Renewable Energy 251 (2025) 123495

Contents lists available at ScienceDirect

Renewable Energy

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

Comprehensive approach to photovoltaic power forecasting using numerical weather prediction data and physics-based models and data-driven techniques

Sara Pereira [a,*], Paulo Canhoto [a,b], Takashi Oozeki [c], Rui Salgado [a,d]

[a] Institute of Earth Sciences, University of Évora, Évora, Portugal
[b] Department of Mechatronics Engineering, University of Évora, Évora, Portugal
[c] Renewable Energy Research Center, National Institute of Advanced Industrial Science and Technology, Koriyama, Fukushima, Japan
[d] Physics Department, University of Évora, Évora, Portugal

ARTICLE INFO

Keywords:
Forecasting
Grid integration
Photovoltaic power
Solar energy
Solar power plant

ABSTRACT

Photovoltaic power forecasting is essential for maintaining electric grid stability and efficiently integrating solar energy power plants into the national power generation system. However, it remains challenging due to the complexity of accurately predicting solar radiation across varying weather conditions and diverse photovoltaic system configurations. This study addresses these challenges by developing a novel integrated forecasting algorithm that includes numerical weather prediction data, physics-based models, and artificial neural networks. The algorithm enhances direct normal irradiance forecasts, computes global tilted irradiance using an improved transposition model, and predicts photovoltaic output with a dynamic thermal-electric model. Losses and inverter efficiency are also incorporated. The algorithm provides 72-h power forecasts with customizable temporal resolution, without the need for on-site observations. Validation against 15-min data from a real photovoltaic plant demonstrated mean bias errors and root mean squared errors of 7.5 W/kWp and 123.7 W/kWp (DC), and 9.3 W/kWp and 121.0 W/kWp (AC), corresponding to relative errors of 1.8 %, 30.0 %, 2.3 %, and 29.9 %. The algorithm is scalable, adaptable to various system configurations, and effective for regions with limited data, thus supporting improved grid operations, enabling better management of photovoltaic generation variability and enhancing energy system efficiency.

1. Introduction

Recently, the integration of solar energy sources in the global energy sector has gained considerable attention as a fundamental strategy for sustainable development and mitigation of the negative impacts of climate change. Among solar energy systems, photovoltaic (PV) energy has emerged as vital component in the energy mix. However, the intermittent and weather dependent nature of solar radiation and photovoltaic power generation present challenges in its integration in the electric grid and energy management.

Accurate forecasting of photovoltaic power generation is essential to address these challenges and optimize the utilization of solar energy resources. The variability of solar radiation due to diurnal and seasonal patterns, as well as due to the dynamic nature of weather conditions, poses substantial difficulties in photovoltaic power forecasting. These difficulties extend to the complex interactions between environmental factors and the performance of photovoltaic systems, requiring comprehensive methods for forecasting.

Integrated solar irradiance and photovoltaic power forecasting models can be divided into physics-based and data-driven. Physics-based models take into account the underlying physical principles and cause-and-effect relationships that govern the response of photovoltaic systems, typically including the effect of the environmental conditions, the properties of photovoltaic materials and the systems configuration. By building a model based on the equations that describe the physical phenomena and the processes in photovoltaics energy conversion, such as the transport of radiation in the atmosphere, its interaction with the surrounding environment and the photovoltaic modules as well as the generation of electrical current, these models aim to provide a more comprehensive and theoretically grounded understanding of the system.

The main steps of a physics-based integrated solar irradiance and

* Corresponding author.
E-mail address: spereira@uevora.pt (S. Pereira).

https://doi.org/10.1016/j.renene.2025.123495
Received 13 September 2024; Received in revised form 3 April 2025; Accepted 18 May 2025
Available online 19 May 2025

Nomenclature

$\vec{c}$	Position vector of the middle point of each segment of photovoltaic panel (m)
D	Distance between rows in the horizontal plane (m)
$\vec{D}_{cs}$	Vector of circumsolar diffuse irradiance values (W/m^2)
DIF	Diffuse horizontal irradiance (W/m^2)
$\vec{D}_{iso}$	Vector of isotropic diffuse irradiance values (W/m^2)
DNI	Direct normal irradiance (W/m^2)
$\boldsymbol{F}$	View-factor matrix (–)
FS	Forecast skill score (%)
GHI	Global horizontal irradiance (W/m^2)
GTI	Global tilted irradiance (W/m^2)
h_0	Vertical distance between the ground and the panel base (m)
h_s	Solar hour angle (°)
$\vec{I}$	Vector of direct horizontal irradiance values (W/m^2)
K_g	Extinction coefficient of the glazing (m^{-1})
Lat	Latitude (°)
L_g	Glazing thickness (m)
MAE	Mean absolute error (W/m^2 or W)
MBE	Mean bias error (W/m^2 or W)
n	Effective refractive index of the solar cell cover (–)
$\boldsymbol{R}$	Reflectance matrix (–)
R^2	Coefficient of determination (–)
rL	Length of the row of modules (m)
rMAE	Relative mean absolute error (%)
rMBE	Relative mean bias error (%)
RMSE	Root mean squared error (W/m^2)
rRMSE	Relative root mean squared error (%)
S	Absorbed irradiance (W/m^2 or W)
$\vec{u}$	Position vector of the midpoint of each segment of ground (m)
$\vec{v}$	Position vector of the midpoint of each segment on the back of front panel (m)

Greek symbols

β	Tilt angle of the photovoltaic panels (°)
$\vec{\gamma}_b$	Vector of incidence angle modifier values for direct irradiance (–)
$\vec{\gamma}_d$	Vector of incidence angle modifier values for diffuse isotropic irradiance (–)
$\boldsymbol{\gamma_r}$	Matrix of incidence angle modifier values for reflected irradiance (–)
δ	Solar declination (°)
θ	Angle of incidence (°)
θ_b	Angle of incidence of direct irradiance (°)
θ_d	Equivalent angle of incidence of diffuse isotropic irradiance (°)
θ_r	Angle of refraction (°)
$\vec{\theta}_{ref}$	Vector of equivalent angles of incidence values for reflected irradiance (°)
$\boldsymbol{\rho}$	Matrix of albedo values (–)
φ	Azimuth of the photovoltaic panels (°)

Acronyms

AC	Alternate Current
ANN	Artificial Neural Networks
CAMS	Copernicus Atmospheric Monitoring Service
DC	Direct Current
ECMWF	European Centre for Medium-range Weather Forecasts
GFS	Global Forecast System
IFS	Integrated Forecasting System
McICA	Monte Carlo Independent Column Approximation
NWP	Numerical Weather Prediction
PV	Photovoltaic
RRTM	Rapid Radiative Transfer Model
STC	Standard Test Conditions

photovoltaic power forecasting model would be the forecasting of the solar irradiance components, their transposition to the plane of a photovoltaic array, the conversion of solar radiation to electric energy and the conversion of DC to AC current for injection in the electric grid.

Physics-based solar irradiance forecasting can be done through various approaches with different temporal and spatial resolutions and forecast horizons, including numerical weather prediction (NWP) models, sky/shadow or satellite imagery. NWP models integrate relations describing the dynamics of the atmosphere and the physical phenomena relevant to weather to predict meteorological variables, including global solar irradiance for the energy balance on the surface. Direct and diffuse solar irradiance are also prognostic variables of currently NWP models, but the accuracy of its prediction is weaker, since for the purposes of weather forecasting, the partition is less relevant than global radiation [1]. The Integrated Forecasting System (IFS) developed at the European Centre for Medium-range Weather Forecasts (ECMWF) is one of the most widely used and evaluated global NWP models in the World. El Alani et al. [2] showed strong accuracy for clear-sky solar forecasts but noted limitations under cloudy conditions. Perdigão et al. [3] observed DNI overestimation and high hourly errors in southern Portugal, proposing a correction index to significantly improve cloudy forecasts. Perez et al. [4] demonstrated ECMWF's superior global performance compared to regional models across multiple regions. Mayer et al. [5] confirmed ECMWF's higher accuracy for irradiance forecasting but noted reduced benefits when applied to regional photovoltaic power predictions. As the model produces forecasts of direct normal irradiation at surface, no separation model for obtaining direct normal and diffuse irradiance components from global horizontal irradiance is required. The use of separation models prior to transposition models typically induces high errors in the prediction of global tilted irradiance and consequently on photovoltaic power forecasting. Gueymard demonstrated significant accuracy degradation due to uncertainties in diffuse-direct separation [6] and reported particularly large errors during cloud enhancement events [7] while Yang and Gueymard [8] emphasized that even advanced separation methods inherently yield considerable uncertainties.

Transposition models are used to compute the global tilted irradiance (GTI), for example the irradiance on the plane of photovoltaic arrays, including a reflected component, based on the Direct Normal Irradiance (DNI) and Diffuse Horizontal Irradiance (DIF). The computation of direct and reflected components tends to be consistent across all models, while the computation of diffuse irradiance differs significantly depending on the modeling approach and the simplifications that are assumed [9]. Several studies compared different transposition models with observations for different locations and tilted surface positions. An extensive analysis of physical photovoltaic forecasting models highlighted the critical impact of transposition model choice on forecast accuracy [10]. The performance of 22 transposition models was evaluated across multiple locations and tilt angles in Libya, emphasizing substantial regional performance differences [11]. A comparative study of 24 transposition models in Palestine revealed particular limitations of existing models under clear-sky conditions [12], while a benchmark study ranking 26 widely used models concluded that no single universal model achieves optimal accuracy for all geographic regions [9].

Recently, improvements were reported by including shading effects and anisotropic diffuse irradiance explicitly, notably enhancing the precision of predictions for photovoltaic arrays arranged in parallel rows [1]. In some models, uniform sky dome radiance (isotropic sky) is assumed, while in others model different regions in the sky dome are considered (anisotropic) which tend to present a better performance when compared to measured data. Evaluating the diffuse component accurately is a challenging task because, along with the different regions in the sky that can be considered – typically the circumsolar, the horizon brightening and the isotropic sky background regions, the complexity and variability of cloud shape and position result in various degrees of models' performance depending on sky conditions [6]. Moreover, the majority of transposition models have been developed for a tilted surface in an open field, but for solar energy power plants with multiple rows of modules, adjustments need to be made to account for the constrained view of the sky dome and ground, as well as potential shading effects between rows. Accurate modeling for rows behind the front row is critical, as shown by studies that introduced adjustments to standard models, considering direct shading effects and diffuse anisotropic irradiance masking from adjacent rows [1]. Corrections were specifically proposed for anisotropic models to account for a reduced sky view angle and adjustments to circumsolar and horizon brightening components for inner rows [13]. A recent work demonstrated how restricted sky views significantly reduce diffuse irradiance at the lower parts of arrays, especially with narrow row spacing and steeper tilt angles [14]. Another analysis further highlighted that simplified shading models, ignoring diffuse irradiance masking, underestimate shading losses in large photovoltaic plants by 50–80 %, emphasizing the need for detailed diffuse shading calculations [15].

Once the irradiance incident on the plane of the photovoltaic array is available, the absorbed energy and electric power output of a photovoltaic system can be determined using a photovoltaic cell/module model. Several photovoltaic models with different inputs, precision, complexity and computational costs have been developed. Gholami et al. [16] presented a comprehensive classification and comparative review of photovoltaic electrical models, detailing differences in parameter extraction methods, computational complexity, and precision. Pereira et al. [17] evaluated photovoltaic thermal models and their integration with electrical models, emphasizing the relevance of temperature modeling accuracy for improved power prediction. The simplest model assumes a linear relationship between solar irradiance and power output, while more advanced models include the effect of cells' temperature and describe the module as an electric circuit. Pereira et al. [17] integrated thermal models with electrical circuit models, showing that temperature corrections increase photovoltaic model realism. Castro [18] compared classical equivalent circuit models with artificial intelligence-based approaches, confirming that models accounting explicitly for temperature perform more accurately in experimental validations. The estimation of parameters for these models often requires assumptions and elaborated analytical or numerical methods due to the limited data provided by the manufacturers. The estimation of parameters is typically conducted for standard test conditions (STC) which, however, differ from real conditions on the field, thus requiring additional adjustments for an accurate estimation of power output. Castro [18] emphasized the challenge of accurately estimating model parameters, noting that artificial intelligence approaches achieve greater accuracy but with higher computational requirements. Chenni et al. [19] presented a four-parameter photovoltaic cell model based on manufacturer datasheets, proving its effectiveness in simulating photovoltaic cell performance under varying irradiance and temperature. Chin et al. [20] reviewed techniques to estimate photovoltaic model parameters from manufacturer data, concluding that sophisticated multi-parameter models offer enhanced accuracy, particularly when environmental conditions vary widely. As the temperature of photovoltaic cells is related and has a critical impact on their efficiency, its modeling is also extremely important. Most models in the literature for predicting the temperature of photovoltaic modules are steady-state and empirical, which may be biased towards different technologies or locations with different climatic conditions. On the other hand, physics-based models that consider energy conservation and dynamic aspects can better describe the thermal response of photovoltaic modules, especially at shorter time steps. Recent works by Li and Wu [21], Perovic et al. [22] and Pereira et al. [17] have developed coupled electrical and thermal models that take into account the relationship between environmental conditions, cell/module temperature and electric power output.

Finally, the inverter introduces losses that are modeled by considering the efficiency of the direct (DC) to alternate current (AC) conversion and the power and AC voltage regulation. On the direct current side, voltage and current are regulated in order to maintain the operation of photovoltaic modules at the maximum power point within the limits of the inverter. However, it is common practice to design systems where the DC power exceeds the nominal power of the inverter leading to clipping losses [23]. The simplest inverter models consist of a constant efficiency value and clipping of the power output, since the required inputs are readily available in the inverter datasheet [10].

Besides the losses referred above, there are other aspects that induce losses in photovoltaic power plants such as DC wiring, bypass diodes and connectors, module mismatch, maximum power point tracking inefficiencies, soiling, degradation induced by the continued exposition to solar radiation and adverse environmental conditions. The computation of these losses in forecasting models depends on the availability of relevant information for the photovoltaic plants of interest.

Considering only physics-based photovoltaic forecasting models, Mayer and Gróf [10] analyzed the performance of all possible combinations of nine direct and diffuse irradiance separation models, ten transposition models for tilted irradiance computation, three reflection losses, five cell temperature models, four photovoltaic module performance models, two shading losses models, and three inverter models for one-year 15-min resolution data of 16 photovoltaic powerplants in Hungary for day-ahead and intraday forecasting time horizons. This study highlighted how model selection affects the accuracy of photovoltaic power forecasting, particularly in the case of separation and transposition models.

Some commercially available tools can also be used for photovoltaic power forecasting, as for example those presented in the study by González-Peña et al. [24], in which five software tools for predicting photovoltaic power generation, namely RETScreen [25], System Advisor Model (SAM) [26], PVGIS [27], PVSyst [28], and PV*SOL [29], were evaluated by comparing predicted data with real field data from three photovoltaic power plants in Castile and Leon, Spain, over a 12-year period.

In opposition to physical models, data-driven models such as linear statistical models and machine learning techniques use historical data to establish relationships between weather or photovoltaic system data and the power output. Bruneau et al. [30] proposed a hybrid-physical model that combines numerical weather prediction data with recurrent neural networks, significantly improving the accuracy of photovoltaic power forecasts compared to purely physical approaches. Cotfas et al. [31] reviewed recent advances in linear statistical and machine learning techniques for photovoltaic power prediction, highlighting their ability to effectively manage the inherent variability and uncertainty of solar radiation forecasts. Pereira et al. [32] developed artificial neural networks specifically for direct normal irradiance forecasting, showing substantial improvements over raw numerical weather predictions, especially in capturing the complex non-linear interactions between atmospheric variables and solar irradiance. While the physical approach requires a detailed understanding of the physical processes and transport phenomena, the data-driven approach relies on a large set of experimental data which will only be available once a specific photovoltaic module or system is under real operational conditions. In this perspective, the physical models approach exhibits a better versatility, since it can be implemented even before the commissioning of the photovoltaic system. This feature renders physical approaches also valuable during the initial stages of photovoltaic projects, as stakeholders leverage them to assess the economic viability. Ahmed et al. [33] reviewed various photovoltaic forecasting methods, highlighting physical models for their effectiveness in capturing the dynamic behavior of solar energy based on weather classification and cloud motion studies. Ohtake et al. [34] conducted a comprehensive review of photovoltaic power forecasting, emphasizing physical models' ability to address unique forecasting challenges posed by factors such as dust and snow accumulation. Ramirez-Vergara et al. [35] specifically assessed photovoltaic forecasting methods in the context of predictive maintenance, underlining the critical role of physically based models in accurately estimating system performance and preventing system failures.

Hybrid forecasting models are a combination of physics-based and data-driven models and has been shown to enhance forecasting performance, especially when incorporating additional data sources, as for example aerosol data [32]. Mathiesen and Kleissl [36] evaluated the performance of different NWP models and applied a stepwise multivariate fourth-order regression for intra-day solar radiation forecasting in seven locations in the continental United States obtaining improvements on the global horizontal irradiance (GHI) predictions. In Ref. [37] a hybrid architecture of recurrent neural networks and shallow neural networks was developed showing improved performance in predicting daily photovoltaic power generation. In Ref. [38], the ECMWF GHI forecasts were used in combination with model output statistics (MOS) to create daily solar energy predictions with reduced root mean square error (RMSE), while in Ref. [39] the authors improved hourly direct normal irradiance predictions. In Ref. [40], satellite-derived data and ECMWF forecasts were integrated with an Artificial Neural Network (ANN) model to improve intra-day solar radiation forecasting. A day-ahead forecasting study using machine learning and the Japanese mesoscale model showed that model performance strongly improved forecast accuracy by effectively addressing seasonal and spatial dependencies [41]. A hybrid approach using numerical weather prediction data combined with artificial neural networks significantly reduced forecasting errors across Brazil's Northeastern region [42]. Physics-informed persistence models incorporating cloud-radiation interactions successfully improved forecast accuracy of direct and diffuse irradiance, particularly for forecasts extending up to 6 h ahead [43]. A combined approach using numerical weather prediction and artificial neural networks for solar resource assessment in southern Portugal considerably enhanced solar irradiance predictions compared to purely numerical models [44]. Expert knowledge in selecting physics-based predictor variables demonstrated clear improvements in photovoltaic power forecasting accuracy and model interpretability [45]. A benchmarking study extensively compared statistical and naïve reference forecasting models, establishing clear performance standards for evaluating solar radiation prediction models [46]. A comprehensive review of machine learning methods for solar radiation forecasting highlighted that hybrid models and ensemble approaches generally achieved superior accuracy compared to individual methods [47]. An extensive review systematically classified forecasting methods by temporal and spatial resolutions, identifying hybrid methods as generally most effective for accurate solar power predictions [48]. Lastly, a broad review on global solar radiation prediction with machine learning emphasized the significant role of feature selection and input data quality in achieving robust and precise forecasts [49].

One aspect of photovoltaic power forecasting that usually is not considered in the literature is the forecast time horizon and temporal resolution. Jung et al. [50] addressed temporal downscaling specifically, converting coarse-resolution solar irradiance forecasts into finer resolutions (e.g., from 3-hourly to hourly), highlighting the necessity and challenges of adapting temporal resolutions for practical forecasting purposes. Yang et al. [51] developed an operational solar forecasting algorithm aligned with real-time market needs, explicitly demonstrating how forecast time horizon and temporal resolution critically impact forecasting accuracy and usability for energy system operators. Solar power forecasting models should provide forecasts over different time frames depending on the energy market requirements and allowing utility companies to make decisions to counteract forecasted shortfalls in solar power output. The time scale of grid load variations shows the need for different forecasting time scales and prediction horizons with higher power fluctuations requiring a higher temporal resolution for accurate analysis [52] while a multiple time-scale data-driven forecast model leveraging spatial and temporal correlations has shown improved performance compared to conventional models [53].

In this context, there is still a need to develop a comprehensive and versatile photovoltaic power forecasting algorithm that can be applied globally for a wide variety of systems, without the need for ground-based observations in the locations of interest and, ideally, only requiring data provided by the manufacturer of the systems and characteristics of the power plant. This will reduce the need for additional sensors and measurements, making the forecasting process more cost-effective and easier to implement. However, this means that purely data-driven models are not suitable.

In this work, a comprehensive photovoltaic power forecasting algorithm is developed. This algorithm combines physics-based and ANNs models to enhance DNI forecasts from NWP models. ANNs are capable of modeling relationships between data sets and were validated using DNI observations from a specific site, proving applicable to a broader region [32,44]. The algorithm provides 72-h photovoltaic power forecasts with customizable temporal resolution, applicable to any fixed crystalline silicon photovoltaic system without requiring on-site observations. It assumes a centralized inverter architecture with single or multiple maximum power point tracking circuits. As mentioned above, this feature reduces costs by eliminating the need for expensive monitoring equipment and maintenance, making it scalable and suitable for regions with difficult access or lacking historical data. Adaptable to different configurations, the algorithm supports continuous, real-time forecasting and integration into grid operations algorithms or procedures, enhancing the management of photovoltaic energy generation variability.

This paper is organized as follows: a description of the different models used in the developed algorithm is presented in Section 2 while its validation against observations, including results and discussion, is presented in Section 3. Finally, the conclusions are drawn in Section 4.

2. Algorithm for photovoltaic power forecasting

The flowchart of the developed integrated forecasting algorithm is presented in Fig. 1, including the different components, namely, the temporal and spatial downscaling of input data, the artificial neural network, the transposition model, the coupled thermal and electric model and the losses and inverter models. The algorithm was developed using MATLAB software [54] takes as inputs hourly forecasts of different meteorological and aerosol variables retrieved from an operational NWP model, in this case the IFS/ECMWF and the Copernicus Atmosphere Monitoring Service (CAMS).

The IFS/ECMWF is the most widely used global NWP model in Europe being its performance attested by various studies such as in Refs. [2–4]. These forecasts are downscaled to the desired temporal resolution and for the location of interest with a temporal horizon of 72 h. The choice of a 72-h forecast horizon ensures the algorithm's adaptability to various operational needs. While 24-h forecasts are standard for grid operators, extending the forecast horizon to 72 h provides valuable insights for medium-term planning, including resource allocation and maintenance scheduling. Additionally, the algorithm is designed to run and update every 24 h, incorporating the latest data to provide rolling forecasts, which ensures that the most recent information is always considered. In the next step, a decision of whether to use or not ANN models is made to obtain improved forecasts of DNI. These models, such

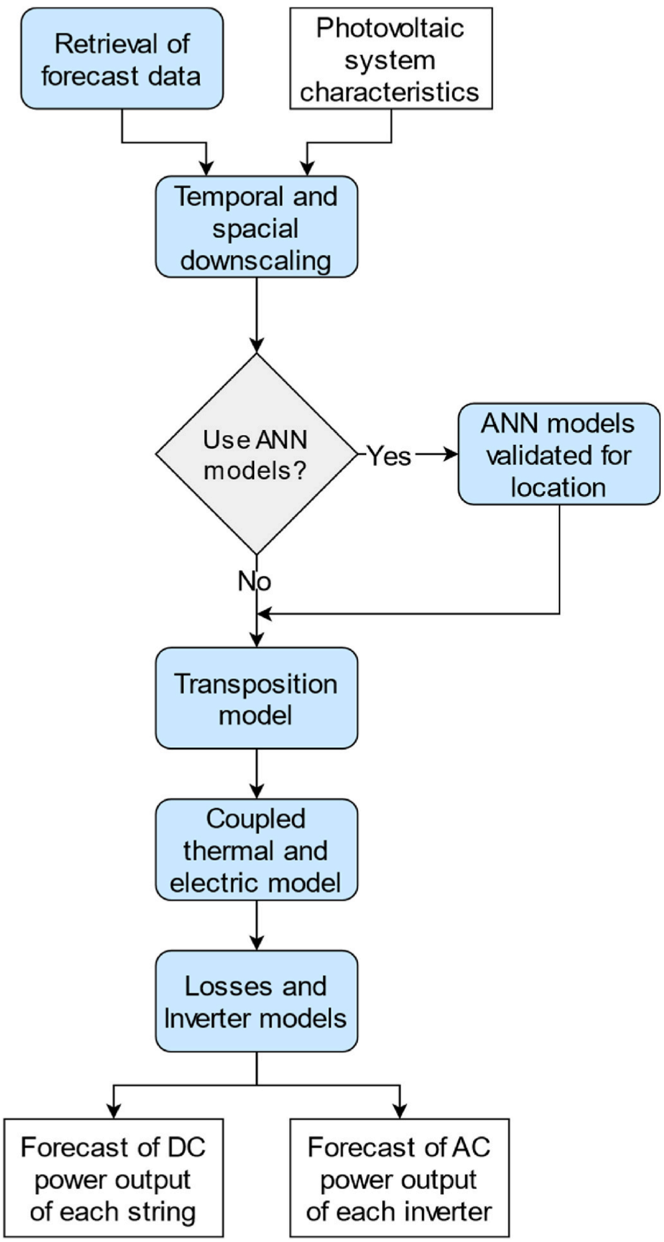

Fig. 1. Flowchart of the developed algorithm for photovoltaic power forecasting using non-observational data. Blue - models; white - inputs/outputs; gray - decisions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

as those developed in Ref. [32], are validated for a given region that should encompass the location of the photovoltaic powerplant. The global tilted irradiance on the surface of the modules is then computed through a transposition model which, together with the downscaled forecasts of air temperature and wind speed and direction, serves as inputs of a dynamic coupled thermal-electric model of the photovoltaic modules in order to obtain power output of each string of the system. The derating factors and conversion from DC to AC are then modeled and the final forecasts of photovoltaic power output are obtained.

In the following sections each main component or model of the algorithm is presented in more detail.

2.1. Input data

The various inputs needed to run the algorithm can be categorized into forecast variables and photovoltaic system variables. In this work, the forecast variables are retrieved from IFS, the operational NWP model of the ECMWF [55], and from CAMS, the Copernicus Service that provides forecasts about constituents such as greenhouse gases, reactive gases, ozone and aerosols [56].

The IFS/ECMWF model incorporates the ecRad radiative scheme [57], which adeptly solves the one-dimensional radiative transfer equation both in the short and long wavelength spectra. This model considers vertical profiles of air temperature and humidity, cloud properties (droplet and ice cloud effective radius), monthly mean climatological data of aerosols, gases such as carbon dioxide and ozone as well as trace gases. Additionally, it considers surface and land cover characteristics, including temperature and albedo and emissivity across different spectral bands and solar zenith angles. The underlying code for this scheme is rooted in the Rapid Radiative Transfer Model (RRTM), leveraging the Monte Carlo Independent Column Approximation (McICA) method to parameterize interactions between radiation and cloud cover [57].

The operational deterministic ECMWF model is executed twice daily, generating forecasts at 00UTC and 12UTC. These forecasts offer hourly predictions extending up to 90 h into the future. Beyond this period, the model provides forecasts at 3-h intervals up to 144 h and at 6-h intervals up to 240 h, all at discrete points across a global grid covering the entire globe with a horizontal spatial resolution of $0.125^\circ \times 0.125^\circ$.

CAMS offers a comprehensive global atmospheric composition forecasting system, building upon the IFS model but incorporating supplementary modules tailored for aerosols, reactive gases, and greenhouse gases. This model considers various emission and transport phenomena, including the emission and transport of trace gases and aerosols, the exchange of these components with vegetation and land or sea surfaces, their removal through dry deposition at the surface and scavenging by precipitation, as well as chemical transformations and aerosol microphysics. It generates a set of prognostic variables related to the atmospheric composition, which includes the aerosol optical depth at various wavelengths, available in a three-dimensional grid, with an horizontal spatial resolution of approximately 40 km and a temporal step of 1 h [58]. Hourly mean total aerosol optical depths are computed daily at 00UTC and 12UTC, extending over a forecasting temporal horizon of 5 days.

IFS and CAMS outputs are made available after 6.2 h from the starting time. Depending on the time zone of the system's location the use of 00UTC or 12UTC forecast runs can be replaced depending on the goal of the user (same-day/day-ahead forecast).

An overview of the variables retrieved from IFS and CAMS is presented in Table 1. The ANN model developed in Ref. [32], which is applicable to any location in the south of Portugal (latitude values below 39.2692°) without need for further procedure of training and validation, is also used in this work. ANN models can be trained and validated for

Table 1
Input variables obtained from numerical prediction systems (* - Variables which are not required if not using the ANN model).

Variables obtained from IFS/ECMWF	Variables obtained from CAMS
Date	Total aerosol optical depth at 670 nm*
Direct normal irradiation (J/m^2, accumulated)	Total aerosol optical depth at 865 nm*
Global horizontal irradiation (J/m^2, accumulated)	Total aerosol optical depth at 1240 nm*
Low cloud cover*	Sea salt aerosol optical depth at 550 nm*
Medium cloud cover*	
High cloud cover*	
Total cloud cover*	
U wind component (m/s)	
V wind component (m/s)	
Air temperature (K)	
Solar zenith angle (°)	

the region of interest through a similar procedure as described in Ref. [32] for generation of improved solar irradiance forecasts for the system location. If the system location is outside the ANN model validation area, it is recommended that this step is excluded being the variables marked with (*) unnecessary for the run of the algorithm.

The photovoltaic system characteristics include the geographical location, type and characteristics of the photovoltaic modules, inverter, mounting and racking and grid connection, since all have an influence on the photovoltaic power output. For an algorithm that only uses readily available data, the photovoltaic system properties can typically be obtained only through the powerplant project and the datasheets of the photovoltaic modules and inverters without needing to deploy and maintain monitoring equipment, sensors or any data collection infrastructure. The variables required as input for the developed integrated algorithm are presented in Table 2.

2.2. Temporal and spatial downscaling

To obtain forecast values for a specific location with higher temporal resolution, spatial and temporal downscaling techniques were employed for all forecast variables [32]. A comprehensive flowchart of the methodology employed is shown in Fig. 2.

The forecast variables were first processed in order to compute hourly mean values for GHI and DNI, expressed in W/m^2, air temperature in °C, wind speed in m/s and wind direction in ° taking the North direction as reference and being East 90°.

Temporal downscaling relies on piecewise cubic hermite interpolation of hourly mean irradiance data. For each subinterval, an hermite interpolating polynomial is specified for the given data points being shape preserving. The slopes at the interpolation points are chosen in such a way that the polynomial preserves the shape of the data and respects monotonicity [32]. Therefore, on intervals where the data is monotonic, so is the polynomial, and at points where the data has a local extremum, so does the polynomial. In this algorithm, the desired time step can be defined by the user.

Spatial downscaling involves bi-linear interpolation by considering the four neighboring grid points surrounding the desired location.

Table 2
Photovoltaic system properties used as input.

Power plant characteristics	Module characteristics	Inverter characteristics
Longitude (°)	Maximum power at STC (W)	Inverter efficiency (%)
Latitude (°)	Voltage at maximum power point for STC (V)	Nominal power of the inverter (W)
Altitude above the m.s.l. (m)	Current at maximum power point for STC (A)	Number of MPPT circuits
Tilt angle of modules (°)	Open circuit voltage for STC (V)	
Azimuth angle of modules (°)	Short circuit current for STC (A)	
Ground albedo	Thermal coefficient of maximum power (%/°C)	
Module orientation (portrait/landscape)	Thermal coefficient of short circuit current (%/°C)	
Length of module rows (m)	Thermal coefficient of open circuit voltage (%/°C)	
Height of module rows (m)	Number of photovoltaic cells in series	
Distance between rows in the horizontal plane (m)	Length (m)	
Vertical distance between the ground and the panel base (m)	Width (m)	
Strings and inverter configuration (series/parallel)		

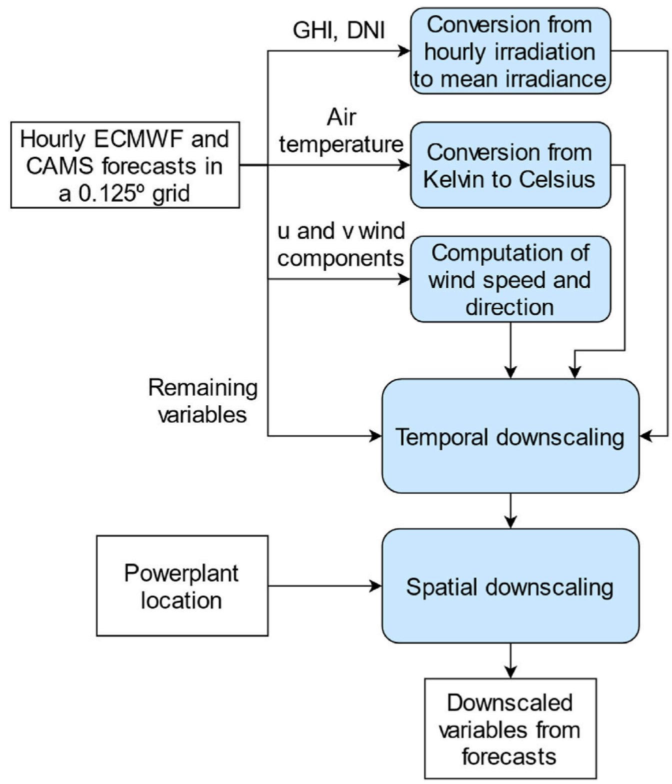

Fig. 2. Flowchart of the temporal and spacial downscaling procedure. Blue - models; white - inputs/outputs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

2.3. ANN models for improved DNI forecasts

In [32], ANN models were developed and optimized to improve DNI forecasts in a spatial and temporal downscaled grid and timestep, respectively. These models were developed for the region of South Portugal (latitude values below 39.2692°), but the same approach can be generalized provided that the ANNs are finetuned for any other region through a similar process of training and validation [32]. In this work these same ANN models are used, and a flowchart of this part of the proposed algorithm is shown in Fig. 3.

The data used as inputs are downscaled forecasts of operational outputs from the ECMWF/IFS and the CAMS models as described in Section 2.1 and 2.2 of the variables presented in Table 1. The downscaled variables are fed into a feed-forward artificial neural network, referred to as ANN model A, including one hidden layer with seven neurons. A backpropagation learning function is used, more specifically the Bayesian regularization backpropagation function, along with a linear layer output using the Nguyen-Widrow initialization algorithm for weights and biases. The network utilizes the hyperbolic tangent sigmoid transfer function and assesses performance using the mean squared error. ANN model A accounts for the nonlinear relationships between atmospheric and aerosol variables and the DNI. Results show improved DNI forecasts at the location of interest and different temporal resolutions defined by the user [32].

Additionally, a second artificial neural network, referred to as ANN model B, was included in the algorithm which takes into account a time series of 12 time steps of the improved DNI forecasts from the ANN model A leading up to the forecast time, in addition to considering seasonality and time of day. This approach captures the temporal variation of DNI, further improving the DNI forecasts obtained through ANN model A. Similar to model A, the ANN model B has a hidden layer, but it incorporates the Levenberg-Marquardt backpropagation algorithm and features eight neurons.

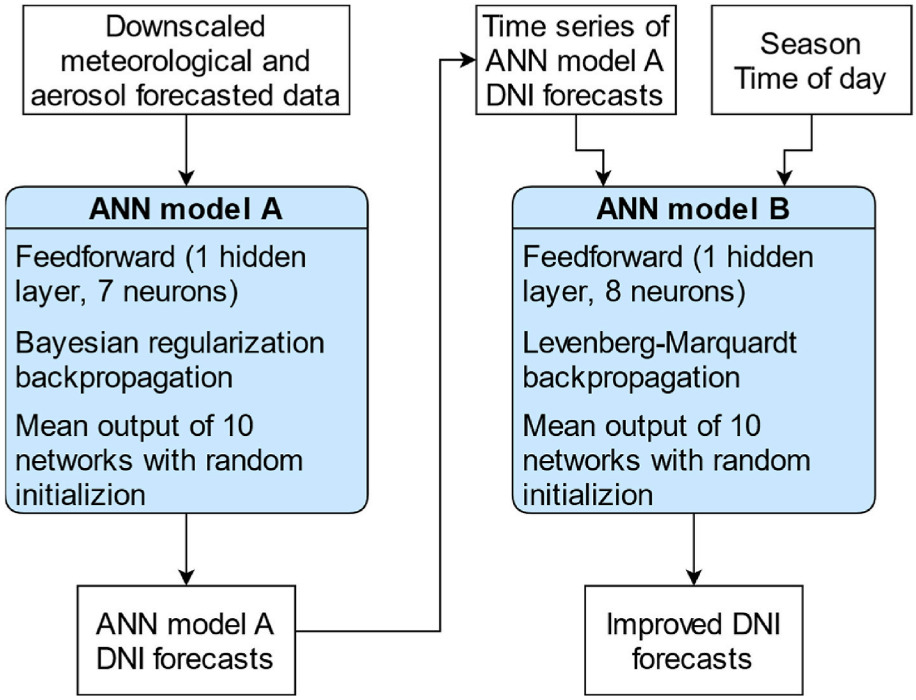

Fig. 3. Flowchart of the ANN models used in the algorithm. Blue - models; white - inputs/outputs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

In both models, a strategy regarding the training and validation procedures involving ten randomly initialized ANNs for each configuration was employed. In this approach, the average output of these ten ANNs is considered as the result for the respective ANN configuration. This methodology aligns with established practices in the field, as previously demonstrated in related studies [32,44].

2.4. Transposition model

For the computation of power output of a solar system the global solar irradiance on its surface is needed. To obtain this, a transposition model was employed that transposes direct and diffuse components into GTI, including the part that is reflected, according to the geometry of the system and albedo of the surfaces [1]. Furthermore the computation of incidence angle modifiers and absorbed irradiance is critical to estimate power output of photovoltaic modules. This aspect was included in this work and Fig. 4 shows this part of the algorithm.

The transposition model from Ref. [1] was selected due to its demonstrated accuracy in estimating global tilted irradiance for photovoltaic systems with multiple rows, validated against high-resolution experimental data and compared to other models available in literature. Unlike other models developed for the first row only, the selected model accounts for shading and sky masking effects on inner rows, considering view factors, circumsolar irradiance obscuration, and reflections from surrounding surfaces. Validation results showed substantial improvements in accuracy, with a decrease in the mean bias error and root mean square error, even under conditions without direct shading.

The transposition model used in this work follows the model presented in Ref. [1]. This model adopts the common representation of panels arranged in rows, where the length of the panels significantly exceeds their height and computes the GTI on the front and inner rows of photovoltaic power plants. The different surfaces, namely the solar module being evaluated, the rear of the front row (for modules being evaluated belonging to rows other than the first) and the ground between the rows are discretized into segments, being the GTI computed for each segment.

The Modified Bugler model [59] is taken as the base model from which the isotropic and circumsolar diffuse fractions are taken [1]. Then, the different obscuring angles are compared with the solar elevation angle projected to the surfaces' azimuth to obtain the direct and circumsolar irradiance shading for all the surfaces in the model. The different view-factors between each of these segments are computed while the albedo is assumed equal to 0.2 for the ground, zero for the panel being considered and 0.92 for the rear surface of the front panel (typically white). Finally the GTI on the panel being evaluated is computed taking also into account the reflected solar irradiance from each corresponding segment of the ground and the back of the front panel if it is present. More detail on this model can be found in Ref. [1].

The transposition model determines the irradiance incident on a photovoltaic panel's surface, but for power output computation, the absorbed irradiance is needed. This algorithm uses available GTI and DNI data, along with the reflected and diffuse irradiance components, to obtain the absorbed irradiance for each segment of the panel. Incidence angle modifiers for direct, diffuse and reflected irradiance components are calculated based on Snell's and Bouguer's laws according to Ref. [60]. The required incidence angles for each panel segment are derived for direct, isotropic diffuse, circumsolar diffuse and reflected irradiance components. The incidence angles for the direct component are calculated using the panel's latitude, solar declination, tilt angle, azimuth, and solar hour angle. Equivalent incidence angles of isotropic and circumsolar diffuse for the front row are determined and adjusted for other rows using an equivalent tilt angle. Equivalent incidence angles of reflected irradiance are computed for reflections from the back of the front row and the ground between rows. With these angles, the incidence angle modifiers are calculated, and the absorbed solar irradiance for each panel segment is obtained. More detail on this approach is presented in Appendix A.

2.5. Coupled thermal-electric model of the photovoltaic module

After computing the irradiance absorbed by photovoltaic modules the next step is to model its conversion into electricity. Various electrical models exist, ranging from simple proportional relationships to more

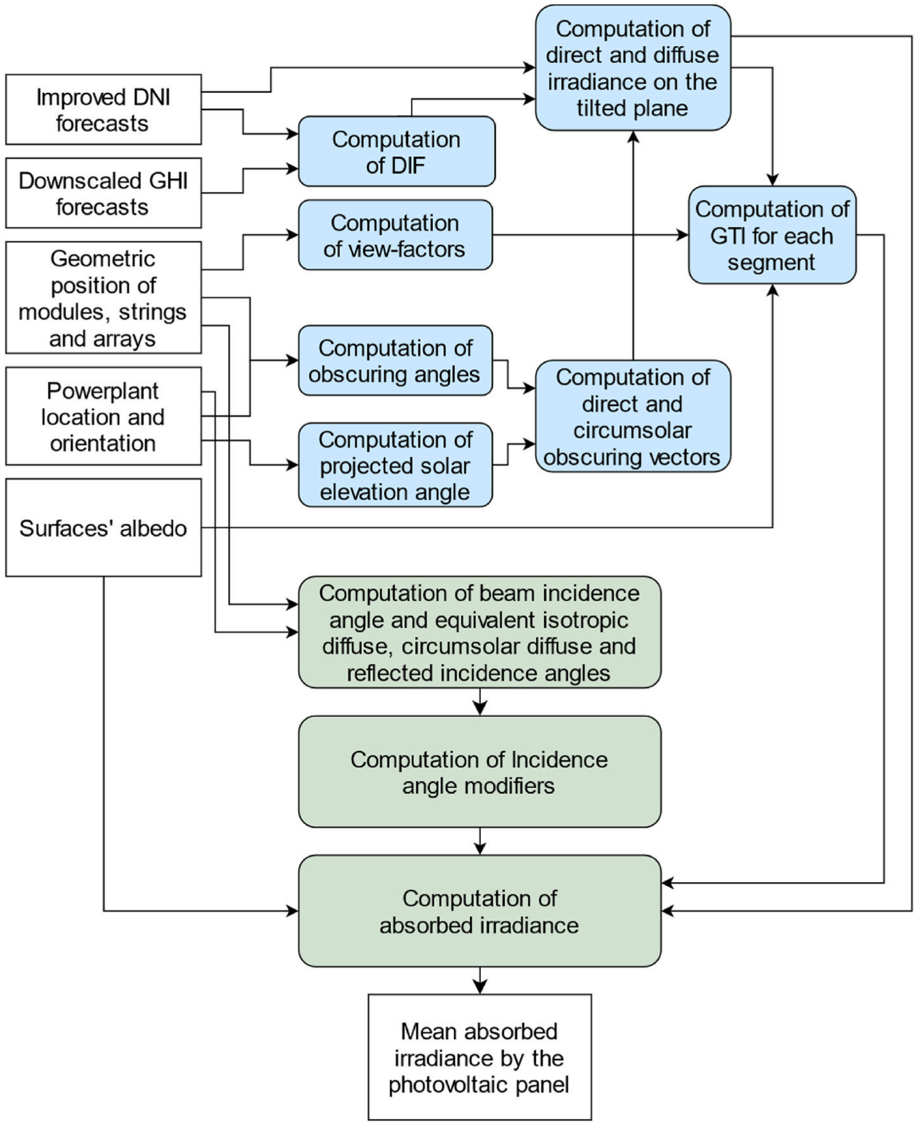

Fig. 4. Flowchart of the transposition model and computation of absorbed irradiance. Blue - models from work of [1]; green - models added in this work; white - inputs/outputs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

complex equivalent circuit models. On the other hand, thermal models are also important since temperature is related to and has a reasonable impact on electric efficiency of photovoltaic conversion. Combining electrical and thermal models provides a comprehensive modeling approach of the response of photovoltaic modules under varying environmental conditions.

In this work, a model as that in the work presented in Ref. [17] is used, which is a dynamic coupled thermal-electric model designed for crystalline silicon cells using readily available information provided by the manufacturers (see Fig. 5).

The thermal model is based on the fundamental principle of energy conservation and on the description of the heat transfer processes that occur in illuminated photovoltaic modules in transient regime. The heat transfer processes by convection and thermal radiation on front and back surfaces of the module are considered, while conduction is assumed negligible since the contact points between the module and supporting structure are small. The influence of wind speed and direction is taken into account when determining the heat transfer coefficients for forced convection, as these factors have a strong impact on the module temperature. The electrical model used in this work is the single diode and five parameters equivalent electrical circuit [17] being the five parameters extracted solely from the information obtained from the datasheet of the modules. This is a dynamic model thus it provides the variation of module temperature and electric power output simultaneously at a given time step defined by the user. For a more detailed explanation of this model please refer to Appendix B and [17]. Taking into account the electric connection between strings and arrays, this model computes the maximum power point of each string considering the configuration of the photovoltaic powerplant.

2.6. Electric losses and inverter model

Fig. 6 shows the flowchart of the algorithm for determining the electric losses and inverter efficiency. Performance losses of photovoltaic systems are typically represented by a derating factor, which scales the power output of photovoltaic arrays to account for real operation conditions in the field. The derating factor accounts for various losses independent of temperature, including DC losses, AC losses, and other such as soiling and shading. The derating factors are usually determined through field measurements or estimations and have a negative impact on the photovoltaic system energy yield. The inverter's efficiency is not included in the derating factors but is considered as a separate input

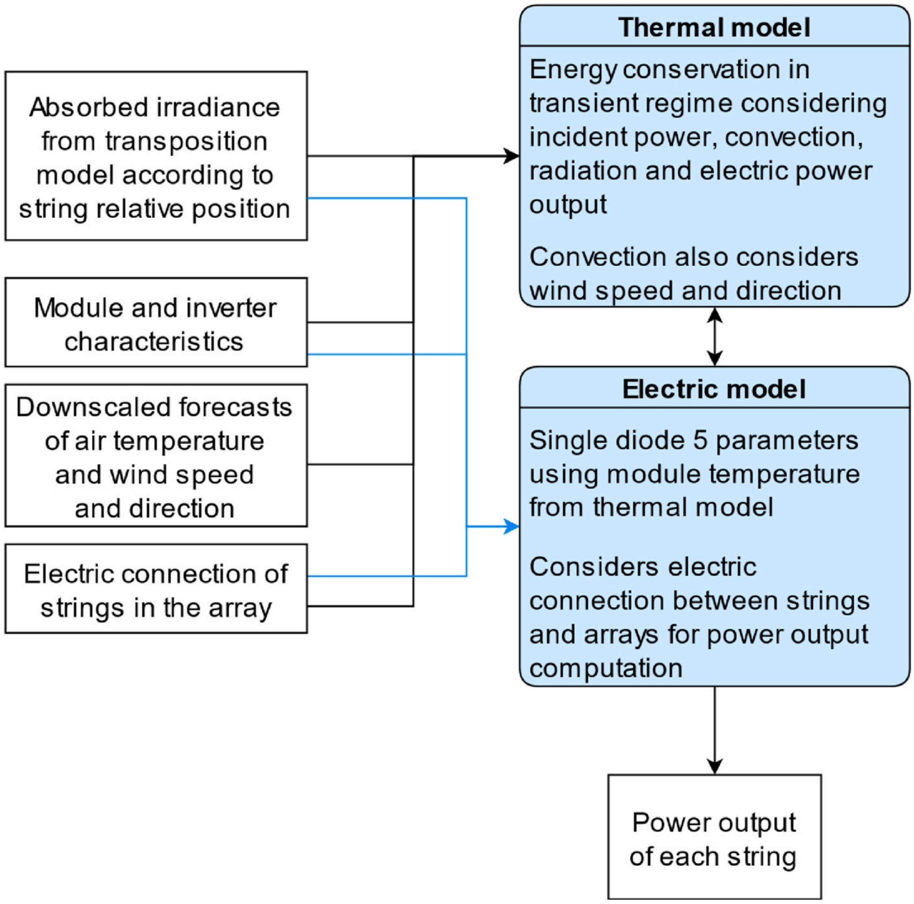

Fig. 5. Flowchart of the coupled thermal-electric model of photovoltaic modules. Blue - models; white - inputs/outputs. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

parameter. Roberts et al. [61] reviewed different derating factors in the literature and obtained typical values for losses in DC wiring, in diodes and connections, module mismatch losses, maximum power point tracking inefficiencies, soiling, degradation induced by the continued exposition to solar radiation and adverse environmental conditions, as shown in Table 3. The user can choose to apply or not these typical derating factors in the algorithm, or to apply different factors that reflect site-specific conditions or experimental data.

As for the conversion from direct to alternate current the efficiency of the inverter and its nominal power are taken into consideration, both values are readily available in the datasheet of the device. To obtain the AC power forecast the DC power of each input to the inverter are summed, clipped if the sum overcomes the nominal power of the inverter multiplied by its efficiency.

2.7. Output

The algorithm was developed to run every day, retrieving data from the NWP models as soon as available. The output comprises the photovoltaic DC power output of each string and AC power output of each inverter from 00UTC or 12UTC of the forecast issue day and with a forecast horizon of 72h in text (ASCII) format. The temporal resolution can be selected by the user but should typically be under 60 min.

The following is an example of an excerpt of an output file defined with 15-min time step:

Year Month Day Hour Minute PowerString1W PowerString2W PowerString3W PowerString4W PowerString5W PowerString6W PowerString7W ACPowerW

(…)

2022 11 10 11 0 4191.4 4216.5 4184.3 4184.6 4191.4 4216.5 4216.6 28930.8

2022 11 10 11 15 4369.7 4394.1 4362.5 4362.8 4369.7 4394.1 4394.1 30156.5

2022 11 10 11 30 4487.8 4512.2 4480.5 4480.8 4487.8 4512.2 4512.2 30969.9

2022 11 10 11 45 4543.3 4568.1 4536.1 4536.4 4543.3 4568.1 4568.1 31353.6

(…)

3. Algorithm results and validation

In this section the developed algorithm is validated against real data from a photovoltaic powerplant.

3.1. Data

For the validation of this algorithm data from a photovoltaic powerplant located in the region of Lisbon, Portugal commissioned by Helexia was used. This powerplant comprises 155 strings of 16 monocrystalline modules each, model SRP-400-BMA-HV PERC [62]. Additionally, data from one inverter model SUN2000-36KTL was used [63], namely current and voltage at the maximum power point of seven different strings (see Fig. 7) as well as the resulting active and reactive power.

Each pair of strings is connected in parallel (1 and 2, 3 and 4, 5 and 6) to one of the four maximum power point tracking circuits of the inverter except string 7 which is connected to its own circuit. These strings were chosen so the developed transposition model used [1] could be

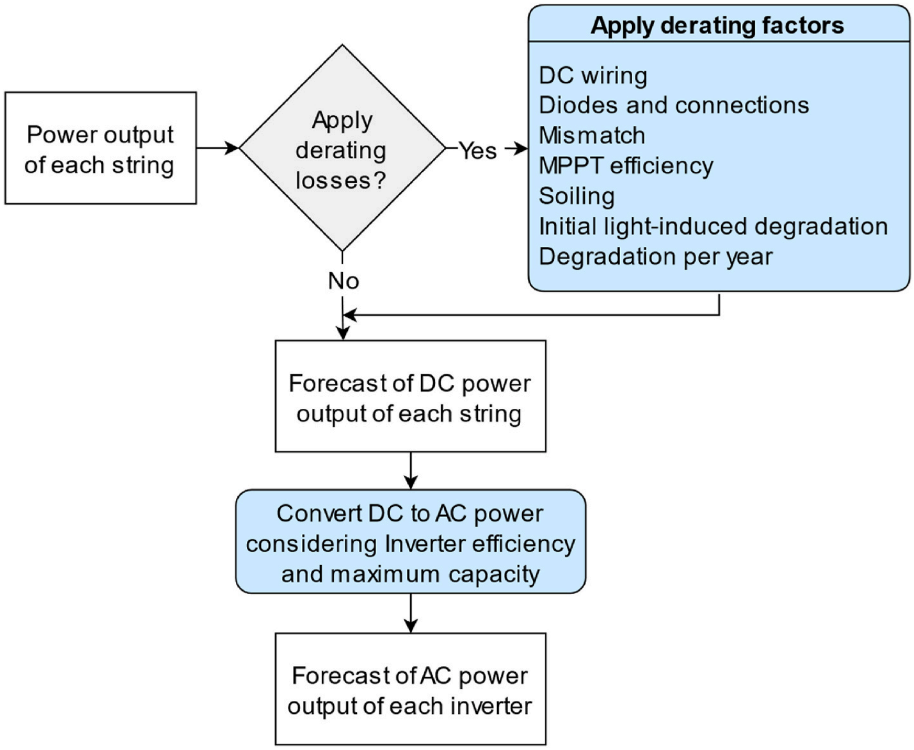

Fig. 6. Flowchart of the electric losses and inverter efficiency models. Blue - models; white - inputs/outputs; gray - decisions. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 3
Typical values of the derating factors [61].

Derating factor	Typical value
DC wiring	0.980
Diodes and connections	0.995
Module mismatch	0.980
Maximum power point tracker efficiency	0.990
Soiling	0.980
Degradation rate per year	0.985
Initial light-induced degradation	0.980

additionally tested for first (string 2, 6 and 7) and inner rows of modules (strings 1, 3, 4 and 5), now for a real power plant in the field. The different input variables related to the powerplant, modules and inverters needed to run the algorithm are presented in Tables 4–6, respectively.

This powerplant was commissioned in 2021 and the data used ranges from April 1st, 2022, and April 30th, 2023, with a timestep of 5 min. Measurements of global tilted irradiance were obtained for the same period and time step from a calibrated silicon irradiance sensor SI-RS485TC-T-MB (calibrated solar cell) [64] on the tilted plane of modules.

These observations were also filtered according to the BSRN quality control procedure considering the extremely rare limits [65] and following the procedure established in other works in this field. A previous study [44] applied similar BSRN quality-control filters to ensure reliable input data for solar resource assessment. Another work [66] used the BSRN guidelines specifically to detect physically impossible or rare values in long-term solar radiation datasets from several locations. Additionally, a detailed methodology [67] was developed using BSRN quality control combined with statistical checks and gap-filling procedures to produce high-quality direct normal irradiance (DNI) datasets. For all observations, negative values in the records were assumed equal to zero, missing values and outliers were discarded and then mean values for a 15 min time step were computed.

According to the developed algorithm, the ECMWF and CAMS forecasts were obtained for the location and period of interest.

For computational time analysis a computer with processor Intel(R) Core(TM) i7-8550U with a base clock speed of 1.80 GHz and 16 GB of RAM was used.

3.2. Results and discussion

The integrated algorithm was run for the observation site and period described in Section 3.1 with a temporal forecast horizon of 72 h and a timestep of 15 min. Validation was carried out through the comparison between predicted and observed values of GTI, output power of each string and output power of the inverter. The metrics used were the correlation coefficient (R^2), mean bias error (MBE), mean absolute error (MAE), root mean squared error (RMSE) and their relative values considering the mean of observations (rMBE, rMAE, rRMSE, respectively). A forecast skill score (FS_{GTI}) representing the impact on the GTI forecasts of using the ANN models for improvement of DNI predictions is also used. This indicator is defined as one minus the ratio of the MAE with the ANN model to the MAE without the ANN model for GTI. The score quantifies the improvement in forecast accuracy, with a higher score indicating a greater positive impact of the ANN model on the GTI forecasts.

The results for the GTI forecasts including the use of the ANN models are presented in Table 7 based on all 395 algorithm runs, each providing data over a 72-h time horizon with 15-min timestep. As expected, the MAE and RMSE increase with forecast horizon, although MBE shows better results for forecast day 1. This improvement is highest for forecast day 0 with a value of 4.1 %.

As an operational example that shows the outputs of a run of this algorithm, a forecast period of three days was selected starting the 10th of November of 2022, being the forecast data used as inputs the ECMWF and CAMS forecasts issued at 00UTC of that day. This period was selected because the first two days are clear sky days, while on the third

Fig. 7. Aerial view of the photovoltaic powerplant (1–7: strings considered in the work).

Table 4
Input values for variables related to the powerplant characteristics.

Powerplant characteristics	Value
Longitude	38.955591°
Latitude	−9.191087°
Altitude above the m.s.l.	276.5 m
Tilt angle of modules	25°
Azimuth angle of modules	0°
Ground albedo	0.2
Module orientation	Portrait
Length of module rows	8.02 m
Height of module rows	4.03 m
Distance between rows in the horizontal plane	6.45 m
Vertical distance between the ground and the panel base	1.00 m
Strings and inverter configuration	Pairs of strings in parallel

Table 5
Input values for variables related to the module characteristics.

Module and string characteristics	Value
Maximum power for STC	400 W
Voltage of maximum power point for STC	41.6 V
Current at maximum power point for STC	9.62 A
Open circuit voltage for STC	49.1 V
Short circuit current for STC	10.10 A
Thermal coefficient of maximum power	−0.36 %/°C
Thermal coefficient of short circuit current	+0.05 %/°C
Thermal coefficient of open circuit voltage	−0.28 %/°C
Number of photovoltaic cells in series	72
Length	2.02 m
Width	1.00 m
Number of strings	7
Number of modules per string	16
String power for STC	6.4 kW

Table 6
Input values for variables related to the inverter characteristics.

Inverter characteristics	Value
Inverter efficiency	0.984
Nominal power of the inverter	40 kW
Number of MPPT circuits	4

day there are some clouds around midday. Fig. 8 shows the GTI results with and without the use of the ANN model and the experimental data. On the third day, the original ECMWF forecasts did not predict this effect of cloud cover. The graph shows the underestimation of both GTI forecasting models as well as the improvement achieved by applying the ANN model.

The computation of the power generation in each string was performed as described in Section 2.5, where the maximum power point tracking of parallel strings (namely strings 1 and 2, 3 and 4 and 5 and 6) was taken into consideration, while string 7 is connected to its own maximum power point tracking circuit. It is important to note that strings 2, 6 and 7 were considered as first row surfaces while strings 1, 3, 4 and 5 were considered as inner-row surfaces differing from each other in the GTI forecasts used, depending on their position in the power plant. The results of this study, including the impact of using the ANN models, are presented in Table 8 for each string and forecast day based on all 395 algorithm runs, each providing data over a 72-h time horizon with 15-min timestep. Here, a forecast skill score (FS_{POW}) is also included representing the improvement of using the ANN models on the power generation prediction. This indicator is defined as one minus the ratio of the mean absolute error (MAE) with the ANN model to the MAE without the ANN model considering DC power generation.

The overestimation of power output is clearly visible since until this point the various losses such as cable, module mismatch and deterioration losses were not considered yet. Similarly to the GTI forecasts, the MAE and RMSE values increase with the forecast horizon while the MBE

Table 7
Metrics of GTI predictions for first row and each forecast day (number of data points: 46076).

Forecast day	R^2	MBE (W/m^2)	rMBE (%)	MAE (W/m^2)	rMAE (%)	RMSE (W/m^2)	rRMSE (%)	FS_{GTI} (%)
0	0.844	25.1	5.7	84.5	19.3	131.0	30.0	4.1
1	0.828	24.6	5.6	88.2	20.2	137.2	31.4	3.3
2	0.807	27.3	6.3	92.4	21.2	145.6	33.4	2.9

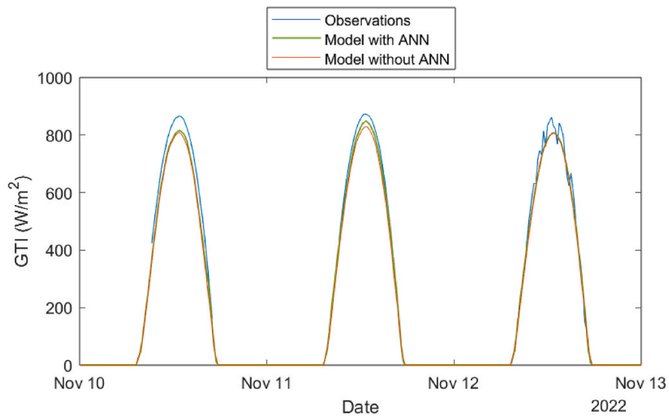

Fig. 8. Observed and predicted values of GTI using the original ECMWF data and the improved DNI and DIF forecasts from the ANN models as input to the transposition model for the operational example.

shows better results for forecast day 1. The contribution of the ANN models for improving results is also higher for smaller forecast horizons with the highest forecast skill score being 7.0 % for string 5 in day 0. The overall results show an overestimation of the power generation being the MBE and RMSE for all strings and the three days of forecast equal to 56.8 W/kWp and 142.2 W/kWp, respectively. This translates to relative values of 13.8 % and 34.5 %, respectively. The forecasts for string 1 typically have the lowest values of MAE and strings 2 and 5 the highest. The reasoning behind this might be the fact that string 2 is a front row and, although string 5 is considered inner row, it is on the edge of the powerplant and part of it behaves as front row. This means that these strings are exposed to more fluctuations in irradiance which are more difficult to predict. The following analysis is for the case in which the ANN models are used.

Fig. 9 shows the results of power generation for the operational example considered. A noticeable trend is the overestimation of power output across all strings, with the forecast curves consistently lying above the measured values. All strings show deviation between forecasted and actual values, particularly in the peak irradiance hours. This overestimation is likely due to higher exposure to irradiance and lack of consideration for certain losses in the model, such as soiling or module degradation, which is also shown in the aggregated results for "All strings" in the bottom-right plot. On the third day since the original forecasts did not consider the presence of clouds these were also not forecasted when applying the algorithm.

Assuming the electrical losses described in Section 2.6, the results in Table 9 are obtained, where the forecast skill ($FS_{POW\text{-}L}$) reflects the improvement gained by incorporating the losses model. This indicator is defined as one minus the ratio of the mean absolute error (MAE) with the ANN and losses models to the MAE with only the ANN model considering the DC power generation. The default/typical values used provide a reference for estimating losses, including the losses due to soiling, and can be adjusted by users if plant-specific data are available. This flexibility ensures the algorithm's adaptability to diverse operating conditions. Furthermore, the power output measurements used for validation inherently include soiling and other derating effects, ensuring that the algorithm's performance metrics accurately reflect real-world conditions.

Again, these values are based on all 395 algorithm runs, each producing data over a 72-h time horizon with a 15-min timestep. The overestimation was greatly reduced for all strings with an overall forecast skill value of 16.4 % regarding MAE, but now strings 4 and 5 show lower values of this metric while the highest are obtained for string 1. Considering the various electric losses of the photovoltaic system, the overall MBE decreases to 7.5 W/kWp and the RMSE to 123.7 W/kWp, being their relative counterparts 1.8 % and 30.0 %. There is a

Table 8
Metrics of photovoltaic power (DC) generation forecasts for each string and each forecast day (number of data points per string: 42689).

Day	String	R^2	MBE (W/kWp)	rMBE (%)	MAE (W/kWp)	rMAE (%)	RMSE (W/kWp)	rRMSE (%)	FS_{POW} (%)
0	1	0.835	42.2	10.0	88.0	20.8	129.3	30.5	6.6
	2	0.829	62.2	15.1	98.1	23.8	138.3	33.6	6.6
	3	0.830	57.0	14.0	96.6	23.7	135.2	33.1	6.9
	4	0.833	53.7	13.0	93.4	22.7	133.0	32.3	6.9
	5	0.833	62.0	15.4	97.8	24.2	136.5	33.8	7.0
	6	0.830	58.7	14.1	95.5	23.0	136.5	32.9	6.5
	7	0.829	57.4	13.8	95.2	22.8	136.4	32.7	6.4
1	1	0.817	41.9	9.9	92.1	21.8	136.1	32.2	6.2
	2	0.811	61.7	15.0	101.7	24.7	144.7	35.2	6.0
	3	0.812	56.7	13.9	100.5	24.6	141.8	34.7	6.4
	4	0.815	53.4	13.0	97.3	23.7	139.6	33.9	6.5
	5	0.814	61.7	15.3	101.5	25.2	143.0	35.5	6.5
	6	0.812	58.2	14.0	99.4	23.9	143.0	34.4	6.0
	7	0.811	56.9	13.7	99.0	23.8	142.9	34.3	5.9
2	1	0.794	44.6	10.6	96.6	22.9	144.9	34.3	5.5
	2	0.787	64.5	15.7	106.6	25.9	153.2	37.3	5.3
	3	0.788	59.4	14.6	105.1	25.8	150.2	36.8	5.8
	4	0.792	56.0	13.6	102.0	24.8	148.2	36.0	5.8
	5	0.791	64.4	16.0	106.3	26.4	151.3	37.6	5.8
	6	0.788	61.0	14.7	104.1	25.1	151.6	36.5	5.2
	7	0.787	59.7	14.3	103.8	25.0	151.5	36.4	5.1
All		0.811	56.8	13.8	99.1	24.0	142.2	34.5	6.1

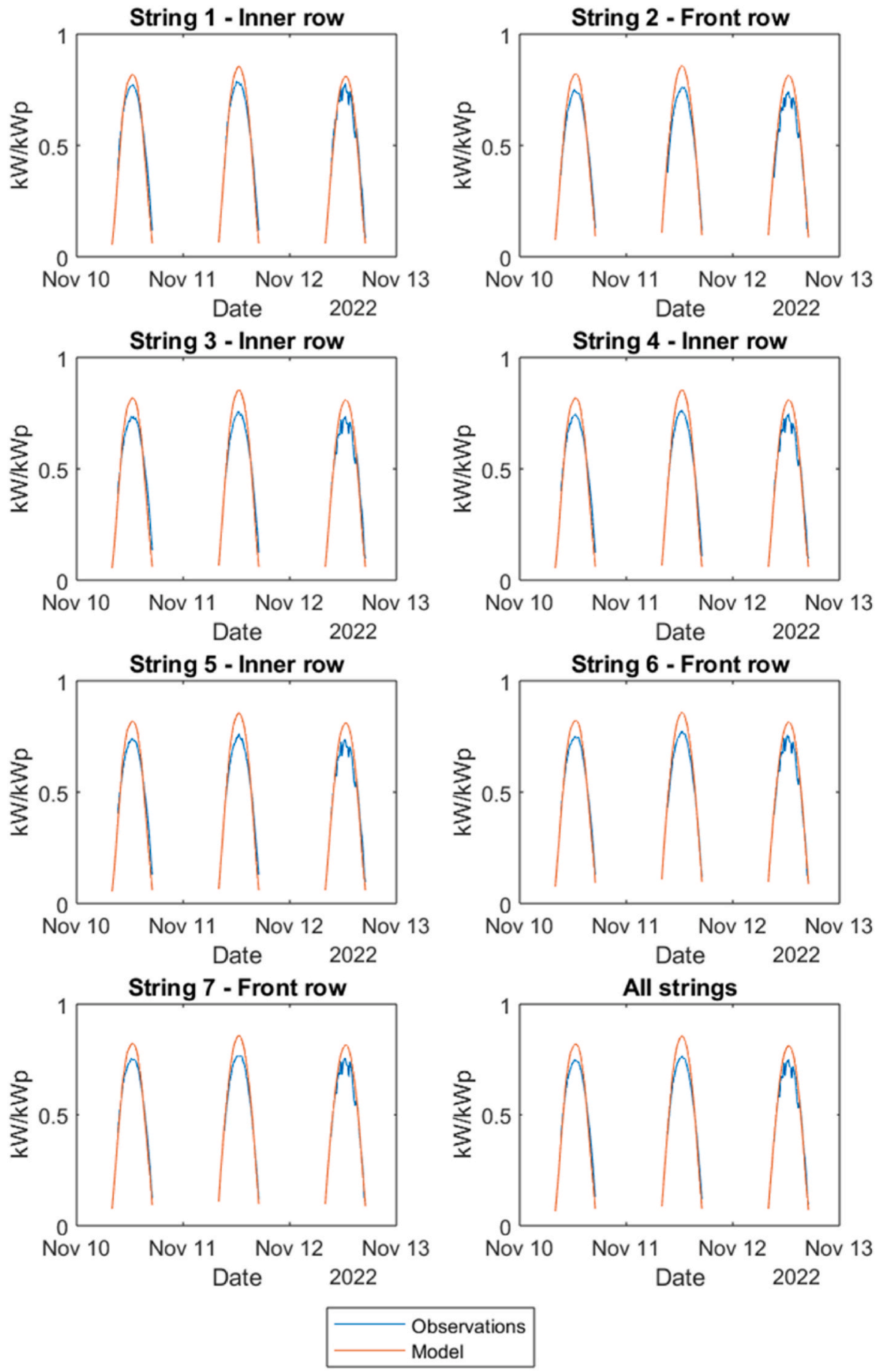

Fig. 9. Power generation (DC) forecasts using the ANN model (orange) and measurements (blue) of each string for the operational example. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

significative decrease of MBE values when including the losses model, which explains the negative MBE for String 1. Without applying the losses, String 1 already has the lowest MBE among all the strings and when the losses are applied as a factor the power output forecasts for all strings are proportionally reduced. However, since String 1's initial MBE was already lower as mentioned above, the additional reduction from the losses model caused it to become a negative value, indicating a slight underestimation of power output. In contrast the MBE of front-row strings, besides being lower, still retain positive values after the loss correction. The fact that string 1 presents lower MBE values than other strings can be attributed to the fact that string 1 is electrically connected in parallel with string 2 (which has a positive MBE) but are apart spatially in different rows, resulting in a potential mismatch in irradiance exposure between these two strings, with string 1 receiving less irradiance due to diffuse sky and reflected irradiance obstruction and shading [1], thus exhibiting a negative bias.

Fig. 10 presents the forecasts using the losses model and observations of power output of each string of the operational example. A clear improvement is visible when comparing Figs. 9 and 10 and thus, the following analysis will consider the use of the losses model. There is

Table 9
Metrics of photovoltaic DC power output forecasts of each string and each forecast day considering typical losses (number of data points per string: 42689).

Day	String	R^2	MBE (W/kWp)	rMBE (%)	MAE (W/kWp)	rMAE (%)	RMSE (W/kWp)	rRMSE (%)	$FS_{POW\text{-}L}$ (%)
0	1	0.835	−6.7	−1.6	81.6	19.3	118.3	28.0	7.3
	2	0.829	12.4	3.0	79.3	19.2	117.7	28.6	19.2
	3	0.830	8.1	2.0	78.7	19.3	115.6	28.3	18.6
	4	0.833	4.8	1.2	77.9	18.9	115.4	28.0	16.6
	5	0.833	13.1	3.3	77.0	19.1	114.6	28.4	21.3
	6	0.830	8.9	2.1	78.9	19.0	117.7	28.3	17.4
	7	0.829	7.6	1.8	79.4	19.1	118.1	28.4	16.5
1	1	0.817	−6.9	−1.6	85.5	20.2	124.6	29.5	7.2
	2	0.811	12.0	2.9	82.8	20.1	124.0	30.1	18.6
	3	0.812	7.9	1.9	82.4	20.2	122.1	29.9	18.0
	4	0.815	4.6	1.1	81.5	19.8	121.8	29.6	16.3
	5	0.814	12.9	3.2	80.8	20.1	121.1	30.0	20.4
	6	0.812	8.5	2.0	82.5	19.9	124.0	29.9	16.9
	7	0.811	7.2	1.7	83.1	20.0	124.5	29.9	16.0
2	1	0.794	−4.4	−1.1	89.4	21.2	132.4	31.3	7.4
	2	0.787	14.5	3.5	87.0	21.2	132.0	32.1	18.4
	3	0.788	10.3	2.5	86.4	21.2	130.0	31.9	17.7
	4	0.792	7.0	1.7	85.7	20.9	129.8	31.6	16.0
	5	0.791	15.3	3.8	84.8	21.1	129.1	32.0	20.2
	6	0.788	11.0	2.7	86.7	20.9	132.0	31.8	16.7
	7	0.787	9.7	2.3	87.3	21.0	132.4	31.8	15.9
All		0.811	7.5	1.8	82.8	20.1	123.7	30.0	16.4

overestimation in string 1, despite the overall negative MBE for this string. This may be due to the complex interplay between partial shading, mismatch losses, and other factors affecting inner-row strings like string 1. Therefore, while the bias for string 1 remains negative due to frequent underproduction, this specific figure reveals occasional instances where the model still overestimates output.

By incorporating inverter efficiency and maximum power clipping, the results shown in Table 10 are obtained. In accordance with the aforementioned findings, the differences tend to increase for higher forecast horizons except for MBE which is lowest for forecast day 1 with a value of 388.7 W. The overall results show an MBE of 9.3 W/kWp and RMSE of 121.0 W/kWp which translates to relative values of 2.3 % and 29.9 %, respectively. Fig. 11 shows the AC power output forecasts and measurements of the operational period taken as example, showing good agreement.

These results, namely the values of Table 10, are of the same order of magnitude and compares well with other work available in the literature. The authors of [68] evaluated seven methods for DC power forecasts of photovoltaic systems, with the best deterministic results achieved using the calibrated ensemble NWP paired with a random model chain (method 3C in Ref. [68]), which resulted in 26.1 % relative MBE and 43.1 % relative RMSE, while in this work values of 20.1 % and 30.0 % were achieved. Although the datasets used by the authors are different, all statistical indicators are normalized to the mean of observations. This evidences how the combination of NWP data with ANN models incorporating aerosol information, together with improved transposition and thermal-electric models, can contribute for generating improved forecasting results.

In this algorithm, specific characteristics of systems with optimizers or microinverters are not explicitly modeled at this stage. Such configurations can influence the system's performance under partial shading, mismatch, or other module-level effects, as these technologies allow for optimization of individual modules. While the presented methodology is adaptable and could potentially be extended to these configurations, at this stage the presented study does not include a module to model the response of these systems. Future work could incorporate detailed modeling of module-level optimization to improve forecast accuracy in systems with optimizers or microinverters.

An important aspect of any photovoltaic forecasting algorithm is its running time. Thus, an analysis based on the selected operational example was performed regarding the computational time that each process of the algorithm takes (Table 11). As explained in Section 2.1, the 00UTC forecast runs can take up to 06:12UTC to become available for the users with the download time dependent on internet connection. The most time-consuming process after the retrieval of the NWP data is the coupled thermal-electric photovoltaic model as expected since it involves an iterative numerical process. The total run time of the algorithm is found to be approximately 12.8 min, thus allowing for the efficient processing and use of the forecast data.

4. Conclusions

The present study presents the development of a novel approach to photovoltaic power forecasting, which has been built utilizing numerical weather prediction forecasts and physics-based models with the option of including data-driven models for a hybrid approach. More specifically, the presented algorithm includes: the retrieval and processing of forecast data from the IFS/ECMWF and CAMS models with the option of using ANN models for DNI forecast improvement; an improved transposition model that computes GTI and absorbed irradiance for first and inner-rows of photovoltaic rows considering inter-row shading and obscuring of direct, circumsolar and isotropic diffuse irradiance and masking of reflected irradiance; a comprehensive dynamic coupled thermal-electric photovoltaic model based on the energy conservation equation, namely the heat transfer to the environment through convection, radiation and the electric power output, taking into account wind speed and direction; losses model (optional) considering typical or user-provided derating factors and inverter model which allows for the computation of DC and AC power output of each string and inverter of the powerplant. This approach is highly versatile, as it can be applied to any fixed crystalline silicone photovoltaic system, without requiring on-site observations, thereby offering significant cost savings by eliminating the necessity for expensive monitoring equipment and infrastructure, as well as maintenance. Moreover, it has the ability to generate 72-h photovoltaic power forecasts, with user-defined temporal resolution.

The algorithm was validated for a temporal resolution of 15 min with approximately 1-year data from a real powerplant with seven photovoltaic strings (4 in front rows and 3 in inner rows) located in the region of Lisbon, Portugal. The overall results showed an MBE of 7.5 W/kWp and RMSE of 123.7 W/kWp for the DC power and an MBE of 9.3 W/kWp and RMSE of 121.0 W/kWp for the AC power output forecasts

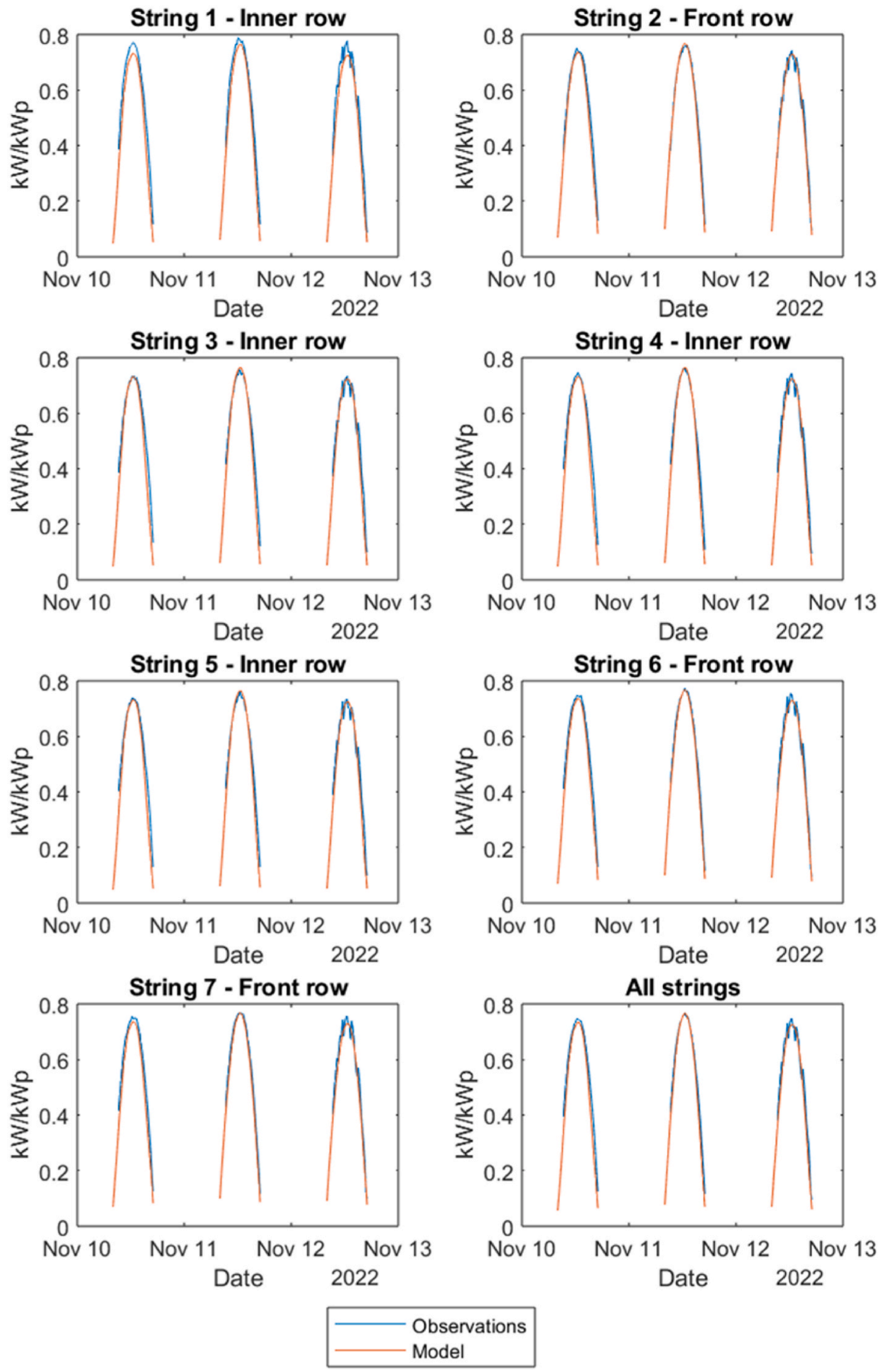

Fig. 10. Power output (DC) forecasts (orange) and measurements (blue) of each string for the operational example including the electric losses. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 10
Metrics of AC power output forecasts for each forecast days (number of data points: 42689).

Day	R^2	MBE (W/kWp)	rMBE (%)	MAE (W/kWp)	rMAE (%)	RMSE (W/kWp)	rRMSE (%)
0	0.832	8.7	2.1	76.5	18.9	114.2	28.2
1	0.814	8.4	2.1	80.1	19.8	120.4	29.8
2	0.790	10.8	2.7	84.2	20.9	128.3	31.8
All	0.812	9.3	2.3	80.3	19.9	121.0	29.9

considering all strings and the 72h forecast horizon. This translates to relative values of 1.8 %, 30.0 %, 2.3 % and 29.9 %, respectively.

Due to its inherent scalability, this algorithm can be effortlessly extended to cover a wider range of installations, without any logistical challenges. It can also be deployed in regions where on-site access is arduous, or in regions where historical data may not be available. Furthermore, it is essential to highlight that this algorithm possesses the adaptability to be configured differently, catering to the unique requirements of various installations providing continuous and real-time forecasting. The integration of this algorithm into grid operations

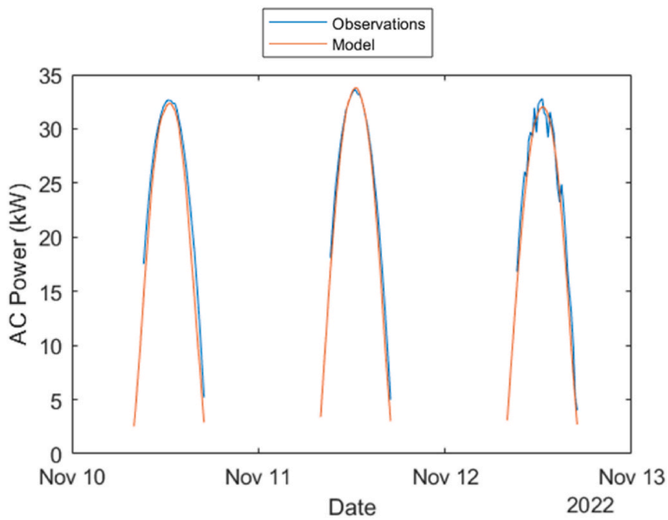

Fig. 11. AC power forecasts (orange) and measurements (blue) for the operational period taken as example. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 11
Running time of the different processes of the forecasting algorithm for the operational period taken as example.

Process	Running time (s)	Time stamp
Availability of forecasts	–	06:12 UTC
Data retrieval	385.201	06:18 UTC
Temporal and spatial downscaling	0.867	06:18 UTC
ANN models	3.401	06:18 UTC
Transposition model	13.153	06:18 UTC
Coupled thermal-electric model	363.988	06:24 UTC
Losses model	0.127	06:24 UTC
Inverter model	0.006	06:24 UTC
Generate output	0.142	06:24 UTC

would result in better management of photovoltaic energy generation variability.

CRediT authorship contribution statement

Sara Pereira: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Paulo Canhoto:** Writing – review & editing, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Takashi Oozeki:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Rui Salgado:** Writing – review & editing, Supervision, Resources, Project administration, Investigation, Funding acquisition, Formal analysis, Conceptualization.

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.

Acknowledgements

This work was funded by National funds through FCT - Fundação para a Ciência e Tecnologia, I.P. (projects UIDB/04683/2020 and UIDP/04683/2020). S. Pereira acknowledges the support of FCT (the Foundation for Science and Technology) through the grant with reference SFRH/BD/145378/2019. The authors are thankful to Afonso Cavaco, Josué Figueira and Samuel Ramos Bárias for the technical support. The authors also gratefully acknowledge Helexia for providing the photovoltaic plant data used in the validation of the models.

Appendix A. Computation of absorbed irradiance

The transposition model presented in Ref. [1] is used to determine the irradiance incident on the photovoltaic panel's surface, however for the computation of the power output of such systems the absorbed irradiance, S, is necessary. In Ref. [17] this is accomplished using only the available GTI and DNI data, yet with this algorithm the reflected and different components of diffuse irradiances are available and so S is obtained through Eq. (1) for each instant and for each segment of the panel being evaluated. Here, $\vec{\gamma}_b$ is the vector with the incidence angle modifiers for direct irradiance for each segment, $\vec{I}$ is the direct irradiance vector, $\vec{D}_{cs}$ is the diffuse circumsolar irradiance vector, $\vec{D}_{iso}$ is the isotropic diffuse irradiance vector, $\vec{\gamma}_d$ is the vector with the incidence angle modifiers for isotropic diffuse irradiance, $\boldsymbol{R}$ is the matrix computed through Eq. (2) which includes the matrixes with the albedo ($\boldsymbol{\rho}$), view-factors ($\boldsymbol{F}$), and incidence angle modifiers for reflected irradiance from all segments ($\boldsymbol{\gamma_r}$).

The incidence angle modifiers are computed through Eqs. (3)–(6) based on the principles of Snell's and Bougher's laws, as outlined in Ref. [60]. Here, θ_r is the angle of refraction in the glazing of the modules, θ is the incidence angle, n is the effective index of refraction of the cell cover assumed to be 1.526, a value close to the typical refractive index of glass, K_g is the glazing extinction coefficient with a value of 4 m^{-1} and L_g is the glazing thickness set at 2 mm, a dimension widely deemed suitable for most photovoltaic cell panels [69].

$$\vec{S} = \vec{\gamma}_b\left(\vec{I} + \vec{D}_{cs}\right) + \vec{\gamma}_d\vec{D}_{iso} + \boldsymbol{R}\overrightarrow{GTI} \tag{1}$$

$$\boldsymbol{R} = \boldsymbol{\rho}\boldsymbol{F}\boldsymbol{\gamma_r} \tag{2}$$

$$\gamma = \frac{\tau(\theta)}{\tau(0)} \tag{3}$$

$$\tau(\theta) = e^{-\left(\frac{K_g L_g}{cos(\theta_r)}\right)}\left[1 - \frac{1}{2}\left(\frac{sin^2(\theta_r - \theta)}{sin^2(\theta_r + \theta)} + \frac{tan^2(\theta_r - \theta)}{tan^2(\theta_r + \theta)}\right)\right] \tag{4}$$

$$\tau(0) = e^{-\left(K_g L_g\right)}\left[1 - \left(\frac{1-n}{1+n}\right)^2\right] \tag{5}$$

$$\theta_r = sin^{-1}\left[\frac{1}{n} sin\ (\theta)\right] \tag{6}$$

The set of incidence angles needed for each segment of the panel being evaluated includes the angles for direct irradiance θ_b, isotropic diffuse irradiance θ_d, and reflected irradiance $\vec{\theta}_{ref}$. The incidence angles of the direct component are straightforward and can be obtained through Eq. (7), where Lat is the latitude, δ the solar declination, β and φ the tilt angle and azimuth of the panel, respectively, and h_s the solar hour angle.

For the isotropic diffuse irradiance incidence angles, an equivalent angle needs to be tailored to the slope of the panel [60]. For first rows Eq. (8) can be used as presented in Ref. [60]. It is worth to mention that while these equivalent angles were initially derived for thermal collectors, they have proven to be a valid option for photovoltaic systems as well.

However, for rows other than the front row an adjustment needs to be made since the portion of sky in the field of view of a given segment varies depending on its relative position to the other surfaces. To overcome this aspect, an approximation was assumed by determining an equivalent tilt angle (Eq. (9)), in which the ground is assumed as being at the same level of the line connecting the middle of each segment and the top of the front row. Here, rL is the row length, D is the distance between rows and $\vec{c}$ is the position vector of the middle point of each segment of photovoltaic panel.

As for the incidence angles of the reflected irradiance, they are computed differently for the irradiance being reflected from the back of the front row $\boldsymbol{\theta_{ref,v}}$ and from the ground between rows $\boldsymbol{\theta_{ref,u}}$. This is done through Eqs. (10) and (11) considering the line connecting the middle of the two segments being evaluated and where h_0 is the height of the panels from the ground while $\vec{v}$ and $\vec{u}$ are the position vectors of the middle points of each segment of the back of the first row and the ground between rows, respectively.

Finally, with these angles, the different incidence angle modifiers can be computed and the absorbed solar irradiance S for each segment of photovoltaic panel is obtained. The mean absorbed irradiance of each panel is then obtained by averaging S over the respective segments.

$$\theta_b = \cos^{-1}\begin{pmatrix} \sin Lat \sin\delta\cos\beta - \cos Lat \sin\delta \sin\beta\cos\varphi + \cos Lat \cos\delta\cos h_s \cos\beta + \\ + \sin Lat \cos\delta\cos h_s \sin\beta\cos\varphi + \cos\delta\sin h_s \sin\beta\sin\varphi \end{pmatrix} \tag{7}$$

$$\theta_d = 59.7 - 0.1388\beta + 0.001497\beta^2 \tag{8}$$

$$\vec{\beta} = \tan^{-1}\left(\frac{rL \times \sin\beta - \vec{c}\sin\beta}{D - rL \times \cos\beta + \vec{c}\cos\beta}\right) \tag{9}$$

$$\boldsymbol{\theta_{ref,v}} = \left|90 - \beta - \tan^{-1}\left(\frac{(\vec{v} - \vec{c})\sin\beta}{D + (\vec{c} - \vec{v})\cos\beta}\right)\right| \tag{10}$$

$$\boldsymbol{\theta_{ref,u}} = 90 - \beta + \tan^{-1}\left(\frac{\vec{c}\sin\beta + h_0}{D + (\vec{c} - \vec{u})\cos\beta}\right) \tag{11}$$

Appendix B. Details on the thermal-electric coupled model

The thermal-electric model employed in this work integrates a dynamic thermal model based on the energy conservation equation with a single-diode five-parameter electrical model, allowing simultaneous computation of photovoltaic module temperature and electrical power output under varying conditions.

The dynamic thermal model relies on the fundamental energy balance, which can be described by:

$$C_{mod}\frac{dT_{mod}}{dt} = Q_{\text{sun}} - Q_{conv} - Q_{rad} - P_e \tag{12}$$

Here, C_{mod} is the equivalent heat capacity of the module, T_{mod} the module temperature, Q_{sun} the absorbed solar irradiance, Q_{conv} convective heat losses, Q_{rad} radiative heat losses, and P_e the electrical power output. Conduction losses are considered negligible.

Heat transfer processes include forced and natural convection, influenced by wind speed and direction, as well as radiative heat transfer to the environment, determined by temperature differences and module emissivity.

The electrical model implemented is the single-diode five-parameter equivalent circuit. It represents the current-voltage (I-V) characteristics of a PV module using five parameters: the photo-generated current, the diode reverse saturation current, the ideality factor of the diode, the series resistance, and the shunt resistance. These parameters are estimated from manufacturer datasheet values such as the short-circuit current, open-circuit voltage, current and voltage at maximum power point, and temperature coefficients. The model allows for accurate prediction of module performance under variable irradiance and temperature conditions by solving the implicit current-voltage relationship.

This coupling explicitly accounts for temperature's influence on PV module efficiency and power output, iteratively resolving the thermal-electrical interactions at user-defined time steps. Numerical integration of the dynamic thermal equation is carried out using the Dorman and Prince version of the Runge-Kutta method.

Additionally, the model considers the configuration of the photovoltaic plant, explicitly modeling the electrical connections between modules, strings, and arrays, as well as the maximum power point tracking (MPPT) systems. This ensures that the overall power output accurately reflects realistic operational conditions and electrical interactions.

Fig. 12 shows the flowchart of the coupled thermal-electric model, clearly illustrating the iterative calculation and interaction between input meteorological data, the thermal model, and the electrical model to yield accurate PV power output forecasts. For a more detailed explanation of this model please refer to Ref. [17].

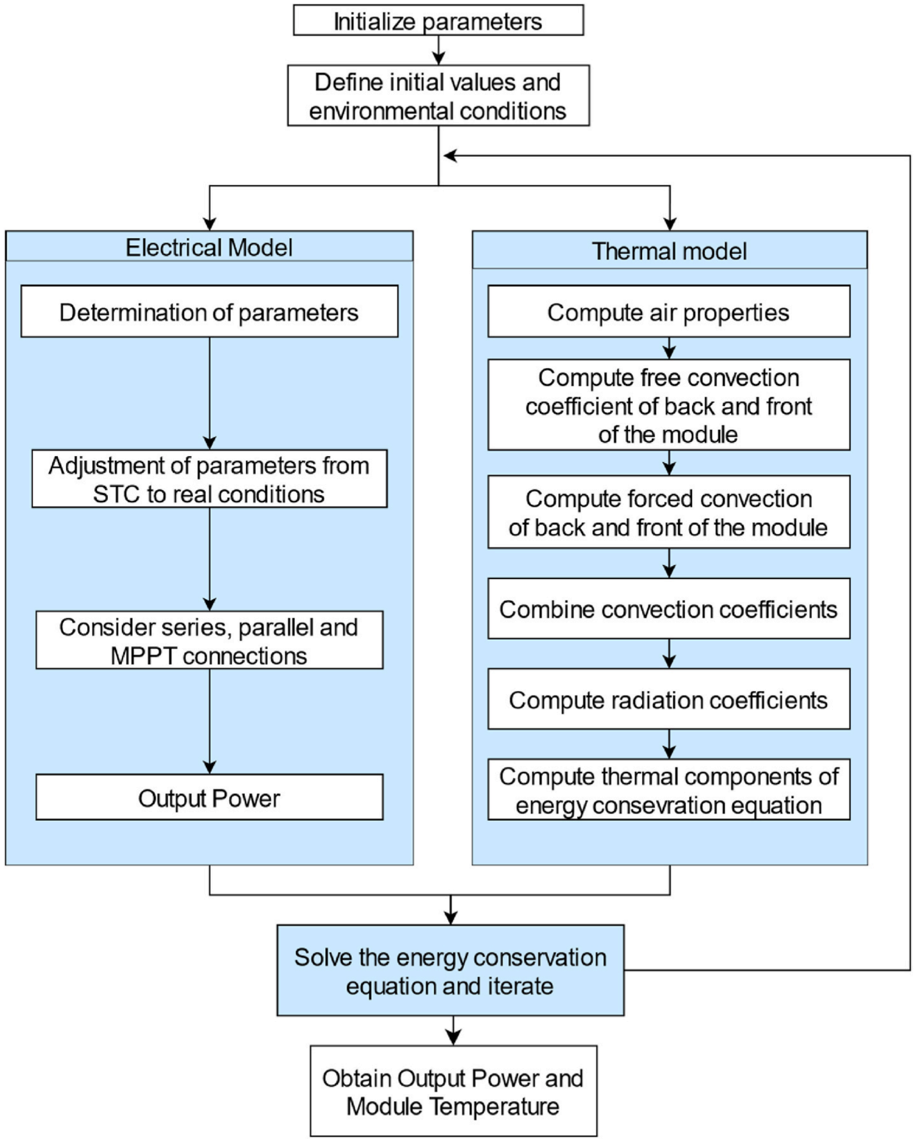

Fig. 12. Flowchart of the thermal and the electric models and its coupling.

References

[1] S. Pereira, P. Canhoto, R. Salgado, Prediction of global solar irradiance on parallel rows of tilted surfaces including the effect of direct and anisotropic diffuse shading, Energies 17 (2024) 3444, https://doi.org/10.3390/en17143444.

[2] O. El Alani, A. Ghennioui, H. Ghennioui, Y.-M. Saint-Drenan, P. Blanc, Evaluation of 24-Hours forecasts of global solar irradiation from IFS, GFS and McClear models, in: AIP Conf Proc, 2020, https://doi.org/10.1063/5.0032744.

[3] J. Perdigão, P. Canhoto, R. Salgado, M.J. Costa, Assessment of direct normal irradiance forecasts based on IFS/ECMWF data and observations in the south of Portugal, Forecasting 2 (2020) 130–150, https://doi.org/10.3390/forecast2020007.

[4] R. Perez, E. Lorenz, S. Pelland, M. Beauharnois, G. Van Knowe, K. Hemker, D. Heinemann, J. Remund, S.C. Müller, W. Traunmüller, G. Steinmauer, D. Pozo, J. A. Ruiz-Arias, V. Lara-Fanego, L. Ramirez-Santigosa, M. Gaston-Romero, L. M. Pomares, Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe, Sol. Energy 94 (2013) 305–326, https://doi.org/10.1016/j.solener.2013.05.005.

[5] M.J. Mayer, D. Yang, B. Szintai, Comparing global and regional downscaled NWP models for irradiance and photovoltaic power forecasting: ECMWF versus AROME, Appl. Energy 352 (2023) 121958, https://doi.org/10.1016/J.APENERGY.2023.121958.

[6] C.A. Gueymard, Direct and indirect uncertainties in the prediction of tilted irradiance for solar engineering applications, Sol. Energy 83 (2009) 432–444, https://doi.org/10.1016/j.solener.2008.11.004.

[7] C.A. Gueymard, Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 2: performance of separation and transposition models for global tilted irradiance, Sol. Energy 153 (2017) 766–779, https://doi.org/10.1016/j.solener.2017.04.068.

[8] D. Yang, C.A. Gueymard, Ensemble model output statistics for the separation of direct and diffuse components from 1-min global irradiance, Sol. Energy 208 (2020) 591–603, https://doi.org/10.1016/j.solener.2020.05.082.

[9] D. Yang, Solar radiation on inclined surfaces: corrections and benchmarks, Sol. Energy 136 (2016) 288–302, https://doi.org/10.1016/j.solener.2016.06.062.

[10] M.J. Mayer, G. Gróf, Extensive comparison of physical models for photovoltaic power forecasting, Appl. Energy 283 (2021) 116239, https://doi.org/10.1016/j.apenergy.2020.116239.

[11] Y.F. Nassar, N.M. Abuhamoud, G.M. Miskeen, H.J. El-Khozondar, S.Y. Alsadi, O. M. Ahwidi, Investigating the applicability of horizontal to tilted sky-diffuse solar irradiation transposition models for key Libyan cities. 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), 2022, pp. 9–14, https://doi.org/10.1109/MI-STA54861.2022.9837500.

[12] Y.F. Nassar, A.A. Hafez, S.Y. Alsadi, Multi-factorial comparison for 24 distinct transposition models for inclined surface solar irradiance computation in the State of Palestine: a case study, Front. Energy Res. 7 (2020) 1–19, https://doi.org/10.3389/fenrg.2019.00163.

[13] J. Appelbaum, Y. Massalha, A. Aronescu, Corrections to anisotropic diffuse radiation model, Sol. Energy 193 (2019) 523–528, https://doi.org/10.1016/j.solener.2019.09.090.

[14] D. Tschopp, A.R. Jensen, J. Dragsted, P. Ohnewein, S. Furbo, Measurement and modeling of diffuse irradiance masking on tilted planes for solar engineering applications, Sol. Energy 231 (2022) 365–378, https://doi.org/10.1016/J.SOLENER.2021.10.083.
[15] N. Varga, M.J. Mayer, Model-based analysis of shading losses in ground-mounted photovoltaic power plants, Sol. Energy 216 (2021) 428–438, https://doi.org/10.1016/J.SOLENER.2021.01.047.
[16] A. Gholami, M. Ameri, M. Zandi, R. Gavagsaz Ghoachani, Electrical, thermal and optical modeling of photovoltaic systems: step-by-step guide and comparative review study, Sustain. Energy Technol. Assessments 49 (2022) 101711, https://doi.org/10.1016/j.seta.2021.101711.
[17] S. Pereira, P. Canhoto, T. Oozeki, R. Salgado, Assessment of thermal modeling of photovoltaic panels for predicting power generation using only manufacturer data, Energy Rep. 12 (2024) 1431–1448, https://doi.org/10.1016/j.egyr.2024.07.039.
[18] R. Castro, Data-driven PV modules modelling: comparison between equivalent electric circuit and artificial intelligence based models, Sustain. Energy Technol. Assessments 30 (2018) 230–238, https://doi.org/10.1016/J.SETA.2018.10.011.
[19] R. Chenni, M. Makhlouf, T. Kerbache, A. Bouzid, A detailed modeling method for photovoltaic cells, Energy (Calg.) 32 (2007) 1724–1730, https://doi.org/10.1016/J.ENERGY.2006.12.006.
[20] V.J. Chin, Z. Salam, K. Ishaque, Cell modelling and model parameters estimation techniques for photovoltaic simulator application: a review, Appl. Energy 154 (2015) 500–519, https://doi.org/10.1016/j.apenergy.2015.05.035.
[21] 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, https://doi.org/10.1016/j.apenergy.2022.119249.
[22] B. Perovic, D. Klimenta, M. Jevtic, M. Milovanovic, A transient thermal model for flat-plate photovoltaic systems and its experimental validation, Elektronika Ir Elektrotechnika 25 (2019) 40–46, https://doi.org/10.5755/j01.eie.25.2.23203.
[23] S. Cossu, R. Baccoli, E. Ghiani, Utility Scale ground mounted photovoltaic plants with gable structure and inverter oversizing for land-use optimization, Energies 14 (2021) 3084, https://doi.org/10.3390/EN14113084, 14 (2021) 3084.
[24] D. González-Peña, I. García-Ruiz, M. Díez-Mediavilla, M.I. Dieste-Velasco, C. Alonso-Tristán, Photovoltaic prediction software: evaluation with real data from northern Spain, Appl. Sci. 11 (2021), https://doi.org/10.3390/app11115025.
[25] Natural resources Canada, RETScreen, (n.d.). https://natural-resources.canada.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465 (accessed January 21, 2025).
[26] National Renewable Energy Laboratory, C. Golden, System avisor Model^{TM}, (n.d.). https://https://sam.nrel.gov (accessed January 21, 2025).
[27] European Commission Joint Research Centre, Photovoltaic Geographical Information System, (n.d.). http://re.jrc.ec.europa.eu/pvgis (accessed January 21, 2025).
[28] PVsyst, PVsyst, (n.d.). https://www.pvsyst.com/(accessed January 21, 2025).
[29] Valentin software, PV*SOL, (n.d.). https://pvsol.software/en/(accessed January 21, 2025).
[30] P. Bruneau, D. Fiorelli, C. Braun, D. Koster, Hybrid-Physical probabilistic forecasting for a set of photovoltaic systems using recurrent neural networks, Int. J. Forecast. (2023) 1–22. http://arxiv.org/abs/2303.08459.
[31] D.T. Cotfas, M. Marzband, P.A. Cotfas, M. Siroux, D. Sera, Editorial: forecasting solar radiation, photovoltaic power and thermal energy production applications, Front. Energy Res. 10 (2022), https://doi.org/10.3389/fenrg.2022.1115096.
[32] S. Pereira, P. Canhoto, R. Salgado, Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data, Energy and AI 15 (2024) 100314, https://doi.org/10.1016/j.egyai.2023.100314.
[33] R. Ahmed, V. Sreeram, Y. Mishra, M.D. 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, https://doi.org/10.1016/J.RSER.2020.109792.
[34] H. Ohtake, T. Takamatsu, T. Oozeki, A review on photovoltaic power forecasting technics, IEEJ Transactions on Power and Energy 142 (2022) 533–541, https://doi.org/10.1541/ieejpes.142.533.
[35] J. Ramirez-Vergara, L.B. Bosman, E. Wollega, W.D. Leon-Salas, Review of forecasting methods to support photovoltaic predictive maintenance, Clean Eng. Technol. 8 (2022) 100460, https://doi.org/10.1016/j.clet.2022.100460.
[36] P. Mathiesen, J. Kleissl, Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States, Sol. Energy 85 (2011) 967–977, https://doi.org/10.1016/j.solener.2011.02.013.
[37] W. Castillo-Rojas, J. Bekios-Calfa, C. Hernández, Daily prediction model of photovoltaic power generation using a hybrid Architecture of recurrent neural networks and shallow neural networks, Int. J. Photoenergy 2023 (2023) 1–19, https://doi.org/10.1155/2023/2592405.
[38] S. Bofinger, G. Heilscher, Solar electricity forecast: approaches and first results. Proceedings of the 21st European Photovoltaic Solar Energy Conference and Exhibition, 2006.
[39] F.M. Lopes, R. Conceição, H.G. Silva, R. Salgado, M. Collares-Pereira, Improved ECMWF forecasts of direct normal irradiance: a tool for better operational strategies in concentrating solar power plants, Renew. Energy 163 (2021) 755–771, https://doi.org/10.1016/J.RENENE.2020.08.140.
[40] L.M. Aguiar, B. Pereira, P. Lauret, F. Díaz, M. David, Combining solar irradiance measurements, satellite-derived data and a numerical weather prediction model to improve intra-day solar forecasting, Renew. Energy 97 (2016) 599–610, https://doi.org/10.1016/j.renene.2016.06.018.
[41] J.G.D.S. Fonseca, F. Uno, H. Ohtake, T. Oozeki, K. Ogimoto, Enhancements in day-ahead forecasts of solar irradiation with machine learning: a novel analysis with the Japanese mesoscale model, J. Appl. Meteorol. Climatol. 59 (2020) 1011–1028, https://doi.org/10.1175/JAMC-D-19-0240.1.
[42] F.J.L.L. Lima, F.R. Martins, E.B. Pereira, E. Lorenz, D. Heinemann, Forecast for surface solar irradiance at the Brazilian Northeastern region using NWP model and artificial neural networks, Renew. Energy 87 (2016) 807–818, https://doi.org/10.1016/j.renene.2015.11.005.
[43] W. Liu, Y. Liu, X. Zhou, Y. Xie, Y. Han, S. Yoo, M. Sengupta, Use of physics to improve solar forecast: Physics-informed persistence models for simultaneously forecasting GHI, DNI, and DHI, Sol. Energy 215 (2021) 252–265, https://doi.org/10.1016/j.solener.2020.12.045.
[44] S. Pereira, E.F.M. Abreu, M. Iakunin, A. Cavaco, R. Salgado, P. Canhoto, Method for solar resource assessment using numerical weather prediction and artificial neural network models based on typical meteorological data: application to the south of Portugal, Sol. Energy 236 (2022) 225–238, https://doi.org/10.1016/j.solener.2022.03.003.
[45] 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, https://doi.org/10.1016/j.solener.2023.01.019.
[46] C. Voyant, G. Notton, J. Duchaud, L.A.G. Gutiérrez, J.M. Bright, D. Yang, Benchmarks for solar radiation time series forecasting, Renew. Energy 191 (2022) 747–762, https://doi.org/10.1016/j.renene.2022.04.065.
[47] C. Voyant, G. Notton, S. Kalogirou, M.L. Nivet, C. Paoli, F. Motte, A. Fouilloy, Machine learning methods for solar radiation forecasting: a review, Renew. Energy 105 (2017) 569–582, https://doi.org/10.1016/j.renene.2016.12.095.
[48] B. Yang, T. Zhu, P. Cao, Z. Guo, C. Zeng, D. Li, Y. Chen, H. Ye, R. Shao, H. Shu, T. Yu, Classification and summarization of solar irradiance and power forecasting methods: a thorough review, CSEE Journal of Power and Energy Systems (2021) 1–19, https://doi.org/10.17775/CSEEJPES.2020.04930.
[49] Y. Zhou, Y. Liu, D. Wang, X. Liu, Y. Wang, A review on global solar radiation prediction with machine learning models in a comprehensive perspective, Energy Convers. Manag. 235 (2021) 113960, https://doi.org/10.1016/j.enconman.2021.113960.
[50] A. hyun Jung, D.-H. Lee, J.-Y. Kim, C.K. Kim, H.-G. Kim, Y.-S. Lee, Spatial and temporal downscaling of solar radiation using statistical techniques, Journal of the Korean Solar Energy Society 40 (2020) 89–94, https://doi.org/10.7836/kses.2020.40.6.089.
[51] D. Yang, E. Wu, J. Kleissl, Operational solar forecasting for the real-time market, Int. J. Forecast. 35 (2019) 1499–1519, https://doi.org/10.1016/j.ijforecast.2019.03.009.
[52] M. Jaszczur, Q. Hassan, A.M. Abdulateef, J. Abdulateef, Assessing the temporal load resolution effect on the photovoltaic energy flows and self-consumption, Renew. Energy 169 (2021) 1077–1090, https://doi.org/10.1016/J.RENENE.2021.01.076.
[53] C. Yang, A.A. Thatte, L. Xie, Multitime-Scale data-driven spatio-temporal forecast of photovoltaic generation, IEEE Trans. Sustain. Energy 6 (2015) 104–112, https://doi.org/10.1109/TSTE.2014.2359974.
[54] MathWorks, MATLAB (R2024a) [Computer software], https://www.mathworks.com/products/matlab/student.html, 2024. (Accessed 21 January 2025).
[55] ECMWF, IFS documentation. https://www.ecmwf.int/en/publications/ifs-documentation, 2021. (Accessed 10 December 2021).
[56] M. Schroedter-Homscheidt, C. Hoyer-klick, N. Killius Armines, M. Lefèvre, L. W. Transvalor, Copernicus Atmosphere Monitoring Service User's Guide to the CAMS Radiation Service, 2017, pp. 1–74.
[57] R.J. Hogan, A. Bozzo, A flexible and efficient radiation scheme for the ECMWF model, J. Adv. Model. Earth Syst. 10 (2018) 1990–2008, https://doi.org/10.1029/2018MS001364.
[58] C.A.M.S. ECMWF, Global atmospheric composition forecast data documentation. https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation, 2021. (Accessed 10 December 2021).
[59] J. Hay, D. McKay, Final report IEA task IX-Calculation of solar irradiances for inclined surfaces: verification of models which use hourly and daily data. https://www.iea-shc.org/Data/Sites/1/publications/Task9-CalculationofSolarIrradiancesforInclinedSurfaces-VerificationofModelswhichUseHourlyandDailyData-June1988.pdf, 1988.
[60] J.A. Duffie, W.A. Beckman, Solar Engineering of Thermal Processes, 2013, pp. 1–885.
[61] J.J. Roberts, A.A. Mendiburu Zevallos, A.M. Cassula, Assessment of photovoltaic performance models for system simulation, Renew. Sustain. Energy Rev. 72 (2017) 1104–1123, https://doi.org/10.1016/j.rser.2016.10.022.
[62] SERAPHIM, SOLARFORCE ENERGY, SII series 390-405 W, (n.d.). https://solarforce-energy.co.za/wp-content/uploads/2021/03/SRP-390-405-BMA-HV_158.75_EN_2020V2.0.pdf (accessed August 26, 2024).
[63] Huawei, SUN2000-36KTL smart string inverter, (n.d.). https://solar.huawei.com/en/download?p=%2F-%2Fmedia%2FSolar%2Fattachment%2Fpdf%2Fmea%2Fdatasheet%2FSUN2000-36KTL.pdf (accessed August 26, 2024).
[64] Meteocontrol, IRRADIANCE SENSORS, (n.d.). https://www.meteocontrol.com/fileadmin/Daten/Dokumente/EN/1_Photovoltaik_Monitoring/2_Zubehoer/1_Sensorik/Irradiance/Silicon_solar_irradiance_sensors/DB_Irradiance_sensor_SI-RS485TC-T_-2T-Tm_-MB_en.pdf (accessed August 26, 2024).
[65] C.N. Long, E.G. Dutton, BSRN Global Network recommended QC tests, V2.0. https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf, 2010. (Accessed 28 May 2024).
[66] E.F.M. Abreu, P. Canhoto, V. Prior, R. Melicio, Solar resource assessment through long-term statistical analysis and typical data generation with different time

resolutions using GHI measurements, Renew. Energy 127 (2018) 398–411, https://doi.org/10.1016/J.RENENE.2018.04.068.

[67] A. Cavaco, P. Canhoto, M. Collares Pereira, Procedures for solar radiation data gathering and processing and their application to DNI assessment in southern Portugal, Renew. Energy 163 (2021) 2208–2219, https://doi.org/10.1016/J.RENENE.2020.10.075.

[68] 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, https://doi.org/10.1016/j.rser.2023.113171.

[69] W. De Soto, S.A. Klein, W.A. Beckman, Improvement and validation of a model for photovoltaic array performance, Sol. Energy 80 (2006) 78–88, https://doi.org/10.1016/j.solener.2005.06.010.