Received: 14 December 2022 | Revised: 12 February 2023 | Accepted: 3 April 2023

IET Renewable Power Generation

DOI: 10.1049/rpg2.12736

The Institution of Engineering and Technology WILEY

REVIEW

A taxonomy of short-term solar power forecasting: Classifications focused on climatic conditions and input data

Ioannis K. Bazionis[1] | Markos A. Kousounadis-Knousen[1] | Pavlos S. Georgilakis[1] | Elham Shirazi[2] | Dimitrios Soudris[1] | Francky Catthoor[3,4]

[1]School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Athens, Greece

[2]Faculty of Engineering Technology, University of Twente, Enschede, The Netherlands

[3]IMEC, Kapeldreef 75, Leuven, Belgium

[4]KULeuven, Kasteelpark Arenberg 10, Heverlee, Belgium

Correspondence
Pavlos S. Georgilakis, School of Electrical and Computer Engineering, National Technical University of Athens (NTUA), Athens, Greece.
Email: pgeorg@power.ece.ntua.gr

Abstract

A review of the state-of-the-art in short-term Solar Power Forecasting (SPF) methodologies is presented in this paper. Over the last few years, developing and improving solar forecasting models has been the main focus of researchers, considering the need to efficiently increase their forecasting accuracy. Forecasting models aim to be used as an efficient tool to help with the stability and control of energy systems and electricity markets. Intending to further comprehend the factors affecting the quality of SPF models, this paper focuses on short-term solar forecasting methodologies since they pose a crucial role in the daily operation and scheduling of power systems, since they focus on forecasting horizons typically ranging from 1 h to 1 day. The reviewed works are classified according to the climatic conditions, technical characteristics, and the forecasting errors of the different methodologies, providing readers with information over various different cases of SPF. Considering the need to improve the SPF efficiency, such classifications allow for important comparative conclusions to be drawn, depending on the location of each case and the meteorological data available. Future directions in the field of short-term solar power forecasting are proposed considering the increasing development of SPF models' architecture and their field of focus.

1 | INTRODUCTION

Two of the most important problems of the modern world are the continuous change in the climatic conditions leading to global warming, and the satisfaction of the constantly increasing global energy needs. Using Renewable Energy Sources (RES) efficiently has proven to be one of the solutions to those problems. Because RES provide a 'green' alternative to power generation for conventional sources of energy (coal, oil), it has been the main focus of researchers in recent years to develop methodologies and efficiently exploit them.

Solar power, thanks to sunlight being an abundant energy source, is one of the most exploited and most important renewable sources of energy [1]. Due to the continuous research and advance in the technology, solar power, via solar photovoltaic (PV) systems, plays a crucial role in the global energy system as well as the global energy markets. Therefore, such renewable sources tend to replace conventional energy resources in the power generation process.

Despite the global economic difficulties caused by Covid-19, the installed solar capacity in Europe was increased by 25.9 GW in 2021, 34% more than 2020, reaching the total capacity of 164.9 GW. On a global level, by the end of 2021, the total installed solar power capacity increased by 151 GW, reaching approximately 942 GW compared to the wind power onshore capacity that increased by 94.3 GW, meeting a 17% decline from 2020 [2]. China remains the country with the most installed solar capacity, followed by the United States, Japan, European countries (like Germany, Italy and France), India, Australia and South Korea [2].

Focusing on solar power as an important part of today's power generation systems, several problems may arise concerning how to properly include solar energy into energy systems. PV production is directly connected to the amount of the incident solar irradiation on photovoltaic panels. However, solar irradiation is limited by various factors [3]. The first factor is time itself, since solar energy can only be obtained during day times. Furthermore, the distribution of the solar irradiance cannot be uniform, not only on a global level but also on a local level. Moreover, PV generation depends on various meteorological factors, such as the temperature on the atmosphere and on the PV modules, the wind speed and direction, the humidity and cloud coverage.

As a result, dealing with all the above problems is of crucial importance in order to maintain the stability of the power grid and the energy markets [4].

In order to tackle those problems, various Solar Power Forecasting (SPF) models have been developed. Over the last few decades, numerous SPF methodologies have been researched, in order to facilitate the control of the continuous increase of the solar power penetration in the global energy systems. Furthermore, such forecasting models aid in the scheduling of power systems and in maintaining their stability and reliability in order to further enhance the control of electricity markets. Short-term SPF models have been in the centre of attention for most researchers considering their importance in the planning and operation of daily electricity markets and management of energy systems. Accurate SPF can reduce the need for backup generation, reduce electricity costs, and ensure that the power grid is operated in a reliable and efficient manner. By integrating short-term solar power forecasts into real-time adjustments to power generation and grid management, system operators can optimally schedule renewable energy resources, reduce the need for fossil fuel-based power plants to meet peak demand, and improve the reliability and stability of the power system.

SPF models developed for solar prediction have mainly focused on creating deterministic forecasting models. Such models have been proposed since the beginning of the penetration of solar power in power systems. These models are able to provide users with expected series of solar power output data by using specific parameters as input data. The outputs are presented as point values. Various deterministic SPF models use different methodologies over the same forecasting problems and thus provide users with different expected outputs, as well as different prediction errors of the forecasted output. Therefore, different methodologies focusing on similar problems can be qualitatively compared.

The work [5] reviewed SPF models, focusing on machine learning and metaheuristic methods, where it was found that hybrid models provided the most accurate predictive results. In [6], the main focus was the review of short-term direct SPF models based on historical data and how the forecasting accuracy is dependent on such data. Study [7] overviews the specifications needed in order to execute an accurate forecast. It further aimed to compare the definition of predictive horizons between solar and wind power forecasting and compare their performances. The work [8] reviewed different techniques used for solar power forecasting and presented important information needed for an accurate forecast. The vast focus on day-ahead forecasting along with the increasing use of NWP was further highlighted. Study [9] reviewed different solar forecasting techniques and identified input data that could potentially improve the forecasting accuracy, identifying solar irradiance, temperature, wind speed and direction, humidity, cloud cover, and aerosol index as the most important variables for SPF. In [10], a review of probabilistic solar power and load forecasting models was the main focus. It further aimed to identify similarities between solar power and load forecasting models. The work [11] focused on highlighting the use of probabilistic SPF into power systems, pointing out the need for further development evaluation and integration of probabilistic SPF into real-life cases. The work [12] presented a review of PV mathematical models and a case study that predicted the performance of a specific PV system. Study [13] presented an in-depth review in state-of-the art methodologies in terms of techniques and optimization. The study further highlighted the importance of the pre-processing of the forecasting process in order to construct the appropriate forecasting model depending on the case at hand, as well as the need of the evaluation process.

TABLE 1 Main focus of review papers on SPF

Review paper	Main focus	Year of publication
[5]	SPF focusing on machine learning and metaheuristic methods.	2019
[6]	Short-term SPF based on historical data.	2018
[7]	Specifications for accurate SPF and comparison between solar and wind power forecasting predictive horizons.	2017
[8]	Different techniques for solar power forecasting and identification of important input for accurate SPF.	2016
[9]	Different SPF techniques and estimation of input data that could improve their forecasting accuracy.	2016
[10]	Probabilistic solar power and load forecasting models and identification of their similarities.	2018
[11]	Integration of probabilistic forecasting methodologies into power systems.	2020
[12]	PV mathematical models and case study of a specific PV system.	2014
[13]	Analysis of the major concerns of SPF models in terms of focus, model architecture, and evaluation.	2020
This paper	Classification and estimation of the forecasting accuracy of SPF models based on the technical characteristics and climatic classification of the forecasting error of the reviewed models.	

The above studies [5–13], published between 2014 and 2020, focus on reviewing recent research efforts. As Table 1 shows, the majority of the review works on SPF focus on presenting state-of-the-art solar forecasting models and important input data for the improvement of the forecasting accuracy.

This paper reviews recent SPF methodologies, published from 2016 to date. As can be seen in Figure 1, the great majority

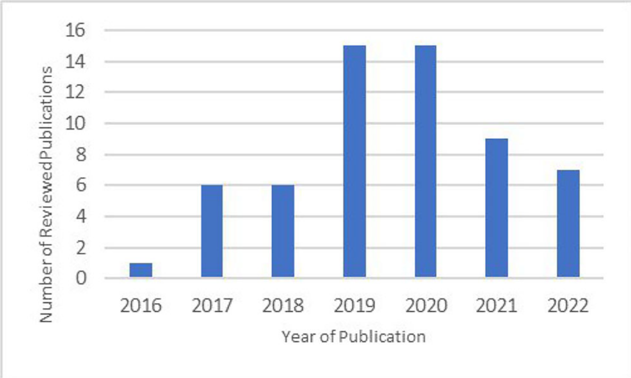

FIGURE 1 Number of works, reviewed in this paper, categorized by year of publication.

of the reviewed works has been published after 2018. Aiming to comprehend the different factors affecting the outcome of the prediction process, the review has further focused on the effect of the different climatic conditions and the geographical location of the reviewed datasets as well as technical characteristics of the forecasting models.

Therefore, the researches being reviewed in this work cover novel, state-of-the-art methodologies and present recent advances in short-term solar power forecasting from the last 5 years. The review process intended to cover a variety of different SPF methodologies and their behaviour towards the forecasting result, considering specifically their climatic conditions that derive from their geographical location, the type of input variables, the capacity of the solar parks, and the evaluation of each SPF methodology.

The contributions of this review paper are manifold:

a. It reviews studies published in the last 5 years, and, as a result, is provides an overview of state-of-the-art models and methodologies for short-term solar power forecasting.
b. It classifies the reviewed studies according to their forecasting accuracy based on climatic and geographical conditions. Given the great variety of meteorological conditions on a global level, datasets from PV farms in different locations could be affected by such conditions and thus could provide different forecasting results.
c. It classifies the reviewed works depending on the type of data used in order to define the most important data used in solar power forecasting as well as to highlight the dependence of the forecasting error values of specific data.
d. It serves as a guide to aid researchers in understanding and identifying novel, state-of-the-art models of SPF.
e. It provides future research directions and presents real solar forecasting challenges that need to be solved via further developing existing SPF models.

The structure of the paper is as follows. Section 2 reviews the state-of-the-art methodologies of short-term SPF. Section 3 presents the classification of the reviewed works based on the technical characteristics and the data type used, as well as a climatic classification of the reviewed models. Section 4 presents the SPF models' complexity and evaluation. Section 5 provides future research directions. Section 6 summarizes the main findings and concludes the paper.

2 | METHODS

2.1 | Pre-processing of multiple data by data fusion

Data fusion methodology aims to deal with multi-source data combination. During the pre-processing of the input variables on an SPF model, different raw data are fused in order to enhance their adaptation to the training process of the SPF model. For example, images from several sources (e.g. different sky imagers located at several locations near to the plant) can be merged into a combined image, which has richer and more accurate content. This data fusion and pre-processing will improve the image clarity and thus improve their understanding and estimation [14].

Data fusion can be a complex process and the input variables collected for the fusion process should be selected reasonable and according to the specific problem at hand in order to avoid data fusion failure.

The main objective of data fusion is to optimize the efficiency of multi-source input variables and after the pre-processing part, the enriched and optimized data are imported to another SPF methodology.

2.2 | Benchmark models

Such models are mainly used as a means of comparison for advanced forecasting models in order to prove their validity and efficiency. Such models are easy to construct and are fast thanks to their low computational cost.

2.2.1 | Persistence model

The persistence model is generally used as a default model that is compared to novel methodologies in order to evaluate their performance. In other words, the persistence model functions as a benchmark model for advanced models [6]. During the application of the persistence model, the predictive values of the historical time series data are calculated, considering that the conditions between time t and future time Δt do not change. For cases where the power output time series are non-stationary data, the persistence model is described as [15]

$$\hat{P}(t + \Delta t) = P(t) \tag{1}$$

where $P(t)$ is the expected power output in clear-sky and $\hat{P}(t + \Delta t)$ is the forecasted value. The persistence forecast accuracy decreases with the increase of the forecasting horizon [16].

2.2.2 | Climatology model

The climatology model provides a simple solution to estimation while lacking advanced tools. By using historical data and following the frequency of specific events, it offers the possibility of estimations of such events in the future.

A climatological forecasting model is typically based on historical data and more specifically the mean or average value of a specific variable. The average is calculated from a valid data sample. Assuming that the said data sample is representative of the total of the considered variable, the mean of the sample could represent the mean of the total of the historical data. A simple climatology model is described as

$$y_i = \bar{x} \tag{2}$$

where y_i is the forecasted value and $\bar{x}$ is the mean of the historical variable x_i and is described as

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{3}$$

The main disadvantage of this model is the fact that it is based on averaging specific climatic conditions for specific periods of time and thus it is not able to provide forecasting variations simply by using these averages [17].

2.3 | Statistical approaches

Statistical methodologies have been around for a long time already and they have interesting properties which have made them attractive for applications where probability distributions are relevant. Such models surpass the benchmark models. They are easier to interpret compared to other approaches, but they are more difficult to train due to their dependence on explanatory variables. We provide a survey of the main approaches here, without having the aim to be exhaustive.

2.3.1 | Auto regressive moving average (ARMA)

The ARMA model has been widely used in SPF as an efficient model in time series forecasting, thanks to its ability to extract useful statistical information [18]. The ARMA model consists of two elementary models, the AutoRegressive (AR) and the Moving Average (MA) models. The combination of the two models is described as

$$X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{4}$$

where X_t is the forecasted solar power at time t, p is the order of the AR model, q is the order of the MA model, φ_i is the ith AR coefficient, θ_j is the jth MA coefficient, and ε is white noise, which functions as an independent variable [19]. ARMA is expressed as ARMA(p, q), where p and q express the orders of AR and MA, respectively. This model is usually applied to auto-correlated time series data.

ARMA is a promising tool to understand and predict the future values of specific time series due to its ability to extract important statistical information during the prediction process. Another important advantage of the ARMA models is their flexibility. By using different orders of the AR and MA models, ARMA is able to constitute different types of time series. However, the necessity of stationary time series is a major disadvantage of the ARMA model [20]. The contributions of the works [21, 22] that propose ARMA-based SPF methodologies are presented in Table 2.

2.3.2 | ARIMA

The ARIMA functions as an extended version of the ARMA model presented in Section 2.2.1 with an added Integrated element. The ARIMA model is very efficient in processing nonstationary time series data [23]. The general equation of successive differences at the dth difference of the forecasted prediction X_t is

$$\Delta^d X_t = (1 - B)^d \ X_t \tag{5}$$

where d is the difference order and is usually 1 or 2, and B is the backshift operator and is defined by

$$BX_t = X_{t-1} \tag{6}$$

When the backshift operator is applied on a value X_t of the time series, it is able to shift the data back one period.

For example, following the previous equation, the successive difference at one time lag is

$$\Delta^1 X_t = (1 - B) X_t = X_t - X_{t-1} \tag{7}$$

The contributions of the works [24–26] that propose ARIMA-based SPF methodologies are presented in Table 2.

2.3.3 | Regression

The regression method is a statistical method that is used to provide information over the relationship between the dependent variables and the explanatory variables. The explanatory variables are essential in order to estimate the dependent variables. Considering solar power forecasting, PV power is the dependent variable, while the explanatory variables can be the meteorological parameters.

TABLE 2 Main contributions of reviewed works

Work	Publication date	Forecasting methodology	Main focus
[22]	July 2016	ARMA	A novel hybrid ARMA-based SPF model along with the wavelet transform was proposed in order to provide accurate solar power predictions.
[48]	September 2017	ANN	A non-parametric ANN-based forecasting model was proposed for accurate day-ahead PV forecasting.
[61]	September 2017	SVM	The evaluation of a deep learning SPF model via NWP data was the main objective of this work, applied to real life, high PV penetration scenario.
[30]	October 2017	ANN	A novel Long-Short Term Memory Recurrent NN (LSTM-RNN)-based model was proposed for PV forecasting, considering the temporal changes in PV power in the architecture of the forecasting model.
[67]	November 2017	ELM	An ELM-based methodology was proposed for accurate predictive results for three grid-connected plants.
[77]	November 2017	Image-based	A sky-imager-based forecasting model was proposed for the error estimation and evaluation of a nowcasting system.
[92]	December 2017	Regression	A quantile regression methodology was applied in order to provide information over the uncertainty of the predictive results.
[68]	June 2018	ELM	An ELM-based methodology was proposed for SPF, focusing on defining the proper weights for the parametrization of the forecasting model.
[88]	June 2018	ELM	An ELM-based methodology was proposed for probabilistic SPF, considering spatio-temporal data.
[47]	September 2018	ANN	ANN-based forecasting models were proposed for accurate hour-ahead solar power forecasting.
[62]	September 2018	SVM	The development and configuration of data-driven algorithms for PV power forecasting focused on individual sites was the main focus of the work.
[49]	December 2018	ANN	Three different NWP-free deep learning-based PV power forecasting models were proposed for accurate short-term SPF.
[27]	December 2018	Regression	A random regression forest model was proposed for the solar power forecasting process.
[31]	January 2019	ANN	ANN-based methodologies implementing different optimization algorithms were proposed in order to estimate the best solar power forecasting accuracy.
[32]	January 2019	ANN	Three different solar power forecasting models were proposed to estimate and compare PV power outputs.
[25]	January 2019	ARIMA	ARIMA and SARIMA SPF models were proposed to provide accurate solar power generation predictions.
[90]	June 2019	ANN	An LSTM-based model was proposed for probabilistic SPF along with providing seasonal information.
[33]	July 2019	ANN	An RNN-based model was proposed for the PV forecasting process in order to estimate the correlation between adjacent days and estimate non-linear information of inter- and intra-day power data.
[34]	July 2019	ANN	An adaptive predictor subset selection model was presented in order to efficiently estimate accurate predictive forecasts.
[59]	September 2019	SVM	A hybrid ensemble power forecasting model was proposed for the prediction of changing weather patterns.
[82]	September 2019	Hybrid	The work proposed a hybrid data-driven, recursive arithmetic average methodology in order to improve the forecasting accuracy via improving the weights of the individual models.
[26]	September 2019	ARIMA	A hybrid forecasting SARIMA-based model was introduced in order to provide information over the effect of the Wavelet Decomposition in solar power forecasting accuracy.
[44]	September 2019	ANN	An LSTM-based ensemble algorithm was proposed to improve the forecasting accuracy via combining the weights of the different implemented LSTM models and by maintaining non-linear statistical information of the time series database.
[59]	September 2019	SVM	An SVM-based model was proposed, focusing on the solar power forecasting based on the changing of weather patterns.
[91]	October 2019	ANN	An improved Markov chain-based methodology was proposed in order to provide accurate predictions of solar power. A rough ser theory was further applied in order to optimize the input variables.
[35]	November 2019	ANN	Different LSTM models were proposed along with the discrete grey model in order to estimate the solar power forecasting accuracy.
[36]	November 2019	ANN	A convolutional NN (CNN)-based solar power forecasting model was proposed, using spatial historical information as input for the prediction process.

(Continues)

TABLE 2 (Continued)

Work	Publication date	Forecasting methodology	Main focus
[87]	November 2019	ANN	A regime-switching process was introduced for solar power forecasting along with the computed cloud regime.
[83]	February 2020	Hybrid	A hybrid deep-learning forecasting model was proposed for one-hour solar power forecasting in order to improve the predictive accuracy.
[37]	March 2020	ANN	A novel solar power forecasting model based on self-attention mechanism was introduced for ultra-short-term forecasting.
[38]	April 2020	ANN	The proposed methodology used a fine-tuned CNN avoiding a time-consuming trial and error process in order to achieve accurate day-ahead forecasts.
[21]	May 2020	ARMA	A BI-LSTM-based model was proposed in order to estimate accurate solar power forecasts for large PV parks.
[39]	June 2020	ANN	Two novel CNNs along with a new data pre-processing process were proposed for PV power forecasting.
[85]	July 2020	Hybrid	A hybrid regression-based model was proposed in order to improve the solar power forecasting accuracy.
[40]	July 2020	ANN	A single dendritic ANN was proposed for the forecasting process along with the wavelet transform for the data decomposition.
[28]	August 2020	Regression	The random forest algorithm was proposed in order to estimate input data of higher importance and a grey ideal value algorithm was used for the optimization process.
[41]	August 2020	ANN	An ANN-based methodology was proposed in order to improve the PV power forecasting accuracy.
[60]	September 2020	SVM	Estimating accurate day-ahead regional PV predictions in a high PV power penetration scenario was the main objective of this work.
[75]	September 2020	Image-based	Accurate SPF results using satellite pictures were the main focus of this work.
[42]	September 2020	ANN	A two-step methodology based on the persistence model and LSTM was proposed to improve the PV forecasting accuracy.
[84]	September 2020	Hybrid	A hybrid PV forecasting model was proposed considering the prior data of adjacent days when constructing the prediction model.
[24]	November 2020	ARIMA	An SARIMA-based SPF model was proposed in order to improve the solar forecasting accuracy.
[72]	July 2021	ANN	A deep reinforcement learning approach was introduced, in order to make the predictive errors of the forecast more compensable.
[50]	February 2021	ANN	A novel AlexNet-based SPF methodology based on CNN structure was proposed in order to provide robust and accurate forecasts.
[43]	September 2021	ANN	An LSTM-based forecasting model was proposed for accurate power forecasts, which was applied in different scenarios in order to estimate the accuracy's dependence on different variables.
[45]	September 2021	ANN	A data-driven ensemble methodology was proposed for accurate hour-ahead PV power forecasting.
[46]	September 2021	ANN	Three different SPF models were applied and evaluated for two different cases in order to produce accurate short-term forecasts.
[51]	September 2021	ANN	An LSTM-based forecasting model was proposed, along with a robust local mean decomposition that explored different hidden properties of solar irradiance and allowed for the appropriate parameters to be included as input, in order to improve the forecasting accuracy.
[63]	September 2021	SVM	Five different day-ahead PV forecasting models based on NWP were proposed and compared for accurate SPF.
[74]	September 2021	Image-based	A segmentation algorithm was introduced for cloud detection and cloud classification via camera images, which was later applied for solar forecasting.
[76]	September 2021	Image-based	Data pre-processing via curtailment detection methods as well as novel machine learning models were proposed for the optimization of PV forecasts.
[52]	October 2021	ANN	A novel parallel pooling CNN-based SPF model was proposed in order to achieve robust and high accuracy.
[71]	January 2022	ANN	An automated reinforcement learning algorithm based on prioritized experience replay was proposed to support a multi-period single-step forecasting model for improving RES prediction accuracy.
[69]	January 2022	ELM	An ELM-based SPF model was proposed, in order to facilitate solar irradiance integration by reducing the forecasting error and thus providing higher forecasting accuracy.
[78]	January 2022	Image-based	The spatio-temporal correlation of different power plants was the main focus of this work, which developed an ultra-short-term SPF model based on the mapping relationship of the cloud characteristics of neighbouring power plants.

(Continues)

TABLE 2 (Continued)

Work	Publication date	Forecasting methodology	Main focus
[54]	February 2022	ANN	Two different feedforward NN-based models were proposed and compared along with different optimization methods in order to accurately forecast the photovoltaic power output.
[53]	March 2022	ANN	An LSTM-based forecasting model was proposed to achieve accurate SPF.
[55]	March 2022	ANN	A hybrid CNN-LSTM model was proposed in order to improve the SPF accuracy.
[56]	April 2022	ANN	A non-linear autoregressive with exogenous inputs model was proposed for accurate SPF and was further optimized via a corrective vector multiplier technique.
[29]	June 2022	Regression	A regression-based methodology, specifically focused on random forest models, followed by LASSO and Ridge regularizations was proposed in order to improve the forecasting accuracy.

The general form of a regression model is given by

$$\mathbf{Y} = f(\mathbf{X}, \beta) \tag{8}$$

where $\boldsymbol{Y}$ is the vector of the dependent variable, $\boldsymbol{X}$ is the matrix of the included independent explanatory variables and β is the regression model's parameters.

Various explanatory variables can be included to a regression model, creating a multiple regression model, which in case of a solar power forecasting regression model can be relevant to solar irradiance. However, simply selecting relevant explanatory variables to the dependent variable could lead to low accuracy and interpretability of the model. To deal with such problems, appropriate selection of explanatory variables is of core importance to improve the efficiency and accuracy of such models.

The contributions of the works [27–29] that propose regression-based SPF are presented in Table 2.

2.4 | Artificial intelligence

Artificial intelligence methodologies have become very popular recently. Such methodologies are easy to train thanks to the advanced technology and various open-source templates. On the other hand, due to their complexity, they can be extremely difficult to interpret. They provide better results and are more accurate compared to the benchmark models. We provide a survey of the main approaches here, again without having the aim to be exhaustive.

2.4.1 | Artificial neural networks (ANN)

Artificial Neural Network methodologies have been developed and researched over the past few years in order to achieve better forecasting accuracy in terms of error metric results. Such methodologies are able to deal with problems of non-stationary data, false data or even incomplete data. The main advantage of the ANN models is their continuous development as well as the flexibility they offer due to their advanced mechanics and the implemented optimization algorithms for their training. However, due to using large number of different types of data, such models are quite complex and difficult to construct in an efficient way.

The standard feed-forward ANN consists of three main layers, the input layer, the hidden layer(s) and the output layer. As its name implies, the input layer is given the input data that will be used for the procedure of the forecasting process. The hidden layer, through the appropriate training of the ANN, analyses the input information. It is possible that there is more than one hidden layer in an ANN model. The output layer gives the output predictive result based on the analysis of the hidden layers. The contributions of the works [30–56] that propose ANN-based SPF are presented in Table 2.

2.4.2 | Support vector machine (SVM)

SVM is a supervised learning technique used for classification by maximizing the separation margin among different classes [57]. It is used to enhance its generalization capability by reducing the empirical risk of the learning machine [58]. SVMs belong to the class of kernel methods. The kernel functions are key features of SVM, which map data into higher dimensional space. SVM has a basic principle of applying non-linear data mapping in some spaces and linear mapping in future space. The main advantage of the SVM is the possibility of determining an error acceptable for validation during the NN's learning process.

The use of an SVM for time series prediction can be expressed as

$$\hat{y}(t+h) = v^T \varphi(x_t) + b \tag{9}$$

where $b \in \mathrm{R}$ is a bias term, $v \in \mathrm{R}^{\mathrm{M}}$ is the weight vector and φ: $\mathrm{R}^{\mathrm{M \bullet D}} \to \mathrm{R}^{\mathrm{M}}$, ($\mathrm{M} \geq \mathrm{m}^{\bullet}\mathrm{D}$) is a non-linear feature map, which transforms the input vector $x_t \in \mathrm{R}^{\mathrm{M \bullet D}}$ to a higher-dimensional vector $\varphi(x_t) \in \mathrm{R}^{\mathrm{M}}$.

The commonly used kernel functions include linear, polynomial, and radial basis function kernels. The non-linear kernel function is defined as

$$k = \exp\left(-\frac{1}{\sigma^2}\|\mathrm{X} - \chi_i^2\|\right) \tag{10}$$

where $\mathbf{X}$ and X_i are the vectors in input space and the vector of features computed from training or test samples, respectively.

The contributions of the works [59–63] that propose SPF based on the SVM methodology are presented in Table 2.

2.4.3 | Extreme learning machine (ELM)

The learning speed of Feed Forward Artificial Neural Networks (FFANNs) decreases due to the slow training algorithms of the ANNs that slow down the learning process of FFANNs [64]. ELM was proposed in [65] in order to increase the computational speed of FFANNs. For a single hidden layer feedforward neural network (SLFNN), the selection procedure of the ELM for the hidden nodes is completely random and determines the SLFNN's output layer's weights. One of the main advantages of the ELM methodology is the simple training process of the NN [64].

Through avoiding specific limitations of conventional NNs, like overtraining and the local minima, it simplifies the training process. As a result, the computational cost of the ELM is greatly reduced and it provides higher generalization capabilities.

For a single hidden layer of ELM, the ith node is defined by

$$h_i(x) = G(a_i, b_i, x) \tag{11}$$

where a_i and b_i represent the parameters of the ith node. The output function of ELM for hidden layer L is

$$f_L(x) = \sum_{i=1}^{L} \beta_i h_i(x) \tag{12}$$

where β_i is the output of the ith hidden node $h(x)$, where

$$h(x) = \left[G(h_i(x) \ldots, h_L(x))\right] \tag{13}$$

is the hidden layer of output mapping of the ELM.

For a training sample N, the output matrix H is

$$H = \begin{bmatrix} h(x_1) \\ \vdots \\ h(x_N) \end{bmatrix} = \begin{bmatrix} G(a_i, b_i, x_1) & \cdots & G(a_i, b_i, x_1) \\ \vdots & \vdots & \vdots \\ G(a_i, b_i, x_N) & \cdots & G(a_i, b_i, x_N) \end{bmatrix} \tag{14}$$

and the training data T of the target matrix is

$$T = \begin{bmatrix} t_1 \\ \vdots \\ t_N \end{bmatrix} \tag{15}$$

The contributions of the works [66–68] that propose ELM-based SPF are presented in Table 2.

2.4.4 | Deep reinforcement learning (DRL)

By combining deep learning with reinforcement learning (RL), the resulting DRL methodologies offer solutions to the complex systems' perception and decision-making issues [69]. The main advantage of DRL is its ability to process large amounts of data, by taking into account complex relationships and dependences between its input variables. However, being a data-dependent approach, such models suffer from increased computational cost, due to the need of many layers and multiple iterations over the RL loop. Hence, this approach is difficult to apply online on an embedded processing platform. In addition, input variables validity can be another main challenge of models based on DRL. In [70], an automated reinforcement learning algorithm based on prioritized experience replay was proposed to support a multi-period single-step forecasting model for improving RES prediction accuracy. A deep reinforcement learning approach was introduced in [71], in order to make the predictive errors of the forecast more compensable. The contributions of the works that propose DRL-based SPF are presented in Table 2.

2.5 | Physical and image-based forecasting models

Clouds have a huge impact on solar surface irradiance and SPF. As a result, estimating clouds appearance at a specific time is of crucial importance in the forecasting process. Wind speed and wind direction affect the cloud motion and strongly influence the cloud structures [72]. Recently, via satellites and ground sky imagers, newly developed forecasting methodologies record the cloud movement. Further processing of sky images allows future cloud cover estimation, which in turn is used for SPF. Such methodologies can be very difficult to train due to the complexity of the data derived from the images as well as their processing. On the other hand, in terms of interpretation, they are really easy to understand and explain. They provide better results than the benchmark models thanks to the novel introduced improved models. The contributions of the studies [73–77] that propose image-based SPF are presented in Table 2.

2.5.1 | Satellite images

The use of satellite images requires the identification of specific stationary features of the clouds. Tracers within the images being processed are selected, corresponding to cloud features that can be easily traced. Convective clouds, which are characterized by low temperatures, large cloud deformations and transformations caused by the cloud motion as well as the cloud movement due to wind speed, are tracers typically being analysed by satellite images. The consistency of those features is being analysed throughout the pre-processing phase. In order to efficiently track those tracers, pattern matching algorithms are used to estimate the future location of each tracker from given images [78].

Satellite images and satellite-derived radiation data are often used in SPF models especially in numerical weather models. However, because of errors in aerosol optical depth and cloud detection problems, data from satellite images may have uncertainties and often show systematic bias. In addition, their resolution is typically in the km range, which is quite limited and hence they are not well suited for very-short-term predictions.

2.5.2 | Ground-based sky imagers

Ground-based sky imagers offer high spatio-temporal resolution. This is in contrast to the satellite images of Section 2.4.1. As a result, such images are suitable for very-short-term and short-term forecasting methodologies where high resolution is required. Especially for intra-hour predictive results, ground-based sky images are preferred, instead of satellite images and Numerical Weather Prediction (NWP) models. Furthermore, their high spatio-temporal resolution offers the ability to locate ramp events that satellite images are unable to estimate.

It should be noted that ground-based imaging, through cloud cover indices, cloud motion vectors and cloud classifications, often lacks in spatial coverage and is thus limited to being used for specific location forecasting applications [79].

2.5.3 | Numerical weather prediction (NWP) models

NWP models provide information of the atmospheric conditions through physical equations. NWP are able to produce forecasts for various types of input variables, i.e., solar irradiance. The forecasting process of NWP models, based on collecting historical data, generates the potential future conditions and error adjustment by processing the forecasts based on past performance [80]. Such models give better results for long-term forecasting than ultra-short and short-term forecasting.

2.6 | Hybrid methods

Hybrid methods refer to the combination of two or more existing methodologies for the forecasting process. The inclusion of more than one methodology aims to tackle existing disadvantages of each method. The above process results in the increase of the forecasting accuracy of the hybrid model and the reliability of the predictive results.

The advantage of the hybrid methods is the improvement of the forecasting accuracy via including the positive aspects of each method as well as an optimization algorithm. However, a major drawback of the hybrid models is the complexity of the final forecasting model. The computational cost, and the programming complexity of the hybrid models are issues that should be taken into account in order to efficiently improve the total forecasting accuracy.

It should be also noted that the forecasting accuracy of the hybrid methodologies highly depends on the accuracy of the individual models as well [6]. Therefore, if one of the included models has low performance and provides low accuracy, the final hybrid model will be affected as well. As a result, hybrid models constitute a trade-off between the optimization of the model's accuracy and model's function cost.

The contributions of the works [81–84] that propose hybrid SPF methodologies are presented in Table 2.

2.7 | Ensemble methods

Ensemble methods are popular in machine learning and are widely used in SPF. The ensemble methodology is based on the use of multiple predictors in order to achieve an aggregated result that is better than the other predictors [85], and thus combine the advantages of multiple models of the same type in order to provide more robust and accurate forecasts. Such models often use NWP as input variables. The pre-processing process of the input variables as well as the training process is really important for such models. The results derive from averaging the results of all the predictors. However, it should be noted that evaluating the data at hand is of vital importance. Maintaining each ensemble as a probability distribution could provide vital information considering the pre-processing state of the forecasting methodology. It could also prove to be a useful aid in the understanding and research of ramp events occurrence and effect on the forecasting process.

2.8 | Probabilistic forecasting

The great majority of real-life SPF problems have been dealt with deterministic power forecasting models. However, deterministic forecasting models lack the ability to deal with the uncertainty parameters of the prediction process.

While deterministic forecasts provide a single time-series view of the possible outcome of the prediction, probabilistic forecasts provide a wider image of the possible outcomes of a prediction, in a form of prediction intervals or distributions.

The contributions of the works [86–91] that focus on probabilistic SPF methodologies are presented in Table 2.

2.9 | Evaluation

The evaluation of the forecasting models is of major importance to the forecasting process. The evaluation operation provides a forecasting accuracy comparison between different methodologies as well as estimation over the efficiency of the forecasting models. Considering that solar power forecasting plays a vital role over the operation and stability of the power grid and the energy markets, the evaluation of existing forecasting models is of crucial importance.

Several evaluation metrics exist that allow the comparison between different models, highlight possible problems in the methodologies under comparison, and indicate possible improvements. Such error metrics are the following:

a. Mean Absolute Error (MAE)
b. Mean Absolute Percentage Error (MAPE)
c. Mean Relative Error (MRE)
d. Mean Bias Error (MBE)
e. Mean Square Error (MSE)
f. Root Mean Square Error (RMSE)
g. normalized Root Mean Square Error (nRMSE)

A more analytical demonstration of the above metrics can be found in the appendix. The above error metrics allow the evaluation of an SPF model as well as the comparison of different SPF models and highlight possible developments. In Table 3, the metrics used in the reviewed works are presented.

TABLE 3 Evaluation metrics used in the reviewed works

Reviewed work	Evaluation metrics
[27]	RMSE, nRMSE, R^2
[28]	RMSE, MAPE, MAE, R^2
[29]	MAE, MRE, R^2
[30]	RMSE
[31]	MAE, MAPE, R^2
[32]	MAE, RMSE
[33]	MAE, MAPE, RMSE
[34]	MAE, MAPE
[35]	RMSE
[36]	nRMSE, MAPE
[37]	MSE, RMSE, MAE
[38]	MAPE, MRE
[39]	MAE, MSE
[40]	RMSE, MAE, MAPE
[41]	RMSE, MAE
[42]	nRMSE, nMAE
[43]	MAE, nRMSE, R^2
[44]	MAE, RMSE
[45]	MAPE, nRMSE
[46]	nRMSE
[47]	MAE, MAPE, RMSE, nRMSE
[48]	nRMSE
[49]	MAE
[50]	RMSE, MAE, MAPE
[51]	MAE, RMSE
[52]	MAE, RMSE
[53]	RMSE, MAE, MAPE, R^2
[54]	nRMSE
[55]	RMSE, MAE, MAPE, R^2
[56]	MSE, RMSE, R^2
[59]	RMSE, MAPE
[60]	RMSE, MAE
[61]	nRMSE, nMAE
[62]	RMSE, MAE, MBE
[63]	MAE, RMSE
[68]	RMSE, MAE, MAPE
[74]	nRMSE
[75]	RMSE
[76]	MAE, RMSE
[77]	RMSE, MAE
[81]	RMSE, MAE, MAPE
[82]	MBE, MAPE, RMSE
[83]	MAE, RMSE, R^2
[84]	MAE, MSE, RMSE, R^2

3 | METHODOLOGIES AND DATA ANALYSIS

3.1 | Forecasting horizon

The continuous penetration of solar power into energy systems has created major problems in terms of stability and control. SPF has been widely used in order to tackle these problems and achieve the stability of energy markets and power systems. It should be noted that different types of problems arise, depending on the forecasting horizon of the models. Such problems vary from electricity markets optimization, management of power systems, daily planning, energy trading and maintenance scheduling.

This work is focused on short-term SPF methodologies which, as can be seen in Table 4, can be further sub-classified into ultra-short-term and short-term SPF models. Table 4 provides a better view of each reviewed work's methodology via classifying them according to the forecasting horizon.

3.1.1 | Ultra-short-term forecasting

Ultra-short-term forecasting or nowcasting uses solar data to estimate photovoltaic power predictive outputs with a forecasting horizon typically ranging from a few seconds to 30 min [92]. Reviewed works [28, 33–40, 42, 46, 49, 71, 73, 76, 83, 84] propose different models and methodologies for ultra-short-term SPF.

3.1.2 | Short-term forecasting

Short-term forecasting provides predictive outputs related to the daily electricity market and system management in terms

TABLE 4 Classification of the reviewed works based on the forecasting horizon

Forecasting horizon	Reviewed works	Type of forecasting horizon
1 min	[49]	Ultra-short-term
5 min	[28, 84]	Ultra-short-term
10 min	[31, 73, 84]	Ultra-short-term
15 min	[28, 33, 40, 42, 46, 76, 83]	Ultra-short-term
30 min	[33, 42, 73, 84]	Ultra-short-term
1 h	[30, 33, 42, 43, 45, 47, 50, 59, 61, 71, 81, 82, 84]	Short-term
2 h	[30, 42, 43, 45, 50, 61, 75, 84]	Short-term
5 h	[30, 42, 43, 45, 61, 84]	Short-term
6 h	[27, 30, 36, 42, 43, 61, 84]	Short-term
16 h	[30, 42, 43, 60, 61, 84]	Short-term
24 h	[30, 32, 34, 38, 39, 41, 42, 43, 44, 48, 61, 62, 63, 84]	Short-term
48 h	[35]	Short-term

of planning and operation. The forecasting horizon typically ranges from 1 h to 1 day [93]. Reviewed works [30, 32–36, 38, 39, 41–45, 47, 48, 59–63, 71, 75, 81, 82, 84] focus on short-term SPF by using different models and methodologies.

3.2 | Climatic conditions

Meteorological conditions play an important role in estimating the predictive error of SPF models. Cloud coverage, wind speed and direction, precipitation or ambient temperature are all climatic aspects that vary based on the location.

In countries closer to the earth's equator higher temperatures are encountered, while in countries closer to the poles, the temperature can decrease even below 0°C. In tropical climates, higher volumes of precipitation are encountered, while in dry climates rainfall as well as wind speed may be limited to extremely low levels.

In Figure 2, a classification of the Earth's climate zones is presented [94] based on the Köppen–Geiger system. It can be seen that several climatic classifications and subclassifications exist based on the location and meteorological conditions.

It should be noted that the Köppen–Geiger system was not intended for PV. Instead, it was proposed in order to present the climatic conditions as experienced by humans and other living creatures across the globe through deep research of the climatic conditions of the different locations. As a result, a complex but life-oriented climatic classification system was proposed.

On the other hand, it was decided that proposing a climatic classification from scratch should not be an appropriate approach. Instead, it should be better to construct and propose a climatic classification focused on PV, based on an existing and widely acknowledged climatic model.

As a result, taking into consideration the importance of the climatic aspect to the SPF process, the Köppen–Geiger climatic subclassifications were clustered, split and combined anew into the climatic zones presented below:

a. Equatorial Zone: this zone is limited to specific zones around the Earth's equator. This zone receives the highest level of solar irradiance since the surface is perpendicular to the incoming sunlight. Those regions are characterized by high humidity levels and high, constant temperatures throughout the year. Rain is an almost daily phenomenon. However, there can exist distinctive drought periods.
   - Humid subequatorial zone, which is characterized by high humidity levels and regular rainfall.
   - Dry subequatorial zone, which is characterized by dry conditions.

From Figure 2, subsets *Af* and *Am* could be included in this category that represent humid climates with extreme conditions, such as rainforests and monsoons as well as *Aw*, representing savanna conditions.

From Figure 2, subsets *BWh* and *BWk* could be included in this category, representing dessert hot and dessert cold climates, respectively.

b. Temperate Zone: compared to the tropical and subtropical zones, the temperatures are lower and more stable, since solar radiation reaches the Earth's surface with a smaller angle and the level of sun irradiance is significantly lower than climate zones closer to the Earth's equator. Extreme climatic events are less frequent and precipitation is regularly distributed in specific seasons. However, due to the fact that the temperate zone covers vast areas in different countries and different continents, different subclassifications of the temperate zone's climate can be encountered, from partly rainy to relatively dry.
   - Humid subtemperate zone, that is mainly characterized by rainfalls and generally humid meteorological conditions.
   - Dry subtemperate zone, that is mainly characterized by dry conditions with rare rainfall presence.

From Figure 2, all subsets from *Csa* to *Cfc* could be included in this category, representing temperate climate conditions from ones with dry hot summers (*Csa*) to ones with cold summers with no dry conditions (*Cfc*). More specifically, *Csa*, *Csb*, *Cwa*, *Cwb*, *Cwc* represent dry conditions and are included in the dry subtemperate zone, while *Cfa*, *Cfb* and *Cfc* represent hot to cold summer zones, without dry conditions and are included in the humid subtemperate zone.

c. Subpolar Zone: this zone refers to regions with long and very cold winters and low temperature summers. The cold winds from this zone typically affect the climate of temperate zone regions during winter.

From Figure 2, all subsets from *Dsa* to *Dfd* could be included in this category, representing subpolar climatic conditions from

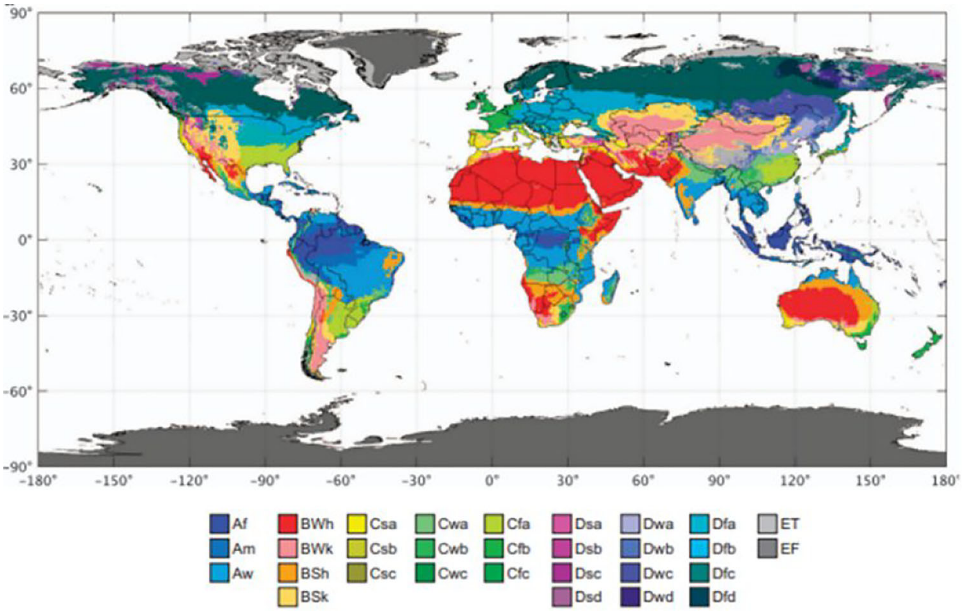

FIGURE 2 Köppen–Geiger global climatic zones [94].

ones with dry summers (*Dsa*, *Dsb*, *Dsc*, *Dsd*), to ones with dry climate throughout the year (*Dwa*, *Dwb*, *Dwc*, *Dwd*), to ones with humid conditions (*Dfa*, *Dsb*, *Dsc*, *Dsd*).

d. Polar Zone: regions in polar zones receive the less solar radiation intensity levels throughout the year. The daylight length varies in those regions. It is characterized by extremely low temperatures and the level of solar irradiance is the lowest possible due to the small angle the sunlight reaches the Earth's surface.

From Figure 2, subsets *ET* and *EF* could be included in this category, representing polar climates with tundra and frost conditions, respectively.

Considering the different climatic conditions, datasets from solar parks located in different countries, in different continents, could affect the forecasting error and thus the forecasting accuracy of SPF models. A classification of the reviewed works concerning the location of the input data used in their proposed model can be found in Table 5 as well as Figure 3 which graphically presents the locations of Table 5.

Tables 5 and 6 present classifications based on the climatic zone on a global level and for European countries. Considering known locations of solar farms on a global level [96], the locations reviewed in our paper align with them and thus are logically included in this review paper. As a result, in order to present the results on a smaller scale, European countries were selected as an example due to the fact that they present a variety in terms of climatic zone appearance.

To further provide information over the climatic conditions' effect on the SPF results, Figures 4 and 5 provide further information over the European continent.

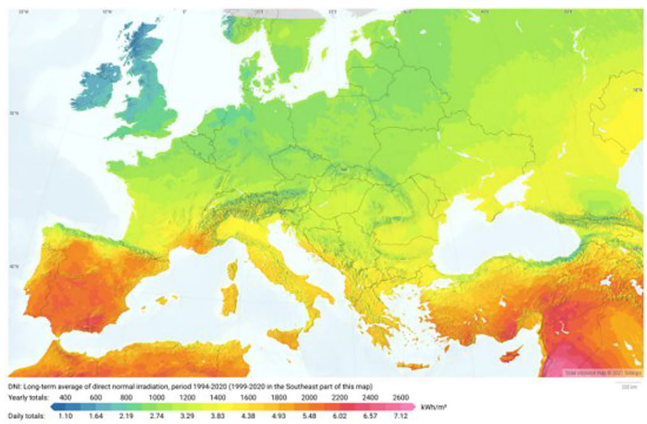

FIGURE 3 Location of the datasets of the reviewed works.

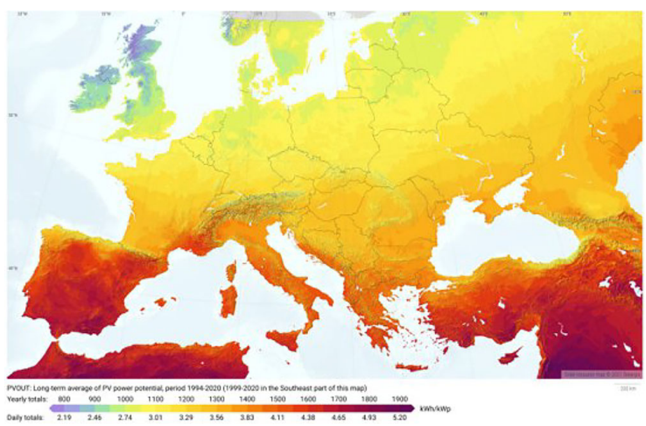

FIGURE 4 European direct normal irradiation map (2021) [96].

TABLE 5 Classification of the reviewed works based on the geographical location of their datasets

Work	Location	City	Climate zone
[27]	Portugal	Faro	Temperate (dry)
[28]	Canada	Desert Gardens	Temperate (humid)
[29]	Taiwan	Zhangbin	Equatorial (humid)
[30]	Egypt	Aswan, Cairo	Tropical
[31]	Turkey	–	Subtropical
[32]	South Korea	Gumi	Temperate (humid)
[33]	Belgium	Flanders	Temperate (humid)
[34]	Finland	Espoo	Temperate (humid)
[35]	China	Shandong	Temperate (humid)
[36]	USA	California, New York, Alabama	Tropical, Temperate (Humid), Subtropical
[37]	North America	–	Subtropical
[38]	Taiwan (south)	–	Equatorial (humid)
[39]	Australia	Alice Springs	Tropical
[40]	China	Gansu	Temperate (dry)
[41]	Jordan	Amman	Tropical
[42]	Italy	Cupertino, Catania	Subtropical
[43]	Greece (northern)	–	Temperate (humid)
[44]	Australia	Queensland	Equatorial (dry)
[46]	Switzerland	Bern, Batterkinden	Temperate (humid)
[47]	Cyprus	Nicosia	Subtropical
[48]	Cyprus	Nicosia	Subtropical
[49]	Japan	Kyoto	Temperate (humid)
[50]	Turkey	Kilis	Subtropical
[51]	Vietnam	–	Equatorial (humid)
[52]	Australia	Alice Springs	Tropical
[53]	Malaysia	Kuala Lumpur	Equatorial (humid)
[54]	Saudi Arabia	Shaqra City	Subtropical
[55]	Abu Dhabi	Sweihan	Subtropical
[56]	Mexico	Temixco	Temperate (humid)
[59]	Australia	Victoria	Subtropical
[60]	Japan	Kyushu	Temperate (humid)
[61]	Japan	Kyushu	Temperate (humid)
[63]	Germany (southern)	–	Temperate (humid)
[68]	Australia	Alice Springs	Tropical
[70]	China	North China	Temperate (dry)
[71]	Belgium	–	Temperate (humid)
[74]	Portugal	–	Subtropical
[75]	China	Qinghai	Temperate (dry)
[77]	China	Jilin Province	Temperate (dry)
[81]	Australia	Alice Springs	Tropical
[82]	Australia	Alice Springs	Tropical

(Continues)

TABLE 5 (Continued)

Work	Location	City	Climate zone
[83]	Austria	Limburg	Temperate (humid)
[86]	USA	California	Tropical
[87]	France	–	Subtropical
[88]	USA	Texas	Subtropical
[89]	Taiwan	–	Equatorial (humid)
[90]	Australia	Alice Springs	Tropical
[91]	Belgium	Flanders, Limburg	Temperate (humid)

TABLE 6 Classification of the reviewed works based on the climatic zone for European countries

Forecasting horizon	Reviewed works
Subtropical	[42, 47, 48, 74, 87]
Temperate dry	[27, 83]
Temperate humid	[33, 34, 43, 46, 63, 71, 91]

FIGURE 5 European photovoltaic power production map (2021) [96].

Figure 4 presents a map of the Direct Normal Irradiance (DNI) over Europe. DNI refers to the amount of solar radiation being received per unit area of a surface in a continuously vertical way from the Sun. Figure 5 presents the photovoltaic power production capabilities of European countries. It can be seen from the two figures that the levels of DNI are consistent with the amount of photovoltaic power production. These figures provide a detailed image of the distribution of solar irradiance in Europe and thus clarify the more promising regions for solar power production. As a result, more advanced SPF models are needed in such regions and it is more likely for researches to be focused on datasets of specific locations.

It can be derived from Figures 4 and 5, as well as Table 6, that the majority of countries in need of most accurate SPF models are countries of central and southern Europe that present the

TABLE 7 Classification of the reviewed works based on the used input data

General data type	Reviewed work
Historical-meteorological data	[21, 22, 24–26, 28, 29, 30–34, 36–38, 40, 41, 43–45, 47, 49, 51–56, 59, 60, 67, 68, 70, 71, 73, 76, 77, 81–84, 86, 89–91]
NWP	[27, 35, 39, 42, 46, 48, 50, 61–63, 66, 74, 75, 87, 88]

TABLE 8 Classification of the reviewed works based on the used input data

Data type	Reviewed works
Wind speed	[28, 35, 36, 39, 41, 42, 47, 48, 50, 54, 81, 82, 87, 90]
Wind direction	[28, 47, 48, 81, 87, 90]
Historical power output	[27, 36, 38, 40, 47, 59, 82, 83, 89]
Relative humidity	[28, 32, 35, 39, 41, 42, 47, 48, 52, 81, 82, 87, 90]
Ambient temperature	[28, 31, 32, 34, 35, 38–42, 47, 48, 50, 52, 53, 54, 59, 62, 82, 87, 89, 90]
Global solar radiation	[27, 28, 29, 31, 32, 34–36, 39–42, 48, 50, 52, 53, 54, 59, 81, 82, 86, 89]
Zenith angle	[42, 48, 62]
Cloud coverage	[32, 34, 35, 38, 42, 77]
Snow depth	[34]
Sunshine duration	[34]
Panel temperature	[27, 31, 40, 53]
Precipitation	[28, 38, 39, 87]
Radiation	[32, 47, 48, 62]

higher levels of radiation. These countries have higher potential of photovoltaic power production. This is connected to their climatic conditions. Due to the abundance of solar irradiance in such climates, those countries keep implementing solar power more and more into their power systems. It should be noted that stable meteorological conditions favour the photovoltaic efficiency. Continuous sunlight incidence, clear sky and normal temperature conditions could optimize the photovoltaic efficiency as well as aid in developing has proven to be the common solution for efficient PV production. This results in limited cases being focused in regions of said climatic zones and thus representative studies for these climates are not found in Table 6.

3.3 | Data type—data dependency

A major aspect affecting the accuracy of solar power forecasting models is the selection of proper input data. Different SPF models consider different types of input data. Table 7 presents the type of input variables selected in each of the reviewed works in terms of historical data and NWP. SPF models have been developed over the years and novel methodologies keep improving their forecasting accuracy.

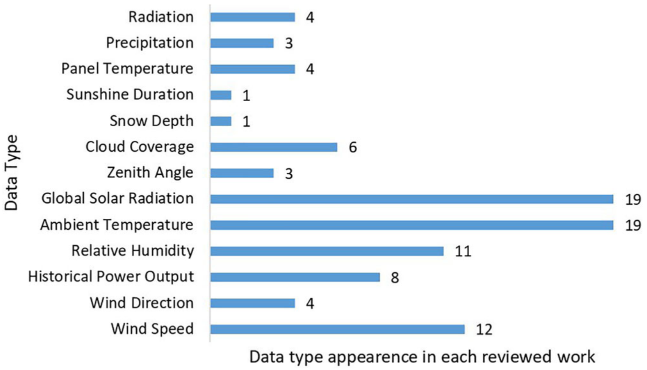

FIGURE 6 Simplified image of Table 8.

By their nature, SPF models are highly dependent on the meteorological and climatic conditions. Historical data of solar power and meteorological data have been used in SPF. Meteorological data like wind speed, temperature and solar radiation have been widely used in various SPF models. Recently, image-based models, taking into account the cloud coverage conditions, have been proposed over the past years and are still being researched and developed. Developing such models has helped in including cloud coverage as an important parameter in the forecasting process and further improving the forecasting accuracy of SPF models.

A classification of the reviewed works concerning the input data used in their proposed model can be found in Table 8 as well as in Figure 6.

3.4 | Size of input data

A crucial aspect in the development of the forecasting models, as well as their efficiency and accuracy, is the number of input data used. Apart from selecting the appropriate data for each SPF case, the number of the selected data could also affect the forecasting process. Estimating the appropriate size of input data in each case could prove to be decisive in improving the accuracy of the proposed forecasting model. On the other hand, the computational burden is an aspect that should always be taken into account. Excessive use of input data could greatly increase the computational cost of the forecasting process and as a result decrease the proposed model's forecasting value.

A classification of the reviewed cases based on the number of data used for their proposed model is presented in Table 9.

3.5 | Capacity of PV farms

In order to further research the correlation between the climatic location and forecasting error as well as the predictive

TABLE 9 Classification of the reviewed works based on the size of input data

Reviewed work	Data resolution	Number of datasets	Number of data per dataset
[61]	1 h	8	210,240
[49]	1 min	1	129,600
[27]	15 min	6	630,720
[32]	1 h	1	16,820
[33]	15 min	2	3072/2976
[81]	5 min	1	165,431
[44]		1	14,620
[59]	1 h	1	6672
[35]	15 min	1	34,560
[82]	5 min	1	210,816
[38]	1 h	1	8760
[84]		1	24,000
[40]	15 min	1	35,040
[28]	5 min	1	105,120
[60]	1 h	1	17,280
[71]	15 min	1	
[74]	15 min	1	35,040
[42]	15 min	2	46,080/34,560
[88]	5 min	1	74,880
[83]	15 min	1	34,560
[46]	15 min	2	69,120/69,120
[75]	15 min	1	40,320
[68]		1	16,368
[89]	1 min	1	
[51]	30 min	4	
[56]	1 h	1	52,560
[54]		2	3566
[53]	5 min	1	2,102,400
[86]	15 min	1	70,080
[87]	15 min	1	48,960
[90]	15 min/60 min	2	105,120/17,520
[91]	15 min	1	35,040

accuracy of SPF models, further enrichment of Table 4 is needed in terms of solar park capacity and SPF model evaluation. The benefit of classifying the solar parks based on their capacity could be twofold. Firstly, it could provide information over solar parks of the same level and thus allow the comparison of similar cases. Secondly, it could prove the correlation of climatic location and forecasting error in smaller and larger photovoltaic installations.

A classification of the reviewed cases based on the number of data used for their proposed model is presented in Tables 5 and 6 for solar parks of smaller and higher capacity respectively.

TABLE 10 Classification of the reviewed works based on the solar park's capacity in kW

Reviewed work	Capacity of solar park (kW)
[34]	4.3
[38]	500
[39]	5
[41]	264
[42]	6.41 (Cupertino), 5.21 (Catania)
[43]	10
[49]	2.5
[52]	23.4
[53]	2, 1.875, 2.7
[59]	3
[62]	662
[82]	26.5
[86]	3.74
[87]	58

4 | CLASSIFICATION

4.1 | Relevant factors for SPF

The reviewed works [18–84] focus on providing accurate short-term SPF models. Those models take into account specific factors in order to achieve accurate forecasts. Such factors are

a. The forecasting horizon, which depends on the problem each reviewed work aims to tackle.
b. The location of the datasets used in each case, which highly affects the type of data, their quality and their complexity.
c. The range of the input data as well as the data resolution.
d. The capacity of the solar parks.
e. The error metrics used for the evaluation process.

Tables 6, 7, 8, 9, 10, and 11 aimed to classify factors *a*–*d*, while Table 3 classifies factor *e*. By combining the information of Table 2 with Tables 6–11, this paper can be efficiently used as a useful guide for the reader through the literature. Possible examples could be the following:

- If the reader is interested in SPF models with the following features: i) focus on ultra-short-term forecasting of 15-min forecasting resolution, ii) focus on subtropical climates, as derived from Tables 4 and 5, the reader should study works [40, 42], and [83].
- If the reader is interested in SPF models with the following features: i) focus on high-capacity solar parks, ii) focus on short-term SPF, as can be seen in Tables 4 and 11, the reader should study works [31, 33, 35, 61, 63, 74, 75], and [83].

TABLE 11 Classification of the reviewed works based on the solar park's capacity in MW

Reviewed work	Capacity of solar park (MW)
[29]	2
[31]	1
[33]	2140
[35]	10
[61]	15,695
[50]	1
[63]	1
[71]	100
[74]	10
[75]	800
[83]	451.82
[88]	100

4.2 | Classification analysis

SPF keeps developing with the advance in modern methodologies. However, the more complex the novel forecasting systems are, the more important becomes their proper parameterization. Understanding the problem at hand, use of the appropriate data, integration of appropriate optimization algorithms as well as selecting the appropriate evaluation metrics are important in order to improve the accuracy of the SPF models.

SPF models aim to solve real-life problems in order to facilitate the solar PV power penetration into energy systems. Each of the reviewed studies is focused on a specific problem, developed over specific datasets, concerning specific meteorological conditions. Moreover, considering the forecasting horizon, different problems may arise. However, it should be noted that real-life SPF problems differ from research-focused SPF cases. While in many research cases, the datasets are optimized and the models are configured to process these datasets, in real-life problems, this is not always the case. Data quality and quantity, forecasting models' architecture, and adaptation to different cases are some of the major challenges to deal with real-life cases. The above result in some SPF models dealing with more complex problems than others and thus require more specific parametrization and accurate optimization.

As can be seen in Table 9, works like [61] use a greater number of data in order to provide accurate forecasts, compared to work [59], which requires a smaller number of data. Larger datasets tend to require more pre-processing stages as well as more time-consuming processing and training. As a result, SPF models dealing with greater amounts of data could be more complex with increased computational cost. Furthermore, as presented in Table 8, cases [28, 35, 36, 39, 41, 42, 81], and [82] use a larger number of input variables than other reviewed works. In should be noted that in each SPF case, the appropriate types of data should be used based on the type of the forecasting problem as well as the location. While using more data could improve the forecasting accuracy, there are cases where using more than the appropriate data could lead to higher computational cost without any significant improvement in the forecasting error. Evaluating the ideal number of input variables as well as optimized pre-processing of the input variables is of core importance. Such procedures could further aid in the parametrization of the forecasting models, such as estimating the ideal number of hidden neurons in cases of ANN-based SPF models, which could improve significantly the forecasting error.

Several conclusions can be drawn from Tables 4 and 5 concerning SPF. Either on a global or on a continental level, the forecasting accuracy of those models is of crucial importance. Considering the complexity of datasets representing non-stable climatic conditions, SPF models focused on subtropical and temperate zones could be more difficult to develop and optimize. Cloud coverage plays a vital role in such conditions. Apart from the appearance of clouds itself, factors like their optical depth, thickness, position, movement, or humidity should also be taken into account in an SPF model. Different wind patterns and wind speed further complicate the forecasting process. This could result in less accurate models and higher forecasting errors. However, thanks to the continuous advance in SPF modelling and the novel optimization algorithms, such problems have been overcome. In northern climates, under normal conditions, PV power forecasting is easier due to normal levels of radiation and temperature. SPF input datasets are less complex and usually follow specific meteorological patterns. While this allows for easy forecasting estimations, SPF models focused on such datasets may not be as complex, or may have been developed for specific problems and specific conditions with no possibility of easily adapting to different climatic conditions, not involving possible ramp events or random extreme conditions.

5 | DISCUSSION AND FUTURE RESEARCH

5.1 | Improvement of the accuracy of existing models

Over the last years, SPF has been in the centre of attention for many researchers. Various SPF forecasting methodologies have been developed, as presented in Section 2. Thanks to the technological advance, those models have been further developed in recent years and have managed to reduce considerably the predictive error of the forecasting process. It should be also taken into account that the predictive error grows significantly with the bigger forecasting horizons. Several aspects are required to identify a well-constructed SPF model. More specifically, accuracy, efficiency, versatility, and computational cost are basic characteristics that determine an appropriate SPF model. Many times, a trade-off between the accuracy and the other aspects is required in order to produce accurate SPF models. Thanks to artificial intelligence and deep learning models, such trade-offs can be avoided.

Further improving the accuracy of existing SPF models is of crucial importance [29, 32, 41]. Focusing on decreasing the predictive error via using comparative error metrics could be achieved by further optimizing existing models. Thanks to such error metrics, a measurable comparison of different forecasting models is possible and thus further decreasing the forecasting error of the models giving worse results can be more focused on specific solutions.

Neural network-based models have been proposed, more and more, over the years, improving the forecasting error of conventional SPF models. Furthermore, novel hybrid models could prove to be more efficient in dealing with data fluctuations as well as further improve individual models' efficiency [37]. Moreover, integration of novel optimization algorithms as well as further data analysis and data pre-processing could further improve the forecasting accuracy of SPF models [52, 53].

It should also be noted that forecasting accuracy could be further enhanced and validated, not only through data analysis and software development, but also via improvement of the hardware features of PV power production. Over the last few years, innovative parts have been included in PV modules in order to improve the data recording process. Such parts include cheap distributed temperature and wind speed monitors that allow the in situ 'capture' of local data accompanied by competitive costs, such innovations should be further included and developed since they could facilitate the pre-processing phase of SPF and thus make the forecasting process less complex.

5.2 | Cloud coverage classification

Recently, thanks to the development in technology and the forecasting models, cloud coverage is being considered and implemented as input into forecasting models. Various algorithms have been proposed for cloud classification and have been effectively used in order to improve the forecasting accuracy. Furthermore, thanks to the implementation of sky imagers, there has been further advance in estimating the fluctuation of solar irradiance on the surface during cloudy weather conditions. However, novel algorithms, constructed for the calculation and quantification of clouds, could be developed in the future [49, 51]. Furthermore, how each type of cloud form affects SPF models should also be of interest for future research. Taking into account the vast randomness of clouds, factors like their movement speed, form, and cloud thickness have a major impact on the forecasting accuracy and estimating that impact should be in the centre of attention for researchers in the future.

5.3 | Investigation of ramp events

By its nature, SPF is mainly based on meteorological factors. Therefore, the performance of the SPF models is highly dependent on the quality of the datasets of such factors. However, ramp events tend to occur in real-life problems that disturb the uniformity of the data, causing stability and grid connection problems. Ramp events are complex to define due to the difficulty in estimating their occurrence. Considering the increase in the use of renewable sources of energy into power systems, understanding and utilizing ramp events into SPF is of crucial importance to further develop SPF models.

In order to define a ramp event more specifically for SPF, if we take into account that in SPF, the most important aspect is sunlight and solar irradiance, a ramp event could be considered a highly unexpected event that interrupts the incidence of solar irradiance on the solar modules. For example, on a larger scale, an unexpected rainfall during a very hot day during summer could be considered a ramp event. On a smaller scale, heavy clouds during the noon hours, when the solar production is at a daily peak, could also be considered a ramp event. Another example of ramp event could be considered a technical failure of the solar production ensemble.

As a result, developing methodologies that investigate the appearance of ramp events as well as their appearance rate should be further researched in the future. Moreover, dependency between ramp events and specific input data of SPF models should be further researched. To deal with this, we would have to add the details about electricity grid connections and the demand profile on that grid.

5.4 | Versatility of SPF models

The majority of the SPF models focus on estimating the forecasting error based on a specific dataset that concerns a specific problem [59]. SPF models developed by research institutes are typically configured to provide predictive results for specific regions in which they perform well. Such models, however, are rarely tested in other regions, either with similar or different climatic conditions. Furthermore, due to limited access on solar data, in various cases specific open access data are being repeatedly used in SPF cases. Such limitations need to be surpassed and the necessity to develop SPF models that are efficient in spite of the input data and the climatic conditions, as well as evaluate them in different regions needs to be further researched. Moreover, it should be noted that the variability of solar power generation is of major importance in estimating the forecasting skill. Better interpretation of the information provided by such variability data could further improve the adaptability of SPF models to different conditions.

5.5 | Probabilistic SPF

The majority of SPF methodologies are focused on developing deterministic forecasting models that provide users with point-forecast output series that offer a specific outcome of the prediction process. However, in recent years, researches have started investigating probabilistic SPF. Contrary to point forecasting models, probabilistic forecasting models offer a wider view of the possible outcome of a prediction in the form of quantiles, prediction intervals (PIs), or distributions. Concerning the need of continuous integration of renewable sources into power systems, probabilistic forecasts could be a valuable tool for the planning and operation of power systems and electricity markets. Furthermore, while deterministic forecasts are widely used in decision-making problems where probability distributions are very relevant, further testing and application of probabilistic SPF models into decision-making cases could provide useful information over such cases [38].

Probabilistic SPF can be classified into two main categories, parametric and non-parametric. Parametric methods interpret the relationship between weather conditions and the predictive outcome of a forecast through historical data and data patterns. On the other hand, non-parametric methods estimate the predictive error without the need of pre-determining the distribution of the data [97] and that is an advantage. Regression models, ensemble methods, deep learning models, or hierarchical models have been developed over the past few years and are still being improved in order to provide accurate predictive results of non-parametric methodologies. There are great possibilities in interpreting probabilistic forecasting into power systems in the future and therefore, further development of probabilistic SPF models should be researched.

5.6 | Input data selection and analysis

With the continuous increase of the penetration of solar power into power systems, optimizing SPF models and improving their accuracy is of major importance. Selection of appropriate input data as well as proposing new pre-processing methodologies could play an important role in improving the predictive error [61]. Further evaluating data fluctuations, noise or errors from different input variables could also affect the forecasting accuracy. Data analysis and pre-processing are of equal importance as constructing an SPF model due to the SPF models' dependency on valid data. As a result, efficient ways to provide accurate and valid data should be further developed in the future.

5.7 | Focus on SPF applications

The need to use renewable sources of energy into power systems keeps increasing. Solar power is one of the most exploited sources of energy. As a result, applying SPF models to real-life problems and datasets is critical [33]. Such problems could be connected to power systems' planning and operation or smart-grid utilities. Moreover, due to the increase in the use of electric vehicles, applying SPF to electric vehicle charging planning problems could also be useful for future research.

5.8 | Data privacy protection

Due to the rapid and continuous improvement and development of forecasting methodologies, acquiring data and more specifically accurate data has become a great challenge. However, while data acquisition has been the main problem and the core focus over the past years, data privacy has been quite neglected. The term 'data privacy' refers to protecting sensitive and confidential information of the forecasting process. Such information may include personal or financial data that may affect the forecasting process. As a result, in modern energy systems, enhancing the security and achieving data privacy is becoming a major challenge [98].

Common data privacy protection problems may include data sharing between organizations, data retention into the forecasting models, data breaches, and cyber-attacks [99], as well as no compliance to the existing data protection laws. Therefore, data privacy protection should be considered in the development and construction of novel SPF models in the future.

5.9 | Spatiotemporal correlation in SPF

Solar power has a high space and time dependence. Over the past few years, more and more researchers have focused on applying a spatiotemporal perspective into developed forecasting models in order to improve their accuracy [100]. Aiming to facilitate the control and reduce the operation cost of power systems, implementation of spatiotemporal information into forecasting models could further enhance this process [101]. Considering solar power, which is highly dependent on meteorological conditions, spatial correlation of data could be highly applicable in probabilistic forecasting models [102]. Including the uncertainty of the prediction into real-life problems and into power systems is one of the most important challenges of the future. Therefore, including spatiotemporal correlation into probabilistic models could prove to highly improve the forecasting accuracy of the uncertainty error.

6 | CONCLUSION

Global energy needs have increased significantly over the last few decades and keep increasing. Considering the environmental consequences of conventional sources of energy, power systems have turned to RES. The penetration of RES plays a vital role in the operation of power systems and energy markets.

Solar power is one of the most important RES due to its abundance on a global level. However, due to its non-stable nature and its increasing penetration into power systems, it has created several problems in terms of stability. The development of SPF models has played an important part in surpassing such problems and efficiently implements solar power into energy systems.

This paper aims to evaluate novel state-of-the-art ultra-short-term and short-term solar power forecasting models. It further classifies the reviewed works from a climatic point of view. Classifications based on the input data and technical characteristics are also provided. Evading to focus specifically on the different forecasting methodologies, this review is a useful guide that provides information over the effect of the different climatic conditions to the forecasting accuracy. Further comparison of similar cases in terms of input variables or technical data offers the possibility of evaluating the efficiency of each forecasting model and indicates the main limitations and directions for further development.

This review work evaluates novel state-of-the-art SPF models. Considering the importance of planning of daily electricity markets and optimized operation of energy systems, short-term SPF models are the main focus. Furthermore, it provides useful classifications over data that affect the accuracy of the SPF models and points out the advantages of the different SPF methodologies. Moreover, it classifies the reviewed works based on different evaluation metrics and serves as a useful guide to readers intending to research short-term SPF cases from different perspectives.

This paper proposed important future possibilities on short-term SPF. More specifically, the proposed future directions aim to highlight the need to further improve the state of SPF in order to deal with real-life problems. Technological advances as well as innovative methodologies, along with proper data selection and pre-processing, could further increase forecasting accuracy, minimizing the forecasting error. In addition, creating forecasting models that could be applied in different problems and meteorological conditions is of vital importance, not only to understand the behaviour of forecasting models in different conditions, but also to compare different forecasting methodologies. In addition, further investing in probabilistic SPF could solve various problems in the future, in terms of planning and operation of power systems and electricity markets.

Despite the continuous advance of solar power forecasting methodologies, the improvement of the forecasting outcome does come with several challenges. The variability of weather conditions remains a great challenge for researchers. Furthermore, availability of sufficient and useable data as well as data security becomes a more and more important problem for researchers, since access to open-source accurate and realistic data is limited. Moreover, when it comes to small and distributed solar power installations, forecasting techniques for solar power frequently have trouble producing precise predictions with high levels of spatial and temporal precision. It is also a phenomenon of recent advanced forecasting models that they are characterized by high complexity and need a large amount of data in order to provide accurate results.

It should be noted that, while the forecasting possibilities are numerous, aiming to solve different every-day energy problems, this paper focuses on very-short-term and short-term forecasting methodologies that are crucial for the daily electricity market and short-term system scheduling and operation in the smart grid context. Therefore, it does not include cases of mid-term and long-term SPF that deal with maintenance scheduling, power generation planning, and capacity expansion cases. Furthermore, this review work aims to study forecasting methodologies based on the climatic and geographical conditions described in each case. As a result, other aspects are less emphasized or even left out to not make the survey overly long. For example, the important technological aspect of the forecasting process, such as the equipment type or the equipment age used for the photovoltaic power production, while important for the quality of the forecasting output and the predictive error, is out of the scope of this review paper.

NOMENCLATURE

ANN	artificial neural network
ARIMA	auto regressive integrated moving average
ARMA	auto regressive moving average
DNI	direct normal irradiance
DRL	deep reinforcement learning
ELM	extreme learning machine
NWP	numerical weather predictions
PI	prediction interval
PV	photovoltaic
RES	renewable source of energy
SPF	solar power forecasting
SVM	support vector machine

AUTHOR CONTRIBUTIONS

Ioannis Bazionis: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing; Markos Kousounadis-Knousen: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing; Pavlos Georgilakis: Conceptualization, Methodology, Supervision, Validation, Writing—review & editing; Elham Shirazi: Data curation, Formal analysis, Writing—review & editing; Dimitrios Soudris: Funding acquisition, Project administration, Writing—review & editing; Francky Catthoor: Conceptualization, Project administration, Supervision, Validation, Writing—review & editing.

ACKNOWLEDGEMENTS

This research has been co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH—CREATE—INNOVATE (project code:T2EDK-00864).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

Data available on request from the authors.

ORCID

Ioannis K. Bazionis https://orcid.org/0000-0002-2693-1044
Pavlos S. Georgilakis https://orcid.org/0000-0001-7231-7662

REFERENCES

1. Jacobson, M.Z., Delucchi, M.A.: Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials. Energy Policy. 39(39), 1154–1169 (2011)
2. IEA – Renewable energy market update outlook for 2022 and 2023. Available: https://www.iea.org/reports/renewable-energy-market-update-may-2022
3. Strzalka, A., Alam, N., Duminil, E., Coors, V., Eicker, U.: Large scale integration of photovoltaics in cities. Appl. Energy. 93, 413–421 (2012)

4. Manz, D., Walling, R., Miller, N., LaRose, B., D'Aquila, R., Daryanian, B.: The grid of the future: Ten trends that will shape the grid over the next decade. IEEE Power Energy Mag. 12(3), 26–36 (2014)
5. Akhter, M.N., Mekhilef, S., Mokhlis, H., Shah, N.M.: Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques. IET RPG. 13(7), 1009–1023 (2019)
6. Das, U.K., Tey, K.S., Seyedmahmoudian, M., Mekhilef, S., Idris, M.Y.I., Van Deventer, W., Horan, B., Stojcevski, A.: Forecasting of photovoltaic power generation and model optimization: A review. Renew. Sust. Energ. Rev. 81, 912–928 (2018)
7. Barbieri, F., Rajakaruna, S., Ghosh, A.: Very short-term photovoltaic power forecasting with cloud modelling: A review. Renew. Sust. Energ. Rev. 75, 242–263 (2017)
8. Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F.: Review of photovoltaic power forecasting. Sol. Energy. 136, 78–111 (2016)
9. Raza, M.Q., Nadarajah, M., Ekanayake, C.: On recent advances in PV output power forecast. Sol. Energy. 136, 125–144 (2016)
10. Van der Meer, D.W., Widen, J., Munkhammar, J.: Review on probabilistic forecasting of photovoltaic power production and electricity consumption. Renew. Sust. Energ. Rev. 81, 1484–1512 (2018)
11. Li, B., Zhang, J.: A review on the integration of probabilistic solar forecasting in power systems. Sol. Energy. 210, 68–86 (2020)
12. Ma, T., Yang, H., Lu, L.: Solar photovoltaic system modelling and performance prediction. Renew. Sust. Energ. Rev. 36, 304–315 (2014)
13. Ahmed, R., Sreeram, V., Mishra, Y., Arif, M.D.: A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. Renew. Sust. Energ. Rev. 124, 109792 (2020)
14. Buwei, W., Jianfeng, C., Bo, W., Shuanglei, F.: A solar power prediction using support vector machines based on multi-source data fusion. In: Proceedings of the 2018 International Conference on Power System Technology, Guangzhou, China (2018)
15. Pedro, H.T.C., Coimbra, C.F.M.: Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol. Energy. 86(7), 2017–2028 (2012)
16. Huang, R., Huang, T., Gadh, R., Li, N.: Solar generation prediction using the ARMA model in a laboratory-level micro-grid. In: Proceedings of the IEEE Third International Conference on Smart Grid Communications (SmartGridComm), Tainan, Taiwan (2012)
17. Palmer, T., Hagedorn, R.: Predictability of Weather & Climate. Cambridge University Press (2006)
18. Box, G.E.P., Jenkins, G.M., Reinsel, G.C.: Time Series Analysis: Forecasting and Control, 4th ed. New York: John Wiley & Sons, Inc. (2008)
19. Wan, C., Zhao, J., Song, Y., Xu, Z., Lin, J., Hu, Z.: Photovoltaic and solar power forecasting for smart grid energy management. CSEE J. Power Energy Syst. 1(4), 38–46 (2015)
20. Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N.: Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. Renew. Sust. Energ. Rev. 27, 65–76 (2013)
21. Sharadga, H., Hajimirza, S., Balog, R.S.: Time series forecasting of solar power generation for large-scale photovoltaic plants. Renew. Energy. 150, 797–807 (2020)
22. Nazaripouya, H., Wang, B., Wang, Y., Chu, P., Pota, H.R., Gadh, R.: Univariate time series prediction of solar power using a hybrid wavelet-ARMA-NARX prediction method. In: Proceedings of the IEEE/PES Transmission and Distribution Conference and Exposition (T&D), Dallas, TX, USA (2016)
23. Boland, J., David, M., Lauret, P.: Short term solar radiation forecasting: Island versus continental sites. Energy. 113, 186–192 (2016)
24. Shadab, A., Ahmad, S., Said, S.: Spatial forecasting of solar radiation using ARIMA model. Remote Sens. Appl.: Soc. Environ. 20, 100427 (2020)
25. Atique, S., Noureen, S., Roy, V., Subburaj, V., Bayne, S., Macfie, J.: Forecasting of total daily solar energy generation using ARIMA: A case study. In Proceedings of the IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC), Las Vegas, NV, USA (2019)
26. Kushwaha, V., Pindoriya, N.M.: A SARIMA-RVFL hybrid model assisted by wavelet decomposition for very short-term solar PV power generation forecast. Renewable Energy. 140, 124–139 (2019)
27. Tato, J.H., Brito, M.C.: Using smart persistence and random forests to predict photovoltaic energy production. Energies. 12(1), 100 (2019)
28. Niu, D., Wang, K., Sun, L., Wu, J., Xu, X.: Short-term photovoltaic power generation forecasting based on random forest feature selection and CEEMD: A case study. Appl. Soft Comput. 93, 106389 (2020)
29. Lateko, A.A.H., Yang, H.-T., Huang, C.-M.: Short-term PV power forecasting using a regression-based ensemble method. Energies. 15(11), 4171 (2022)
30. Abdel-Nasser, M., Mahmoud, K.: Accurate photovoltaic power forecasting models using deep LSTM-RNN. Neural. Comput. Appl. 31, 2727–2740 (2019)
31. Demirdelen, T., Ozge Aksu, I., Esenboga, B., Aygul, K., Ekinci, F., Bilgili, M.: A new method for generating short-term power forecasting based on artificial neural networks and optimization methods for solar photovoltaic power plants. Solar Photovoltaic Power Plants. 165–189 (2019)
32. Lee, D., Kim, K.: Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. Energies. 12(2), 215 (2019)
33. Li, G., Wang, H., Zhang, S., Xin, J., Liu, H.: Recurrent neural networks based photovoltaic power forecasting approach. Energies. 12(13), 2538 (2019)
34. Eseye, A.T., Lehtonen, M., Tukia, T., Uimonen, S., Millar, R.J.: Adaptive predictor subset selection strategy for enhanced forecasting of distributed PV power generation. IEEE Access. 7, 90652–90665 (2019)
35. Gao, M., Li, J., Hong, F., Long, D.: Day-ahead power forecasting in a large-scale photovoltaic plant based on weather classification using. LSTM. Energy. 187, 115838 (2019)
36. Jeong, J., Kim, H.: Multi-site photovoltaic forecasting exploiting space-time convolutional neural network. Energies. 12(23), 4490 (2019)
37. Ju, Y., Li, J., Sun, G.: Ultra-short-term photovoltaic power prediction based on self-attention mechanism and multi-task learning. IEEE Access. 8, 44821–44829 (2020)
38. Aprillia, H., Yang, H.-T., Huang, C.-M.: Short-term photovoltaic power forecasting using a convolutional neural network–salp swarm algorithm. Energies. 13(8), 1879 (2020)
39. Zang, H., Cheng, L., Ding, T., Cheung, K.W., Wei, Z., Sun, G.: Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. Int. J. Electr. Power Energy Syst. 118, 105790 (2020)
40. Zhang, T., Lv, C., Ma, F., Zhao, K., Wang, H., O'Hare, G.M.P.: A photovoltaic power forecasting model based on dendritic neuron networks with the aid of wavelet transform. Neurocomputing. 397, 438–446 (2020)
41. Al-Dahidi, S., Louzazni, M., Omran, N.: A local training strategy-based artificial neural network for predicting the power production of solar photovoltaic systems. IEEE Access. 8, 150262–150281 (2020)
42. Zhang, Y., Qin, C., Srivastava, A.K., Jin, C., Sharma, R.: Data-driven day-ahead PV estimation using autoencoder-LSTM and persistence model. IEEE Trans. Ind. Appl. 56(6), 7185–7192 (2020)
43. Kothona, D., Zamanidou, A., Panapakidia, I., Christoforidis, G.: Day-ahead PV generation forecasting based on deep learning approach. In: Proceedings of the 38th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC) (2021)
44. Emamian, M., Milimonfared, J., Eskandari, A., Aghaei, M., Abardeh, R.H., Vidal de Oliveira, A.K.: Solar power forecasting with LSTM network ensemble. In: Proceedings of the 36th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Marseille, France (2019)
45. Theocharides, S., Makrides, G., Theristis, M., Georghiou, G.: Novel intra-day photovoltaic production forecasting algorithm using deep learning ensemble models. In: Proceedings of the 38th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC) (2020)
46. Carrillo, R., Schubnel, B., Simeunovic, J., Langou, R., Alet, P.-J.: Spatio-temporal machine learning methods for multi-site PV power forecasting. In: Proceedings of the 38th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC) (2020)
47. Theocharides, S., Makrides, G., Theristis, M., Almonacid, F., Fernandez, E., Georghiou, G.: Short-term photovoltaic power forecasting based on artificial neural networks: A numerical weather prediction-free approach.

In: Proceedings of the 35th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Brussels, Belgium (2018)
48. Theocharides, S., Makrides, G., Venizelou, V., Kaimakis, P., Kyprianou, A., Georghiou, G.: PV production forecasting model based on artificial neural networks (ANN). In: Proceedings of the 33rd European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Amsterdam, The Netherlands (2017)
49. Zhang, J., Verschae, R., Nobuhara, S., Lalonde, J.-F.: Deep photovoltaic nowcasting. Solar Energy. 176, 267–276 (2018)
50. Korkmaz, D., Acikgoz, H., Yildiz, C.: A novel short-term photovoltaic power forecasting approach based on deep convolutional neural network. Int. J. Green Energy. 18(5), 525–539 (2021)
51. Huynh, A.N.-L., Deo, R.C., Ali, M., Abdulla, S., Raj, N.: Novel short-term solar radiation hybrid model: Long short-term memory network integrated with robust local mean decomposition. Appl. Energy. 298, 117193 (2021)
52. Korkmaz, D.: SolarNet: A hybrid reliable model based on convolutional neural network and variational mode decomposition for hourly photovoltaic power forecasting. Appl. Energy. 300, 117410 (2021)
53. Akhter, M.N., Mekhilef, S., Mokhlis, H., Almohaimeed, Z.M., Muhammad, M.A., Khairuddin, A.S.M., Akram, R., Hussain, M.M.: An hour-ahead PV power forecasting method based on an RNN-LSTM model for three different PV plants. Energies. 15, 2243 (2022)
54. Talaat, M., Said, T., Essa, M.A., Hatata, A.Y.: Integrated MFFNN-MVO approach for PV solar power forecasting considering thermal effects and environmental conditions. Int. J. Electr. Power Energy Syst. 135, 107570 (2022)
55. Michael, N.E., Mishra, M., Hasan, S., Al-Durra, A.: Short-term solar power predicting model based on multi-step CNN stacked LSTM technique. Energies. 15(6), 2150 (2022)
56. Heras, E.R., Camacho, C.A., Calderón, E.C.: Short-term forecasting of energy production for a photovoltaic system using a NARX-CVM hybrid model. Energies. 15(8), 2842 (2022)
57. Cortes, C., Vapnik, V.: Support vector networks. Mach. Learn. 20(3), 273–297 (1995)
58. Quej, V.H., Almorox, J., Arnaldo, J.A., Saito, L.: ANFIS, SVM and ANN softcomputing techniques to estimate daily global solar radiation in a warm subhumid environment. J. Atmos. Sol. Terr. Phys. 155, 62–70 (2017)
59. VanDeventer, W., Jamei, E., Thirunavukkarasu, G.S., Seyedmahmoudian, M., Soon, T.K., Horan, B., Mekhilef, S., Stojcevski, A.: Short-term PV power forecasting using hybrid GASVM technique. Renew. Energy. 140, 367–379 (2019)
60. Fonseca, J.G., Udagawa, Y., Oozeki, T., Ogimoto, K.: Characteristics of day-ahead residual demand, PV power and demand forecasts in a scenario of large penetration of PV. In: Proceedings of the 37th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Lisbon, Portugal (2020)
61. Fonseca, J.G., Uno, F., Oozeki, T., Ogimoto, K.: Intra-day forecasts of PV power with numerical weather prediction data and machine learning in Kyushu, Japan. In: Proceedings of the 33rd European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Lisbon, Portugal (2020)
62. Boegli, M., Pierro, M., Moser, D., Alet, P.-J.: Machine learning techniques for forecasting single-site PV production. In: Proceedings of the 35th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Brussels, Belgium (2018)
63. Starosta, A., Kaushik, K., Jhaveri, P., Munzke, N., Hiller, M.: A comparative analysis of forecasting methods for photovoltaic power and energy generation with and without exogenous inputs. In: Proceedings of the 38th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC) (2021)
64. Hossain, M., Mekhilef, S., Danesh, M., Olatomiwa, L., Shamshirband, S.: Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. J. Clean. Prod. 167, 395–405 (2017)
65. Huang, G.-B., Zhu, Q.-Y., Siew, C.-K.: Extreme learning machine: A new learning scheme of feedforward neural network. In: Proceedings of the IEEE International Joint Conference on Neural Networks, Budapest, Hungary (2004)
66. Hossain, M., Mekhilef, S., Danesh, M., Olatomiwa, L., Shamshirband, S.:Application of extreme learning machine for short term output power forecasting of three grid-connected PV systems. J. Clean. Prod. 167395–405 (2017)
67. Behera, M.K., Majumder, I., Nayak, N.: Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. Eng. Sci. Technol. Int. J. 21(3), 428–438 (2018)
68. Acikgoz, H.: A novel approach based on integration of convolutional neural networks and deep feature selection for short-term solar radiation forecasting. Appl. Energy. 305, 117912 (2022)
69. Dairi, A., Harrou, F., Sun, Y., Khadraoui, S.: Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach. Appl. Sci. 10(23), 8400 (2020)
70. Li, Y., Wang, R., Yang, Z.: Optimal scheduling of isolated microgrids using automated reinforcement learning-based multi-period forecasting. IEEE Trans. Sustain. Energy. 13(1), 159–169 (2021)
71. Jeong, J., Kim, H.: DeepComp: Deep reinforcement learning based renewable energy error compensable forecasting. Appl. Energy. 294, 116970 (2021)
72. Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N.: Review of solar irradiance forecasting methods and a proposition for small scale insular grids. Renew. Sust. Energ. Rev. 27, 65–76 (2013)
73. Melliti, D.E., Long, D., Marechal, P., Ghalila-Sevestre, S.: Cloud segmentation and classification to improve the short-term PV forecasting using all sky imager. In: Proceedings of the 38th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC) (2021)
74. Esteves, J., Pinho da Silva, N., Pestana, R., Cao, Y.: Solar power forecast using satellite pictures. In: Proceedings of the 37th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC), Lisbon, Portugal (2020)
75. Holland, N., Pang, X., Herzberg, W., Bor, J., Lorenz, E.: Combination of physics based simulation and machine learning for PV power forecasting of large power plants. In: Proceedings of the 38th European Photovoltaic Solar Energy Conference and Exhibition (EU PVSEC) (2021)
76. Kuhn, P., Nouri, B., Wilbert, S., Prahl, C., Kozonek, N., Schmidt, T., Yasser, Z., Ramirez, L., Zarzalejo, L., Meyer, A., Vuilleumier, L., Heinemann, D., Blanc, P., Pitz-Paal, R.: Validation of an all-sky imager-based nowcasting system for industrial PV plants. Prog. Photovolt. 26(8), 608–621 (2017)
77. Wang, F., Lu, X., Mei, S., Su, Y., Zhen, Z., Zou, Z., Zhang, X., Yin, R., Duić, N., Shafie-Khan, M., Catalão, J.P.S.: A satellite image data based ultra-short-term solar PV power forecasting method considering cloud information from neighboring plant. Energy. 238, 121946 (2022)
78. Law, E.W., Prasad, A.A., Kay, M., Taylor, R.A.: Direct normal irradiance forecasting and its application to concentrated solar thermal output forecasting – a review. Sol. Energy. 108, 287–307 (2014)
79. Tapakis, R., Charalambides, A.G.: Equipment and methodologies for cloud detection and classification: A review. Sol. Energy. 95, 392–430 (2013)
80. Larson, V.E.: Chapter 12 – Forecasting solar irradiance with numerical weather prediction models. In: Kleissl, J. (ed.), Solar Energy Forecasting and Resource Assessment. Academic Press, New York, pp. 299–318 (2013)
81. Liu, L., Zhan, M., Bai, Y.: A recursive ensemble model for forecasting the power output of photovoltaic systems. Sol. Energy. 189, 291–298 (2019)
82. Li, P., Zhou, K., Lu, X., Yang, S.: A hybrid deep learning model for short-term PV power forecasting. Appl. Energy. 259, 114216 (2020)
83. Li, G., Xie, S., Wang, B., Xin, J., Li, Y., Du, S.: Photovoltaic power forecasting with a hybrid deep learning approach. IEEE Access. 8, 175871–175880 (2020)
84. Massaoudi, M., Refaat, S.S., Abu-Rub, H., Chihi, I., Wesleti, F.S.: A hybrid bayesian ridge regression-CWT-catboost model for PV power forecasting. In: Proceedings of the IEEE Kansas Power and Energy Conference (KPEC), Manhattan, KS, USA (2020)
85. Opitz, D., Maclin, R.: Popular ensemble methods: An empirical study. J. Artif. Intell. Res. 11, 169–198 (1999)

86. Ramakrishna, R., Scaglione, A., Vittal, V., Dall'Anese, E., Bernstein, A.: A model for joint probabilistic forecast of solar photoboltaic power and outdoor temperature. IEEE Trans. Signal Process. 67(24), 6368–6383 (2019)
87. Agoua, X.G., Girard, R., Kariniotakis, G.: Probabilistic model for spatio-temporal photovoltaic power forecasting. IEEE Trans. Sustain. Energy. 10(2), 780–789 (2018)
88. Doubleday, K., Jascourt, S., Kleiber, W., Hodge, B.M.: Probabilistic solar power forecasting using bayesian model averaging. IEEE Trans. Sustain. Energy. 12(1), 325–337 (2020)
89. Huang, C.J., Kuo, P.H.: Multiple-input deep convolutional neural network model for short-term photovoltaic power forecasting. IEEE Access. 7, 74822–74834 (2019)
90. Chen, B., Li, J.: Combined probabilistic forecasting method for photovoltaic power using an improved Markov chain. IET Gener. Transm. Distrib. 13(13), 4364–4373 (2019)
91. Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H., Liu, W.: Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. Energy Convers. Manag. 153, 409–422 (2017)
92. Kumar, D.K., Gandhi, O., Rodriguez-Gallegos, C.D., Srinivasan, D.: Review of power system impacts at high PV penetration part II: Potential solutions and the way forward. Sol. Energy. 210, 202–221 (2020)
93. Ren, Y., Suganthan, P.N., Srikanth, N.: Ensemble methods for wind and solar power forecasting-a state-of-the-art review. Renew. Sust. Energ. Rev. 50, 82–91 (2015)
94. Peel, M.C., Finlayson, B.L., McMahon, T.A.: Updated world map of the Köppen-Geiger climate classification. Hydrol. Earth Syst. Sci. 11, 1633–1644 (2007)
95. Dunnett, S., Sorichetta, A., Taylor, G., Eigenbrod, F.: Harmonised global datasets of wind and solar farm locations and power. Sci. Data. 130 (2020)
96. Solar resource map © 2021 Solargis. Available: https://solargis.com
97. Bazionis, I., Karafotis, P., Georgilakis, P.: A review of short-term wind power probabilistic forecasting and a taxonomy focused on input data. IET Renew. Power Gener. 16(1), 77–91 (2022)
98. Li, Y., Wang, R., Li, Y., Zhang, M., Long, C.: Wind power forecasting considering data privacy protection: A federated deep reinforcement learning approach. Appl. Energy. 329, 120291 (2023)
99. Li, Y., Xinhao, W., Yuanzheng, L., Zhaoyang, D., Mohammad, S.: Detection of false data injection attacks in smart grid: A secure federated deep learning approach. IEEE Trans. Smart Grid. 13(6), 4862–4872 (2022)
100. Pan, X., Zhou, J., Sun, X., Cao, Y., Cheng, X., Farahmand, H.: A hybrid method for day-ahead photovoltaic power forecasting based on generative adversarial network combined with convolutional autoencoder. IET Renew. Power Gener. 1–15 (2022)
101. Chen, Y., Zhang, Z., Chen, H., Zheng, H.: Robust UC model based on multi-band uncertainty set considering temporal correlation of wind/load prediction errors. IET Gener. Transm. Distrib. 14(2), 180–190 (2019)
102. Zhang, Z., Chen, Y., Ma, J., Liu, X., Wang, W.: Two-stage robust security constrained unit commitment considering the spatiotemporal correlation of uncertainty prediction error. IEEE Access. 7, 22891–22901 (2019)

How to cite this article: Bazionis, I.K., Kousounadis-Knousen, M.A., Georgilakis, P.S., Shirazi, E., Soudris, D., Catthoor, F.: A taxonomy of short-term solar power forecasting: Classifications focused on climatic conditions and input data. IET Renew. Power Gener. 17, 2411–2432 (2023). https://doi.org/10.1049/rpg2.12736

APPENDIX

a. The MAE is the average value of the absolute error of the N forecasted error values and is defined as [29]

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\left|\hat{y}_i - y_i\right| \tag{1}$$

where N is the total number of samples, $\hat{y}_i$ is the forecasted value and y_i is the real value.

b. The MAPE represents the average value of absolute percentage errors and is calculated by [29]

$$MAPE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{y}_i - y_i\right|}{y_i} \times 100\% \tag{2}$$

c. The MRE represents the ratio of the absolute error of a variable to the variable's range and is defined as [29]

$$MRE = \frac{1}{N}\sum_{i=1}^{N}\frac{\left|\hat{y}_i - y_i\right|}{y_{total}} \times 100\% \tag{3}$$

where y_{total} is the range of the output power.

d. The MBE is used to capture the average bias in a prediction and is calculated by [82]

$$MBE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right) \tag{4}$$

e. The MSE is the average value of the squared error, which is the squared difference between the actual and predicted values, and is defined as [56]

$$MSE = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2 \tag{5}$$

f. The RMSE represents the squared root of the quadratic mean of the difference between the actual and predicted values and is calculated by [56]

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2} \tag{6}$$

g. The nRMSE relates the RMSE to the observed range of the given variable and is defined by [36]

$$nRMSE = \left(\sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\hat{y}_i - y_i\right)^2}\right) \times 100\% / y_{i(max)} \tag{7}$$

h. The R^2 is used as a means of measurement of how well the predictive outcome is reproduced by the forecasting model, based on the proportion of the total variation of the predictions [29]:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}\left(\hat{y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(\bar{y}_i - y_i\right)^2} \tag{8}$$