Solar Energy 284 (2024) 113044

Contents lists available at ScienceDirect

Solar Energy

journal homepage: www.elsevier.com/locate/solener

Photovoltaic power estimation and forecast models integrating physics and machine learning: A review on hybrid techniques

Leticia de Oliveira Santos [a], Tarek AlSkaif [b,*], Giovanni Cordeiro Barroso [a,c], Paulo Cesar Marques de Carvalho [a]

[a] *Department of Electrical Engineering, Federal University of Ceará, Campus do Pici, 60455-760, Fortaleza, Ceará, Brazil*
[b] *Information Technology Group, Wageningen University, Wageningen, 6708 PB, The Netherlands*
[c] *Department of Physics, Federal University of Ceará, Campus do Pici, 604400-900, Fortaleza, Ceará, Brazil*

ARTICLE INFO

Keywords:
Photovoltaic power
Hybrid model
Solar power forecasting
PV power estimation

ABSTRACT

Photovoltaic (PV) models are essential for energy planning and grid integration applications. The models used for PV power conversion typically adopt physical, data-driven, or hybrid approaches. Different hybrid techniques, combining elements of physical and data-driven methods, have been effectively developed for PV power estimation and forecasting, leveraging the strengths of both native methods. The data-driven approach allows models to account for unmodeled uncertainties and nonlinearities that purely physical models cannot capture. Meanwhile, the guidance of scientific theory makes the models less dependent on data, thereby improving generalization, interpretability, and accuracy. This review paper provides a comprehensive overview of hybrid methodologies for PV power estimation and forecasting. The main available hybridization methods are summarized and discussed under a novel methodological classification into three primary categories: physics-informed machine learning models, optimized physical models, and physics-guided models. Furthermore, these hybrid models are compared in terms of methodology, applications, and elucidating the merits and demerits of different techniques. By offering insights into these hybrid models, this review lays a foundation for researchers in this field and contributes to the advancement of PV power estimation and forecasting methodologies.

Contents

* Corresponding author.
E-mail address: tarek.alskaif@wur.nl (T. AlSkaif).

https://doi.org/10.1016/j.solener.2024.113044
Received 18 June 2024; Received in revised form 20 September 2024; Accepted 23 October 2024
Available online 10 November 2024
0038-092X/

Nomenclature

AB	Ada-Boost Regression
AC	Alternating Current
AML	Automatic Machine Learning
ANN	Artificial Neural Network
ANFIS	Adaptive Network based Fuzzy Inference System
ARIMA	AutoRegressive Integrated Moving Average Prediction
ARIMAX	ARIMA with exogenous Variable
ARCH	AutoRegressive Conditional Heteroskedasticity Artificial Neural Network
ARMA	AutoRegressive Moving Average
ARX	AutoRegressive eXogenous
BP-ANN	Backpropagation Artificial Neural Network Optimization
BR	Bagging Regression
CNN	Convolutional Neural Network
CSPE	Clear Sky Persistence Ensemble
DC	Direct Current
DE	Differential Evolution
DT	Decision Tree
ELM	Extreme Learning Machine
GA	Genetic Algorithm
GBR	Gradient Boosting Regressor
GHI	Global Horizontal Irradiance
KDE	Kernel Density Estimation
KNN	k-Nearest Neighbors
LR	Linear Regression
LSTM	Long Short-Term Memory
ML	Machine Learning
MLP	Multilayer Perceptron
MLR	Multiple Linear Regression
MOS	Model Output Statistic
NWP	Numerical Weather
OLS	Ordinary Least Squares
PHANN	Physical Hybrid
PMC	Physical Model Chain
POA	Plane-of-Array
PSO	Particle Swarm
PV	Photovoltaic
PVP	PV Plant
QR	Quantile Regression
QRA	Quantile Regression Averaging
QRF	Quantile Regression Forest
QkNN	Quantile k-Nearest Neighbors
RF	Random Forest
RHNN	Relative Humidity Neural Network
RL	Reinforcement Learning
SDM	Single-Diode Model
SPAR	Semi-Parametric Auto-Regressive Model
SVM	Support Vector Machine
TCN	Temporal Convolutional Network

1. Introduction

The emergence of energy communities, microgrids, and virtual power plants requires precise power generation models. These models play a crucial role in simulating various scenarios and enhancing power forecasting for integration with the grid. Solar photovoltaic (PV) forecasting has attracted researchers from different fields such as meteorology, data sciences, and engineering, focusing on accurately estimating solar irradiance and converting it to electricity. Despite significant advancements in PV forecasting aimed at increasing reliability, current practices are still far from standardization [1]. The conversion of irradiance to power modeling typically follows two primary approaches: the physical and the data-driven methods. The former relies on a so-called physical model chain (PMC), a sequence of physical or empirical models characterizing irradiance to PV power, the latter explores the relationship between the PV power and irradiance from historical data.

Physical modeling requires no historical data on the PV system and is appropriate for any term forecast. However, it has limitations: first, these physics-based models require detailed parameters designed specifically for a particular PV plant (PVP) and location, therefore, they are usually used for modeling MW-level PV farms and are less applied to residential rooftop PV systems [2]. Also, regarding their complexity, they require more domain knowledge to operate than data-driven approaches [3], as they are based on mathematical equations that describe the PV conversion. The algorithms employed involve the solution of complex differential equations which can require a large amount of computer power and time to give accurate predictions. Although the physical phenomena involved in PV conversion is nowadays well understood and modeled in sufficient fidelity, even mature and well-studied models introduce errors [4]. These models do not account for malfunction of the system or local effects on the irradiance, such as intra-hour power fluctuations caused by cloud movements, partial shading, or snow cover, which could be an argument for data-driven methods [5]. However, physical models are important tools when extended horizons are needed, as they usually outperform data-driven models for long-range forecasts [6].

Nowadays data-driven techniques have received significant attention to model the PV power. The advantage of this modeling approach is that it offers an alternative to conventional physical modeling, requiring no knowledge of internal system parameters. Instead of complex mathematical routines or physical representations, these techniques can learn the key information patterns from the data. Regression methods are frequently employed in this context, especially specific computing techniques based on kernel, neural networks, and ensemble methods have been successfully employed for intrahour to days-ahead horizons of PV power forecast [7,8]. Although these data-driven methods demonstrate relatively strong forecast performance, they neglect the underlying physical principles governing PV conversion. As a result, their use is restricted to scenarios similar to datasets, presenting limited generalization applicability, and accuracy at the turning point of weather changes [9].

Comparisons of the physical and data-driven approaches have shown that the best model depends on the application, sometimes pointing out that physical forecasting models are better [10], sometimes that machine learning (ML) outperforms PMCs [11], or even that it varies depending on the training data quality and length [3]. The fusion between physics-based and data-driven models, the so-called hybrid modeling, has the potential to achieve both their advantages and assess individual deficiencies in estimation and forecasting. Physical models are applied to irradiance — PV power conversion or to adjust weather variables. Then, data-driven methods are used to improve the prediction accuracy or PV power estimation based on physics information [12]. The data-driven approach allows the model to account for unmodeled uncertainties and nonlinearities not captured by purely physical models, while the guidance of scientific theory makes the model less dependent on the data [13]. Different hybrid techniques have been effectively introduced for PV power forecast. These models are capable of joining the forces of the native methods by mitigating weaknesses of the singles to produce an optimized forecast for a range of time periods [14].

Recent literature reviews have extensively cataloged various methods for PV power modeling, with a predominant focus on more recent techniques based on ML and deep learning forecasting methods as in [7,15–17]. In other reviews such as [18,19], that present physical, data-driven, and hybrid models, the main focus on the individual methodologies makes the hybrid models receive scant attention. Despite the hybrid models limited coverage in published reviews compared with the individual approaches, authors in [15,18,20] point out their potential to increase the accuracy of individual models for PV power forecast, and in [6,21–23] highlight the benefits of these models. Also, these reviews differ in terms of hybrid model definition. The hybrid models are sometimes defined as a “combination of physical and statistical/ML methods” [24]. In other works, they are defined as a “combination of two or more forecasting models,” including ensemble ML models [7,15]. Furthermore, in terms of terminology used to refer to these models, they are called hybrid-physical, ensemble methods, knowledge-guided ML, and others. In this paper, the term “hybrid models” is used to address models based on physical and data-driven approaches, distinct from ensemble ML models which we categorize as data-driven techniques.

Our review provides the first comprehensive overview dedicated solely to the existing hybrid methodologies for PV power estimation and forecasting. We systematically summarize the main available hybridization methods, comparing them in terms of applications, presenting a novel methodological classification, and elucidating the merits and demerits of different techniques. Key factors influencing models performance are also discussed, such as input variables, preprocessing steps, and benchmark models. By offering insights into hybrid PV power estimation and forecasting models development, our review lays a foundation for researchers interested in this field, contributing to the advancement of PV power modeling methodologies.

This article is divided into six sections, alongside the introduction, Section 2 details the review methodology and includes a bibliometric analysis of selected manuscripts. Hybrid methods are categorized into three groups based on their modeling approach. These are discussed in Sections 3, 4 and 5, respectively: physics-informed ML models (ML models incorporating physically relevant regressors); optimized physical models (physical equations with optimization algorithms for parameter determination); and physics-guided models (collaborative approaches between physical models and ML techniques). Finally, Section 6 presents the conclusions.

2. Review methodology

We employ the methodology of a systematic literature review followed by a bibliometric analysis to identify the publications in the area. First, we define the search terms and search engines. Then, we select and analyze the relevant publications and present the results. The methodology flowchart is shown in Fig. 1. To carry out the publications search, we use keywords and boolean constraints to identify the most representative publications. The search was conducted on December 28, 2023, using the search engines IEEE Explore and Scopus, with the following search string in the ‘title, keywords, and abstract’ fields: “($photovoltaic$ OR $photovoltaics$ OR PV) AND ($forecasting$ OR $forecast$ OR $estimation$ OR $prediction$) AND ($physics$ OR $physical$ OR $physic$)”. The search was conducted using English terms, and we considered the publications in English, Spanish, and Portuguese languages, keeping 97.5% of the publications. No time or document type restrictions were applied. Additionally, we excluded publications from the following areas: health, chemistry, and biology. Conference proceedings books were also excluded, as the conference papers related to the topic also appeared individually in the search. Review articles were not considered since they do not include implementation of techniques or methodological analysis. Considering our focus on hybrid methodologies for time-series forecasting/estimation of PV power, we excluded publications with focus on thermal or PV–thermal solar technologies, predictive control or failure detection, and with methodologies based on image processing. Finally, publications that mention PV module modeling without including estimation or forecasting of PV power were discarded. By reading the publications, those that did not fit into the review scope or that did not meet the inclusion criteria were removed, resulting in a final literature set of 67 publications analyzed in depth in this systematic review.

The bibliometric parameters used to analyze publications on the research topic are related to the publication patterns by year, type of publications, journals, and countries. The publications on this topic span from 2013 to 2023, as shown in Fig. 2(a). Hybrid methodologies for time-series forecasting/estimation of PV power have been widely investigated in the past ten years, and the number of publications has grown over the years. Compared to publications in the area of purely physical models for PV power modeling, this is a relatively young approach. This increasing tendency can be a consequence of the popularity of ML techniques and their potential to improve the already established physical models. The final obtained set of 67 publications covers journal articles (64.18%), conference papers (34.33%), and book chapters (1.49%). From the 43 journal articles found from the bibliometric analysis, 23 are published in 7 different journals, where “Solar Energy” and “Energies” concentrate most of the publications, 5 each; followed by “Renewable Energy”, “Renewable and Sustainable Energy Reviews”, and “IEEE Transactions on Smart Grid”, with 3 publications each; and the “Journal of Solar Energy Engineering” and “Energy”, both with 2 publications. The countries with the highest number of publications are Italy (15), China (12), the USA (5), Germany (4), India (3), and Hungary (3), as shown in Fig. 2(b). Most of the publications from Italy are concentrated on a single technique, consisting of a hybridization of a clear-sky model and artificial neural networks (ANN), commonly referred to as physic-hybrid ANN (PHANN). The majority of these publications are concentrated between 2014 and 2020 (13 publications), since then, publications on ML models with physical features have started to include a wider variety of physical inputs, not just the clear sky index as in the Italian PHANN technique. Meanwhile, the research from China, representing a significant number of publications on this topic, is based on different techniques of hybridization for time-series forecasting/estimation of PV power. As it is not possible to identify a specific research focus in China, and given the significant growth in installed solar PV capacity there in recent years, the increasing number of studies in this area may be linked to a greater emphasis on the planning and integration of solar PV electricity into the grid.

A second analysis is conducted concerning methodology patterns (deterministic or probabilistic approach, time horizon, and modeling approach and goal — forecast, estimation, or performance analysis). Most of the studies are based on deterministic techniques (52 publications), with some publications (7) implementing both deterministic and probabilistic models, and only 4 studies focus on probabilistic approaches (Fig. 3(a)), indicating that probabilistic hybrid techniques have the potential to be further explored. Additionally, most publications are focused on PV power forecasting (62%), followed by PV power estimation (30%) and performance prediction/description (5%). A smaller number of studies test the proposed models for both forecast and estimation (3%).

Among the studies implementing hybrid techniques for PV power forecasting, the majority focus on short-term forecasts (73.4%), as

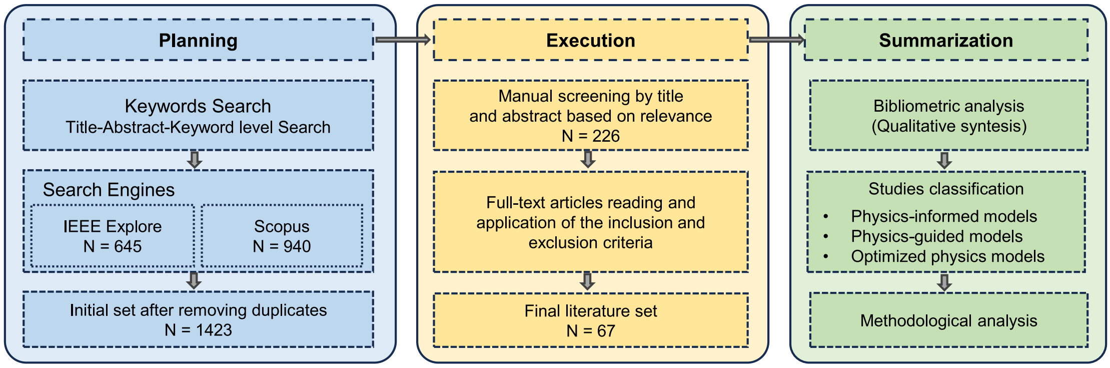
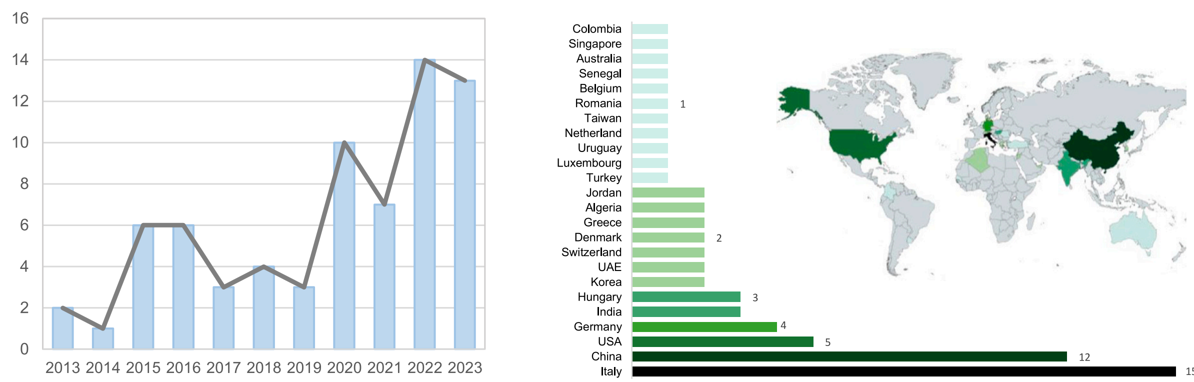


Fig. 1. Flowchart depicting the review steps: search process, execution, and analysis of research.



(a) Publications per year. (b) Publications per country.

Fig. 2. Distributions of the publications implementing hybrid models for PV power per (a) year, and (b) country.

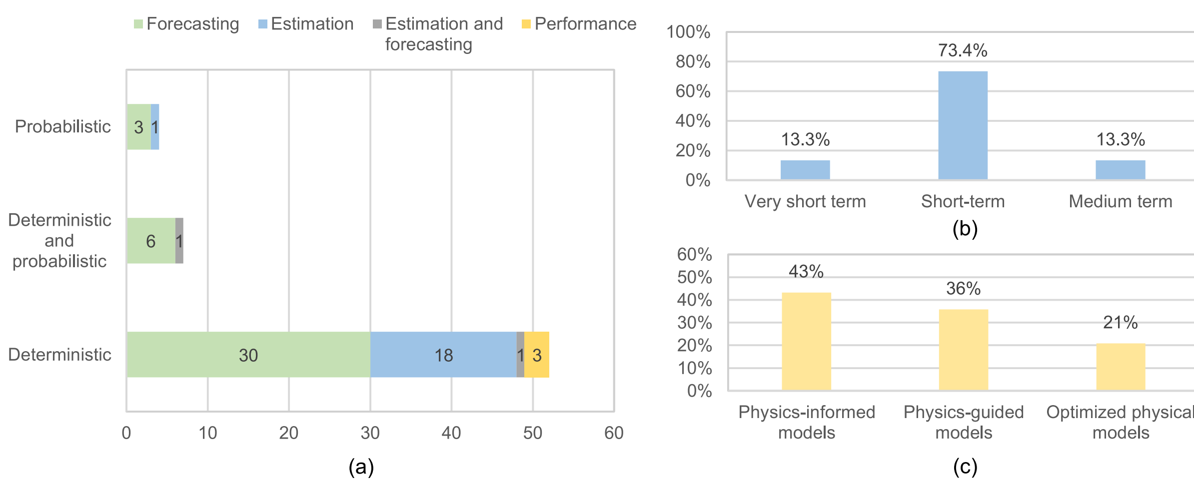


Fig. 3. Distributions of the publications per (a) methodology and objective, (b) forecast horizon, and (c) category.

shown in Fig. 3(b). Short-term forecasts, ranging from 1 h to 24 h ahead, include studies aiming at intraday and day-ahead predictions. These models are particularly useful for providing decision support in real-time dispatching. A smaller number of publications focus on very short-term and medium-term forecasts, 13.3% each. Very short-term forecast models, covering timeframes from milliseconds to 1 h ahead, include nowcast and intrahour forecasts, which are useful for power management. Meanwhile, medium-term forecast models, those applied for 24 h to one week ahead, generally serve similar purposes as short-term models, operational security, and for the electricity market. Notably, none of the selected studies implement hybrid models for long-term forecasts, which span from one week to months or even longer. This time horizon is useful for PV plants global management, sizing, investments analyses, and payback time calculations. Common models for long-term estimations are mainly physical models, widely adopted in PV software due to their reliability and generalizability. Given the lack of studies employing hybrid techniques for long-term forecasts in our review, their potential benefits or challenges for this time horizon remain unknown, indicating that hybrid techniques have the potential to be further explored for long-term forecasts. As evidenced, most studies in this field focus on very short-term and short-term forecasting, particularly for applications related to the dispatch and integration of PV electricity with the grid. This discrepancy, compared to the smaller number of studies on medium- and long-term forecasting models, is

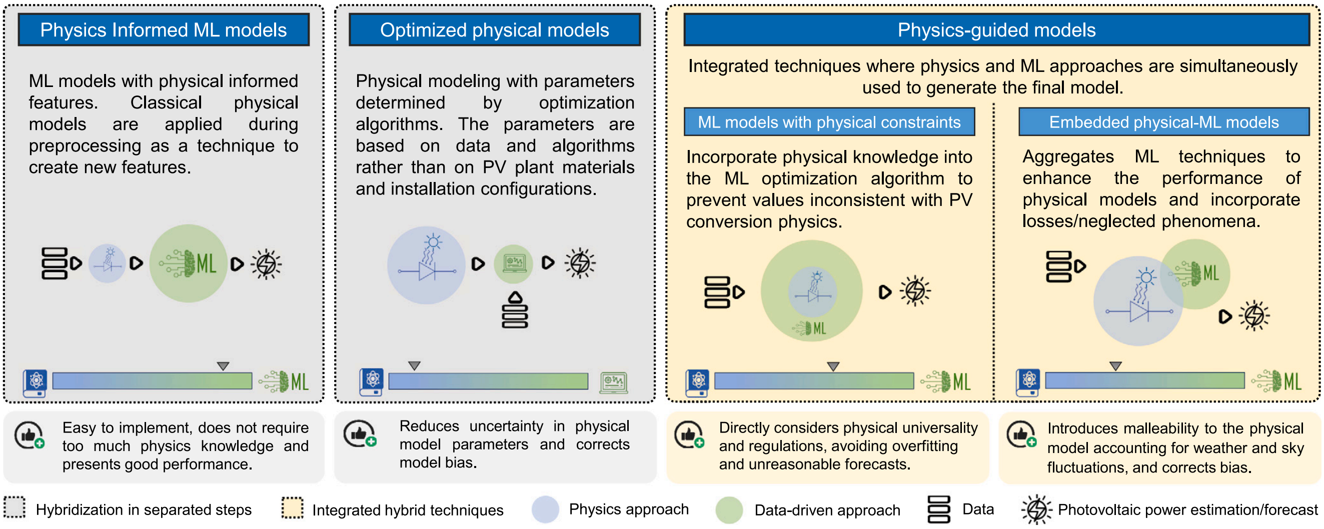


Fig. 4. Overview of distinct hybrid methods for PV power, including definition and advantages, illustrating the hybridization framework, and highlighting the varying degrees of reliance on physical and data-driven approaches — represented by the size of the blue and green circles. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

likely due to the increasing attention being given to the integration of intermittent energy sources, particularly solar PV, which is more variable and needs to be predicted on a finer time scale.

After determining the elements to be included in the bibliometric analysis, the obtained publications were carefully analyzed to characterize the hybrid methodologies in the current literature. Our analysis led to a categorization of the hybrid methods used according to their modeling approach, which resulted in three main categories:

1. Physics-informed ML models: ML models with theoretically formulated or physically relevant regressors;
2. Optimized physical models: models mainly based on physical equations, with optimization algorithms for parameter values determination;
3. Physics-guided models: methodologies considering analytical physical models and ML techniques in a collaborative approach.

The distribution of studies by category is represented in Fig. 3(c). Physics-informed ML and the optimized physical models both use physics and data-driven techniques, but in separate steps. The first approach employs physical models to expand ML model features, instead of using mathematical routines or physical representations, these techniques focus on learn the key information patterns from the data provided, showing strong forecast performance. Meanwhile, the optimized physical models are mainly physical but use data to refine the model, this approach mitigates model bias and uncertainty in physical model parameters, determining them based on data rather than based on PV plant materials and installation configuration. The third group, physics-guided models, refers to more integrated techniques where physical and ML approaches are simultaneously used to generate the final model. This group is divided into two subgroups: ML models integrated with physical constraints, where the learning procedure is not only driven by data but also by domain knowledge avoiding overfitting and unreasonable forecasts; and embedded physical-ML models, a collaborative approach primarily using ML techniques to enhance the performance of physical models by incorporating losses and all other unknown and neglected physical phenomena not described by traditional models, integrating phenomena like seasonality and accounting for weather fluctuations. The studies by category are illustrated in Fig. 4, summarized in Table 1, and discussed individually in the following sections.

3. Physics-informed ML models

The methodologies classified in this category consist mainly of ML models with regressors generated by PV power physical equations or knowledge. The physical models are applied as a pre-processing technique before the ML models' training/testing, i.e. the features dataset of the ML model is expanded with physics-informed metrics. The called physics-informed features are outputs or components of one or more models involved in PV modules PMC, which aims to describe: the effective irradiance absorbed by the PV module (involving solar position, and optical phenomena), and the percentage of this absorbed irradiance that is converted in heat (thermal models), electrical energy (PV conversion models), and energy lost (as the loses caused by the inverter). This procedure for dataset expansion combines basic meteorological features widely available to compute new physics-informed ones, allowing the ML model to learn the physical relationships among the different features. The works classified in this category are presented in Table 2. From the 30 studies identified for this category, most of the contributions present deterministic techniques, a smaller number present probabilistic techniques [25–27], or both approaches [28–31]. Among the forecasting works, the temporal resolution ranges from intra-hour to 72 h ahead. Short-term forecast studies predominate in this category (62%), with very short and medium-term studies being exceptions.

The next two sections discuss separately the features engineering process and ML techniques adopted in the publications classified in this category, including the physical approaches used for new features generation.

3.1. Data source and physics-informed features

3.1.1. Pre-processing techniques

An accurate forecast mainly depends on good raw data quality; hence, in most of the reviewed publications, data have been preprocessed, making them coherent and improving the learning procedure. A common preprocessing strategy is discarding nighttime values to decrease the scale of the dataset, considering that there is no generation during this time. This has been achieved through different approaches, such as filtering a range of hours, for instance between 6:00 A.M. to 6:00 P.M. as in [26,36]; filtering for positive clear sky irradiance values [28]; or by analyzing the moment the Sun is above the horizon, i.e., when the solar zenith angle is $< 90°$ [3,27]. The PVlib Python

Table 1
Comparative analysis of hybrid approaches for PV power forecasting.

Criteria	Physics-informed ML models	Optimized physical models	Physics-guided models	
			ML models with physical constraints	Embedded physical-ML models
Primary approach	ML model	Physical model	ML model	Physical model
Required physical knowledge	Basic domain	High domain	Basic to medium	High domain
Required data science knowledge	High domain	Basic domain	Depends on the technique	Depends on the technique
Historical data	Required	Required	Required	Required
PV system specifications	Required	Optional	Optional	Optional
Computation power & time	High	Low–medium	Medium–high	Medium–high
Generalization ability	Limited	Good	Limited	Good
Captures hidden temporal/spatial phenomena	Yes	No	Yes	Yes
Application	Intra-hour to days-ahead PV power forecast	Power estimation, performance analysis, and any term forecast	From very short- to long-term forecasting, aggregated forecasts, and digital twins.	From very short- to long-term forecasting, aggregated forecasts, and digital twins.

Table 2
Overview of physics-informed ML methodologies demonstrating physical and ML models implemented.

Ref.	Physical models for dataset expansion					ML techniques	Site	Horizon ahead	Data length
	Irrad. Adjust.	Reflect.	Temp.	Elect.	Loss*				
[27]^{P}	✓	✓	✓	✓	✓	Calibrated PMC & QR	8 PVPs in Hungary	24–48 h	2 years
[3]^{D}	✓	✓	✓	✓	✓	Optimized PMC & MLP-ANN	14 PVPs in Hungary	1 day	3 years
[28]^{DP}	✓		✓	✓		Persistence, MLR, RF, QRF, QR & CSPE	2 PVP: Netherlands & Italy	36 h	3 years
[32]^{D}	✓		✓	✓		CNN, ANN, KNN & SVM	5 PVPs in Australia	–	> 1 year
[33]^{D}	✓		✓	✓	✓	RF, SVM, SPAR, CNN, LSTM, CNN-LSTM, Persistence & ARIMA	1 PVP Denmark	24 h	15 months
[20]^{D}	✓		✓	✓	✓	RF, SVM, SPAR, CNN, LSTM, CNN-LSTM & Persistence	1 PVP Denmark	5 h	15 months
[34]^{D}	✓			✓		Persistence, SVM, RF & GBR	4 PVPs in China	Intra-hour	± 1 year
[35]^{D}	✓		✓	✓		ANN	1 PVP in Germany	12 h	1 year
[36]^{D}			✓	✓		AML, GA, DT, LR, K-NN, RF, AB, GBR, BR, PSO, LSTM & PSO-SVM	6 PVPs in Northern Japan	1 day	2 years
[37]^{D}	✓		✓			KNN, SVM, ANN & Weighted KNN	3 PVPs in Australia	1 day	2 years
[38]^{D}			✓			MLR & ANN	1 PVP in Jordan	–	3.5 years
[39]^{D}			✓			ANN	1 PVP in Germany	–	6 months
[30]^{D}	✓					Polynomial regression	3 PVPs in Korea	–	5 months
[25]^{P}	✓					QR	1 PVP in USA	1 day	4 years
[26]^{P}	✓					QR, QRA, RF, MLR, ARCH, QRF & QkNN	3 PVPs in USA	1 day	1–2 years
[40]^{DP}	✓					LSTM, DeepAR	119 PVPs in Luxemburg	0–72 h	2 years
[31]^{DP}	✓					Persistence, SVM, ELM, RT, GBR-DT, CNN, LSTM, ARIMAX & KDE	2 PVPs in Belgium	2, 3, 4 & 12 h	3 years
[41]^{D}	✓					ANN, Persistence & PHANN	1 PVP in Italy	1 day	2 years
[42]^{D}	✓					PHANN	1 PVP in Italy	1 day	–
[43]^{D}	✓					PHANN	2 PVPs in Italy	24 h	–
[44]^{D}	✓					Persistence, RHNN & Ensemble of ANNs	1 PVP in Italy	1 day	4 years
[29]^{DP}	✓					PHANN	1 PVP in Italy	24 h	10 months
[45]^{D}	✓					PHANN & ANN	1 PVP in Italy	1 day	240 days
[46]^{D}	✓					PHANN, ARX-Wavelet-NN & ANFIS	1 PVP in Italy	–	6 years
[47]^{D}	✓					PHANN	3 PVPs in Italy	1 day	1 year
[48]^{D}	✓					PHANN	4 PVPs in Italy	1 day	1 year
[49]^{D}	✓					PHANN & ANN	1 PVP in Italy	24 h	10 months
[50]^{D}	✓					BP-ANN, Elman-ANN & RF-ANN	1 PVP in China	–	345 days
[51]^{D}	✓					BP-ANN	1 PVP in Jordan	–	3 years

Model type: ^{D}deterministic; ^{P}probabilistic; ^{DP}deterministic and probabilistic.
*Including inverter, shadow, soiling, and other models to account for losses.

package is used in [34] to calculate the time of sunrise and sunset to eliminate data points at night. Removal of the daytime timesteps with no PV production, which occur during a PV plant shutdown or malfunction, has been also implemented [3,25,27,51].

Since historical observations are often vulnerable to data quality problems, such as outliers, abnormal and missing data, discontinuous time-series, etc., data cleaning is vital. Outlier detection and subsequent correction are accomplished in the reviewed publications using a B-spline smoothing approach [52] or an improved sliding window prediction-based approach [25,26]. The $3-\sigma$ principle is used in [34] to find outliers in power data, and the average value is used to replace them. Also, negative values recorded in solar radiation are typically set to zero [38,51]. To obtain a uniform time-series, missing values are updated with recent values in [25,26,52], using a suitable interpolation function in [36,42,50], or the average value in [36]. In some publications, days with a number of missing data exceeding a threshold established by the authors are excluded [34]. Data validation can be performed to ensure data quality; in [49], this is implemented on the basis of comparisons between solar radiation, PV power, and theoretical radiation predicted by a mathematical model [49]. Days with misleading power profiles, when the power is really close to zero despite an appreciable level of radiation, are discarded in [42].

Considering PV data, the input data can have a great difference in the range of magnitude and data variation due to different units. To eliminate the effect of magnitude between indicators, data standardization processing, i.e., normalization, is required. After normalization, the indicators become in the same order of magnitude, which is suitable for comprehensive comparative evaluation [34,35,50]. The input dataset values are typically scaled to be in the range of [0–1] [28,38,51]. PV power is normalized to the reported installed AC capacity in [28].

Meteorological variables, typically obtained from weather monitoring systems, are common inputs used of the reviewed models in this category. The most common is solar irradiance — directly responsible for the PV conversion [3,20,28–31,33–35,37,38,41,42,46–48,51], followed by ambient temperature [3,20,28–31,33,35–38,41,42,46–48,50,51], and wind speed [3,20,28–30,33,35,36,38,42,47,48,50,51], both affecting solar energy loss as heat from the PV module to the environment. Other meteorological variables that affect the PV performance but are considered "second-order effects" — because PV conversion can still be physically described with certain degree of fidelity without them — are also used as features in a minor number of works. These include wind direction [29,33,35–37,47,48], which influences PV module convective heat loss from the PV module; relative humidity [20,30,33,35,42,47,48], which can contribute to dew formation on the PV panel surface; atmospheric pressure [28,29,33,47,48]; precipitation level [28–30,36,47,48]; and cloud cover [28,29,42] or clear sky index [31].

An analysis of the Pearson correlation coefficient between the features and PV power is performed in [3,20,28,31,33,34,36,38,50,51] and can be a helpful approach in quantifying the value of predictor variables for PV power estimation and forecasting. However, an alternative feature selection method is studied in [20,51], comparing features choice based on Pearson coefficients and on physical relevance, showing that the latter demonstrates a more significant improvement in ML model performance.

Finally, another strategy that can make the subsequent ML process accurate and efficient is to create new variables to expand the dataset. This has been done by clustering analysis of weather features in [32], grouping the data by weather or day type in [3,32,33,35,37,44,48,49,51], or by season in [3,26,31–33,37,38,51]. The common approach of the 30 works classified in this category is the dataset expansion with physically relevant metrics, discussed in more detail in Section 3.1.2.

3.1.2. Physical approaches for dataset expansion

The physical modeling approach relies on a PMC consisting of a sequence of physical or empirical models jointly characterizing how irradiance is converted to PV power. The PMC includes models for separation and transposition to estimate the effective irradiation on the PV module surface, an optical model to account for reflection losses and portion of irradiance absorbed by the PV cells, a temperature model to estimate the PV cell operating temperature, an electrical model for PV conversion, an inverter model and optional PMC models to account for losses caused by shadowing and soiling. This approach requires certain domain knowledge and, in several recent works, the idea of PMC-based PV power forecasting has been modernized [3]. A popular use of the PMC is as a pre-processing step, where the output of the separated physical models is used as features in ML models, as illustrated in Fig. 5 called physics-informed features. The introduction of these new variables significantly improves the performance of the estimation and forecasting models [28]. As advantage, this technique facilitates the learning procedure. These physics-informed features capture the relationship between weather and PV operational state, while keeping strong correlation towards the intrinsic feature. This allows the ML models to learn about the physical interdependence of different features, potentially yielding a higher accuracy than conventional methods [20]. For example, instead of solely provide data of ambient temperature and wind speed and expect the machine learning to understand their impact on the PV power, we can use both values to calculate the module temperature and include this temperature as extra input, the last presents a better coefficient of correlation with PV power than both wind and ambient temperature individually [28]. In this way, ML needs lees efforts to learn the nonlinearities of the irradiance-to-power conversion, and can focus on identifying and correcting recurring error patterns [3].

As shown in Table 2, the models used to adjust the incident irradiance are the most common. Transposition and separation models are highly adopted to create new variables. According to the literature, these are the two most important physical modeling steps; the rest of the physical PV simulation can be learned by ML in hybrid models without a significant increase in errors [3]. According to [28], the variables generated by the transposition model are even more valuable. These physical models use the location of PV stations and the tilt angle of PV panels to convert the measured irradiance into plane-of-array (POA) irradiance. Meanwhile, the ML approach is able to correct bias errors caused by any physical approximation during the physical model development, providing better irradiance transposition on the PV plane [44].

The clear sky irradiance computed according to the PVP geographical coordinates by separation models is the most common physical feature [28,29,31,33,44–49]. Common approaches employed are the solar spectral model of Richard Bird [53] to generate direct-normal and diffuse horizontal irradiance, commonly used in the PHANN works [45,49]; the McClear method [54] or the Hottel model [55] to obtain the direct irradiance on the horizontal plane, and the GHI [31,48]; or open-source tools like PVlib, which provides different functions for irradiance separation, such as the Erbs model [28,34,56]. However, there are many mature clear-sky models that can be used for this task, more than 80 clear-sky models are presented in [57].

For irradiance transposition, the models project horizontal irradiance components to those on an inclined surface, determining the POA irradiance [20,28,29,33,37,41,42,47,51]. Common models include the Liu–Jordan isotropic transposition model [44,58], and the most common one, the Perez model [59], which is included in PVlib and one of the two models offered by PVsyst [28,34]. Unless the irradiance transposition can be simply calculated analytically [30,32], as well as for separation, there are several different models in the literature [60]. M. Mayer works are the only ones in this review that test more than one separation and transposition model and compare their performances in creating physical features [3,27].

A strategy to avoid the physical models complexity and assist the ML models in understanding the physical phenomena involving effective irradiance determination is to use the input of the separation and/or transposition models as features. In this case, some studies use physical models only to guide the choice of new features, avoiding directly applying the equations. Thus, as the transposition and separation models depend on time, day of the year, geographical coordinates, etc., it is understood that those are factors correlated with the effective irradiance and used as features without the need to put them in a physical model. Day of the year is included directly as features in [3,38,40,47,48,51], hour of the day in [20,27,33,35,38,40,47,48,51], and geographical coordinates in [29,34,35,45]. This information is used for the calculation of solar position and included as input in [36,49,51], and for the solar azimuth and zenith angle in [25,28,33,35]. The sine function of solar elevation angle and of solar azimuth angle, and cosine function of the incident angle of solar radiation are used as features in [25,26].

After irradiance determination, temperature models are the most commonly adopted in the studies identified in this review [20,28,33,36–38,47], as shown in Table 2. The module operating temperature is highly related to its efficiency; hence, the use of this temperature as a feature is a good strategy to incorporate the temperature effect on the ML models. There are many temperature models in the literature [61]; some of these models used to generate ML features are: linear temperature estimation by meteorological factors [35]; Faiman model [32]; Standard, Skoplaki, Kurtz, Koehl, Mattei, and Homer models [38]; and most commonly, the Sandia model, available in PVlib and used in [3,20,27,28,33,34]. PV operating temperature is estimated by means of developed dynamic temperature models in [33,39]. These more detailed models present better performance when compared with real PV temperature data than the common steady-state models used in the previously described publications at the beginning of this paragraph [62]. Temperature models have been compared in [3,27,38];

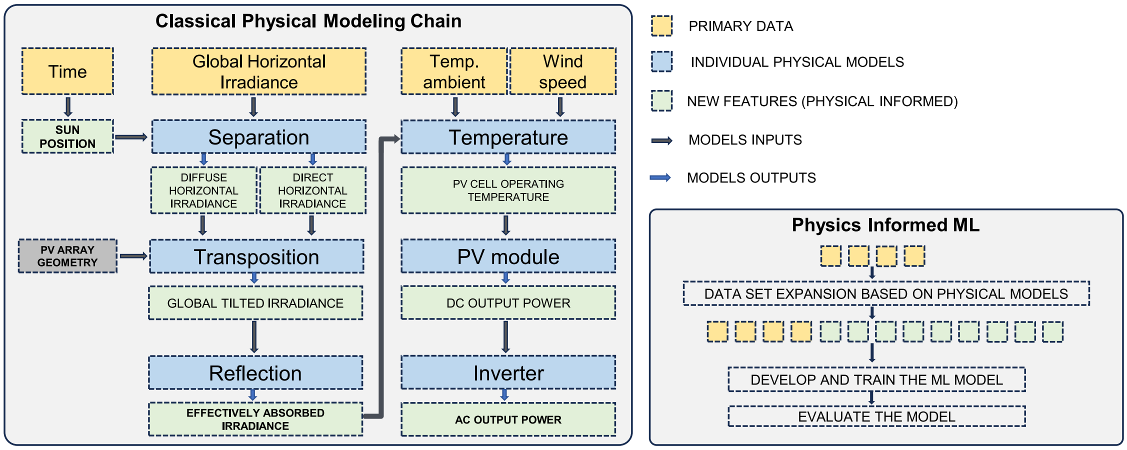


Fig. 5. Classical physical model chain and its use with ML models.

however, the best model for each PVP depends on several factors. Hence, if temperature data are available, it is important to compare different models before applying them.

In some research, the electrical conversion is left to the ML model, while others also use the output power from physical models as features, which may help in the learning procedure. In such cases, two kinds of physical models are implemented: electrical models, in which the output is the DC power [20,28,33,46], and inverter models, focusing on the AC power [3,20,27,33]. PVWatts model [63], considering the temperature impact on the yield, is the most commonly employed [28,34–37]; the same model, including a pollution coefficient, is adopted in [32], and including a normalized coefficient for temperature and power in [20,33]. Different models are tested for DC power estimation in [3,34], including the single diode model. AC power is calculated by multiplying the DC power by the inverter efficiency [20,33], or using different inverter models [3,27].

3.2. Common ML and benchmark models in physics-informed ML

Considering probabilistic approaches, quantile regression (QR) is the most commonly used model, implemented in [25–28,52], followed by quantile random forest (QRF), implemented in [26,28,52]. Besides these, quantile regression averaging (QRA) and quantile k-Nearest Neighbors (QkNN) are also used in [26]; clear sky persistence ensemble (CSPE) in [28]; an optimized QRF in [52]; and a DeepAR is implemented in [40]. Meanwhile, a novel framework is proposed in [31] for short-term PV power forecast using a deep attention Convolutional Long Short-Term Memory (LSTM) network and kernel density estimation (KDE). Finally, PHANN methods with probability densities of the maximum hourly errors are used in [29]. Models that have been used to benchmark probabilistic approaches are CSPE, QRF, QkNN, and basic QR [26,28,52]. Benchmark against persistence [26,28,31], and ML models without the physics informed regressors in [28,31] shows that these inputs contribute to the models' performance enhancement.

Among studies presenting deterministic approaches, ANNs have significant prominence compared to other ML models (see the summary in Table 1), being used with predictors derived from a Clear Sky model, known as PHANN in [29,41–43,45–49], and also with various other physical features in [3,20,31–33,35–39,41,44–46,49–51].

PHANN is a well-explored and established hybrid method based on ANN, with a Clear Sky solar radiation model [49]. A new version of this method is proposed in [45], implementing a next-day mobile window forecast. In both works, PHANN is compared with a standard ANN method and shown to be more accurate. PHANN is compared with a nonlinear AutoRegressive exogenous (ARX) model in [46], which uses a wavelet network as its nonlinearity estimator, and an approach based on Adaptive Network-based Fuzzy Inference System (ANFIS), both using the same structure as PHANN. These hybrid methods are also benchmarked against both a pure physical model of the PVP and a pure ML method, a Multi-Layer Perceptron (MLP) ANN. Results demonstrate that the PHANN approach allows for a more precise estimation than the other four models [46]. In [41,47], the authors focused on finding the best neural network configuration for the PHANN model. A tolerance threshold is included in the error assessment, and a corrective factor has been applied to the output of the model to minimize the final error in [47]. PHANN accuracy is assessed in [48] according to the day type (sunny, partially cloudy, and cloudy) and the number of days employed in the training. In [29], PHANN is compared with two well-established physical models, the three and five-parameter electric equivalent circuit; PHANN achieves the best results. In [43], a PHANN is applied to provide the 24 hour ahead PV power forecast for microgrids. In [42], a PHANN is implemented for computing the PV power forecast on different time horizons and resolutions. In summary, PHANN is a well-defined method in the literature and has been tested in different arrangements, time scales, and horizons, and successfully outperformed the performance of pure physical and stochastic models.

Other publications propose hybrid techniques with new physical features being utilized to build ANN models. A Relative Humidity Neural Network (RHNN) model is proposed in [44]. The RHNN uses the relative humidity of different vertical levels to forecast the clear sky performance index and forecast PV power. The RHNN performance is benchmarked against a persistence model and a PMC presenting enhanced accuracy. A Backpropagation Artificial Neural Network (BP-ANN) with physically relevant features is presented in [51]; the approach has shown superior performance compared to models employing the old features and other features extracted using principal component analysis-selected features. Authors reconstruct the input of a Convolutional Neural Network (CNN) in [32] by using key weather features from analytical physical modeling, improving its performance. Irradiance components are used as input into a BP-ANN, Elman ANN, and RF ANN in [50], results show that all three ML models have higher accuracy with the extra physical features. Hybrid methods that involve physically calculated predictors and an ANN are compared to physical and ML models without any physical considerations in [3,35]; both works show that the combination provides the most accurate forecast values. A detailed physics-based electro-thermal model is validated along with other state-of-the-art models in [39]. Dynamic modeling results are transferred to an ANN model, further increasing the accuracy up to six times better than the considered parametric solutions. In [38], two main prediction models, namely Multiple linear regression (MLR) and ANNs are tested substituting ambient temperature with PV cell temperature and comparing the results: ANN with module temperature as predictor presents the best performance metrics.

Besides ANNs, the most used models are Support Vector Machine (SVM) in [20,31–34,37], Random Forest (RF) in [20,28,33,34,36], LSTM in [20,31,33,40], KNN in [32,36,37], and Gradient Boosting Regressor (GBR) in [31,34,36]. ML models are implemented in [28] using diverse groups of physically informed variables, results show that these inputs contribute to the models' performance enhancement. Seven different types of traditional prediction models based on feature engineering and grid search parameter optimization and three forecast models (Particle Swarm Optimization (PSO)-LSTM, PSO-BI-LSTM, and PSO-SVM) with parameter optimization-based PSO algorithm are implemented in [36]. An attention mechanism is introduced into ConvLSTM in [31] to adjust the weights of the PV data; experiments show that the proposed method can realize the optimal fusion of the historical data and clear-sky prior knowledge, and significantly improve the forecast accuracy in all seasons. Single-point forecast models performance is benchmarked mostly using persistence [20,28,31,33,34,40,41,44] and other conventional ML methods [20,31–34,36]. In studies implementing PHANN, ANNs are commonly used for benchmark [3,32,35,41,45,46,49]. Other authors use a physical model as a benchmark, such as the 3 and 5 parameters PV circuit models, PVSyst, and PVlib [28,29,35,39].

3.3. Discussions

In conclusion, the publications classified as physics-informed ML are those consisting of ML models with physics knowledge included in the form of features. The models used to generate these new features are physical equations classically employed in PMCs. The use of such classical models in ML models is a modern and promising strategy for PV power forecasting. As the development of physical models involves a series of physical approximations, the ML models can help incorporate the effects disregarded in the physical models, such as seasonality, soiling, and shadow cycles. However, some points about these models are worth mentioning:

1. Optical models for accounting for reflection losses have been neglected in most of the works. As shown in Table 2, these models are only directly considered in [3,27]. No studies addressing the potential benefits of integrating optical models to generate features for ML models in PV power forecasting were found in this literature review.
2. Most of the publications in this area implement simplified physical models for feature determination; only [33,39] bring new models to generate physical features, both presenting better results when compared to the use of classical simplified equations.
3. Different physical models have not been well explored. Exceptions for that are the works in [3,27], and [38], where different models are implemented and compared for each feature. There are a large number of physical models for each modeling step, and the impact of the model choice is currently unknown. In most of these works, authors implement the basic models of common libraries such as PVlib and do not compare other approaches for feature estimation. There is no general model for PV estimation, as the best model can differ depending on several factors, such as the model generalizability in terms of PV technology and local. Even though the ML model can learn even with some systematic error in the physical model, the impact of an improvement in the physical model used to generate the features on the ML model performance is not well addressed in the current literature.
4. Another point is the benchmark for these hybrid models. Most of the publications compare these hybrid models with physical features with persistence models or classical ML models such as SVM and ANN without any physics information. The results show that the incorporation of physical features improves the model results. However, now that it is already known, the question is: is it fair to continue comparing ML models that do not receive the same inputs?
   On the other hand, some authors use purely physical models as benchmarks. This decision can also be controversial, as the great advantage of physical models is their generality, explainability, and good performance for long-term estimation; while most of the works referred to here are based on a single PVP and with a short-term horizon, where ML models usually show a better performance.
   As there is no established hybrid model for benchmarking, it is a difficult challenge to address. The PHANN model is the only hybrid technique found in our review that has been well explored, tested with different physical and ML models, in different PVP and forecast horizons. Hence, PHANN can be a candidate as a benchmark model in this area. Another option is to compare new hybrid models with already published ones that use the same physical features.

4. Optimized physical models

The methodologies categorized here primarily focus on physical models describing the irradiance to power conversion. These model parameters are refined or determined by optimization algorithms and thus applied for PV power estimation. Fig. 6 presents a visual illustration of the process. By integrating a data-driven approach hybridized with classical physical models, these methods can address unmodeled uncertainties and nonlinearities that are not captured when using physical models only [64]. Although there are several publications exploring this approach, it is predominantly applied for PV performance analysis and control rather than for PV power estimation and forecasting. In this section, we explore the 15 publications found that implement optimized physical models selected according to our review criteria, summarized in Table 3.

These methodologies are mainly characterized by deterministic techniques, as probabilistic approaches are notably absent. Moreover, the majority are focused on PV power estimation (80%) [65,67–75,77,78]. The publications aiming to forecast PV power (20%) present forecast horizons between 24 to 120 h [64,66], with some applications also targeting now-cast forecasting [76]. Since these models rely mostly on physical modeling, involving a detailed design of the panel's installation and configuration, they are typically applied for a single PVP. Exceptions include [75,76], both based on the PMC proposed by the Sandia Laboratory in [79] and known for their good generalization ability. Due to the emphasis on physical modeling, most of the publications here barely, or simply do not, explain the data used and the preprocessing steps. According to Table 3, most studies in this category are based on data spanning approximately one year, although some are conducted over a few days [77]. It is important to mention that even the methodologies based on limited data can demonstrate good local generalizability, given their reliance on fundamental PV system physics, which is universal across locations, avoiding "regional bias".

The optimized physical models range from single equations for electrical conversion to more complex coupled models in PMCs. As shown in Table 3, the single-diode equivalent circuit model (SDM) is the most frequently employed physical model in this category. This model's parameters have a nonlinear relationship with PV power and may vary over time due to aging, necessitating optimization for accurate power estimation [70]. Diverse SDM parameters have been optimized or determined, including physical parameters and their variation with irradiance, temperature, built-in voltage, and non-physical parameters such as the ideality factor n, α, and γ [65–72]. Furthermore, variations of the SDM are also explored, in [70] the five SDM parameters are converted into 13 parameters optimized under various weather conditions.

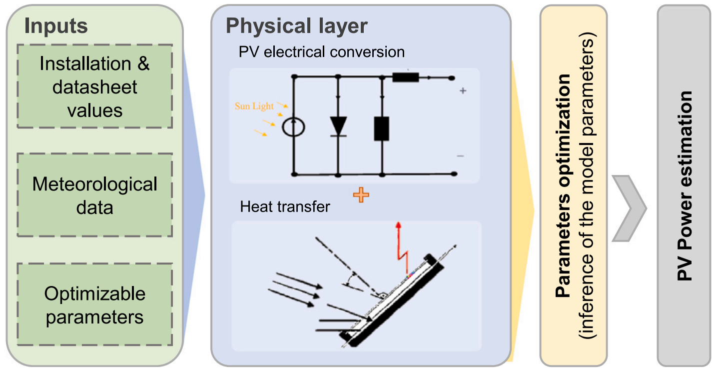


Fig. 6. Example of an optimized (thermo–electrical) physical model for PV power estimation.

Table 3
Overview of methodologies incorporating physical models and optimization techniques.

Ref.	Physical model	Optimization technique	Site	Data length
[65][D]	Five-parameter SDM	Least-squares method	Different PV panels in Belgium	–
[66][D]	Five-parameter SDM	Social Network Optimization	1 PVP in Italy	1 year
[67][D]	Five-parameter SDM	Trust-Region Dogleg Method	1 PVP in Senegal	1–2 years
[68][D]	Five-parameter SDM	Guaranteed Convergence PSO (GCPSO)	2 different PV modules in USA	–
[69][D]	Five-parameter SDM	Artificial Bee Colony algorithm, DE and PSO	1 PVP in Algeria	–
[70][D]	Modified five-parameter SDM	Hybrid charged system search, DE, and PSO	1 PVP in Taiwan	–
[71][D]	Analytical equation inspired by SDM	Least-Squares Method	1 PVP in Italy	1 year
[72]	Five-parameter SDM + dynamic thermal model	Algorithm based on the Newton–Raphson method	1 PVP in USA	1 year
[73][D]	Sandia PV-PMC	PSO	1 PVP in Germany	1 year
[74][D]	Sandia PV-PMC	Not specified	14 PVPs in USA	–
[75][D]	Sandia PV-PMC	Pattern Search (PS) and PSO	1 PVP in UAE	40 days
[64][D]	Sandia PV-PMC	Genetic algorithm optimization	16 PVPs in Germany	2 years
[76][D]	Optimized PV-PMC	Not specified	1 PVP in Italy	–
[77][D]	PMC	Newton–Raphson and Runge–Kutta methods	1 PVP in Turkey	4 days
[78][D]	Sandia PV model	Least-squares method	1 PVP in Algeria	5 months

[D] deterministic models.

Additionally, in [71], a compact semi-empirical model inspired by the SDM is proposed. Apart from the SDM model, calibration is applied to optimize the Sandia array performance model parameters in [74,78]. Coupled physical models are also optimized and applied for PV power estimation and forecasting. In [72,77], the SDM is integrated with a thermal transient model based on heat transfer PDE to model both PV electrical and thermal behaviors. Similarly, [73] compares different coupled models, including three electrical models (three, four, and five-parameter SDM) and two thermal models for PV cell temperature estimation (NOCT and Sandia). Beyond these thermo–electrical models, parameter determination techniques are also applied to other physical models in PMCs, most commonly the Sandia PV-PMC [64,75,76].

Although this category of models is primarily physical, findings from [73] underscore that the accuracy of these models is more influenced by the calibration data and method. The optimization methods employed to tune the above-mentioned physical models include classic approaches such as the least-squares method [65,71,78], and techniques based on the trust-region Dogleg method [67]. As well as meta-heuristic algorithms, such as the Artificial Bee Colony (ABC) [69], Pattern Search (PS) [75], Hybrid Charged System Search (HCSS) [70], social network optimization [66], Genetic Algorithm (GA) optimization [76], Differential Evolution (DE) and Particle Swarm Optimization (PSO) [68–70,73,75]. Additionally, methodologies such as those in [72,77] incorporate the Newton–Raphson method alongside the physical models to optimize system parameters.

To assess the effectiveness of optimizing the physical model parameters, these methods have been benchmarked with different models in literature, including purely physical, purely ML, and other hybrid models. The most basic comparison involves using the same physical model but with the manufacturer parameters, significant improvements in estimation results have been observed after the optimization [64–66,70,75,77]. Other physical models, such as those based on maximum power point calculation [66,68,69] and the PVsyst model [65], have also been used as benchmarks. Also, some publications use ML models as benchmarks, as in [66,67], where the optimized SDM is compared with an ANN. In all these cases, a lower power estimation error is achieved by using the optimized physical model, indicating the effectiveness of hybridization compared to the single physical or ML models. Considering comparisons with other hybrid modeling approaches, it typically involves using the same physical model but with different optimization approaches. In [69,70], PSO and DE methods are also used for SDM optimization and compared with the proposed methodologies (HCSS algorithm in [70] and the ABC algorithm in [69]), both demonstrating better results in terms of local minimum avoidance and accuracy compared to PSO and DE. Moreover, in [66], the optimized physical model is compared with two physics-informed ML models: a PHANN and an ANN with the power calculated by a physical model as an extra feature, revealing comparable performance among the three hybrid methods.

4.1. Discussions

The models classified as optimized physical models are primarily based on physical models, with parameters optimized or determined through optimization techniques. Across the examined publications, the modeling methodology remains consistent: choosing a physical model, usually the SDM or a well-established PMC, preceded by optimization of its physical parameters to enhance the accuracy of PV power estimation. This approach's main advantages are mitigating uncertainty in physical model parameters and determining these parameters based on data rather than based on PV plant materials and installation configuration, other than that, it shares the same strengths and limitations as purely physical models. Some points regarding the publications found in this category are worth highlighting:

1. Only deterministic techniques have been published in this category for PV power estimation and forecast. These methodologies are characterized by estimation techniques, a few studies focus on forecasting. However, the publications aiming to forecast PV power highlight the potential of this hybrid approach for such applications [64,66,76].
2. Most of the publications in this category consist of electrical models' parameters optimization, while studies incorporating thermal or multiphysical modeling predominantly rely on a limited set of simplified classical equations. Explorations into diverse physical models are sparse, only [72,77], and [73] bring more detailed models presenting better results when compared to the utilization of classical equations.
3. Benchmarking against different models in literature, including purely physical, purely ML, and hybrid models, reveals the effectiveness of this hybridization over single physical or ML models. However, findings from [66] suggest that these hybrid methodologies, consisting of optimizing the physical model parameters, have comparable performance to the physics-informed ML models. Notably, this hybrid approach, primarily physical, can be more complex and demand a deeper understanding of physical concepts compared to the physics-informed ML models. Therefore, additional investigation can be conducted in this line to effectively determine how well the ML models, through the features, can learn and describe the underlying physics of PV systems compared to the purely physical and well-established models.

5. Physics-guided models

The methods in this category blend physical and ML approaches. Much of the research here involves integrating physical models or knowledge as constraints into the ML optimization algorithm, as discussed in 5.1. In other publications, these methodologies adopt a collaborative approach, as outlined in 5.2. In some cases, ML models are used to improve the physical models by predicting the error, applying loss terms, or estimating the model uncertainty to generate probabilistic forecasts. Alternatively, ML techniques may be applied first to adjust inputs or cluster data before estimating PV power using a physical model. Additionally, some publications initially apply both ML and physical models independently and then weigh their outputs to produce the final PV power. Since the physics-guided methodologies directly incorporate physical modeling and use ML as a strategy to improve the model, they are commonly applied for PVPs' digital twins [80–82]. These methodologies for PV power forecasting/estimation are characterized by deterministic techniques (79.2%), a limited number of publications implement probabilistic (12.5%) [83–85], or both (8.3%) [86,87] approaches. Also, the focus is on PV power forecasting (78.3%) with forecast horizons ranging from 100 ms to several days ahead. However, some publications aim to estimate PV power (21.7%) [80–83,85,88].

5.1. ML models with physical constraints

The research belonging to this sub-category is classified as ML models with physical constraints, which are summarized in Table 4. These methodologies incorporate domain-specific knowledge about PV in the form of integrated physical constraints, which guide the ML model alongside data, or are integrated as a post-processing technique to correct illogical predictions after the model inference [43,49,84,87–92]. Derived from fundamental physics laws or natural science knowledge of PV, these constraints enhance the model predictive accuracy, generalization ability, and robustness [88]. This approach serves to prevent illogical predictions, such as negative power output, non-zero PV power at night, or values inconsistent with PV conversion physics — like extremely low or high power output for a given solar irradiance level — by setting boundary conditions for the PV power output.

To ensure zero PV power during nighttime, a constraint can be introduced by either setting PV power to zero after sunset based on the timestamp [88–90] or employing a non-zero field check, which nullifies any non-zero PV power when zero Global Horizontal Irradiance (GHI) is detected [87]. Another constraint that is commonly implemented, aimed at maintaining a realistic profile, is the positive constraint. This constraint restricts the models' output to positive or zero values [49,88–90,92]. Additionally, boundary constraints, which involve defining maximum and minimum PV power for a given level of irradiance or set of input variables, are also commonly used [43,84,88–92]. These boundaries are determined based on historical time series data [84,89,90] or using K nearest samples [91]. In some cases, only lower or upper boundaries are determined based on factors such as maximum installed capacity [92] or the minimum amount of power required to operate the inverter [43].

To perform the constraints, the rectified linear unit (ReLU) function is employed to ensure positive or zero PV power outputs in a physics-constrained LSTM (PC-LSTM) model in [90,91], a transformer network in [89], and an ensemble CNN-LSTM model in [88]. The ReLU function returns zero for negative values and preserves the original value for positive ones. Additionally, in these publications, upper and lower boundaries are utilized to restrict the output by modifying the model's loss function. This loss function includes the penalty of errors, activating whenever the model output violates the boundary constraints. In [92], positive and upper boundary constraints are incorporated into the optimization problem formulation of the SVM model, modifying the SVM primal function to ensure a minimum bound of zero. Post-processing techniques, on the other hand, adjust outputs by enforcing similar upper and lower limits after the ML model has generated predictions, allowing models to remain unconstrained during training and inference but still adherent to physical laws. This approach requires less ML knowledge to implement, embedding no constraints directly into the model itself, offering flexibility but can require additional computational steps after model output prediction, adding complexity to the forecasting workflow.

Considering the probabilistic hybrid models for PV power incorporating physical constraints, an analog ensemble method with a non-zero field check for PV power and GHI is presented in [87]. Meanwhile, in [84], a hybrid model for PV power forecast is introduced integrating the SDM to estimate PV power AC output, a converter regression model for AC–DC conversion, along with k-means clustering to define prediction intervals. Upper and lower quantiles are extracted by the representative clusters, establishing prediction bounds for very short-term forecasting. Furthermore, the model incorporates error modeling for uncertainty assessment and addresses setpoint uncertainty based on internal control laws.

In general, ML models that incorporate physical constraints are benchmarked against conventional ML and statistical methods such as persistence, ANN, AutoRegressive Integrated Moving Average (ARIMA), KNN, CNN, LSTM, and CNN-LSTM [43,49,89–92]. These constraint-based ML models demonstrate stronger forecasting capability, pointing out that the constraints hold. The approach is also compared to other ML models with constraints, a constraint Ordinary Least Squares (OLS) in [92], and a constraint LSTM in [91], increasing the PV forecast quality. Additional advantages of such models include their suitability with sparse data [89], and enhanced forecasting under different sky conditions [91]. However, according to [49], the accuracy of the model with constraints in comparison to an ANN method on cloudy days is just slightly different. Besides, the accuracy of these methods is strictly related to the preprocessing of historical data and the accuracy of historical weather forecasts used during the training phase.

5.2. Embedded physical-ML models

Embedded physical-ML models incorporate ML and physics in a collaborative approach, primarily using ML techniques to enhance the

Table 4
Overview of ML models with physical constraints. The constraints adopted are represented by a checkmark (✓).

Ref.	Physical knowledge guided constraints				ML techniques	Site	Horizon ahead	Data length
	Night-zero $P_{out}^{Night} = 0$	Positive values $P_{out} \geq 0$	Lower boundary[a] $P_{out} \geq P_{min}$	Upper boundary $P_{out} \leq P_{max}$				
[88]^D	✓	✓	✓	✓	CNN-LSTM, LSTM, ANN, GBR, SVM	8 PVPs: Asia (area A)	–	1 year
[89]^D	✓	✓	✓	✓	LSTM, CNN, CNN-LSTM, Transf. ANN	1 PVP in Korea	Day ahead	1.5 year
[90]^D	✓	✓	✓	✓	ARMA, KNN, FCNN, LSTM, PC-LSTM	2 PVPs in Australia	Day ahead	–
[84]^P			✓	✓	Hybrid PMC, k-means	1 generic PVP	100 ms to 5 min	6–12 days
[91]^D			✓	✓	ARIMA, KNN, FCNN, LSTM, CNN-LSTM	2 PVPs in Australia	Day ahead	27 months
[92]^D		✓		✓	SVM & ordinary least squares estimator	1 PVP in Germany	–	1 year
[49]^D		✓			ANN & PHANN	1 PVP in Italy	Next 24 h	10 months
[87]^{DP}	✓				RF, GBR, ANN, and regression tree	1 PVP in Kuwait	15 min to day ahead	2 years
[43]^D			✓		PHANN	1 Microgrid in Italy	Next 24 h	–

Model type: ^{D}deterministic; ^{P}probabilistic; ^{DP}deterministic and probabilistic.

[a] Although the "Positive values" and "Lower boundary" constraints may seem similar, "lower boundary" represents the minimum output for a given set of inputs, e.g. minimum PV power for a given level of irradiation and temperature.

Table 5
Overview of embedded physical-ML methodologies incorporating physical models and ML techniques.

Ref.	Physical model	ML model	Local	Forecast horizon	Data length
[80]^D	SDM	LSTM	1 PVP in UAE	15 min	2.5 years
[93]^D	PMC including POA Irradiance, optimal losses, temperature and power conversion models	Optimization based on physically guided loss function	2 PVPs in Switzerland	–	1 year
[94]^D	PMC including POA Irradiance and power conversion models	Regression MOS	11 PVPs in Uruguay	0–6 h	1 year
[13]^D	PVlib PMC	ANN & RL	1 PVP in Greece	24 h	9 months
[95]^D	Empirical polynomial equation	ARIMA	1 PVP in Romania	1–30 min	–
[96]^D	PVWatts calculator	KNN	1 PVP in USA	–	10 months
[97]^D	PMC including POA Irradiance and power conversion models	Linear regression	Regional (158 PVPs) in USA	Up to 2 days	6 months
[81]^D	SDM & thermal model based on the energy balance	LSTM	–	–	Few days
[98]^D	NREL model	Exponential Smoothing, MAPA, & THETA methods	12 simulated PVPs in USA	Day ahead	1 year
[82]^D	PVP digital physical model (SDM e Faiman temperature model)	CNN-LSTM	1 PVP in China	–	2 years
[99]^D	SDM	SNO & PHANN	1 PVP in Italy	24 h	2 years
[2]^D	PVlib PMC	LR, RF, SVR, MLP, LSTM, CNN, VGG-8, and GARNN	Regional (95 PVPs) in USA	1–6 h	11 months
[83]^P	A variation of SunDance physical model	K-means and Bayesian estimator	Regional (70 PVPs) in Australia	1 h	–
[86]^{DP}	PMC including separation, transposition, reflection, thermal, electrical, and losses models	Optimization + QR	14 PVPs in Hungary	–	2 years
[85]^{DP}	SDM, NOCT temperature and efficiency models	Monte Carlo method	1 PVP in Colombia	–	–

Model type: ^{D}deterministic; ^{P}probabilistic; ^{DP}deterministic and probabilistic.

performance of physical models or combining various hybridization techniques discussed previously into a unified framework. The core idea of these models is to use ML to incorporate losses and all other unknown and neglected physical phenomena not described by traditional physical models. This section discusses such approaches and their blending processes. The publications implementing these hybrid models included in our review are summarized in Table 5.

A common approach in this category involves first implementing physical modeling for PV power conversion, followed by the application of ML techniques to introduce a loss factor or correct prediction bias in the estimation. A SDM is used for PV power estimation in [80], the model is adjusted by introducing a loss factor, and an LSTM network is employed to forecast PV power 15 min ahead. In [93], an algorithm is presented for improving the performance of a PV PMC and accounting for systematic errors. The approach consists of evaluating the parameters of a PV model to maximize the likelihood that simulations match power measurements. However, the algorithm is a little sensitive to outliers and performs suboptimally when the PV module is shaded. In [94], a model Output Statistic (MOS) is introduced, consisting of a hybrid regression-physics-based model. Initially, irradiance values are converted from GHI to POA by a PMC. Then, the MOS model is calibrated and implemented for irradiance to power conversion. MOS technique significantly improves the forecast results for cloudy and partially cloudy conditions, and also shows improvement under clear sky conditions.

Beyond incorporating a loss factor, ML techniques are also used to account for weather and sky conditions, thereby complementing the physical models. Reinforcement Learning (RL), ANN, and ARIMA models are implemented in [13,95] not only to correct the physical model error but also to account for cloud transmittance when total cloud coverage exceeds a certain threshold. The approach considers a statistical predictor for the sunshine number, a binary indicator that determines whether the sun is shining. In cloudy scenarios, the estimated PV power is adjusted accordingly. In [96,97], model calibration is performed regionally. The work in [96] presents a KNN algorithm that includes a weather condition classification process is presented to estimate the physical model residuals. Meanwhile, [97] divides the region under investigation into clusters, developing a representative "virtual" PVP for each cluster. These hybrid approaches accounting for on-site weather and/or sky conditions, show enhanced forecast accuracy compared to individual physical and ML models [13,95–97].

Hybridization can also be achieved by applying both ML and physical models independently at first and then weighing their results via error minimization to generate the final output. The advantages of these integrated weighted models include implementation simplicity compared to other hybrid methodologies, outperformance compared to the individual forecasts, and mitigation of the weaknesses of individual methods [98]. The physical model can reflect the physical operational mechanisms, while the ML model captures hidden temporal and spatial features [82]. A hybrid approach weighting the estimations of a physical and an LSTM model is presented in [81].

This approach computes the weights of a linear combination of their individual predictions, assigning a higher weight to the physical model (0.66) than to the LSTM model (0.34), indicating a greater influence on the PV power estimation. In [98], PV power forecasts from different methods, including physical and time series models, are combined. The weights are determined via minimization of a loss function (mean square error) to specify each method's contribution to the forecast. Another combination method is proposed in [82], integrating a physical model and a parallel CNN-LSTM model for PV power forecasting. These components are combined using a fusion formula, a linear combination of both models' results, to achieve PV power prediction. The weighted models are compared with purely physical and ML models individually, as well as with other hybrid approaches in [81,82,98]. In all cases, the combination methods contribute to eliminating the limitations of individual methods, resulting in more accurate forecasts.

Other combined approaches to embed physical and ML are also found in the literature for PV power forecast. A physics-based model — PVlib — and a data-driven fluctuation — cloud-related — forecasting model are hybridized in [2]. The forecasting framework includes three Temporal Convolutional Networks (TCN): the first TCN blends input data, the second identifies correlated neighbors to form the detector network for intra-hour PV forecasting, and the third reconciles the estimation from both models. Comparative analysis against CNN-LSTM, Visual Geometry Group model with 8 layers (VGG-8), and Graph Attention Recurrent Neural Network (GARNN) demonstrates superior performance of the reconciliation model. In [99], another combined approach is introduced to estimate PV power. It involves using both thermal and electrical physical models with parameters determined by SNO. The estimated value serves as input in a PHANN. This method enhances the PHANN forecasting capability by combining it with the performing optimization characteristics of SNO, using both techniques discussed in the previous sections, an "optimized physical model" and a "physics-informed ML model". SNO-based model errors are comparable with persistence but less accurate than a PHANN.

Considering probabilistic approaches, a hybrid method using smart meter and local weather data is proposed in [83]. The combined PV model is based primarily on physical modeling and a data clustering method. After that, a Bayesian framework estimates the physical model parameters for hourly PV power at both customer and feeder levels. In [86], a hybrid model that converts ensemble NWP quantiles into PV power by a physical model chain is proposed, calibrating the resulting ensemble by a linear QR. Also, the probabilistic forecast is turned into a deterministic forecast and the method is also evaluated for its deterministic forecasting performance. The methodology developed in [85] comprises a PV plant analytical model, accounting for losses caused by the modules series–parallel connection, and determination of the inverter efficiency. This method allows determining the probability density functions representing the behavior of the radiation and temperature from real measurements in a statistically reliable way.

5.3. Discussions

The models classified as 'physics-guided' in our review incorporate physical and ML approaches working collaboratively. This category of models focuses more on physical modeling compared to the 'physics-informed ML models' that focus on incorporating new physical features into the ML approach. The 'physics-guided' models directly consider the impacts of physical universality and regulations, avoiding overfitting and unreasonable forecasts. The learning procedure is not only driven by data but also by domain knowledge, which can assist the ML model in obtaining better accuracy, robustness, and interpretability. It is well known that pure physical models can achieve stable hourly irradiance-power conversion for longer forecasting horizons than data-driven models. However, they cannot effectively capture intra-hour power fluctuations. Unlike the 'optimized physical models' that are strictly restricted to the physical model, optimizing their parameters, the 'physics-guided' models blend the physical and ML in a more modern way, incorporating ML techniques to introduce malleability to the physical model — integrating phenomena like clouds, seasonality, and soiling —, correct bias in the estimation, and account for weather and sky fluctuations.

Given the advantages of these models, they can be applied for different applications, from very short to long-term forecasting, single and aggregated PVP forecasts, and even for PVP digital twins, assuming it as a virtual representation of the real system from the physical point of view. The methodologies implemented have been benchmarked against hybrid models, conventional physical, ML, and statistical methods, showing strong forecasting capability, indicating that the physical and ML knowledge holds. Some points about these models are worth mentioning:

1. The use of physical constraints has huge potential to avoid unreasonable forecasts. The physical knowledge and equations provide a good reference for maximum and minimum PV power per irradiance level and can be a potential strategy to define lower and upper boundaries apart from defining these boundaries based on data. More detailed studies can be developed in this line to improve the assumptions in generating the constraints.
2. Physical models can achieve good performance in longer forecasting horizons, but ML models can more effectively capture intra-hour power fluctuations. As the 'physics-guided' methodologies consider both the advantages of physical and ML models, the performance of these models can be analyzed in terms of forecast horizon. To the best of our knowledge, no publication with this focus is found in the literature.
3. A common approach in PV forecast is to remove night data or measurements with low values of irradiance. However, in some of the publications presented here, the models are evaluated based on 24 hour data, and constraints are used to automatically estimate night data as zero, forcing the error to be zero in these intervals. This strategy reduces the final error in performance metrics, and care should be taken when comparing these models' results with purely ML models or with the mean error values that do not consider night data.

6. Conclusion

This review provides the first comprehensive overview dedicated solely to the hybrid methodologies for PV power estimation and forecasting. The main available hybridization methods are summarized and compared in terms of model application, presenting a novel methodological classification, and elucidating the merits and demerits of different techniques. The hybrid models are grouped into three main categories according to their approach, namely: physics-informed ML models, optimized physics models, and physics-guided models. Most of the found papers implement a physics-informed ML approach, followed by physics-guided methodologies and optimized physical models.

The three identified categories include papers that focus on both estimation and forecasting. However, although most of the papers found (62%) propose forecast models, in the category of optimized physical models, the majority aim to estimate PV power. This category is the closest to pure physical modeling among the hybrid models, characterized by typical physical model applications, mostly aimed at estimations or, if used for forecasting, applied with longer time horizons. Among all the reviewed manuscripts targeting PV power forecasting, the majority focus on short-term forecasts (73.4%). A smaller number of publications focus on very short-term and medium-term forecasts. None of the selected studies implement hybrid models for long-term forecasts, an area dominated by physical models; hence, there is a potential study field for hybrid models in this time horizon. Additionally, most studies in this area are based on deterministic techniques, so probabilistic hybrid techniques for PV power have the potential to be further explored.

Methods classified as physics-informed ML are based on ML models, whereas physics-optimized models focus on physical approaches. Physics-informed models integrate physical knowledge into ML models through features derived from classical physical models used in PMCs. Optimized physical models are based primarily on physical models, refining or determining their parameters via optimization algorithms. This method reduces the uncertainty of physical model parameters and bases these parameters on data rather than PV plant materials and installation configurations. Publications in both categories have explored more ML than physical modeling techniques, and the explored physical models rely on a limited set of simplified classical equations, indicating a tendency in the current literature to avoid complex physical models. Different physical models, as well as more detailed physical modeling techniques such as dynamic models, can be better explored in this area. It is already known that pure physical models generally exhibit better generalization and long-term forecasting performance, while pure ML models excel in short-term and local forecasts. Consequently, further investigation comparing both hybrid approaches over different time horizons can effectively determine the efficacy of ML models in learning and describing the underlying physics of PV systems compared to optimized physical models. Such research could reveal whether ML models can incorporate the robustness of physical models through the features.

Physics-guided models incorporate both physical and ML approaches collaboratively, divided into two subgroups: ML models with physical constraints, primarily ML models; and embedded physical-ML models, primarily physical. Unlike physics-informed ML, which uses physical modeling as a pre-processing step, and optimized physical models, which use data solely to determine physical model parameters, physics-guided models implement physical and ML modeling techniques together collaboratively. Works integrating physical models or knowledge as constraints help to prevent illogical predictions, such as negative PV power output, non-zero PV power at night, or values inconsistent with PV conversion physics, by setting boundary conditions for the power output. As the constraints are usually incorporated into the ML optimization algorithm, these models focuses more on physical modeling compared to physics-informed ML models, as it directly considers the impacts of physical universality and regulations, thereby avoiding overfitting and unreasonable forecasts. The learning procedure is not only driven by data but also by domain knowledge, which can assist the ML model in achieving better accuracy, robustness, and interpretability. In other publications, models incorporate ML techniques to enhance the performance of physical models by incorporating losses and all other unknown and neglected physical phenomena not described by traditional physical models. Unlike optimized physical models, that strictly adhere to physical equations, physics-guided ML models blend physical and ML approaches in a more modern way, incorporating ML techniques to introduce malleability to the physical model — integrating phenomena like clouds, seasonality, and soiling — correcting bias in estimation, and accounting for weather and sky fluctuations.

In summary, hybrid models connect two powerful methodologies in a modern and promising strategy. The data-driven approach, hybridized with physical models, can address unmodeled uncertainties and nonlinearities not captured by purely physical models while leveraging the advantages of both approaches in terms of robustness and accuracy. By providing detailed insights into hybrid models for PV power estimation and forecasting, as well as reflecting the current state of the art in these techniques, our review contributes to paving the way for the development of hybrid models and provides a roadmap for researchers interested in this field. For future works in this area, in addition to the possibilities of testing new physical approaches and the performance of hybrid models in terms of forecast horizon, care should be taken when choosing a model for benchmarking to avoid unfair comparisons. Most of the publications compare hybrid models with classical ML models and purely physical models, which may not have the same applicability or receive the same information.

CRediT authorship contribution statement

Leticia de Oliveira Santos: Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Tarek AlSkaif:** Writing – review & editing, Validation, Supervision, Conceptualization. **Giovanni Cordeiro Barroso:** Writing – review & editing, Validation, Methodology. **Paulo Cesar Marques de Carvalho:** Writing – review & editing, Validation, Supervision, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, *Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico* (FUNCAP) and the *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq) - Brazil.

References

[1] D. Yang, S. Alessandrini, J. Antonanzas, F. Antonanzas-Torres, V. Badescu, H.G. Beyer, R. Blaga, J. Boland, J.M. Bright, C.F. Coimbra, et al., Verification of deterministic solar forecasts, Sol. Energy 210 (2020) 20–37.
[2] Y. Li, L. Song, S. Zhang, L. Kraus, T. Adcox, R. Willardson, A. Komandur, N. Lu, A TCN-based hybrid forecasting framework for hours-ahead utility-scale PV forecasting, IEEE Trans. Smart Grid (2023).
[3] M.J. Mayer, Benefits of physical and machine learning hybridization for photovoltaic power forecasting, Renew. Sustain. Energy Rev. 168 (2022) 112772.
[4] D. Pylorof, H.E. Garcia, Uncertainty-aware photovoltaic generation estimation through fusion of physics with harmonics information using Bayesian neural networks, in: 2023 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference, ISGT, IEEE, 2023, pp. 1–5.
[5] D. Koster, D. Fiorelli, P. Bruneau, C. Braun, Single-site forecasts for 130 photovoltaic systems at distribution system operator level, using a hybrid-physical approach, to improve grid-integration and enable future smart-grid operation, Solar RRL 7 (8) (2023) 2200652.
[6] T.C. Carneiro, P.C.M. de Carvalho, H. Alves dos Santos, M.A.F.B. Lima, A.P.d.S. Braga, Review on photovoltaic power and solar resource forecasting: Current status and trends, J. Sol. Energy Eng. 144 (1) (2022) 010801.
[7] R. Ahmed, V. Sreeram, Y. Mishra, M. Arif, A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization, Renew. Sustain. Energy Rev. 124 (2020) 109792.
[8] L. Oneto, F. Laureri, M. Robba, F. Delfino, D. Anguita, Data-driven photovoltaic power production nowcasting and forecasting for polygeneration microgrids, IEEE Syst. J. 12 (3) (2017) 2842–2853.
[9] Y. Zhi, T. Sun, X. Yang, A physical model with meteorological forecasting for hourly rooftop photovoltaic power prediction, J. Build. Eng. 75 (2023) 106997.
[10] M. Parida, M.K. Behera, N. Nayak, Combined EMD-ELM and OS-ELM techniques based on feed-forward networks for PV power forecasting, in: 2018 Technologies for Smart-City Energy Security and Power, ICSESP, IEEE, 2018, pp. 1–6.
[11] L. Visser, T. AlSkaif, W. van Sark, Operational day-ahead solar power forecasting for aggregated PV systems with a varying spatial distribution, Renew. Energy 183 (2022) 267–282.
[12] S.E. Haupt, T.C. McCandless, S. Dettling, S. Alessandrini, J.A. Lee, S. Linden, W. Petzke, T. Brummet, N. Nguyen, B. Kosović, et al., Combining artificial intelligence with physics-based methods for probabilistic renewable energy forecasting, Energies 13 (8) (2020) 1979.
[13] C. Timplalexis, N. Bezas, A.D. Bintoudi, L. Zyglakis, V. Pavlopoulos, A.C. Tsolakis, S. Krinidis, D. Tzovaras, A hybrid physical/statistical day-ahead direct PV forecasting engine, in: The 12th Mediterranean Conference on Power Generation, Transmission, Distribution and Energy Conversion, MEDPOWER 2020, IET, 2021, pp. 258–263.
[14] K. Bellinguer, R. Girard, G. Bontron, G. Kariniotakis, A generic methodology to efficiently integrate weather information in short-term photovoltaic generation forecasting models, Sol. Energy 244 (2022) 401–413.
[15] Y.-K. Wu, C.-L. Huang, Q.-T. Phan, Y.-Y. Li, Completed review of various solar power forecasting techniques considering different viewpoints, Energies 15 (9) (2022) 3320.

[16] A. Mellit, An overview on the application of machine learning and deep learning for photovoltaic output power forecasting, in: Proceedings of the 2nd International Conference on Electronic Engineering and Renewable Energy Systems: ICEERE 2020, 13-15 April 2020, Saidia, Morocco, Springer, 2021, pp. 55–68.

[17] A. Mellit, S.A. Kalogirou, A survey on the application of artificial intelligence techniques for photovoltaic systems, in: McEvoy's Handbook of Photovoltaics, Elsevier, 2018, pp. 735–761.

[18] A. Mellit, A. Massi Pavan, E. Ogliari, S. Leva, V. Lughi, Advanced methods for photovoltaic output power forecasting: A review, Appl. Sci. 10 (2) (2020) 487.

[19] L.R. Visser, E. Lorenz, D. Heinemann, W.G. Van Sark, Ch. 1.11-Solar power forecasts, in: Comprehensive Renewable Energy, second ed., in: Solar photovoltaic energy, Vol. 1, Elsevier, 2022, pp. 213–233.

[20] D.V. Pombo, P. Bacher, C. Ziras, H.W. Bindner, S.V. Spataru, P.E. Sørensen, Benchmarking physics-informed machine learning-based short term PV-power forecasting tools, Energy Rep. 8 (2022) 6512–6520.

[21] C.-L. Huang, Y.-K. Wu, Y.-Y. Li, Deterministic and probabilistic solar power forecasts: A review on forecasting models, in: 2021 7th International Conference on Applied System Innovation, ICASI, IEEE, 2021, pp. 15–18.

[22] P. Li, K. Zhou, S. Yang, Photovoltaic power forecasting: Models and methods, in: 2018 2nd IEEE Conference on Energy Internet and Energy System Integration, EI2, Ieee, 2018, pp. 1–6.

[23] O.E. Hassan, A.K. Abdelsalam, New time horizon based classification of PV power generation forecasting techniques, in: 2020 30th International Conference on Computer Theory and Applications, ICCTA, IEEE, 2020, pp. 88–95.

[24] S. Sobri, S. Koohi-Kamali, N.A. Rahim, Solar photovoltaic generation forecasting methods: A review, Energy Convers. Manage. 156 (2018) 459–497.

[25] D.S. Tripathy, B.R. Prusty, D. Jena, Probabilistic forecasting of daily PV generation using quantile regression method, in: 2020 IEEE India Council International Subsections Conference, INDISCON, IEEE, 2020, pp. 260–265.

[26] D.S. Tripathy, B.R. Prusty, K. Bingi, Reliable prediction intervals of PV generation using quantile regression averaging approach, in: 2020 3rd International Conference on Energy, Power and Environment: Towards Clean Energy Technologies, IEEE, 2021, pp. 1–6.

[27] M.J. Mayer, D. Yang, Probabilistic photovoltaic power forecasting using a calibrated ensemble of model chains, Renew. Sustain. Energy Rev. 168 (2022) 112821.

[28] L. Visser, T. AlSkaif, J. Hu, A. Louwen, W. van Sark, On the value of expert knowledge in estimation and forecasting of solar photovoltaic power generation, Sol. Energy 251 (2023) 86–105.

[29] E. Ogliari, A. Dolara, G. Manzolini, S. Leva, Physical and hybrid methods comparison for the day ahead PV output power forecast, Renew. Energy 113 (2017) 11–21.

[30] S. Kim, Y. Kim, Y. On, J. So, C.-Y. Yoon, S. Kim, Hybrid performance modeling of an agrophotovoltaic system in South Korea, Energies 15 (18) (2022) 6512.

[31] M. Bai, Y. Chen, X. Zhao, J. Liu, D. Yu, Deep attention ConvLSTM-based adaptive fusion of clear-sky physical prior knowledge and multivariable historical information for probabilistic prediction of photovoltaic power, Expert Syst. Appl. 202 (2022) 117335.

[32] Y. Zhou, S. Pan, L. Qin, Z. Yuan, W. Huang, H. Bai, J. Lei, Short-term solar power forecasting based on convolutional neural network and analytical knowledge, Int. Trans. Electr. Energy Syst. 31 (11) (2021).

[33] D.V. Pombo, M.J. Rincón, P. Bacher, H.W. Bindner, S.V. Spataru, P.E. Sørensen, Assessing stacked physics-informed machine learning models for co-located wind–solar power forecasting, Sustain. Energy Grids Netw. 32 (2022) 100943.

[34] K. Wang, W. Dou, H. Wei, K. Zhang, Intra-hour PV power forecasting based on multi-source data and PSC-SVR model, in: 2022 41st Chinese Control Conference, CCC, IEEE, 2022, pp. 3202–3207.

[35] M. Schmelas, T. Feldmann, J. da Costa Fernandes, E. Bollin, Photovoltaics energy prediction under complex conditions for a predictive energy management system, J. Sol. Energy Eng. 137 (3) (2015) 031015.

[36] W. Zhao, H. Zhang, J. Zheng, Y. Dai, L. Huang, W. Shang, Y. Liang, A point prediction method based automatic machine learning for day-ahead power output of multi-region photovoltaic plants, Energy 223 (2021) 120026.

[37] J. Wang, H. Zhong, X. Lai, Q. Xia, Y. Wang, C. Kang, Exploring key weather factors from analytical modeling toward improved solar power forecasting, IEEE Trans. Smart Grid 10 (2) (2017) 1417–1427.

[38] S. Al-Dahidi, S. Al-Nazer, O. Ayadi, S. Shawish, N. Omran, Analysis of the effects of cell temperature on the predictability of the solar photovoltaic power production, Int. J. Energy Econ. Policy 10 (5) (2020) 208.

[39] D. Anagnostos, H. Goverde, F. Catthoor, D. Soudris, J. Poortmans, Systematic cross-validation of photovoltaic energy yield models for dynamic environmental conditions, Sol. Energy 155 (2017) 698–705.

[40] D. Koster, D. Fiorelli, P. Bruneau, C. Braun, Single-site forecasts for 130 photovoltaic systems at distribution system operator level, using a hybrid-physical approach, to improve grid-integration and enable future smart-grid operation, Solar RRL 7 (8) (2023) 2200652.

[41] E. Ogliari, A. Nespoli, Photovoltaic plant output power forecast by means of hybrid artificial neural networks, in: A Practical Guide for Advanced Methods in Solar Photovoltaic Systems, Springer, 2020, pp. 203–222.

[42] S. Leva, A. Nespoli, S. Pretto, M. Mussetta, E.G.C. Ogliari, PV plant power nowcasting: A real case comparative study with an open access dataset, IEEE Access 8 (2020) 194428–194440.

[43] A. Nespoli, M. Mussetta, E. Ogliari, S. Leva, L. Fernández-Ramírez, P. García-Triviño, Robust 24 hours ahead forecast in a microgrid: A real case study, Electronics 8 (12) (2019) 1434.

[44] M. Pierro, F. Bucci, M. De Felice, E. Maggioni, A. Perotto, F. Spada, D. Moser, C. Cornaro, Deterministic and stochastic approaches for day-ahead solar power forecasting, J. Sol. Energy Eng. 139 (2) (2017) 021010.

[45] A. Dolara, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, A physical hybrid artificial neural network for short term forecasting of PV plant power output, Energies 8 (2) (2015) 1138–1153.

[46] G. Graditi, S. Ferlito, G. Adinolfi, G.M. Tina, C. Ventura, Energy yield estimation of thin-film photovoltaic plants by using physical approach and artificial neural networks, Sol. Energy 130 (2016) 232–243.

[47] E. Ogliari, A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, Neural forecasting of the day-ahead hourly power curve of a photovoltaic plant, in: 2016 International Joint Conference on Neural Networks, IJCNN, IEEE, 2016, pp. 654–659.

[48] A. Dolara, S. Leva, M. Mussetta, E. Ogliari, PV hourly day-ahead power forecasting in a micro grid context, in: 2016 IEEE 16th International Conference on Environment and Electrical Engineering, EEEIC, IEEE, 2016, pp. 1–5.

[49] A. Gandelli, F. Grimaccia, S. Leva, M. Mussetta, E. Ogliari, Hybrid model analysis and validation for PV energy production forecasting, in: 2014 International Joint Conference on Neural Networks, IJCNN, IEEE, 2014, pp. 1957–1962.

[50] S. Chen, J. Jiang, S. Li, H. Li, Photovoltaic power prediction considering direct and scattered radiation effect, in: 2023 7th International Conference on Smart Grid and Smart Cities, ICSGSC, IEEE, 2023, pp. 481–486.

[51] S. Al-Dahidi, J. Adeeb, O. Ayadi, M. Alrbai, L. Al-Ghussain, A feature transformation and extraction approach-based artificial neural network for an improved production prediction of grid-connected solar photovoltaic systems, Energy Sources A 44 (4) (2022) 9232–9254.

[52] D.S. Tripathy, B.R. Prusty, D. Jena, M.K. Sahu, Multi-time instant probabilistic PV generation forecasting using quantile regression forests, in: 2020 IEEE 9th Power India International Conference, PIICON, IEEE, 2020, pp. 1–6.

[53] R.E. Bird, A simple, solar spectral model for direct-normal and diffuse horizontal irradiance, Solar Energy 32 (4) (1984) 461–471.

[54] M. Lefèvre, A. Oumbe, P. Blanc, B. Espinar, B. Gschwind, Z. Qu, L. Wald, M. Schroedter-Homscheidt, C. Hoyer-Klick, A. Arola, et al., McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions, Atmos. Meas. Tech. 6 (9) (2013) 2403–2418.

[55] H.C. Hottel, A simple model for estimating the transmittance of direct solar radiation through clear atmospheres, Solar Energy 18 (2) (1976) 129–134.

[56] D. Erbs, S. Klein, J. Duffie, Estimation of the diffuse radiation fraction for hourly, daily and monthly-average global radiation, Solar Energy 28 (4) (1982) 293–302.

[57] X. Sun, J.M. Bright, C.A. Gueymard, X. Bai, B. Acord, P. Wang, Worldwide performance assessment of 95 direct and diffuse clear-sky irradiance models using principal component analysis, Renew. Sustain. Energy Rev. 135 (2021) 110087.

[58] B.Y. Liu, R.C. Jordan, The interrelationship and characteristic distribution of direct, diffuse and total solar radiation, Solar Energy 4 (3) (1960) 1–19.

[59] R. Perez, R. Seals, P. Ineichen, R. Stewart, D. Menicucci, A new simplified version of the perez diffuse irradiance model for tilted surfaces, Solar Energy 39 (3) (1987) 221–231.

[60] M. Hofmann, G. Seckmeyer, Influence of various irradiance models and their combination on simulation results of photovoltaic systems, Energies 10 (10) (2017) 1495.

[61] L. de Oliveira Santos, P.C.M. de Carvalho, C. de Oliveira Carvalho Filho, Photovoltaic cell operating temperature models: A review of correlations and parameters, IEEE J. Photovolt. 12 (1) (2021) 179–190.

[62] L.O. Santos, F.A.A. Souza, C.O. Carvalho Filho, P.C.M. Carvalho, T. AlSkaif, R.I.S. Pereira, Hybrid modeling for photovoltaic module operating temperature estimation, IEEE J. Photovolt. 14 (3) (2024) 488–496, http://dx.doi.org/10.1109/JPHOTOV.2024.3372328.

[63] C.R. Osterwald, Translation of device performance measurements to reference conditions, Solar Cells 18 (3) (1986) 269–279.

[64] N. Al-Messabi, C. Goh, Y. Li, Heuristic grey-box modelling for photovoltaic power systems, Syst. Sci. Control. Eng. 4 (1) (2016) 235–246.

[65] S. Ramesh, A. Tuomiranta, A. Hajjiah, M. Meuris, B. Vermang, J. Poortmans, Physics-based electrical modelling of CIGS thin-film photovoltaic modules for system-level energy yield simulations, npj Flex. Electron. 6 (1) (2022) 87.

[66] A. Niccolai, A. Dolara, E. Ogliari, Hybrid PV power forecasting methods: A comparison of different approaches, Energies 14 (2) (2021) 451.

[67] G.J.-P. Tevi, D. Diouf, M.E. Faye, M. Sene, A.S. Maiga, et al., A performance model using an analytical-numerical method and artificial neural networks for a PV plant output forecast, in: 2021 International Conference on Electrical, Computer and Energy Technologies, ICECET, IEEE, 2021, pp. 1–6.

[68] J. Zhang, Y. Zhang, A novel method for parameter identification of PV modules by using effective irradiance and temperature, in: 2023 IEEE 6th International Electrical and Energy Conference, CIEEC, IEEE, 2023, pp. 2636–2641.

[69] E. Garoudja, K. Kara, A. Chouder, S. Silvestre, Parameters extraction of photovoltaic module for long-term prediction using artifical bee colony optimization, in: 2015 3rd International Conference on Control, Engineering & Information Technology, CEIT, IEEE, 2015, pp. 1–6.
[70] C.-M. Huang, S.-J. Chen, S.-P. Yang, Parameter estimation of a single-diode PV model using a hybrid charged system search algorithm, in: 2020 IEEE 29th International Symposium on Industrial Electronics, ISIE, IEEE, 2020, pp. 947–952.
[71] S. Vinco, L. Bottaccioli, E. Patti, A. Acquaviva, M. Poncino, A compact PV panel model for cyber-physical systems in smart cities, in: 2018 IEEE International Symposium on Circuits and Systems, ISCAS, IEEE, 2018, pp. 1–5.
[72] F. Li, W. Wu, Coupled electrical-thermal performance estimation of photovoltaic devices: A transient multiphysics framework with robust parameter extraction and 3-D thermal analysis, Appl. Energy 319 (2022) 119249.
[73] A. Dolara, S. Leva, G. Manzolini, Comparison of different physical models for PV power output prediction, Solar Energy 119 (2015) 83–99.
[74] P. Danner, H. de Meer, Location and solar system parameter extraction from power measurement time series, Energy Inform. 4 (2021) 1–20.
[75] N. Bashir, D. Chen, D. Irwin, P. Shenoy, Solar-TK: A data-driven toolkit for solar PV performance modeling and forecasting, in: 2019 IEEE 16th International Conference on Mobile Ad Hoc and Sensor Systems, MASS, IEEE, 2019, pp. 456–466.
[76] D.E. Guzman Razo, H. Madsen, C. Wittwer, Genetic algorithm optimization for parametrization, digital twinning, and now-casting of unknown small-and medium-scale PV systems based only on on-site measured data, Front. Energy Res. 11 (2023) 1060215.
[77] K. Yaman, G. Arslan, A detailed mathematical model and experimental validation for coupled thermal and electrical performance of a photovoltaic (PV) module, Appl. Therm. Eng. 195 (2021) 117224.
[78] F. Hadjrioua, N. Belhaouas, A. Aissaoui, F. Mehareb, K. Bakria, Outdoor PV module characterization and Sandia model calibration in local climate, in: 2022 International Conference of Advanced Technology in Electronic and Electrical Engineering, ICATEEE, IEEE, 2022, pp. 1–6.
[79] D.L. King, J.A. Kratochvil, W.E. Boyson, Photovoltaic array performance model, vol. 8, Sandia Nat. Lab. (SNL), Albuquerque, NM, Livermore, CA, USA, 2004, http://dx.doi.org/10.2172/919131, Rep. SAND2004-3535, Aug. 2004. [Online].
[80] O.A. Al-Isawi, L.H. Amirah, O.A. Al-Mufti, C. Ghenai, Digital twinning and LSTM-based forecasting model of solar PV power output, in: 2023 Advances in Science and Engineering Technology International Conferences, ASET, IEEE, 2023, pp. 1–8.
[81] F. Delussu, D. Manzione, R. Meo, G. Ottino, M. Asare, Experiments and comparison of digital twinning of photovoltaic panels by machine learning models and a cyber-physical model in modelica, IEEE Trans. Ind. Inform. 18 (6) (2021) 4018–4028.
[82] X. Zhang, Y. Li, T. Li, Y. Gui, Q. Sun, D.W. Gao, Digital twin empowered PV power prediction, J. Mod. Power Syst. Clean Energy (2023).
[83] N. Mahdavi, D. Weeraddana, Y. Guo, Probabilistic estimation of PV generation at customer and distribution feeder levels using net-demand data, IEEE Trans. Smart Grid 14 (3) (2022) 1974–1984.
[84] E. Scolari, L. Reyes-Chamorro, F. Sossan, M. Paolone, A comprehensive assessment of the short-term uncertainty of grid-connected PV systems, IEEE Trans. Sustain. Energy 9 (3) (2018) 1458–1467.
[85] J. Hernandez, G. Gordillo, W. Vallejo, Predicting the behavior of a grid-connected photovoltaic system from measurements of solar radiation and ambient temperature, Appl. Energy 104 (2013) 527–537.
[86] M.J. Mayer, D. Yang, Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting, Renew. Sustain. Energy Rev. 175 (2023) 113171.
[87] S.E. Haupt, T.C. McCandless, S. Dettling, S. Alessandrini, J.A. Lee, S. Linden, W. Petzke, T. Brummet, N. Nguyen, B. Kosović, et al., Combining artificial intelligence with physics-based methods for probabilistic renewable energy forecasting, Energies 13 (8) (2020) 1979.
[88] J. Du, J. Zheng, Y. Liang, Q. Liao, B. Wang, X. Sun, H. Zhang, M. Azaza, J. Yan, A theory-guided deep-learning method for predicting power generation of multi-region photovoltaic plants, Eng. Appl. Artif. Intell. 118 (2023) 105647.
[89] D.H. Tran, I.B.K.Y. Utama, H. Nguyen, B. Chung, Y.M. Jang, et al., A method of PV power generation forecasting using constrained transformer network, in: 2023 International Conference on Artificial Intelligence in Information and Communication, ICAIIC, IEEE, 2023, pp. 188–191.
[90] X. Luo, D. Zhang, X. Zhu, Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge, Energy 225 (2021) 120240.
[91] X. Luo, D. Zhang, X. Zhu, Combining transfer learning and constrained long short-term memory for power generation forecasting of newly-constructed photovoltaic plants, Renew. Energy 185 (2022) 1062–1077.
[92] M. Hildmann, A. Rohatgi, G. Andersson, Constrained support vector machines for photovoltaic in-feed prediction, in: 2013 1st IEEE Conference on Technologies for Sustainability, SusTech, IEEE, 2013, pp. 23–28.
[93] Y.-M. Saint-Drenan, S. Bofinger, R. Fritz, S. Vogt, G. Good, J. Dobschinski, An empirical approach to parameterizing photovoltaic plants for power forecasting and simulation, Sol. Energy 120 (2015) 479–493.
[94] C. Porrini, A. Gutiérrez, G.C. Boezio, G. Hermida, D. Oroño, M. Puppo, Development of a model output statistic and implementation of an operational solar photovoltaic energy forecast model based in WRF, in: 2015 IEEE PES Innovative Smart Grid Technologies Latin America, ISGT LATAM, IEEE, 2015, pp. 248–253.
[95] M. Paulescu, N. Stefu, C. Dughir, A. Sabadus, D. Calinoiu, V. Badescu, A simple but accurate two-state model for nowcasting PV power, Renew. Energy 195 (2022) 322–330.
[96] Z. Liu, Z. Zhang, Solar forecasting by K-nearest neighbors method with weather classification and physical model, in: 2016 North American Power Symposium, NAPS, IEEE, 2016, pp. 1–6.
[97] Y. Li, Q. Sun, B. Lehman, S. Lu, H.F. Hamann, J. Simmons, J. Black, A machine-learning approach for regional photovoltaic power forecasting, in: 2016 IEEE Power and Energy Society General Meeting, PESGM, IEEE, 2016, pp. 1–5.
[98] G.M. Yagli, D. Yang, D. Srinivasan, et al., Using combinational methods for forecast improvement in PV power plants, in: 2018 IEEE Innovative Smart Grid Technologies-Asia, ISGT Asia, IEEE, 2018, pp. 540–545.
[99] E. Ogliari, A. Niccolai, S. Leva, R.E. Zich, Computational intelligence techniques applied to the day ahead PV output power forecast: PHANN, SNO and mixed, Energies 11 (6) (2018) 1487.