A. Koca et al., 2011 [17]	Turkey	Latitude, longitude, month, cloud cover, wind speed	ANN	-	RMSE, R <sup>2</sup>	Solar radiation	RMSE = 0.0358 R <sup>2</sup> = 0.9974
F. Wang et al., 2012 [26]	China	Third order difference of SI, T <sub>m</sub> day, T <sub>a</sub>	ANN	24-72 h ahead	MAPE, RMSE, MABE	Solar radiation	MAPE = 9.09-26.7% RMSE = 42.29-84.65 (W/m <sup>2</sup> ) MABE = 31.10-64.6 (W/m <sup>2</sup> )
R. Hossain et al., 2012 [27]	Australia	Wind speed, wind direction, relative humidity, meteorological variables	ANN	3 h	R	WE, SE	R <sup>2</sup> = 0.96399
S. Bhardwaj et al., 2013 [28]	Gurgaon	AT, RH, AP, WS, WD, SR	HMM + GFM	N/A	MAPE, RMSE	Solar radiation	RMSE = 7.9 MAPE = 3.4
S. X. Chen et al., 2013 [29]	Singapore	T, SKI, SR	Fuzzy+ANN	1-h day	MAPE	PV power output	MAPE = 6.03%-9.65%
I. Colak et al., 2015 [30]	N/A	Time series	ARMA & ARIMA	1-3 step ahead	MAPE	Solar power	MAPE_ARMA = 72.67% MAPE_ARIMA = 33.07%
A. W. Itun et al., 2016 [31]	Thailand	Global solar radiation, daily solar radiation, clear sky index	Markov transition matrix	Hourly	RMSE	Global solar radiation	Second order > first order
L. M. Aguiar et al., 2016 [32]	Spain	Meteorological data, NWP, satellite data	Neural network	1-6 h	RMSE	GHI	RMSE = 83.58%-147.88%
N. Kumar et al., 2017 [33]	India	WS, RH, P, SD, T <sub>max</sub> , T <sub>min</sub>	ANN + unity feedback RBF + LR	N/A	MAPE	Daily global solar radiation	MAPE = 14.84%-16.32%

(continued)

**Table 3** (continued)

Author (year)	Location	Input value	Prediction method	Time horizon	Evaluation metrics	Forecast variables	Results
S. Li et al., 2017 [34]	FortPeck, Montana	Time series data	Discrete Markov chain	Seasonal	% error	TSRY	% Error lies between 10% to 6%
C. R. Chen et al., 2017 [35]	Taiwan	Temp., relative humidity, wind speed, wind direction	Neural network	60 min	MBE, RMSE	GSI	RMSE = 232 (W/m <sup>2</sup> ) MABE = 32 (W/m <sup>2</sup> )
M. Bou-Rabee et al., 2017 [36]	Kuwait	Time series data	GD algorithm, LM algorithm	N/A	MAPE	Solar radiation	MAPE = 86.3%
M., Guermoui et al., 2018 [37]	NEA, Singapore	Maximum elevation, declination, angle day duration	WGPR-CFA WGPR-PFA	N/A	RMSE, r <sup>2</sup>	GHR, DHR	RMSEGHR = 3.18 (MJ/m <sup>2</sup> ) r <sup>2</sup> GHR = 85.85% RMSE DHR = 5.23 (MJ/m <sup>2</sup> ) r <sup>2</sup> DHR = 56.21%
H. Verbois et al., [38] 2018	Singapore	NWP	NWP + PCA	1-day	RMSE, rMSE, MAE, rMAE	SI	RMSE = 169 (W/m <sup>2</sup> ) rRMSE = 35.7% MAE = 193 (W/m <sup>2</sup> ) rMAE = 28.1%
L. Benali et al., 2019 [39]	France	Time series data	SP,ANN,RF	1–6 h ahead	Nrmse, MSE, MAE, nMAE	PV power	nRMSE can be improved by 20%
Persson et al., 2017 [40]	Japan	Historical data, NWP data	Gradient boosted regression tress	1-h	nRMSE	Solar power generation	Nrmse = 0.1–0.137 kW
L. Cornejo-Bueno et al., 2019 [41]	Spain	Clear sky, irradiance, CI	ELM	1-h ahead	RMSE	GSR	RMSE = 60.60 W/m <sup>2</sup>

A. T. Eseye et al., 2018 [42]	China	Temperature, cloud cover, relative humidity, wind speed	Wavelet-PSO-SVM	1 day	MAPE	Solar PV power	MAPE = 4.22%
K. R. Kumar et al., 2018 [43]	Andhrapradesh, India	Temperature, pressure, wind speed	ANN, ANFIS	-	% error	PV power	ANN > ANFIS
F. Rodriguez et al., 2018 [44]	Euskalmet	Atmospheric pressure, relative humidity, temperature	ANN	N/A	RMSE	Solar irradiation	RMSE = 0.03%
D. Liu et al., 2019 [25]	GEFcom 2014	Meteorological data	Random Forest+ PCA	1-3 step ahead	RMSE, MAE	PV power	RMSE = 8.88-9.82% MAE = 4.76-5.80%
W. Vandeventer et al., 2019 [45]	Australia	Ambient temperature, solar irradiation	GASVM	Hourly	RMSE, MAPE	PV power	RMSE = 11.226WMAPE = 1.7052%
A. Shadab et al., 2019 [46]	New Delhi	Time series	SARIMA	Month	MPE	Solar radiation	MPE = 1.402
S. Ghimire et al., 2019 [23]	Australia	Past global solar radiation	CNN + LSTM	30 min	MAPE, APE, rRMSE	Solar radiation	MAPE = 4.672% APE = 1.233% Rrmse = 1.515%
Y. Liu et al., 2020 [21]	China	Sunshine ration, average temperature, relative humidity	SVM Empirical model	-	RMSE, MABE, MBE, R <sup>2</sup>	Solar radiation	SVM > CNQR > empirical model
J. Dong et al., 2020 [47]	Oak Ridge National Laboratory	Past PV	Uncertainty bias +kalman filter	1-50 min	MAPE, nRMSE	Solar radiation PV power	MAPE = 5.72-25.75% Nrmse = 7.43-26.13%

### ***10.4 Night Hour and Normalization***

The solar irradiance is not available in the night hours. But energy providers required PV production continuous at all times. The bulk of the test took place during the day time and omitted the night time hours. To avoid the effects of inaccurate readings, the time just after sunrise and just before sunset were also excluded from the data collection [53].

### ***10.5 Preprocessing Techniques***

The quality of input data plays crucial role in the enhancement of forecasted model. The data collection from various sites mostly available in raw format and does not have a significant characteristics to provide appropriate accuracy. So, the data has to be process before processing with the model called preprocessing stage. Here, the preprocessing means scale up or down the input measurements, clean up and define the input data accordingly to the specifications. There are number of preprocessing techniques available in the literature such as: wavelet transforms kalman filter, empirical mode decomposition, self organization map, normalization, trend free time series which were used before the model learning [54].

### ***10.6 Training and Testing Period***

The training and testing cycle is also one of the factors which affect the accuracy of the model. Various studies have shown that the large collection of training data set enhance the learning capacity and also improve the accuracy. B. Sivaneasan et al. used 4 months data set to train the model and 1 month data set is used to test the model [55] whereas Mohammed Bou-Rabee et al., used 3 years data to train and 1 year data to test the model [36].

### ***10.7 Geographical Location***

The behavior of the model varies according the geographical location. The model performance directly affected by the area or locations having certain/uncertain climatic conditions like Leh, India where the cold desert receives the enormous amount of solar radiation may perform better than the area having most of the cloud in the sky [14].



## 11 Solar Forecasting Evaluation Metrics

Various evaluation metrics have used by researchers to predict solar irradiation value. The aim of the evaluation metrics is to compare the actual observed value with the forecasted value. Different performance metrics have different units; for example, the statistical error of solar radiation is measured in  $W/m^2$ , whereas power is measured in KW or MW. The forecast evaluation provides a forecaster with:

- The ability of selecting correct forecasting model so, that the maximum prediction accuracy can be achieved as comparison to others.
- Forecasters analyze forecasting error and utilizing it for improving performance of forecasting model.

Forecasting model accuracy is the primary concern for the forecaster and it can be evaluated by using the following Conventional Statistical Assessment Metrics:

- **Normalized Error:**

It is indicate by Ne and is used to identify outliers in a set of data used. Mathematically it is represented as [26]

$$nE = \frac{R_{prediction} - R_{real}}{\max(R_{prediction})} \quad (11)$$

- **Mean Bias Error:**

This metric is used to measure the system's or model's average bias [56].

$$MBE = \frac{1}{n} \sum_{i=1}^n (R_{prediction,i} - R_{real,i}) \quad (12)$$

The MBE positive value indicates that the model is overestimation whereas the negative value represents the underestimation.

- **Mean Absolute Error (MAE):**

It provides uniform forecasting error. This metric provides a difference between two set of data [57].

$$MAE = \frac{1}{n} \sum_{i=1}^n |R_{prediction,i} - R_{real,i}| \quad (13)$$

- **Standard Deviation Error (SDE):**

This metric is used to assess the deviation from the average [42].

$$SDE = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{prediction,i} - R_{real,i} - MBE)^2} \quad (14)$$

- **Root Mean Square Error (RMSE):**

It is a metric for determining the largest expected error in the forecasted data [58].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (R_{prediction,i} - R_{real,i})^2} \quad (15)$$

- **Mean Absolute Percentage Error (MAPE):**

It is a metric for uniform forecasting error expressed as a percentage [56]

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{R_{prediction,i} - R_{real,i}}{R_{real,i}} \right| \quad (16)$$

- **Mean deviation Absolute Percentage Error (Md-APE):**

Outliers have less of an effect on this metric than they do on the MAPE [59].

$$MdAPE = median \left( \left| 100 \cdot \frac{R_{forecast} - R_{real}}{R_{real}} \right| \right) \quad (17)$$

- **Relative root mean square error (Rrmse):**

It is a metric for determining the largest expected error in the forecasted data set [59]

$$Rrmse = \frac{RMSE}{R_{real}} \times 100 \quad (18)$$

- **Correlation Coefficient**

This metric is used for representation a connection between two set of data.

Forecasted Model's ability is better if the value of correlation coefficient is high. The optimal correlation coefficient value is 1 [60]

$$\rho = \frac{(Conv(R_{real}R_{prediction}))^2}{Var} \quad (19)$$

where  $R_{real}$  represent real radiation value and  $R_{prediction}$  represent predicted radiation value.

- **Determination Coefficient**

It is used to derive knowledge about the association between predicted and actual values and this metric is denoted by  $R^2$  [61]

$$R^2 = 1 - \frac{\text{var}(R_{real} - R_{prediction})}{\text{var}(R_{prediction})} \tag{20}$$

- **Clear Sky Index**

It is defined as the proportion of measure radiation to the clear sky radiation

$$K_t = \frac{R_{real}}{R_{real(CKS)}} \tag{21}$$

### 11.1 Contemporary Statistical Metrics

The MAPE, MAE and RMSE cannot distinguish the two different data sets with the same mean and standard deviation but having varying consistency or skewness distributions and Kurtosis. However, traditional metrics are required to measure the system but other parameters such as skewness, kurtosis may affect the real time procedure.

- **Kolmogorov-Smirnov test integral (KSI) and OVER metrics**

The Kolmogorov-Smirnov test is used to distinguish the relationship between two data sets. The distinction between two CDFs is represented as [62].

$$D = \max |F_{(ni)} - \widehat{F}_{(ni)}| \tag{22}$$

F represent the actual data set for solar power generations and  $\widehat{F}$  represents the predicted solar power generation data set. D statistics define the disparity between one sample and the reference sample is smaller than the target value ( $V_c$ ). The target value depends on the amount of points in the estimation of the data series, measured at a confidence level of 99% [62].

$$V_c = \frac{1.63}{\sqrt{N}} N \geq 35 \tag{23}$$

The distinction between the two CDFs of real and forecasted energy is specified for each phase

$$D_j = \max |F_{ni} - \widehat{F}_{ni}| \text{ where } j = 1, 2, 3, \dots, m \tag{24}$$

Where  $P_i \in [P_{\min} + (j - 1)d, p_{\min} + jd]$ .

The period difference  $d$  is calculated as follows:

$$d = \frac{P_{\max} - P_{\min}}{m} \quad (25)$$

The KSI factor is represent as the distinction between two CDFs calculated as

$$KSI = \int_{x_{\max}}^{x_{\min}} D_n dx \quad (26)$$

The actual value and the predicted value are identical when the KSI is lower [62]  $D_n$  represent the distinction between the two CDFs.

- **OVER**

It is used to define the difference between the cumulative distribution function of real and predicted solar value [59].

$$OVER = \int_{x_{\max}}^{x_{\min}} T dt \quad (27)$$

Where  $x_{\min}$  and  $x_{\max}$  represent the minimum and maximum radiation value and  $t$  is represented as

$$T = \begin{cases} T_j - U_c & \text{if } D_j > U_c \\ 0 & \text{if } D_j < U_c \end{cases}$$

$U_c$  represent the critical value.

- **Skewness and Kurtosis**

The assessment of incongruity in a probability distribution is skewness [55].

$$\gamma = E \left| \left( \frac{e - \mu_e}{\sigma_e} \right)^2 \right| \quad (28)$$

$e$  = difference between the forecast solar power and real solar power  $\mu_e$  indicate mean error and  $\sigma_e$  represent the standard deviation error.

Kurtosis: It is a metric used for assessing the magnitude of the distribution

$$K = \frac{\mu_4}{\sigma_e^4} - 3 \quad (29)$$

$K$  is the kurtosis,  $\mu_e$  represent the mean and  $\sigma_e$  denote standard deviation error.

- **Uncertainty Quantification**

Renyi entropy of solar forecast error: The Renyi entropy is used to quantify the degree of uncertainty in solar prediction and expressed as [63, 64]

$$H_a(x) = \frac{1}{1-a} \log_2 \sum_{i=1}^n P_i^a \quad (30)$$

$a$  represent the scale of Renyi entropy and  $P_i$  represent the probability distribution function. Larger Renyi entropy value indicate more ambiguities present in the expected outcome.

- **Metrics for Ramp Characterization**

The main priority associated with grid operators is to maintain a constant solar power output because a number of fluctuations occurred in the solar output due to variability of weather events. Solar ramps also influenced by different time and geographic factors, they can be up ramp or down ramp. The accurate solar forecasting help to overcome these types of uncertainties [65].

In case of Ramp Characterization Florita et al. developed a signal compression algorithm in which used to extract ramp interval into a sequence of power cycle by specifying the beginning and finishing point of each ramp [66].

- **Ramp Detection Index (RDI)**

This metric is used to measure the caliber of a model to predict ramps in a short time frame [67].

$$RDI = \frac{N_{hit}}{N_{hit} + N_{miss}} \quad (31)$$

Where  $N_{hit}$  represent the total number of strike counts

$N_{hit} + N_{miss}$  represent the cumulative number of times a ramp appears.

- **Ramp Magnitude (RM)**

It is used to measure the difference between radiation value at current time and after small time with respect to the clear sky radiation value of the current time. Chu et al. study the concept of ramp magnitude in their research paper to explore the caliber to predict ramps [68]

$$RM = \frac{R_h(t) - R_h(t + \Delta t)}{R_{csk}(t)} \quad (32)$$

## 12 Conclusion

This study conducts on several statistical, physical and ensemble methods. NWP models, satellite based models and cloud Imagery are studied in the case of physical method. These models are used for long-term forecasting horizons ranging from a few hours to several days and are ideal for circumstances where no other information is available. The only downside of the physical approach is that they are suffering from spatial and temporal resolution. Various time series and learning model discussed in the statistical model. In the time series method, the observation is measured over time. AR, MA, ARMA, ARIMA included in the time series model. Learning model include Markov chain, artificial neural network, support vector machine which provide excellent information about the solar irradiance when enough historical data is available. Nowadays, a hybrid method is used to overcome the shortcomings of individual model. These techniques also reduce the forecasting error. For evaluating the performance of prediction model various error metrics are discussed. Solar prediction error assessments allow understanding the model and re-evaluating it in case of high error.

## References

1. Kleissl, J. (2013). *Solar energy forecasting and resource assessment*. Academic.
2. Kostylev, V., & Pavlovski, A. (2011). Solar power forecasting performance towards industry standards. In *Proceedings of 1st International Workshop on Integration of solar power into power systems*, Aarhus, Denmark.
3. Mentzer, J. T., & Hoon, M. A. (2004). *Sales forecasting management: a demand management approach*. Sage.
4. Diagne, H. M., David, M., Lauret, P., Bolan, J., & Schmutz, N. (2013). Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renewable and Sustainable Energy Reviews*, 27, 65–76.
5. Mentzer, J. T., & Hoon, M. A. (2004). *Sales forecasting management: a demand management approach*. Sage.
6. Bacher, P., Madsen, H., & Nielsen, H. A. (2009). Online short-term solar power forecasting. *Solar Energy*, 83(10), 1772–1783.
7. *Econometric Modeler App Overview-MATLAB & Simulink-MathWorks India*. <https://in.mathworks.com/help/econ/econometric-modeler-overview.html>.
8. Hunand, R., Hunag, T., Gadh, R., & Li, N. (2012). Solar generation prediction using the ARMA model in a laboratory-level Micro-grid. In *IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, Tainan (pp. 528–533).
9. Xie, Y. (2011). Values and limitations of statistical models. *Research in Social Stratification and Mobility*, 29(3), 343–349.
10. Haykin, S. (2008). *Neural networks and learning machines*. Prentice Hall.
11. Khosravi, A., Nunes, R. O., Assad, M. E. H., & Machado, L. (2018). Comparison of artificial intelligence methods in estimation of daily global solar radiation. *Journal of Cleaner Production*, 194, 342–358.
12. Koca, A., Oztop, H. F., Varol, Y., & Koca, G. O. (2011). Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey. *Expert System with Applications*, 38(7), 8756–8762.

13. Jahani, B., & Mohammadi, B. (2019). A Comparison between the application of empirical and artificial neural network methods for estimation of daily global solar radiation in Iran. *Theoretical and Applied Climatology*, 137(1–2), 1257–1269.
14. Premalatha, N., & Arasu, A. V. (2016). Prediction of solar radiation for solar system by using artificial neural network (ANN) models with different back propagation algorithm. *Journal of Applied Research Technology*, 14(3), 206–214.
15. Voyant, C., et al. (2017). Machine learning method for solar radiation forecasting: A review. *Renewable Energy*, 105, 569–582.
16. Zendejboudi, A., Baseer, M. A., & Saidure, R. (2018). Application of support vector machine models for forecasting solar and wind energy resources: A review. *Journal of Cleaner Production*, 199, 272–285.
17. Hargreaves, G. H., & Samani, Z. A. (1982). Estimation potential Evapo-transpiration. *Journal of the Irrigation and Drainage Division*, 108(3), 225–230.
18. Shi, J., Lee, W. J., Liu, Y., Yang, Y., & Wang, P. (2012). Forecasting power output of photovoltaic systems based on weather classification and support vector machine in. *IEEE Transactions on Industry Applications*, 48(3), 1064–1069.
19. Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., & Hu, Z. (2016). Photovoltaic and solar forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems*, 1(4), 38–46.
20. Sanjari, et al. (2017). Probabilistic forecast of PV power generation based on higher order Markov chain. *IEEE Access*, 32, 2942–2952.
21. Liu, Y., Zhou, Y., Chen, Y., Wang, D., Wang, Y., & Zhu, Y. (2020). Comparison of support vector machine and copula-based nonlinear quantile regression for estimating the daily diffuse solar radiation: A case study in China. *Renewable Energy*, 146, 1101–1112.
22. Quansah, E., et al. (2014). Empirical models for estimating global solar radiation over the Ashanti region of Ghana. *Journal of Solar Energy*, 14, 1–6.
23. Ghimire, S., Deo, R. C., Downs, N. J., & Raj, N. (2019). Global solar radiation prediction by ANN integrated with European centre for medium range weather forecast fields in solar rich cities of Queensland Australia. *Journal of Cleaner Production*, 216, 288–310.
24. Melit, A., et al. (2020). Advanced methods for photovoltaic output power forecasting. A review. *Applied Science*, 10, 487.
25. Liu, D., & Sun, K. (2019). Random forest solar power forest based on classification optimization. *Energy*, 187, 115940.
26. Wang, F., Mi, Z., Su, S., & Zhao, H. (2012). Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters. *Energies*, 5(5), 1355–1370.
27. Hossain, R., Oo, A. M. T., & Ali, A. B. M. S. (2012). Historical weather data supported hybrid renewable energy forecasting using Artificial Neural Network (ANN). *Energy Procedia*, 14, 1035–1040.
28. Bhardwaj, S., et al. (2013). Estimation of solar radiation using a combinations of Hidden Markov Model and generalized Fuzzy model. *Solar Energy*, 93, 43–54.
29. Chen, S. X., Gooi, H. B., & Wang, M. Q. (2013). Solar radiation forecast based on fuzzy logic and neural networks. *Renewable Energy*, 60, 195–201.
30. Colak, I., Yesilbudak, M., Genc, N., & Bayindir, R. (2016). Multiperiod predictions of solar radiation using ARMA and ARIMA models. In *Proceeding-2015 IEEE 14th International conference on Machine Learning and Applications, ICMLA 2015* (pp. 1045–1049).
31. Wibun, A., & Chaiwiwatworakul, P. (2016). An estimation of Thailand's hourly solar radiation using Markov transition matrix method. *Applied Mechanics and Materials*, 839, 29–33.
32. Aguiar, L. M., Pereira, B., Lauret, P., Diaz, F., & David, M. (2016). Combing solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting. *Renewable Energy*, 97, 599–610.
33. Kumar, N., Sinha, U. K., Sharma, S. P., & Nayak, Y. K. (2017). Prediction of daily global solar radiation using Neural Networks with improved gain factors and RBF Networks. *International Journal of Renewable Energy Research*, 7(3), 1235–1244.

34. Li, S., Ma, H., & Li, W. (2017). Typical solar radiation year construction using k-means clustering and discrete time Markov chain. *Applied Energy*, 205, 720–731.
35. Chen, C. R., & Kartini, U. T. (2017). k-nearest neighbor neural network models for very short term global solar irradiance forecasting based on meteorological data. *Energies*, 10(2), 186.
36. Bou-Rabee, M., Sulaiman, S. A., Saleh, M. S., & Marafi, S. (2017). Using artificial neural networks to estimate solar radiation in Kuwait. *Renewable and Sustainable Energy Reviews*, 72, 434–438.
37. Guermoui, M., Melgani, F., & Danilo, C. (2018). Multi-step ahead forecasting of daily global and direct solar radiation: A review and case study of Ghardaia region. *Journal of Cleaner Production*, 201, 716–734.
38. Verbois, H., Huva, R., Rusydi, A., & Walsh, W. (2018). Solar irradiance forecasting in the tropics using numerical weather prediction and statistical learning. *Solar Energy*, 162, 265–277.
39. Benali, L., Notton, G., Fouilloy, A., Voyant, C., & Dizene, R. (2019). Solar radiation forecasting using artificial neural network and random forest methods: Application to normal beam, horizontal diffuse and global components. *Renewable Energy*, 132, 871–884.
40. Persson, C., Bacher, P., Shiga, T., & Madsen, H. (2017). Multi site solar power forecasting using gradient boosted regression trees. *Solar Energy*, 150, 423–436.
41. Cornejo-Bueno, L., Casanova-Mateo, C., Sanz-Justo, J., & Salcedo-Sanz, S. (2019). Machine learning regressor for solar radiation estimation from satellite data. *Solar Energy*, 183, 768–775.
42. Eseye, A. T., Zhang, J., & Zheng, D. (2018). Short-term photovoltaic solar power forecasting using a hybrid wavelet-PSO-SVM model based on SCADA and meteorological information. *Renewable Energy*, 118, 357–367.
43. Kumar, K. R., & Kalavathi, M. S. (2018). Artificial intelligence based forecast models for prediction solar power generation. *Materials Today: Proceedings*, 5(1), 796–802.
44. Rodriguez, F., Fleetwood, A., Galarza, A., & Fontan, L. (2018). Predicting solar energy generation through artificial neural networks using weather forecasted for microgrid control. *Renewable Energy*, 126, 855–864.
45. Van Deventer, W., et al. (2019). Short-term PV power forecasting using hybrid GASVM technique. *Renewable Energy*, 140, 367–379.
46. Shadab, A., Said, S., & Ahmad, S. (2019). Box-Jenkins multiplicative ARIMA modeling for prediction of solar radiation: A case study. *International Journal of Energy Water Resources*, 3, 305–318.
47. Dong, J., et al. (2020). Novel stochastic methods to predict short term solar radiation and photovoltaic power. *Renewable Energy*, 145, 333–346.
48. Voyant, C., & Notton, G. (2018). Solar irradiation now casting by stochastic persistence: A new parsimonious simple and efficient forecasting tool. *Renewable and Sustainable Energy Reviews*, 92, 343–352.
49. Jiang, F., et al. (2017). Artificial intelligence in healthcare: Past, present and future. *Stroke and Vascular Neurology*, 2(4), 230–243.
50. Bowing, H., Xiong, J. N., & Yinzhaoh, C. (2015). The mid-term forecast method of solar radiation index. *Chinese Astronomy and Astrophysics*, 39(2), 198–211.
51. Yang, D. (2019). A guideline to solar forecasting research practice: Reproducible, operational, probabilistic or physically based ensemble and skills (ROPES). *Journal of Renewable and Sustainable Energy*, 11(2), 022701.
52. Behrang, M. A., Assareh, E., Ghanbarzadeg, A., & Noghrehabadi, A. R. (2010). The potential of different artificial neural network techniques in daily global solar radiation modeling based on meteorological data. *Solar Energy*, 84(8), 1468–1480.
53. Paulescu, M., & Paulscu, E. (2019). Short term forecasting of solar irradiance. *Renewable Energy*, 143, 985–994.
54. Monjoly, S., Andre, M., Calif, R., & Soubdhan, T. (2017). Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach. *Energy*, 119, 228–298.



55. Sivaneasan, B., Yu, C. Y., & Goh, K. P. (2017). Solar forecasting using artificial neural network with fuzzy logic pre-processing. *Energy Procedia*, *143*, 727–732.
56. Willmott, C. J., & Matsuura, K. (2005). Advantage of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climatic Research*, *30* (1), 79–82.
57. Zhang, J., Hodge, B. M., Florita, A., Lu, S., Hamann, H. F., & Banunarayanan, V. (2013). Metrics for evaluating the accuracy of solar power forecasting. In *3rd International Workshops on Integration of Solar Power into Power Systems* (Vol. 17436, pp. 1–10).
58. Lauret, P., Voyant, C., Soubdhan, T., David, M., & Poggi, P. (2015). A benchmarking of machine learning technique for solar radiation forecasting in an insular context. *Solar Energy*, *112*, 446–457.
59. Sobri, S., Koohi-Kamali, S., & Rahim, N. A. (2018). Solar photovoltaic generation forecasting methods: A review. In *Energy conversion and management* (Vol. 156, pp. 459–497). Elsevier.
60. Hyndman, R. J., & Koehler, A. B. (2006). Another look at measure of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688.
61. Bouzgou, H., & Gueymard, C. A. (2019). Fast short term global solar irradiance forecasting with wrapper mutual information. *Renewable Energy*, *133*, 1055–1065.
62. Espinar, B., Ramirez, L., Drews, A., Bayer, H. G., Zarzalejo, L. F., Polo, J., & Martin, L. (2009). Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. *Solar Energy*, *83*(1), 118–125.
63. Bessa, R. J., Miranda, V., Botterud, A., & Wang, J. (2011). Good or bad wind power forecast: A relative concept. *International Journal on Wind Energy*, *14*(5), 625–636.
64. Hodge, B. M., Orwig, K., & Milligan, M. (2012). Examining information entropy approaches as wind power forecasting performance metrics. In *12th International Conference on Probabilistic Methods Applied to Power Systems*, Istanbul, Turkey (pp. 1–6).
65. Mills, A., & Wiser, R. (2010). *Implications of wide area geographic diversity for short-term variability of solar power*. Report on Environment Energy Technologies Division, Ernest Orlando Lawrence Berkeley National Laboratory.
66. Florita, A., Hodge, B. M., & Orwig, K. (2013). Identifying wind and solar ramping events. In *IEEE Green Technologies Conference* (pp. 147–152).
67. Caldas, M., & Alonso-Suarez, R. (2019). Very short term solar irradiance forecast using all-sky imaging and real time irradiance measurements. *Renewable Energy*, *143*, 1643–1658.
68. Chu, Y., Pedro, H. T. C., Li, M., & Coimbra, C. F. M. (2015). Real-time forecasting of solar irradiance ramps with smart image processing. *Solar Energy*, *114*, 91–104.