Universidade de Évora - Instituto de Investigação e Formação Avançada

Programa de Doutoramento em Engenharia Mecatrónica e Energia

Área de especialização | Energia

Tese de Doutoramento

Solar radiation and photovoltaic power forecasting using numerical weather prediction models

Sara Raquel Alves Pereira

Orientador(es) | Paulo Canhoto
Rui Paulo Salgado
Takashi Oozeki

Évora 2025

Universidade de Évora - Instituto de Investigação e Formação Avançada

Programa de Doutoramento em Engenharia Mecatrónica e Energia

Área de especialização | Energia

Tese de Doutoramento

Solar radiation and photovoltaic power forecasting using numerical weather prediction models

Sara Raquel Alves Pereira

Orientador(es) | Paulo Canhoto
Rui Paulo Salgado
Takashi Oozeki

Évora 2025

A tese de doutoramento foi objeto de apreciação e discussão pública pelo seguinte júri nomeado pelo Diretor do Instituto de Investigação e Formação Avançada:

Presidente	João Manuel Figueiredo (Universidade de Évora)
Vogais	Cláudio Domingos Martins Monteiro (Universidade do Porto - Faculdade de Engenharia) Diana Pereira Neves (Universidade de Lisboa - Instituto Superior Técnico) Luís Filipe Moreira Mendes (Universidade de Lisboa - Instituto Superior Técnico) Mouhaydine Tlemcani (Universidade de Évora) Paulo Canhoto (Universidade de Évora) (Orientador) Pedro Miguel Matos Soares (Universidade de Lisboa - Faculdade de Ciências)

Évora 2025

To the memory of my great grandmother

Maria José Jesus who taught me how to be resilient

Summary

This work presents a comprehensive approach for photovoltaic (PV) power forecasting, integrating various models and data sources to improve predictions accuracy and reliability. It combines numerical weather prediction (NWP) models with artificial neural networks (ANNs) aiming to assess and predict solar resource and PV power generation at different spatial and temporal scales. First, this approach was used to improve NWP solar irradiance data for a typical meteorological year, reducing errors of global horizontal (GHI) and direct normal irradiance (DNI) estimates to 2.3 % and 3.4 %, respectively, compared to data from eight stations in southern Portugal. Weather and aerosol data are also combined into an ANN-based model for DNI forecasting, reducing root mean squared error (RMSE) by 24.2 W/m² compared to original NWP forecasts, using data from six stations over four years as reference. An improved solar irradiance transposition model is also proposed that accounts for shading effects from adjacent PV rows, resulting in an improvement of 224.4 W/m² in RMSE for global tilted irradiance using data from an experimental setup. Thermal and electrical models for predicting PV module temperature and power output were assessed against three years of data from four systems, and the best performing empirical models were identified. A coupled thermal-electric model, using the single diode 5-parameter electric model, achieves RMSE values of 3.6 °C for temperature and 84.2 W for power output. Finally, an integrated algorithm combining these models was set up to provide accurate PV power forecasts across various systems and locations. Validation with real power plant data demonstrates its accuracy and robustness, achieving a RMSE of 121.0 W/kWp for AC power output forecasts, proving its value in managing solar energy systems.

Resumo

Previsão de radiação solar e produção de energia em sistemas fotovoltaicos com base em modelos numéricos de previsão do tempo

Este trabalho apresenta um método para previsão de potência fotovoltaica (PV), integrando vários modelos e fontes de dados para aumentar a precisão e fiabilidade das previsões. Combina modelos de previsão numérica do tempo (NWP) com redes neuronais artificiais (ANNs) com o objetivo de avaliar e prever o recurso solar e a geração de energia PV em diferentes escalas espaciais e temporais. Esta abordagem foi inicialmente usada para melhorar as previsões de modelos NWP para um ano meteorológico típico, reduzindo os erros da irradiância horizontal global (GHI) e normal direta (DNI) para 2,3% e 3,4%, respetivamente, em comparação com os dados de oito estações no sul de Portugal. Foram incorporados dados meteorológicos e de aerossóis num modelo baseado em ANNs para previsão de DNI, reduzindo a raiz do erro quadrático médio (RMSE) em 24,2W/m^2 em comparação com previsões NWP, usando como referência dados de quatro anos de seis estações no sul de Portugal. Foi melhorado um modelo de transposição da irradiância solar, considerando os efeitos do sombreamento de filas de módulos PV adjacentes, resultando numa melhoria de 224,4W/m^2 no RMSE da irradiância no plano inclinado, utilizando dados experimentais. Foram avaliados modelos térmicos e elétricos para previsão da temperatura dos módulos PV e da potência gerada com dados de três anos e quatro sistemas, identificando-se os modelos empíricos com melhor desempenho. Um modelo acoplado térmico-elétrico, baseado num modelo elétrico de um díodo e 5 parâmetros, apresentou um RMSE de 3,6°C para a temperatura e 84,2W para a potência. Finalmente, foi construído um algoritmo integrado combinando estes modelos para fornecer previsões de potência PV para vários sistemas e localizações. A validação com dados reais de uma central fotovoltaica demonstrou a sua precisão e robustez, apresentando um RMSE de 121,0W/kWp para previsões de potência AC, provando a sua utilidade na gestão de sistemas solares.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Paulo Canhoto. His unwavering patience, constant motivation, and dedication have been invaluable throughout this journey. His insightful ideas, constructive comments, and thoughtful suggestions have played a crucial role in shaping the research and findings presented in this thesis.

I extend my sincere thanks to my co-supervisor, Professor Rui Salgado, for his insightful comments, invaluable suggestions and encouragement, which have greatly contributed to the development and refinement of this research.

I am also deeply grateful to my co-supervisor, Takashi Oozeki, whose support was pivotal during my international stay. His readiness to assist and provide guidance at every step has been immensely helpful.

I gratefully acknowledge the support of the Portuguese National Science Foundation (FCT - Fundação para a Ciência e a Tecnologia) through Grant No. SFRH/BD/145378/2019. Additionally, I am deeply thankful for the backing provided by the Institute of Earth Sciences (ICT), funded by national funds through FCT (projects UIDB/04683/2020 and UIDP/04683/2020). I would also like to extend my deepest gratitude to the PV System and Application Team, as well as the Fukushima Renewable Energy Institute (FREA, AIST), for their invaluable support and contributions to this research.

I extend my heartfelt thanks to André Albino, Afonso Cavaco, Edgar Abreu, Josué Figueira, Maksim Iakunin, and Samuel Ramos Bárias for their invaluable support in obtaining experimental data. Additionally, I would like to offer special thanks to my colleagues at ICT for their companionship and support throughout these years.

Lastly, my deepest appreciation goes to my family, especially my mother, Carmen Pereira, for their unwavering encouragement throughout this journey. Their constant belief in my abilities and their emotional and moral support have been the foundation of my success. Profound gratitude for their patience, understanding, and sacrifices, which have allowed me to dedicate myself fully to this work.

Contents

List of Publications

This thesis includes the following papers:

I. Pereira, S.; Abreu, E.F.M.; Iakunin, M.; Cavaco, A.; Salgado, R.; Canhoto, P. 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. *Solar Energy* **2022**, *236*, 225–238, doi:10.1016/j.solener.2022.03.003. [Chapter 2]

II. Pereira, S.; Canhoto, P.; Salgado, R. Development and Assessment of Artificial Neural Network Models for Direct Normal Solar Irradiance Forecasting Using Operational Numerical Weather Prediction Data. *Energy and AI* **2024**, *15*, 7000, doi:10.1016/j.egyai.2023.100314. [Chapter 3]

III. Pereira, S.; Canhoto, P.; Salgado, R. Prediction of Global Solar Irradiance on Parallel Rows of Tilted Surfaces Including the Effect of Direct and Anisotropic Diffuse Shading. *Energies* **2024**, *17*, 3444, doi:10.3390/en17143444. [Chapter 4]

IV. Pereira, S.; Canhoto, P.; Oozeki, T.; Salgado, R. Assessment of Thermal Modeling of Photovoltaic Panels for Predicting Power Generation Using Only Manufacturer Data. *Energy Reports* **2024**, *12*, 1431–1448, doi:10.1016/j.egyr.2024.07.039. [Chapter 5]

V. Pereira, S.; Canhoto, P.; Oozeki, T.; Salgado, R. Comprehensive approach to photovoltaic power forecasting using numerical weather prediction data and physics-based models and data-driven techniques. *Renewable Energy* **2025** (under review). [Chapter 6]

Other publications related with the subject of the present thesis, but not included, are:

i. Pereira, S.; Canhoto, P.; Salgado, R.; Costa, M.J. Development of an ANN Based Corrective Algorithm of the Operational ECMWF Global Horizontal Irradiation Forecasts. *Solar Energy* **2019**, *185*, 387–405, doi:10.1016/j.solener.2019.04.070.

ii. Pereira, S.; Abreu, E.; Iakunin, M.; Canhoto, P.; Salgado, R. Prediction of Solar Resource and Photovoltaic Energy Production through the Generation of a Typical Meteorological Year and Meso-NH Simulations: Application to the South of Portugal. 2019 IEEE 2nd International Conference on Renewable Energy and Power Engineering, REPE, Toronto, Canada, 2 – 4 November 2019, 182–186, doi:10.1109/REPE48501.2019.9025118.

iii. Pereira, S.; Abreu, E.; Iakunin, M.; Canhoto, P.; Salgado, R. Estimativa Do Potencial Fotovoltaico No Sul De Portugal Através De Simulações Do Modelo Meso-NH Para Um Ano Meteorológico Típico. III Congresso Luso-Extremadurense De Ciências E Tecnologia, Évora, Portugal, 25 – 26 November 2019, doi.org/10.13140/RG.2.2.19961.83045

iv. Iakunin, M.; Abreu, E.F.M.; Canhoto, P.; Pereira, S.; Salgado, R. Impact of a Large Artificial Lake on Regional Climate: A Typical Meteorological Year Meso-NH Simulation Results. *International Journal of Climatology* **2022**, *42*, 1231–1252, doi:10.1002/joc.7299.

v. Pereira, S.; Abreu, E.; Iakunin, M.; Canhoto, P.; Salgado, R. Improved Method for Solar Resource Assessment Using Simulations from the

Numerical Weather Prediction Model Meso-NH and Artificial Neural Networks. PVSEC-33; Nagoya, 13 – 17 November 2022, doi:10.13140/RG.2.2.19508.87684.

vi. Pereira, S.; Canhoto, P.; Salgado, R.; Oozeki, T. Development and Validation of Coupled Thermal-Electric Transient Model of a Photovoltaic System. EU PVSEC; Lisbon, 18 – 22 September 2023, doi:10.13140/RG.2.2.35512.93440.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Preliminary remarks

In recent years, the use of renewable energy sources has seen a significant increase, driven by climate and socioeconomic changes. At the same time, electricity is increasingly being used as a key energy vector across various sectors, enhancing the integration of renewable resources into the energy mix. These resources are inherently variable, making the continuous monitoring of electrical energy production by companies and electric grid operators crucial for ensuring the efficient performance of energy conversion and distribution systems. Reliable assessments and accurate forecasts of these renewable energy sources, along with the energy generation from conversion systems, enable operators to make timely adjustments to equipment and informed decisions that enhance overall efficiency.

Solar energy, in particular, has more recently been the main focus of investment promoters in this area, with more than 5% of the world's electricity generation covered by photovoltaic energy in 2023 [1] and its continued growth expected in the coming years. As photovoltaic (PV) energy integration into the grid increases, the importance of accurately forecasting solar radiation and PV power output becomes paramount. Solar PV generation depends on a variety of atmospheric factors, including solar radiation, air temperature and humidity, cloud cover, and wind speed all of which can fluctuate over time and significantly impact energy output. Given these complexities, accurately modeling and forecasting these processes is essential for optimizing PV performance and ensuring reliable integration into the grid.

Weather and climate involve complex systems with many interacting variables. Traditional dynamic physical models, while useful, are often computationally expensive and require approximations. In contrast, statistical models, particularly those using machine learning techniques, can analyze vast amounts of data and capture the intricate interactions among variables, leading to accurate predictions. Thus, in this thesis, the goal is to answer the question: How can the accuracy of solar resource assessment and short-term photovoltaic power forecasting be improved across different locations and technologies, by evaluating the performance of existing data sources and developing adaptable, physically-informed correction models that do not require on-site measurements?

1.2 Solar radiation and solar energy resource

Solar radiation consists of the energy emitted by the Sun within the ultraviolet, visible and infrared regions of the electromagnetic spectrum, ranging from 0.15 to 0.3 μm [2]. The amount of solar radiation incident at the top of Earth's atmosphere fluctuates with the varying distance between the Earth and the Sun and the activity of the Sun. The solar radiation value at the Earth's surface varies even more significantly due to complex phenomena in the atmosphere and the interaction with the surface, including scattering, reflection back into space, reflection on the surface and absorption.

An important distinction in solar energy research is between solar irradiance and solar irradiation. Solar irradiance refers to the radiant power incident on a surface per unit area (W/m^2), while solar irradiation denotes the radiant energy incident on a surface per unit area (J/m^2 or Wh/m^2).

According to [2], solar irradiance received on a horizontal plane on the ground, whether total or spectral, is called global horizontal radiation. This is the sum of the direct and diffuse components. The direct component can be defined as the energy per unit time and area received on the plane within a solid angle under which the sun is centered. Its exact definition may vary based on the field of application or the measuring instruments used, which may have different aperture angles. The diffuse component is the sum of all the energy per unit time and area received by the plane coming from the sky due to scattering process in the atmosphere. If the plane is tilted, it also receives part of the radiation reflected from the surrounding ground, which is called the reflected component. Another important variable is the direct normal irradiance (DNI) which is the amount of solar radiation coming from a small solid angle centered at the sun's disk received per unit area by a surface that is perpendicular to the sun's rays and is particularly important for concentrating solar power plants.

Measuring and physically modeling these variables are essential for understanding and studying solar radiation and solar energy applications.

1.2.1 Solar radiation measurement and resource assessment

Solar radiation data are critical for various applications, requiring different types of equipment to measure the instantaneous and long-term integrated values of direct, diffuse, and global radiation on a surface. The process of radiation measurement is known as radiometry, wherein radiometric detectors convert irradiance into measurable signals through thermal processes (thermopile) or the photovoltaic effect [3].

To measure the DNI component of solar radiation, a pyrheliometer is typically used. This instrument tracks the sun's position in the sky and collects all light within the solid angle subtended by the solar disk, which has approximately 0.5° total solid angle, while blocking photons from the rest of the hemisphere [4]. Pyrheliometers have an aperture angle greater than the limit angle of the solar disk, contributing circumsolar radiation to direct normal irradiance measurements.

A pyranometer, on the other hand, measures global radiation on a surface. When shaded from direct radiation by a shading sphere or disk that tracks the sun, it measures diffuse radiation. Positioned on a horizontal surface, it measures global horizontal irradiance (GHI) or, if shaded, diffuse horizontal irradiance (DIF). When placed on a tilted surface, the data observed are termed global tilted irradiance (GTI), which includes also the reflected component.

Solar resource assessment involves characterizing the solar irradiance and/or irradiation available for energy conversion in a region or specific location over a given period. Reliable resource assessment requires calibrated measuring instruments and a prolonged period of quality-controlled observations. The Baseline Surface Radiation Network (BSRN) is an international project initiated by the World Meteorological Organization (WMO) to provide high-quality ground-based radiation measurements, following WMO guidelines [5,6]. However, the distribution of high-quality measurement stations worldwide is uneven and uncoordinated, primarily due to the costs associated with procurement, installation, operations, maintenance, data collection, quality assessment, archiving, and the impact of political climates [7].

If observations of solar radiation and other relevant meteorological variables for solar energy systems, such as air temperature and humidity, and wind speed, are available for a specific location, a typical meteorological year (TMY) can be constructed [8]. The recommended period for generating a TMY generally ranges from 10 to 25 years. A 10-year period is commonly used and considered sufficient for many applications [9,10], while a 25-year period can offer enhanced accuracy for long-term weather representation [11]. This TMY was developed for energy efficiency analysis in buildings and consists of twelve calendar months selected from specific years of a long time series, concatenated to form a complete year, and serves as a representative dataset of long-term typical weather data for that location, enabling the estimation of the typical solar radiation resource.

When planning and designing solar energy systems, extended periods of observations for the specific site of interest are often unavailable, or the site might initially be unknown. In such cases, solar resource maps provide a valuable general overview of solar radiation availability. Large-scale solar resource maps are typically based on numerous climatological databases and geographical interpolation models. Regional solar resource maps can be derived from TMY data obtained from observations at representative locations of the regional climate [12]. Ideally, these maps would be built using observations from an extensive network throughout the region over long periods. However, due to the expense and difficulty of maintaining such a network, alternative approaches, such as satellite data [13–15], numerical weather prediction models including reanalysis data [16] and local adaptation models, are often employed for developing regional solar resource maps.

In addition to solar resource mapping, the accurate measurement of aerosols is a critical factor in determining solar radiation levels, as aerosols can significantly affect the amount of solar energy reaching the Earth's surface. The AERONET (Aerosol Robotic Network) [17] provides valuable data in this regard, offering a comprehensive, long-term, and globally distributed network of ground-based observations. By measuring aerosol optical properties, AERONET data helps improve the accuracy of solar radiation models and forecasts, particularly by refining estimates of DNI and other solar components.

1.2.2 Solar irradiance on tilted surfaces

Solar radiation is typically measured on a horizontal plane or normal to the sun’s rays in the case of DNI. For the design, simulation and performance assessment of different solar energy applications, such as the design and energy analysis of buildings or photovoltaic power output estimation, for example, knowing the solar irradiance on the respective surfaces is critical [18]. In these cases, a transposition model is used to estimate the irradiance on a tilted surface by taking measured and/or modelled values of the solar radiation components on a horizontal surface and knowing the geometric parameters. This procedure includes the computation of the direct, diffuse and reflected radiation incident on the surface.

There are several models available in the literature, and most of those models agree on the computation of the direct and ground reflected irradiance while the modeling of diffuse radiation remains challenging [19]. On the other hand, the distribution of the diffuse component over the sky dome is not isotropic and depends on cloud cover and the scattering phase function of atmospheric particles, which results in phenomena such as circumsolar radiation and increased brightness near the horizon, both of which are highly variable. The isotropic sky transposition model, developed by Liu and Jordan [20], was the first to be published and assumes that all diffuse radiation is uniformly distributed over the sky dome and that ground reflection is diffuse. Since then, many models – both physical and empirical – have been developed, considering the anisotropic nature of sky irradiance and increasing in complexity [21].

Empirical transposition models are based on observational data, which can generate model biases towards the specific location depending on climate variables and/or orientation of the surface, depending on the experimental setup. Anisotropic physical models typically compute the diffuse component differently for two to three regions of the sky dome: i) the isotropic region, which consists of the background area of the sky dome remaining after excluding the other regions considered, from which radiation is received uniformly; ii) the circumsolar diffuse region, where forward scattering of solar radiation dominate and is defined as a circular corona area surrounding the sun disk; iii) the horizon brightening region, commonly defined as a band along the horizon and most evident under clear skies [22]. Global tilted irradiance can also be computed using ray-tracing software, which simulates the propagation of light rays to determine how irradiance is distributed on surfaces [23,24]. The Monte Carlo ray tracing method, in particular, is widely used for global illumination calculations [25].

The performance of transposition diffuse models tends to be better under clear-sky conditions when the sky diffuse irradiance is lower compared to the higher direct component of irradiance and its short-term variation over time is minimal. Clear-sky conditions are also characterized by a more homogeneous radiance over the sky dome compared to other sky conditions. Although developed transposition models consider the anisotropic aspect of diffuse radiance, they still make assumptions that induce errors, especially under cloudy conditions. The complexity of the scattering process, the rapidly varying shape and position of clouds, and their uneven distribution across the sky make an accurate short-term evaluation of the diffuse irradiance on a tilted surface exceedingly difficult [26].

Isotropic and anisotropic models were developed for unobstructed tilted surfaces, however, in many cases, the sky dome can be obscured, or surfaces can be shadowed by surrounding objects. Depending on the application of the transposition model, this aspect can be critical. For example, in photovoltaic power plants, the front rows of modules typically obscure the sky dome and surrounding ground surface and sometimes even shadow the modules behind them. Given that the largest part of a photovoltaic powerplant consists of rows of tilted panels that are not the first, considering this aspect when designing and modeling is extremely important.

Similarly, in small to medium-scale photovoltaic systems, including those integrated into buildings (Building-Integrated Photovoltaics or BIPV), and in agrivoltaic systems where solar panels are installed on agricultural land, shading issues are also significant. In BIPV applications, panels are often installed on building facades or rooftops, where nearby structures and other architectural features can cast shadows. In agrivoltaic systems, the interaction between solar panels and crops or livestock can affect shading patterns. In these contexts, accurately modeling and accounting for shading and the transposition of solar radiation are crucial for optimizing energy production.

1.3 Solar radiation forecasting

A forecast is a statement about what is expected to happen in the future based on the information currently available. Solar radiation forecasts estimate these variables using present data and various forecasting techniques. The approaches for computing these forecasts can be categorized into image-based models, numerical weather prediction models, statistical models, and hybrid models. These forecasting methods can also be classified based on the time scale of the predictions: short-term, mid-term, and long-term. Short-term forecasts, typically ranging from minutes to a few hours ahead, often rely on image-based models, such as satellite or sky camera data, and are crucial for real-time grid management. Mid-term forecasts, which cover several hours to a few days, usually involve numerical weather prediction models that simulate atmospheric conditions. Long-term forecasts, extending from weeks to months, often involve a combination of NWP and non-linear statistical models. Hybrid models, which integrate both NWP and statistical techniques, have been shown to provide more accurate predictions for time horizons of 2-3 weeks, as they leverage the strengths of both approaches [27]. These models are essential for planning and decision-making in energy management and infrastructure development.

1.3.1 Image based models

Image-based models estimate solar radiation by analyzing satellite and/or ground-based sky imagery obtained from a total sky imager (TSI). These models typically identify cloud structure and dynamic through the observations and extrapolate their motion vector fields to forecast cloud movement short-term and, consequently, solar radiation [28]. Satellite-based models tend to perform best for short temporal horizons (up to 6 hours) but exhibit greater errors under low irradiance conditions, such as when the sun is low in the sky (high zenith angles) and in locations with high spatial variability [29]. TSI-based models offer higher spatial and temporal resolutions, but the forecast horizon viability depends on the monitored spatial extent and the velocity of the clouds, usually ranging from 5 to 25 minutes [30]. More detailed information on this topic can be found in [28].

1.3.2 Numerical Weather Prediction models

Numerical Weather Prediction (NWP) models solve the physical constitutive and state equations that describe atmospheric processes obtaining estimates of the evolution of atmospheric conditions over time [31]. For both the atmosphere and ocean, motion is described by fluid dynamic equations applied to air and water, derived from fundamental principles like the conservation of mass, linear momentum and energy and constitutive and thermodynamic state laws. Ideally, these equations would perfectly describe the Earth's dynamics, but their strongly non-linear nature impose approximate numerical solutions. To achieve this, space is discretized into grid points and time into steps.

NWP models can be global, encompassing the entire Earth's atmosphere, or regional, requiring lateral boundary conditions in addition to surface information. While using a grid facilitates the numerical solving of equations of motion and energy, it inherently limits the model's ability to resolve features smaller than the grid spacing. Thus, finding the optimal discretization of these equations on a given grid spacing is crucial, balancing accuracy and computational effort. Efficient numerical methods enable using finer grids and higher model resolution, but they can also induce numerical errors, such as discretization errors and round-off errors, that may lead to instabilities or inaccurate forecasts. These errors become particularly significant when attempting to resolve small-scale features of the Earth system, such as individual clouds, which are often too small to be accurately represented on the commonly used grid spacings.

To represent those features, sub-grid-scale parametrization schemes can be employed using information on the main grid nodes to approximate sub-grid features. This process is important but can always only be a rough approximation of the actual physical process. In NWP modeling, parameterization schemes are needed in order to represent small-scale dynamics of clouds, convection, land, water, urban and sea surfaces, and small turbulent eddies [32].

The grids which are commonly used in state-of-the-art global NWP models have hundreds of millions of 3D grid-points. To perform simulations with these models a supercomputer is needed and forecast centers such as the European Centre for Medium-range Weather Forecasts (ECMWF) host supercomputers which are among the fastest computers in the world [33].

These models are also dependent on the quality of the information used for the initial state [34]. This information comes from ground based and/or remote observations of three-dimensional fields of pressure, wind, temperature and water species contents. The initial states are called meteorological analyses and are obtained through processes known as data assimilation. The principle of data assimilation is to get the result based on the prediction, compare the predicted result with real time data at the specified time, and reflect the uncertainty back to the next time step and make it more accurate. Data assimilation combines observations with forecasting models in an iterative process in which the model results are corrected on the basis of observations at subsequent instants. The accuracy of the NWP forecasts relies on the quality of data assimilation. After several years of continuous research, the data assimilation process attained a mature stage in currently operational NWP [27].

Additionally, the prediction of aerosols has become an integral part of NWP models, as aerosols can significantly influence weather and climate by affecting radiation balance and cloud formation. The Copernicus Atmosphere Monitoring Service (CAMS) [35] provides forecast data on aerosol concentrations, which can be used to improve the accuracy of weather forecasts by accounting for the impact of aerosols on atmospheric processes.

Many weather forecasts today are based on ensemble simulations from numerical weather prediction models [31]. Ensemble forecasts consist of typically around 10 – 50 individual simulation runs of a numerical weather prediction model with perturbed initial conditions and alterations in the model parameterizations, resulting in a collection of possible scenarios for the future developments of the weather, generating probabilistic forecasts, from which several key products can be derived such as probabilistic forecasts, most probable forecast scenario, prediction intervals and extreme event forecasts [31].

1.3.3 Statistical and machine learning models

Statistical and machine learning forecast models are based on historical data. Statistical linear or time-series methods derive relations between the variables used as input and the variable to be predicted through statistical analysis. Examples of statistical models are the persistence model or naïve predictor commonly used as a reference model [36], the Autoregressive Moving Average (ARMA) and the Autoregressive Integrated Moving average

(ARIMA) [37].The persistence model assumes that the next value of a time series is calculated under the assumption that nothing changes between the current time and the forecast step time and thus the model's accuracy decreases greatly with forecast duration. Several studies with respect to direct time series modeling have been performed [38,39].

Machine learning models are non-linear models [40]. Today, machine learning and artificial intelligence are often used interchangeably, however, machine learning is a set of algorithms that improve their performance on a set task through experience. Often this is achieved by a combination of statistical methods and numerical optimization to incrementally improve the machine learning algorithm and gain insight into the task that generalizes to future variations of that same task [40].

The most known type of machine learning is supervised learning [41]. In supervised learning, a dataset and labels or output data for this dataset is used and then a training process is carried out to map the samples in the dataset to the according labels or output data. Tasks within this type of machine learning are classification tasks, such as detecting clouds in satellite images, or regression tasks, including predicting the temperature from observations at weather stations, for example.

Machine learning modeling is different from classical numerical models, where the data and the rules are known, and the answers are obtained from the model. In supervised machine learning the data and answers are provided to derive the rules (albeit those rules are often implicitly stated).

Artificial neural networks (ANNs) are machine learning algorithms inspired by the structure and function of the human brain, specifically the neurons in the brain. An artificial neural network is made up of layers of interconnected neurons that perform simple calculations on their input and pass the result to the next layer. Typical inputs include historical solar radiation data, along with key meteorological variables like temperature, humidity, wind speed, and cloud cover. Temporal factors such as time of day, day of the year, and geographical location are also included, as they account for natural variations in solar radiation [42]. The input data is fed through the network, layer by

layer, until it reaches the output layer, where the final decision or prediction is made. During the training process, the network is presented with a large set of input-output pairs of experiments or observations and makes predictions based on these training examples. Whenever the predictions are incorrect, the network adjusts its internal weights and biases to better fit the data. This process is repeated over and over, until the network can accurately predict the best output for a given input.
A more in-depth review of the different machine learning models including their advantages and disadvantages for solar radiation forecasting can be seen in [43].

1.3.4 Hybrid models

Hybrid forecasting models combine two or more models with different forecasting approaches typically outperforming individual models [40]. Some examples are the use of statistical models for post-processing outputs from numerical weather prediction models or the use of machine learning techniques for estimating irradiance from satellite data.
Model Output Statistics (MOS) is applied to outputs of NWP models to derive regression functions for localized bias reduction for future forecasts based on historical data of forecasts and ground-based irradiance measurements. This can also be achieved by using artificial neural networks or other types of machine learning models for a non-linear approach to bias correction and it can be applied not only to the output of NWP models but also to other physical models, namely image based models.

1.4 Solar photovoltaic system modeling

Solar photovoltaic systems convert part of the incident solar radiation directly into electricity through the photovoltaic effect and are made of a junction of doped semiconductor materials [22]. This junction, known as the p-n junction, is formed by doping the semiconductor on one side with a material that has fewer electrons in its valence layer to create p-silicon, which has a deficiency of electrons, and doping the other side with a material that has more electrons in its valence layer to form n-silicon, which has an excess of electrons. When a semiconductor absorbs a photon with sufficient energy, an electron from the outer layer of an atom is freed, creating a hole-electron pair within the structure. In the case of silicon, which is an indirect bandgap semiconductor, photon absorption alone is not sufficient to promote an electron from the valence to the conduction band. The process also requires the involvement of a phonon to conserve momentum within the crystal lattice. When the two materials that make up a photovoltaic cell are joined, the surplus electrons from the n-type material fill the holes in the p-type material, and the gaps from the p-type diffuse into the n-type material. Consequently, the n-side of the junction becomes positively charged, while the p-side becomes negatively charged. The negative charges on the p-side inhibit the movement of additional electrons from the n-side, while the movement of additional electrons from the p-side is easier due to the positive charges at the junction on the n-side. Thus, the p-n junction behaves like a diode. These two materials are connected through an external circuit, allowing electrons (i.e., electric current) to flow through.

The amount of solar irradiance absorbed by the photovoltaic cell significantly impacts its power output. Additionally, cell temperature is a crucial factor, as higher temperatures can negatively impact the semiconductor material in photovoltaic cells. Specifically, increased temperature lowers the bandgap energy of the semiconductor, which is the energy required to free an electron from its atomic bond. This reduction in bandgap energy leads to a higher intrinsic carrier concentration and increased recombination rates, both of which reduce the cell's power output and efficiency. Essentially, as the temperature rises, the semiconductor becomes less effective at converting sunlight into electricity [44].

The most prevalent photovoltaic technologies use silicon and include monocrystalline silicon, polycrystalline silicon, and amorphous silicon doped with materials such as phosphorus (for type n) and boron (for type p). Monocrystalline silicon cells are composed of a single continuous crystal lattice structure, making them more efficient but also more expensive due to the production process. The record efficiency for monocrystalline silicon cells is 26.1 % [45] under Standard Test Conditions (STC), which involves a cell temperature of 25°C, an irradiance of 1000 W/m² and air mass of 1.5. Polycrystalline silicon cells, produced using multiple grains of silicon, have lower efficiency [46], with the record being 24.4 % [45]. Amorphous silicon, also known as thin-film photovoltaic technology, consists of silicon in a thin, homogeneous layer that can be deposited on a variety of substrates. These cells generally have lower efficiencies, with the record being 21.24% [45]. In addition to these traditional silicon-based cells, there are several other types of photovoltaic technologies. Second-generation cells, like cadmium telluride (CdTe) and copper indium gallium selenide (CIGS), are also thin-film technologies with record efficiencies of 22.4 % and 23.6 %, respectively [45]. Third-generation cells include advanced concepts such as multi-junction cells, which stack multiple layers of photovoltaic materials to capture a broader spectrum of sunlight and achieve efficiencies above 30% [45]. Perovskite solar cells, an emerging technology, have shown rapid improvements in efficiency, reaching up to 26.1% in laboratory conditions [45], while organic photovoltaics, though still less efficient, offer potential benefits in flexibility and low-cost production. Each of these technologies has its own advantages and trade-offs regarding efficiency, cost, and application suitability.

Photovoltaic systems are typically modular and can be built to virtually any size. Photovoltaic cells are grouped into modules, connected in series and parallel, which are encapsulated with various materials to protect the cells and electrical connections from the environment. These modules can be combined to form photovoltaic strings (groups of modules in series) and arrays (groups of modules or strings in parallel), thereby increasing the power output. Photovoltaic equipment usually has no moving parts, requiring minimal maintenance and offering long-term reliability. However, some systems may include a tracking mechanism to follow the apparent movement of the sun in the sky, thereby increasing the collected solar irradiation. Most solar photovoltaic power plants do not use tracking mechanisms.

For the design and performance optimization of photovoltaic systems, trustworthy modeling before and during their operation is essential [47]. Models simulate the response of these complex systems, providing clear insights into multivariate and changing conditions. Depending on the conditions and limitations of each problem, the first step is to select an appropriate model.

1.4.1 Modeling photovoltaic conversion

Various models have been proposed in the literature for estimating photovoltaic power output of solar cells. The most basic models employ simplified linear relationships that consider the efficiency of the PV cell or module, as well as the impact of incident solar irradiance and sometimes the cells temperature.

More complex physical models are based on modeling an electric circuit that can be considered equivalent to the photovoltaic cell or module through assumptions and approximations.

Equivalent electric circuit models typically consist of a diode representing the p-n junction and a current source representing the electric current generated due to the photovoltaic effect. Additional electric resistors are also included, namely a series resistance representing the resistance inside the cell, a shunt or parallel resistance associated with the internal resistance of the p-n junction (diode), and sometimes additional diodes to account for recombination losses at low irradiance levels, the impact of grain boundaries, and leakage currents due to proximity effects [47].

These models rely on several parameters defining the electric circuit, which are typically not measured or disclosed by photovoltaic cell or module manufacturers. Therefore, before the model can be applied, these parameters must be extracted using available information or by performing experiments. The more complex the model, the more information is needed, that is, more parameters need to be determined. Another crucial aspect for applying these models in real conditions is accounting for the irradiance and temperature dependence of the photovoltaic cell's efficiency and power output. This requires adjusting the different parameters to the actual conditions.

The process of extracting and adjusting model parameters can be time-consuming. Many approaches have been proposed, ranging from purely empirical models to physical models, either using iterative or non-iterative methods, as well as machine learning techniques and hybrid methods [48].

1.4.2 Modeling photovoltaic cell temperature

The temperature of a photovoltaic cell or module under illumination is determined by modeling the energy balance in the cell [44]. The solar energy absorbed by the cell is converted in a large part to thermal energy, which is transferred to the environment through a combination of heat transfer mechanisms: conduction-convection, and thermal radiation. Given the significant impact of cell temperature on electrical energy generation, accurately determining this temperature that arises from the energy balance is crucial for the application of photovoltaic electric power output models. In photovoltaic powerplants, this variable is typically measured by attaching a temperature sensor to the back of photovoltaic modules; however, this may not always provide an accurate measurement of the actual photovoltaic cell temperature.

Several models have been developed to compute photovoltaic cell temperature, many of which are simple empirical correlations involving air temperature and wind speed. Other approaches consider the energy balance of the photovoltaic module, which can be modeled as either multiple layers with different properties or a single layer with weighted average thermal properties. These models can also be categorized as steady-state or dynamic models (transient) and may be one-, two- or three-dimensional.

Another aspect to consider is whether the thermal and electrical photovoltaic models are coupled, as well as the type of solver used (numerical or analytical) since temperature and power output are dependent [49]. It is important to note the trade-off between performance and complexity when selecting the most appropriate model for a specific application [50].

1.4.3 Modeling electric power output

To connect a solar photovoltaic system to the electric grid, a conversion from direct current (DC) to alternating current (AC) is necessary. Power conversion systems (PCS) used in photovoltaic powerplants typically include inverters and a controller with maximum power point tracking (MPPT). The maximum power point corresponds to the point on the current-voltage (I–V) curve where the product of current and voltage is maximized. It is closely related to the fill factor (FF), which is defined as the ratio between the maximum power output and the product of open-circuit voltage and short-circuit current, providing an indicator of the cell's electrical quality and efficiency. The MPPT is a high-efficiency DC-to-DC converter that functions as an optimal electrical load for the solar panel or array to operate at maximum power output [44].

This conversion process relies on various electronic circuit devices, and it inherently involves losses, which can be quantified by the efficiency of the PCS. When using only datasheet information, the modeling of these devices often focuses on their efficiency and power capacity, while some models consider efficiency as a function of input power, provided there is sufficient data [51]. Typical equipment datasheets usually offer a constant efficiency value and the nominal power rating.

1.4.4 Other losses

In solar photovoltaic systems, other types of losses can impact overall performance depending on the specific characteristics of the modules, electric connections and environmental conditions. One significant loss is due to the reflection of solar radiation on photovoltaic modules, typically accounted for through a function that considers the incidence angle, known as the incidence angle modifier [52]. Shading of modules and the obscuring of the sky dome also contribute to losses, which can be factored in when calculating the irradiance on tilted surfaces. Cable losses, another important origin of losses, can be computed based on the nominal voltage drop for the specific system design [53]. Additional losses include those from soiling, light-induced degradation, and module mismatch [53]. For systems incorporating electric batteries, energy storage losses must also be taken into account.

1.5 Applications and importance of solar forecasting

Solar energy forecasting, namely photovoltaic power, is essential for integrating solar energy systems into the power grid. Additionally, it is very useful for optimizing photovoltaic system operations and managing financial risks associated with solar energy projects. Accurate forecasts enable grid operators to balance supply and demand, enhancing grid stability and reducing reliance on backup power sources. They assist in scheduling maintenance and deploying energy storage systems, maximizing energy production and efficiency. Reliable forecasts help investors and stakeholders assess the economic feasibility of projects, secure financing, and develop pricing strategies. Additionally, reliable solar energy forecasts contribute to stabilize energy markets, promote competitive pricing, and support environmental sustainability by reducing greenhouse gas emissions and promoting renewable energy transition.

The need for integrated PV forecasting models is increasingly evident as these models play a vital role across various time scales. Long-term forecasts, which involve solar resource mapping, are essential during the design and planning phases of solar power plants. These forecasts help determine the optimal location and configuration of PV systems, ensuring that the long-term energy output justifies the investment. Mid-term forecasts, ranging from days to weeks, are critical for the operational management of PV installations. They allow for better planning of maintenance activities, energy storage management, and grid integration, ensuring that energy production is optimized, and downtime is minimized.

In the short term, forecasts covering minutes to hours are indispensable for real-time grid management and the immediate operational control of PV systems. These short-term predictions enable rapid responses to changes in weather conditions, helping to maintain grid stability and prevent power outages. By integrating forecasting across these different time scales, PV systems can be designed, operated, and managed more effectively, ensuring that solar energy contributes reliably and efficiently to the overall energy mix. This comprehensive approach not only supports the technical and economic viability of solar projects but also reinforces their role in the broader transition to renewable energy.

1.6 Objectives

The goal of this work was to improve the accuracy of solar resource assessment and short-term photovoltaic power forecasting across different locations and technologies, by developing an integrated and transferable model which incorporates physical modeling, data-driven approaches and inputs from Numerical Weather Prediction and aerosol models.

To fulfill this main objective, the following partial objectives were pursued: i) assess NWP models outputs for solar resource mapping and solar radiation forecasting; ii) develop ANN models to improve solar radiation modeling and forecasting; iii) evaluate and improve transposition models for converting solar irradiance components to tilted surface irradiance; iv) develop coupled models for estimation of photovoltaic temperature and power output; v) research inverter and typical losses models; vi) integrate different models into a global forecast model and validate with experimental data for accurate and reliable PV power prediction.

By achieving these objectives, the model aims to provide a robust tool for forecasting the power output of PV systems, thereby supporting the efficient operation and management of solar energy resources.

1.7 Outline of the thesis

This thesis is organized into seven chapters. The first chapter provides an introduction to the background and significance of the study, offering an overview of solar modeling and forecast, of photovoltaic systems and the importance of global forecasting models. It introduces key concepts such as solar radiation, solar photovoltaic systems, and the different models used for forecasting solar energy. After this, the following four chapters correspond to published papers in international peer-review journals while the fifth one is under review. The second chapter focuses on the development of ANN models aimed at improving solar irradiance forecasting using NWP and aerosol data as input. Details on the process of optimizing ANN configurations, including the selection of input variables and training functions are presented, as well as the evaluation of model performance against the traditional numerical weather prediction results. The use of the proposed method for mapping solar resource for a typical meteorological year is also presented. Building on this, the third chapter extends the work by validating the developed ANN models using operational NWP data. It discusses the application of these models in predicting direct normal irradiance and their integration with existing weather prediction data within a mid-term forecast horizon, highlighting the improvements in forecasting accuracy. The fourth chapter shifts focus to the prediction of global solar irradiance on tilted surfaces, considering the effects of shading and anisotropic diffuse radiation. This chapter explores the development of models designed to enhance the accuracy of power predictions in real photovoltaic powerplants. The fifth chapter delves into the development and validation of coupled thermal-electric models for photovoltaic systems. It addresses the challenges associated with predicting the temperature and power output of photovoltaic systems, considering the effect of various environmental factors. In the sixth chapter, the study introduces the model for forecasting photovoltaic power output by integrating the previously presented models. It is designed to operate without on-site observations, allowing for broad application and reduced operational costs. It provides 72-hour power output forecasts with flexible temporal resolution. The chapter highlights the algorithm's versatility and its potential to improve PV system management and grid integration, even in locations with limited historical data. Finally, the seventh chapter presents the conclusions of the work by summarizing the key findings, discussing the impact of the research carried out, and suggesting directions for future work. This chapter reflects on the contributions of the study to the fields of solar energy forecasting and photovoltaic system optimization.

References

1. International Energy Agency Renewables 2023. Analysis and Forecast to 2028.; Paris, 2028;
2. World Meteorological Organization WMO, International Meteorological Vocabulary; 2nd, Ed.; Geneva, 1992;
3. Coulson, K. Solar and Terrestrial Radiation; Academic Press: New York, New York, USA, 1975;
4. Myers, D.R. SOLAR RADIATION Practical Modeling for Renewable Energy Applications; Elsevier, 2013; ISBN 978-1-4665-0327-4.
5. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2.0 Available online: https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 28 May 2024).
6. World Meteorological Organization WMO, International Meteorological Vocabulary; 2nd, Ed.; Geneva, 1992;
7. Stoffel, T.; Renne, D.; Myers, D.; Wilcox, S.; Sengupta, M.; George, R.; Turchi, C. Concentrating Solar Power: Best Practices Handbook for the Collection and Use of Solar Resource Data (CSP); Golden, CO (United States), 2010;
8. Lhendup, T.; Lhundup, S. Comparison of Methodologies for Generating a Typical Meteorological Year (TMY). Energy for Sustainable Development 2007, 11, 5–10, doi:10.1016/S0973-0826(08)60571-2.
9. Janjai, S.; Deeyai, P. Comparison of Methods for Generating Typical Meteorological Year Using Meteorological Data from a Tropical Environment. Appl Energy 2009, 86, 528–537, doi:10.1016/J.APENERGY.2008.08.008.

10. Skeiker, K. Comparison of Methodologies for TMY Generation Using 10 Years Data for Damascus, Syria. Energy Convers Manag 2007, 48, 2090–2102, doi:10.1016/J.ENCONMAN.2006.12.014.
11. Chan, A.L.S.; Chow, T.T.; Fong, S.K.F.; Lin, J.Z. Generation of a Typical Meteorological Year for Hong Kong. Energy Convers Manag 2006, 47, 87–96, doi:10.1016/J.ENCONMAN.2005.02.010.
12. Pereira, S.; Abreu, E.F.M.; Iakunin, M.; Cavaco, A.; Salgado, R.; Canhoto, P. 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. Solar Energy 2022, 236, 225–238, doi:10.1016/j.solener.2022.03.003.
13. Sørensen, B. GIS Management of Solar Resource Data. Solar Energy Materials and Solar Cells 2001, 67, 503–509.
14. Martins, F.R.; Pereira, E.B.; Abreu, S.L. Satellite-Derived Solar Resource Maps for Brazil under SWERA Project. Solar Energy 2007, 81, 517–528, doi:10.1016/J.SOLENER.2006.07.009.
15. Polo, J.; Bernardos, A.; Navarro, A.A.; Fernandez-Peruchena, C.M.; Ramírez, L.; Guisado, M. V.; Martínez, S. Solar Resources and Power Potential Mapping in Vietnam Using Satellite-Derived and GIS-Based Information. Energy Convers Manag 2015, 98, 348–358, doi:10.1016/J.ENCONMAN.2015.04.016.
16. Jiménez Pedro A. and Lee, J.A. and K.B. and H.S.E. Solar Resource Evaluation with Numerical Weather Prediction Models. In Solar Resources Mapping: Fundamentals and Applications; Polo Jesús and Martín-Pomares, L. and S.A., Ed.; Springer International Publishing: Cham, 2019; pp. 199–219 ISBN 978-3-319-97484-2.
17. NASA AERONET AEROSOL ROBOTIC NETWORK Available online: https://aeronet.gsfc.nasa.gov/ (accessed on 27 August 2024).
18. Kalogirou, S.A. Solar Energy Engineering: Processes and Systems; Academic Press Inc., 2009;
19. Hassan, Q.; Jaszczur, M.; Teneta, J.; Abbas, M.K.; Hasan, A.; Al-Jiboory, A.K. Experimental Investigation for the Estimation of the Intensity of

Solar Irradiance on Oblique Surfaces by Means of Various Models. Energy Harvesting and Systems 2022, 9, 227–237, doi:10.1515/ehs-2021-0087.

20. Liu, B.Y.H.; Jordan, R.C. The Long-Term Average Performance of Flat-Plate Solar-Energy Collectors. Solar Energy 1963, 7, 53–74, doi:10.1016/0038-092X(63)90006-9.
21. Yang, D. Solar Radiation on Inclined Surfaces: Corrections and Benchmarks. Solar Energy 2016, 136, 288–302, doi:10.1016/j.solener.2016.06.062.
22. Duffie, J.A.; Beckman, W.A. Solar Engineering of Thermal Processes. 2013, 1–885.
23. Ayala Pelaez, S.; Deline, C. Bifacial_radiance: A Python Package for Modeling Bifacial Solar Photovoltaic Systems. J Open Source Softw 2020, 5, 1865, doi:10.21105/joss.01865.
24. Russell, A.C.J.; Valdivia, C.E.; Bohemier, C.; Haysom, J.E.; Hinzer, K. DUET: A Novel Energy Yield Model With 3-D Shading for Bifacial Photovoltaic Systems. IEEE J Photovolt 2022, 12, 1576–1585, doi:10.1109/JPHOTOV.2022.3185546.
25. Ershov, S.; Voloboy, A.; Zhdanov, D.; Zhdanov, A.; Frolov, V. Direct Calculation of the Optimal Weight for MIS. In Proceedings of the CPT2020 The 8th International Scientific Conference on Computing in Physics and Technology Proceedings; Bryansk State Technical University, November 19 2020; pp. 256–259.
26. Gueymard, C. An Anisotropic Solar Irradiance Model for Tilted Surfaces and Its Comparison with Selected Engineering Algorithms. Solar Energy 1987, 38, 367–386, doi:10.1016/0038-092X(87)90009-0.
27. Krishnan, N.; Kumar, K.R.; Inda, C.S. How Solar Radiation Forecasting Impacts the Utilization of Solar Energy: A Critical Review. J Clean Prod 2023, 388, 135860, doi:10.1016/J.JCLEPRO.2023.135860.
28. Lorenz, E.; Heinemann, D. Prediction of Solar Irradiance and Photovoltaic Power. Comprehensive Renewable Energy 2012, 1, 239–292, doi:10.1016/B978-0-08-087872-0.00114-1.

29. Rigollier, C.; Lefèvre, M.; Wald, L. The Method Heliosat-2 for Deriving Shortwave Solar Radiation from Satellite Images. Solar Energy 2004, 77, 159–169, doi:10.1016/j.solener.2004.04.017.
30. Chow, C.W.; Urquhart, B.; Lave, M.; Dominguez, A.; Kleissl, J.; Shields, J.; Washom, B. Intra-Hour Forecasting with a Total Sky Imager at the UC San Diego Solar Energy Testbed. Solar Energy 2011, 85, 2881–2893, doi:10.1016/j.solener.2011.08.025.
31. Warner, T.T. Numerical Weather and Climate Prediction; Cambridge University Press: New York, New York, USA; ISBN 978-0-521-51389-0.
32. McTaggart-Cowan, R.; Vaillancourt, P.A.; Zadra, A.; Chamberland, S.; Charron, M.; Corvec, S.; Milbrandt, J.A.; Paquin-Ricard, D.; Patoine, A.; Roch, M.; et al. Modernization of Atmospheric Physics Parameterization in Canadian NWP. J Adv Model Earth Syst 2019, 11, 3593–3635, doi:10.1029/2019MS001781.
33. ECMWF Bologna Press Kit. 2021,.
34. Zhang Fuqing and Routray, A. Data Assimilation: Comparison and Hybridization between Ensemble and Variational Methods. In Advanced Numerical Modeling and Data Assimilation Techniques for Tropical Cyclone Prediction; Mohanty U. C. and Gopalakrishnan, S.G., Ed.; Springer Netherlands: Dordrecht, 2016; pp. 361–383 ISBN 978-94-024-0896-6.
35. ECMWF CAMS: Global Atmospheric Composition Forecast Data Documentation Available online: https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation (accessed on 10 December 2021).
36. Yang, D. Making Reference Solar Forecasts with Climatology, Persistence, and Their Optimal Convex Combination. Solar Energy 2019, 193, 981–985, doi:10.1016/J.SOLENER.2019.10.006.
37. Belmahdi, B.; Louzazni, M.; Bouardi, A. El One Month-Ahead Forecasting of Mean Daily Global Solar Radiation Using Time Series

Models. Optik (Stuttg) 2020, 219, 165207, doi:10.1016/J.IJLEO.2020.165207.

38. Bacher, P.; Madsen, H.; Nielsen, H.A. Online Short-Term Solar Power Forecasting. Solar Energy 2009, 83, 1772–1783, doi:10.1016/j.solener.2009.05.016.
39. Reikard, G. Predicting Solar Radiation at High Resolutions: A Comparison of Time Series Forecasts. Solar Energy 2009, 83, 342–349, doi:10.1016/j.solener.2008.08.007.
40. Mellit, A.; Kalogirou, S. Handbook of Artificial Intelligence Techniques in Photovoltaic Systems: Modeling, Control, Optimization, Forecasting and Fault Diagnosis; 1st ed.; Academic Press, 2022; ISBN 9780128206416.
41. Gupta, V.; Mishra, V.K.; Singhal, P.; Kumar, A. An Overview of Supervised Machine Learning Algorithm. In Proceedings of the 2022 11th International Conference on System Modeling & Advancement in Research Trends (SMART); 2022; pp. 87–92.
42. Pang, Z.; Niu, F.; O'Neill, Z. Solar Radiation Prediction Using Recurrent Neural Network and Artificial Neural Network: A Case Study with Comparisons. Renew Energy 2020, 156, 279–289, doi:10.1016/J.RENENE.2020.04.042.
43. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting: A Review. Renew Energy 2017, 105, 569–582, doi:10.1016/j.renene.2016.12.095.
44. Smets, A.H.; Jäger, K.; Isabella, O.; Swaaij, R.A. van; Zeman, M. Solar Energy The Physics and Engineering of Photovoltaic Conversion, Technologies and Systems; UIT Cambridge, Ed.; 2016;
45. NREL Best Research-Cell Efficiencies Available online: https://www.nrel.gov/pv/interactive-cell-efficiency.html (accessed on 27 August 2024).

46. McEvoy, A.; Markvart, T.; Castaner, L. Practical Handbook of Photovoltaics - Fundamentals and Applications; Academic Press, Ed.; 2nd ed.; 2012; ISBN 9780123859341.
47. Xiao, W. Photovoltaic Power System: Modeling, Design, and Control; Wiley, Ed.; Wiley, 2017; ISBN 9781119280347.
48. Gholami, A.; Ameri, M.; Zandi, M.; Gavagsaz Ghoachani, R. Electrical, Thermal and Optical Modeling of Photovoltaic Systems: Step-by-Step Guide and Comparative Review Study. Sustainable Energy Technologies and Assessments 2022, 49, 101711, doi:10.1016/j.seta.2021.101711.
49. Kalogirou, S.A. Solar Energy Engineering: Processes and Systems; Academic Press Inc., 2009;
50. Mayer, M.J.; Yang, D. Pairing Ensemble Numerical Weather Prediction with Ensemble Physical Model Chain for Probabilistic Photovoltaic Power Forecasting. Renewable and Sustainable Energy Reviews 2023, 175, 113171, doi:10.1016/j.rser.2023.113171.
51. Xiao, W.; Edwin, F.F.; Spagnuolo, G.; Jatskevich, J. Efficient Approaches for Modeling and Simulating Photovoltaic Power Systems. IEEE J Photovolt 2013, 3, 500–508, doi:10.1109/JPHOTOV.2012.2226435.
52. Duffie, J.A.; Beckman, W.A. Solar Engineering of Thermal Processes. 2013, 1–885.
53. Roberts, J.J.; Mendiburu Zevallos, A.A.; Cassula, A.M. Assessment of Photovoltaic Performance Models for System Simulation. Renewable and Sustainable Energy Reviews 2017, 72, 1104–1123, doi:10.1016/j.rser.2016.10.022.

Nomenclature

AC	Alternate Current
ANN	Artificial Neural Network
ARIMA	Autoregressive Integrated Moving Average

ARMA	Autoregressive Moving Average
DC	Direct Current
DIF	Diffuse Horizontal Irradiance
DNI	Direct Normal Irradiance
ECMWF	European Centre for Medium-range Weather Forecasts
FF	Fill Factor
GHI	Global Horizontal Irradiance
GTI	Global Tilted Irradiance
MOS	Model Output Statistics
MPPT	Maximum Power Point Tracker
NWP	Numerical Weather Prediction
PCS	Power Conversion Systems
PV	Photovoltaic
TMY	Typical Meteorological Year
TSI	Total Sky Imager

Chapter 2

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†

Abstract

In this work a method for regional solar resource assessment based on numerical weather prediction (NWP) and artificial neural network (ANN) models is presented. The method was developed using typical meteorological and solar radiation data and applied to the location of Évora, Portugal with the goal of assessing solar global horizontal (GHI) and direct normal (DNI) irradiations with 1.25 km of horizontal resolution in the south of Portugal. The NWP model used was the research model Meso-NH and a site adaptation model was developed based in ANNs and using as inputs the simulated meteorological variables from Meso-NH and aerosol data from Copernicus Atmospheric Monitoring Services (CAMS) for the observation site. The resulting annual relative mean bias errors for GHI and DNI at Évora and typical meteorological year are of 0.55 % and 0.98 %, respectively, while the values for the original Meso-NH simulations are of 8.24 % for GHI and 31.71 % for DNI. The developed site adaptation model is applied to the region for the purpose of solar radiation assessment and validated using data from a network of solar radiation measuring stations scattered throughout the

---

†Sara Pereira[a], Edgar F.M. Abreu[a], Maksim Iakunin[a], Afonso Cavaco[a,b], Rui Salgado[a,c] and Paulo Canhoto[a,d], 2022. 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, 225–238. https://doi.org/10.1016/j.solener.2022.03.003

[a] Institute of Earth Sciences, University of Évora, Évora, Portugal
[b] Renewable Energies Chair, University of Évora, Évora, Portugal
[c] Physics Department, University of Évora, Évora, Portugal
[d] Department of Mechatronics Engineering, University of Évora, Évora, Portugal

south of Portugal, showing relative mean bias errors of 2.34 % for GHI and 3.41 % for DNI, while the original Meso-NH simulations presents relative mean bias errors of 8.50 % and 29.54 %, respectively. These results allowed the generation of improved solar resource availability maps which are a very useful tool in solar resource assessment, the study of shortwave radiative climate, as well as project planning and solar system design and operation.

Keywords
Solar radiation; Solar energy; Site adaptation; Typical meteorological year; Numerical weather prediction; Artificial neural network

2.1 Introduction

The solar energy industry has been growing since the 1990s mainly in response to environmental issues related to the use of fossil fuels. At the end of 2020, the total installed power worldwide was of 707 GW for photovoltaic technologies and 6.5 GW for solar concentration systems [1]. The number of new solar energy systems is expected to grow in the future as countries aim for low-carbon alternatives for electricity generation. Thus, a reliable assessment of the solar resource and solar radiation forecast tools are crucial as they are helpful not only for the understanding of the climate of the Earth but also for project planning and energy systems design and operation.

Solar radiation comprises two components, the direct normal irradiation (DNI), which is the direct beam on a plane perpendicular to the rays of the sun, and the diffuse irradiation, usually measured on a horizontal plane (DHI - diffuse horizontal irradiation). The total irradiation on a horizontal surface is known as Global Horizontal Irradiation (GHI). If the surface of interest is tilted, a photovoltaic panel or a solar thermal collector for instance, then a third component must be considered, which is the solar radiation reflected on the ground and surrounding objects. On the other hand, if the focus is solar concentration systems the component of interest is DNI, including circumsolar irradiance.

Ground-based measurements of solar irradiance are usually made by pyranometers for the global or diffuse components and by pyrheliometers for the direct normal component and can be used to assess the solar resource in a specific location or in a region using observations from different locations [2,3]. For regional resource assessment, it would be ideal to have observations made throughout a long period of time in multiple sites and in a somewhat gridded layout. Such network, however, is difficult to install and maintain.

Once long-term data series are available for a given location, a way to assess solar resource is generating a Typical Meteorological Year (TMY) by concatenating typical meteorological months of real measurements that reproduce the long-term statistics. These are useful for energy simulation of solar power plants and buildings, either to determine their thermal response under different environmental conditions or to design and simulate integrated renewable energy systems. A TMY generated for solar resource assessment is composed of solar radiation and other meteorological variables during a period of one year [4] usually including mean and extreme values of temperature, wind and relative humidity.

The first version of TMY was developed using the Sandia method proposed by I. J. Hall et al. [5]. The National Renewable Energy Laboratory (NREL) has developed a second (TMY2) [6] and third version (TMY3) [7] by slightly modifying the process of selecting the typical months. Regardless of the method used, the TMY has been recently used in applications related to solar energy as reported in the literature [8–11].

In addition to observations, satellite data [12,13] and numerical weather prediction (NWP) models [14–16] have been used to study the solar resource in more extensive regions instead of in discrete locations as is the case of TMY. NWP models solve the physical constitutive and state equations that describe atmospheric processes obtaining estimates of the evolution of atmospheric conditions over time. Usually, the main goal of these NWP models is weather forecast, so the detailed prediction of the different components of solar radiation is not essential for that purpose, namely the DNI and DHI, since only GHI is considered relevant for the computation of the energy balance of the surface and atmosphere. More recently, due to the need for solar energy system development, attention has been drawn to this aspect and to the improvement of the accuracy regarding solar radiation specific variables [17]. Soukissian et al. [18] used the recently released ERA5 reanalysis dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF) for the study of offshore wind and solar resource in the Mediterranean Sea.

The existence and location of clouds as well as the concentration and type of aerosols in the atmosphere strongly affect solar radiation at the surface, which is difficult to accurately describe in NWP models. In the case of aerosols, for example, monthly-mean aerosol climatologies are typically used to reduce the computational effort of the operational models. Modeling and prediction of aerosols has been recently studied [19–21] and its application to NWP models tested. Breitkreuz et al. [22] used aerosol predictions and experimental data to forecast solar irradiance, which was compared with the output of the ECMWF model. There was a decrease in the mean squared error (MSE) of hourly GHI for the second and third forecast days from 11.5 % to 7.2 % for clear sky conditions.

In another approach, site adaptation models are used to compare and correct modelled data against observations made at a specific site. This term is mainly used for the bias correction of satellite data but many other techniques can be used to correct modelled data [23], as is the case of machine learning models. Alfadda et al. [24] compared hourly GHI, DNI and DHI forecasting ability of various machine learning models, including Artificial Neural Networks (ANN), using radiation and aerosol measurements as well as wind and aerosol forecasts in Saudi Arabia as inputs, showing that the ANN model performs better than the remaining tested models. Fonseca et al. [25] developed a forecasting model for day-ahead solar irradiation using support vector machine learning algorithm that has as input NWP data from the mesoscale model of the Japan Meteorological Agency, achieving a reduction of 16 % of the root mean squared error (RMSE) on the regional scale taking measurements as reference.

ANN models consist of input and output layers, as well as one or more hidden layers containing units defined as neurons. Such computing systems progressively improve task performance by considering examples, generally without task-specific programming. Many authors have used ANN in the field of renewable energy, e.g. for correction of wind speed or solar irradiance forecasts [26–28], and for simulation of photovoltaic power [29–37]. Ghimire et al. [38] assessed daily solar energy potential for five cities of Australia using different machine learning algorithms such as ANN, support vector machine, gaussian process machine learning and genetic programming using reanalysis data from ECMWF as inputs. The results show that the ANN model outperforms all other machine learning models as well as the deterministic models used for benchmarking.

In this work, a method is proposed to assess direct normal and global horizontal irradiations at a regional scale using an artificial neural network model for site adaptation of simulations from a numerical weather prediction model based on typical meteorological data and including aerosol information. This approach combines the advantage of TMY in identifying the typical periods that better represent the long-term climatology, the ability of NWP model to simulate solar radiation and meteorological variables in a regional scale and the capability of ANN models for data correction. Experimental meteorological and solar radiation data from Évora, Portugal, was used for model development while the validation of the developed method also used data from a network of solar radiation measuring stations dispersed throughout a region centered in Évora.

In Section 2, the experimental data and quality control tests are presented, and the generation of the typical meteorological year (TMY) is described. In Section 3 the typical months of the TMY are simulated using the NWP model Meso-NH and analysis data from ECMWF for initialization, generating atmospheric and solar radiation data for the desired region. The simulated variable of study (in this work there are two: GHI and DNI) for the selected location (Évora) is analyzed against observations and other simulated meteorological variables. In Section 2.4, a procedure for the development and optimization of a site adaptation model based on ANN is presented using not only the outputs from the Meso-NH simulations but also aerosol related variables from the Copernicus Atmosphere Monitoring Service (CAMS). After the ANN model is optimized for the selected location, it can be applied to the desired region allowing for a better assessment of the resource as shown in the results and discussion section of this work (Section 2.5). Finally, the conclusions are presented in Section 2.6.

2.2. Experimental data and typical meteorological year generation

To develop this work, long-term data series of observed meteorological and solar radiation variables are needed. In the following, a description of these data series and data processing procedure are presented as well as the generation of a typical meteorological year.

2.2.1 Experimental data

A long and complete series of daily meteorological data is required to determine a TMY. A long-term data series of 16 years (2003 to 2018) of observations made at Évora – Verney (Fig. 2.1) is available for the method development. However, these are subject to instrumentation failures, thus a careful analysis and processing of observations are necessary for the data series to be correct and complete. Hence, observations that are beyond the appropriate physical limits are identified and removed while data gaps are filled through linear interpolation or using adequate correlations with other observations made in nearby meteorological stations.

The first step is the identification and removal of data that are beyond the physical limits that can be admitted for each variable measured in Évora – Verney station (38.567811, -7.911459). The variables studied were GHI (obtained from global horizontal irradiance observations made with a Black & White Pyranometer from EPLAB model 8-48), average, maximum and minimum air temperature, and relative humidity (Hydro-Thermo Transmitter from Thies CLIMA, model 1.1005.50) and mean wind speed and wind gust (A100R contact closure anemometer from Vector Instruments until 2014 inclusive, and WindSonic ultrasonic wind sensor from Gill Instruments after 2014).

Adjustments were made for relative humidity and wind speed. Over time, values of daily relative humidity show a small linear drift (decreasing values) until a new calibration is made. To take this into account, slopes were determined for each period of observations between calibrations and corrections were applied. A restraint was applied limiting relative humidity to 100 %.

Adequate wind observations were only available in the Évora - Mitra meteorological station from ICT (38.525388, -8.016602, not shown in Fig. 2.1), approximately 10 km southwest of Évora-Verney (Fig. 2.1). However, these observations were not measured at the same height throughout time. From 2003 to 2014 they were obtained at 6 m height and from 2015 to 2018 at 10 m height. The measurements at 6 m were extrapolated through the power law of wind speed profile near the ground using the exponents determined for 12 wind direction sectors according to [39] and a correction due to sensor degradation was applied using the software Windographer™, after which the daily values were computed.

Correlations were found between the observations of each variable (except wind) performed in Évora – Verney and two other meteorological stations nearby for data gap filling, namely the Évora – Mitra station and a station located in Portel (38.306528, -7.689500). If the use of correlations is not possible, then linear interpolation is performed. After this procedure, a complete data series of daily GHI, mean, maximum and minimum air temperature, mean, maximum and minimum relative humidity, mean wind speed and wind gust for the period since 01/01/2003 to 31/12/2018 were obtained.

DNI observations made at the Évora – Verney station were also needed for this work, being obtained from global horizontal irradiance and diffuse horizontal irradiance observations made with two Black & White Pyranometers from EPLAB (model 8-48) and a timestep of 10 min until 2016

[40]. After 2016, direct normal irradiance and global horizontal irradiance data were obtained from 1-minute timestep observations from a Kipp & Zonen CHP1 pyrheliometer and a CM6B pyranometer, respectively.

The solar radiation data were subject to quality control procedures, namely those of the Baseline Surface Radiation Network (BSRN) [41] which account for physically impossible values, extremely rare values and values that are not reliable when considering the ratio between diffuse and global horizontal irradiance. To account for instrument and/or data acquisition system malfunctions, a procedure similar to that presented above was implemented to fill the missing global horizontal irradiance values using correlations with nearby meteorological stations, as presented in [42].

For the validation of the method developed in this work, observation data of direct normal irradiance and global horizontal irradiance from stations of the DNI-A project network were used [43]. The DNI-A network is a solar radiation measurement network which comprises 13 stations installed in the south of Portugal. These stations are equipped with Solys2 sun trackers with one CHP1 pyrheliometer, two CMP11 pyranometers to measure horizontal global and diffuse irradiance. The instruments were calibrated in accordance with ISO 9059:1990 and 9847:1992 and the observations are corrected regarding their zero offset, filtered according to the BSRN quality control procedure and gaps filled according to the method developed in [43].

The solar radiation observations used for the method validation were retrieved from the following stations of the DNI-A network: Évora – PECS (38.5306, -8.0112), Évora – EMSP (38.5289, -8.0053), Portalegre (39.2692, -7.4428), Beja (38.0249, -7.8672), Lisboa (38.7734, -9.1779), Sines (37.9576, -8.8473) and Enercoutim (37.4431, -7.7409). The location of these stations can be seen in Fig. 2.1 including the Évora – Verney station which data was used for the development of the method.

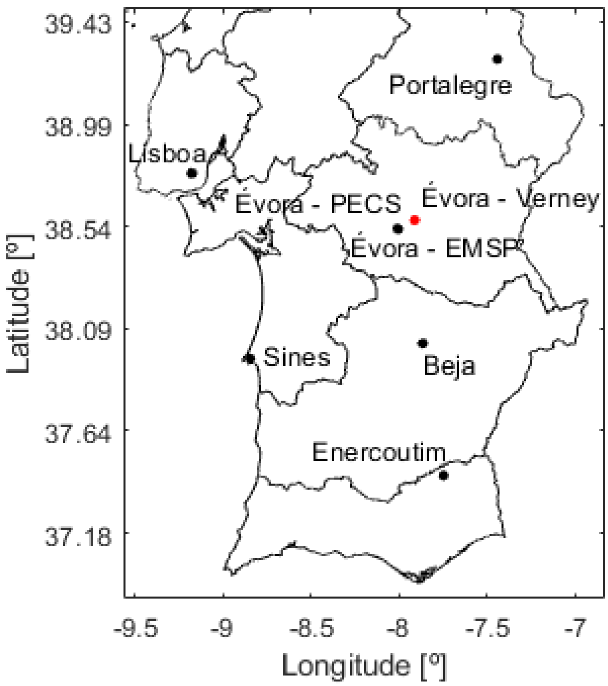


Fig. 2.1 - Location of solar radiation measuring stations from the DNI-A network.

2.2.2 Generation of a typical meteorological year

Sixteen years (2003 to 2018 inclusive) of daily observations of GHI, average, maximum and minimum air temperature and relative humidity and average wind speed and wind gust in Évora – Verney (Fig. 2.1), in a total of nine variables, were used to determine a TMY. The TMY is obtained by concatenating twelve typical meteorological months (TMMs) which are selected through a sequence of steps starting with the determination of the Finkelstein-Schaffer (FS) statistics [44] of the daily values for all variables, according to the Sandia method [5]. Next, the cumulative distribution function (CDF) for each variable in each month of the long-term series is computed and then a weighted sum of CDFs is determined for each month based on the weights shown in Table 2.1.

Table 2.1 - Statistical weights of the variables used on the determination of the TMY.

Variable	**Air temperature**			**Relative humidity**			**Wind speed**		**GHI**
	Mean	**Max.**	**Min**	**Mean**	**Max.**	**Min**	**Mean**	**Max.**	
Weight	2/24	1/24	1/24	2/24	1/24	1/24	2/24	2/24	12/24

The five months (identified by their respective year, from here on termed month/year) for each calendar month which weighted CDFs are closest to the long-term CDF are selected as candidates for TMM. Next, the Sandia method with the slight alterations included by Pissimanis et al. [45] and used in other studies [5,42], is used. This procedure consists in the determination of the root mean square difference (RMSD) of the hourly global horizontal irradiation (GHI) and then the selection of the months/years that show a RMSD lower than min{RMSD} + 0.02 kWh/m^2 for each calendar month. Then, their FS values are computed and the month/year showing a FS lower than min{FS} + 0.03 is finally selected. If this procedure does not result in only one month/year for a given calendar month, the month/year that shows lowest FS values for mean air temperature is selected. The resulting months/years selected for each calendar month of the generated TMY are shown in Table 2.2. This TMY was validated by evaluating the monthly TMY means against long-term means (16 years) as shown for the example of global horizontal irradiance in Fig. 2.2.

Table 2.2 - Selected months/years for each calendar month of the generated TMY for Évora.

Jan.	Feb.	Mar.	Apr.	May	Jun.	Jul.	Aug.	Sep.	Oct.	Nov.	Dec.
2018	2006	2017	2008	2005	2010	2014	2007	2011	2008	2004	2018

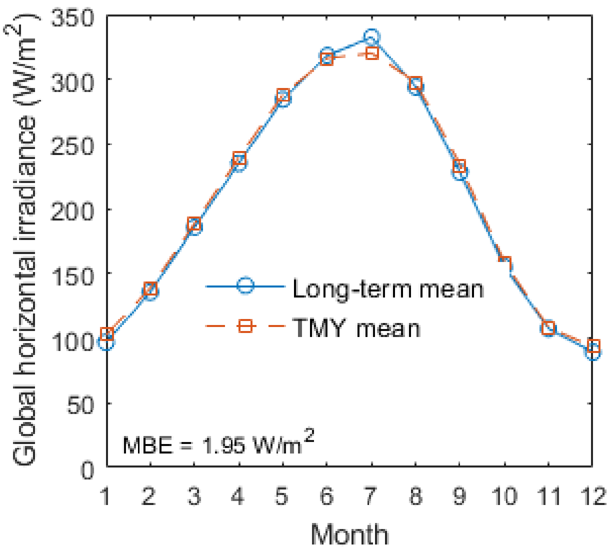


Fig. 2.2 - Comparison between the TMY and long-term monthly mean values of daily mean global horizontal irradiance.

The GHI, DNI and DHI values of the generated TMY can be seen in Fig. 2.3, 2.4 and 2.5, respectively.

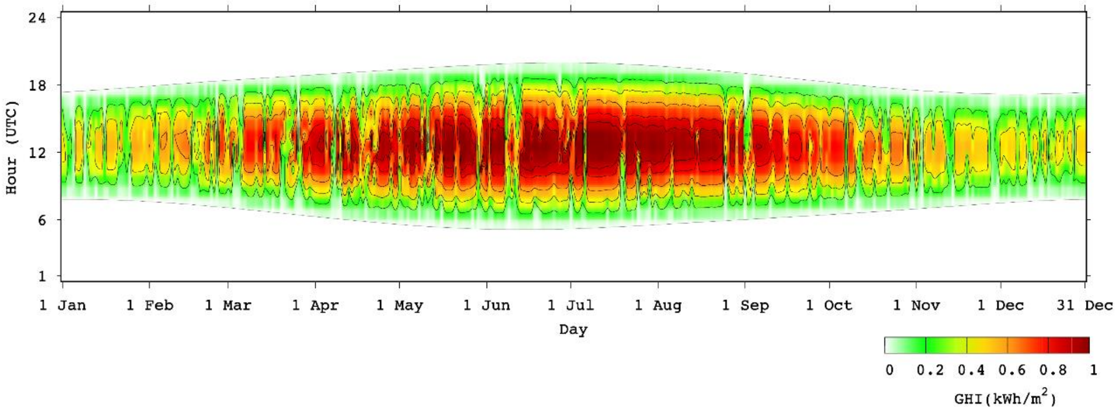


Fig. 2.3 - GHI observed at Évora-Verney station for the generated TMY.

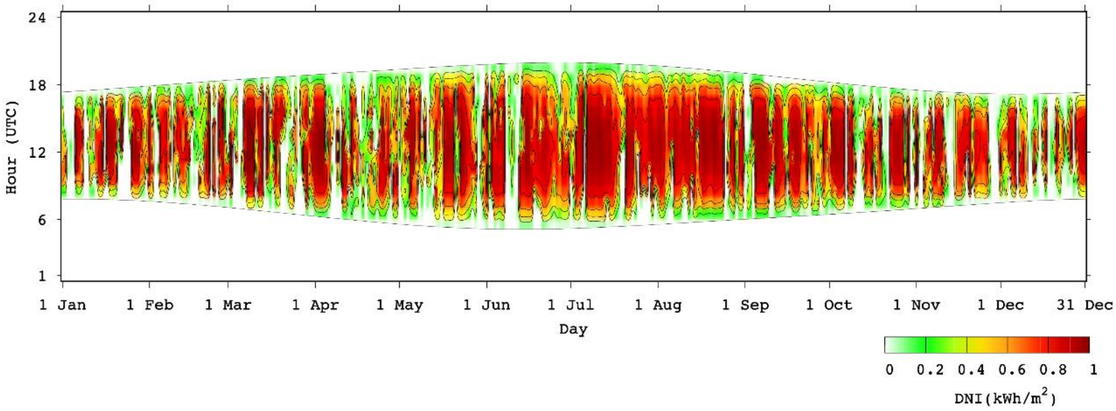


Fig. 2.4 - DNI observed at Évora-Verney station for the generated TMY.

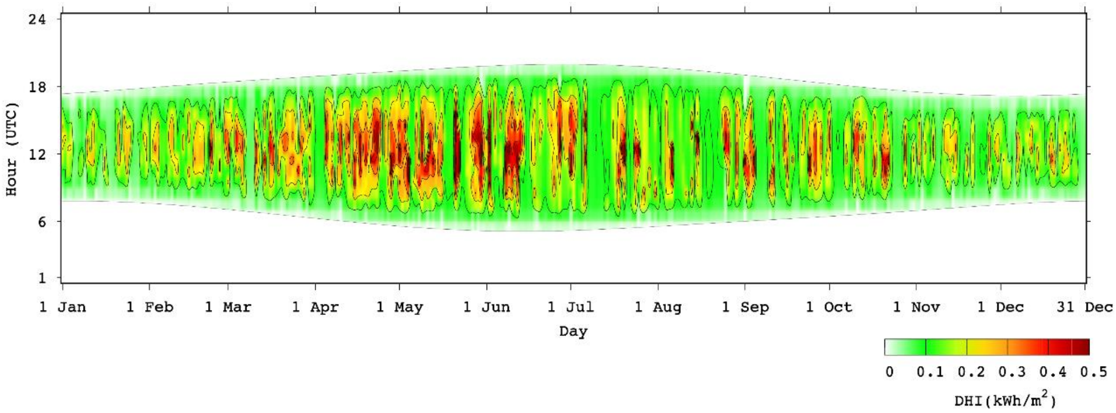


Fig. 2.5 - DHI observed at Évora-Verney station for the generated TMY.

2.3 Solar radiation results from the NWP model

The NWP model chosen to simulate the atmospheric conditions during the TMY was the research model Meso-NH [46]. A brief description of the model and procedure for the simulation of the TMY is made in this section as well as an analysis of the model output data.

2.3.1 Meso-NH model simulations setup

The Meso-NH model is a mesoscale atmospheric model jointly developed by the Laboratoire d'Aérologie (CNRS/University of Toulouse) and by the Centre National de Recherches Météorologiques, CNRM (CNRS/Météo-France) that aims to simulate the state of the atmosphere by incorporating a system of non-hydrostatic momentum equations and computing balances over a wide range of scales. It has a set of physical parameterizations with high detail in the representation of clouds and precipitation [46,47]. Surface-atmosphere interactions are represented by the SURFEX surface platform of schemes [48] and the ECOCLIMAP database, which contains surface parameters with 1 km resolution [49]. The Meso-NH model is a cloud-resolving and large eddy simulation model that was extensively validated and inter-compared with other state-of-art models (see Lac et al. [46], and the references herein or Capecchi [50]). A detailed and comprehensive scientific and technical documentation is available on the Meso-NH web site (mesonh.aero.obs-mip.fr, last access: 11 January 2022), where a list of hundreds of publications can be accessed. The model has long been used successfully in Portugal for studies related to phenomena that are relevant to solar radiation forecast, namely convection, clouds and precipitation [51,52], fog [53] and dust atmospheric transport [52,54].

The radiation transport in the atmosphere is modelled through the most recent ecRad scheme developed by the ECMWF. This scheme is based on the Rapid Radiative Transfer Model (RRTM) [55,56], which takes into account cloud parameterization, liquid water content, ice and snow, using the Monte Carlo Independent Column Approximation (McICA) method that allows for a more accurate representation of the interaction between radiation and the clouds.

The ECMWF developed the Integrated Forecasting System (IFS) which consists in several global operational models, namely an atmospheric general circulation model, an ocean wave model, a land surface model, an ocean general circulation model and perturbation models for the data assimilation and generation of forecast ensembles. This system generates forecasts that encompass different time ranges from the intermediate range (up to 15 days) to seasonal intervals (up to 7 months) with a horizontal resolution of 0.125° and temporal resolution of 1 h. Analyses are produced by combining short-range forecast data with observations to produce the best fit between them. These data are available twice per day and since this work falls under the scope of the development of an operational solar energy and photovoltaic power forecasting algorithm the operational forecasts of the ECMWF were retrieved and used for initialization of the Meso-NH model.

The simulations of the TMY were performed in series of 3 days with a 6-hour spin-up period (every model run covered 6 + 72 h) for the region of the south of Portugal as in Fig. 2.1, with 1.25 km of horizontal resolution and generating several meteorologic near surface variables with the timestep of 1 min. This numerical experiment was also used for the evaluation of the impact of the Alqueva lake on the regional climate as described in detail in [57] where the specific schemes and parameters used can be found as well as a validation of the simulations against several meteorological observations.

2.3.2 Analysis of solar radiation data output from Meso-NH

After integrating the simulated 1 min global horizontal and direct normal irradiances into 10-minute values and interpolating these for the location of Évora – Verney station through a bi-linear interpolation of the values from the 4 neighboring grid-points, these were paired with the available observations for the whole TMY resulting in Fig. 2.6. The model tends to overestimate both variables, which is also visible when comparing the annual values of the modelled and observed variables for this site, which are

approximately 1.97 MWh/m² and 1.82 MWh/m², respectively, for GHI and 2.70 MWh/m² and 2.05 MWh/m², respectively, for DNI. This results in a relative mean bias error (MBE) of 8.24 % for GHI and of 31.71 % for DNI.
As expected, the simulations of GHI tend to be better than DNI. This happens due to the increased uncertainty associated with DNI resulting from its higher variability and dependency of clouds and aerosols in the atmosphere which are particularly difficult to correctly simulate.

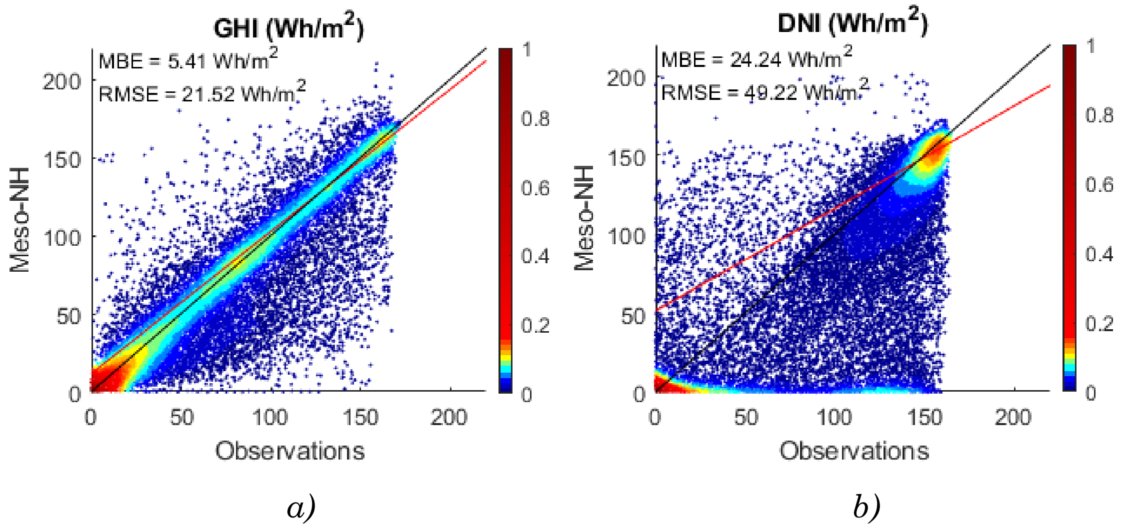


Fig. 2.6 - Comparison between 10-minute simulated (using Meso-NH) and observed a) GHI and b) DNI at the station of Évora – Verney for the TMY (Wh/m²).

Figs. 2.7 and 2.8 show the daily values of simulated and observed GHI and DNI, respectively, in Évora – Verney as well as the daily differences between model and measurements, for the generated TMY.
Once more, the overestimation of these two variables by the NWP model is visible. This overestimation happens not only on cloudy or partially cloudy days, when the model has difficulties in correctly simulating the effects of clouds on solar radiation, but also on clear sky days when the usage of monthly climatologies of aerosols by the NWP model affects the simulation accuracy, especially regarding DNI. This was the main motivation for the development of a correcting ANN model that could improve the simulation results.

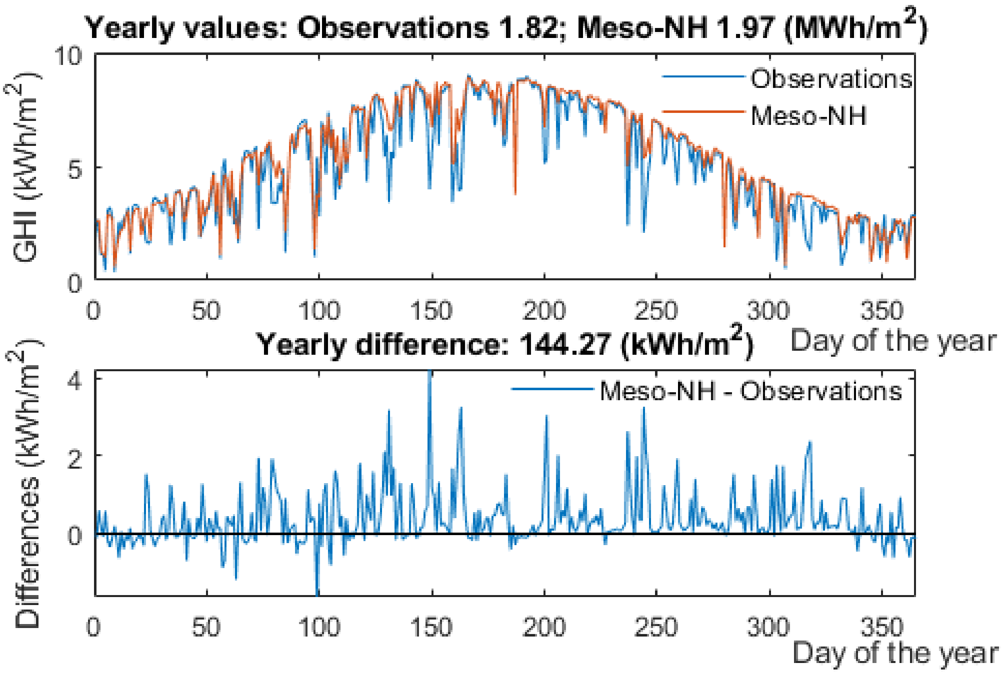


Fig. 2.7 - Daily GHI values simulated using the Meso-NH model and observed at the station of Évora – Verney (top) and respective difference (bottom).

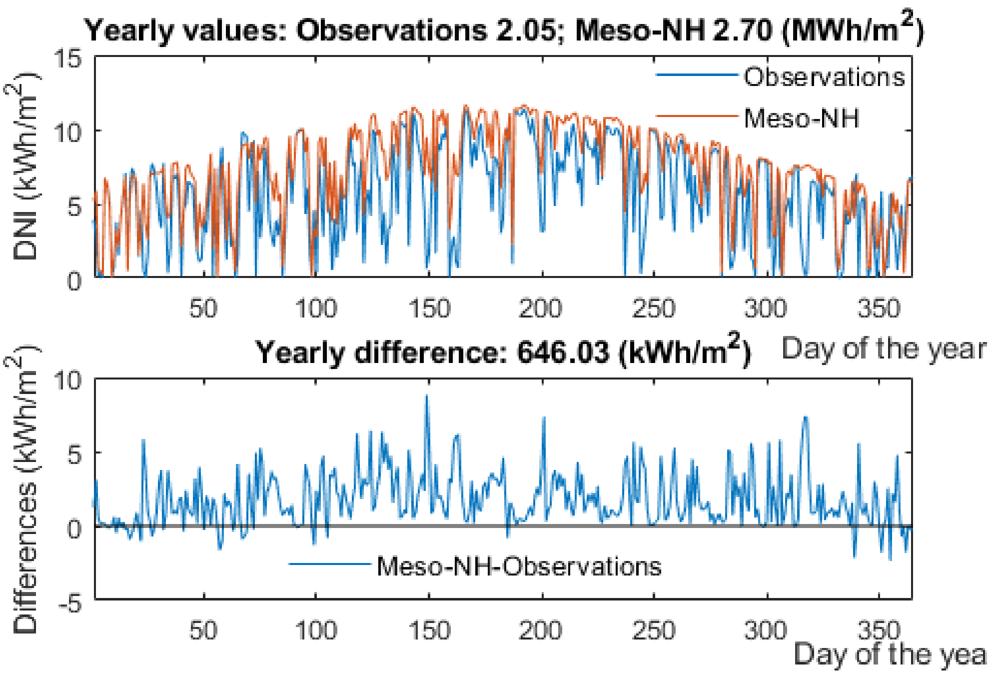


Fig. 2.8 - Daily DNI values simulated using the Meso-NH model and observed at the station of Évora – Verney (top) and respective difference (bottom).

2.4 ANN model development

Two ANN models, one for GHI and another for DNI, were developed using the neural network toolbox of MATLAB for site adaptation of the Meso-NH simulations using various atmospheric related variables as inputs and 10-minute GHI and DNI observed at Évora – Verney station as targets. The developed ANNs are networks that were trained to map the inputs given to the desired targets. There are various parameters and specifications that can define an ANN thus, tests were performed to select the best configuration.

Some of the ANN parameters were pre-defined such as using a *fitnet* (which is a fitting network function), the initialization function *initnw* (which initializes the weights and biases of each layer according to the Nguyen-Widrow initialization algorithm), the hyperbolic tangent sigmoid transfer function *tansig* and the performance function *mse* (mean squared error). The input and output data were treated using *removeconstantrows*, which removes rows with constant values and *mapstd* that rescales the distribution of values so that the mean of each variable is 0 and the standard deviation is 1. The input data was randomly divided in 70% of data for training and 30% for validation, while the random seed was kept constant for comparison between different ANN configurations. The training functions tested were *trainlm* (uses Levenberg-Marquardt backpropagation) and *trainbr* (uses Bayesian regularization backpropagation).

As inputs, the meteorological variables obtained from the Meso-NH simulations were considered, namely GHI, DNI, DHI, the solar zenith angle (Zen), mean air temperature (T), wind speed (WS) and direction (WD) and average (avgCF) and maximum (maxCF) cloud fraction.

To account for the effect of aerosols on solar radiation, aerosol variables obtained from the CAMS analysis data set were also considered to be used as inputs for the ANN models. The CAMS analysis data set is the most recent set of global reanalysis data for atmospheric composition produced by the ECMWF. It has atmospheric composition related variables, including aerosols, in a three-dimensional grid with approximately 80 km of horizontal resolution and 3 h of temporal resolution since 2003. The aerosol model of CAMS is a hybrid bulk-bin scheme with 12 prognostic tracers (three bins for sea salt, three for dust, hydrophilic and hydrophobic organic matter and black carbon, sulphate aerosol and a gas-phase sulfur dioxide precursor) [58]. The total aerosol is corrected by the assimilation of total AOD observations which is done through a 4D-Var data assimilation system [59].

The aerosol related variables retrieved from CAMS were aerosol optical depth at 469 nm (AOD469), 550 nm (AOD550), 670 nm (AOD670), 865 nm (AOD865) and 1240 nm (AOD1240), sea salt aerosol optical depth at 550 nm (SSAOD), organic matter aerosol optical depth at 550 nm (OMAOD) and dust aerosol optical depth at 550 nm (DUAOD). These variables were obtained in a grid with 0.125° of horizontal resolution and temporal resolution of 3 h.

The procedure for correction of simulations was the same for GHI and DNI resulting in the generation of an ANN model for each variable. After obtaining the 10-minute data by performing bi-linear interpolation of the 4 grid-points surrounding the site and matching the variables to the observed values (of GHI or DNI) for the observation site, in this case Évora – Verney, the correlations between the modelled variables and the observations (see Fig. 2.19 in Appendix A) are determined. The Pearson's linear correlation coefficients are used to establish the order in which the different variables are added to the input matrix of the ANN for testing. The Pearson's linear correlation coefficients obtained in this work for each variable of study are presented in Table 2.3.

To find the optimal configuration in terms of training function, inputs, and number of neurons, these parameters are tested by computing the mean squared error (MSE) of the ANN and the original Meso-NH simulations for the variable of interest and applying a relative difference computed through Eq. 2.1. The parameters chosen are the ones that result in the highest value of relative MSE difference (MSE_{rd}), thus resulting in the best improvement (in terms of MSE) with respect to the original simulations.

Table 2.3 - Pearson's linear correlation coefficients of the simulated variables with respect to the GHI and DNI observations in Évora – Verney for the generated TMY.

Simulated Variables	**Observed GHI**	**Observed DNI**
GHI	0.91	0.62
DNI	0.74	0.73
DHI	0.34	0.00
Zen	-0.87	-0.48
T	0.49	0.37
WS	-0.04	-0.09
WD	0.02	-0.11
avgCF	-0.32	-0.43
maxCF	-0.35	-0.52
AOD469	-0.03	-0.23
AOD550	-0.04	-0.24
AOD670	-0.06	-0.26
AOD865	-0.08	-0.27
AOD1240	-0.12	-0.28
OMAOD	-0.06	-0.13
DUAOD	0.03	-0.08
SSAOD	-0.18	-0.29

$$MSE_{rd} = \frac{MSE_{Meso-NH} - MSE_{ANN}}{MSE_{Meso-NH}} \times 100\% \tag{2.1}$$

For each training function (*trainlm*, *trainbr*), the number of neurons in the hidden layer is made to vary between 1 and 100 and for each of these combinations, the variables used as inputs are changed by adding them to the input matrix of the ANN, one by one, in the order defined by the Pearson's linear correlation coefficients. After adding one variable to the inputs, the results of the ANN are evaluated and if the addition of this variable improves the MSE_{rd}, that variable is kept as input and the next is added, if not, the variable is removed from the input matrix and the next variable is added.

In Table 2.4, the ANN configurations that result in the highest values of MSE_{rd} are presented, with two ANN models being produced, one for GHI and the other for DNI. The next best combinations and its results can be consulted in Table 2.6 of Appendix A.

Table 2.4 - Configuration parameters of the ANN models for correction and site adaptation of GHI and DNI simulations from the NWP model.

Variable of interest	MSE_{rd} (%)	Function	Number of neurons	Inputs
GHI	67.60	*trainbr*	100	GHI, Zen, DNI, T, maxCF, DHI, avgCF, SSAOD, AOD1240, AOD865, AOD670, OMAOD, AOD550, WS, AOD469, DUAOD, WD
DNI	82.41	*trainbr*	97	DNI, GHI, maxCF, Zen, avgCF, T, SSAOD, AOD1240, AOD865, AOD670, AOD550, AOD469, OMAOD, WD, WS, DUAOD, DHI

The results of the selected ANN models obtained a MSE_{rd} of 67.60% for GHI and 82.41% for DNI showing that they improve the Meso-NH simulations for the station of Évora – Verney. The higher improvement in DNI simulations when compared to the GHI model is explained by the fact that the NWP model is better at simulating GHI than DNI.

When comparing the 10-minute data obtained through these ANN models with the Meso-NH simulations and observations made in Évora-Verney, as is shown in Fig. 2.9 for the DNI as example, small oscillations in the ANN corrected values are detected which are not in accordance with observations. The solution found was to train 100 ANNs with the parameters defined above but without a fixed random seed (which means initialization values of each ANN model are different and randomized) and use as output of the model the mean of the results of those 100 ANNs (more details on this procedure can be found in Appendix A).

Fig. 2.9 shows the improvement that results from this procedure for the DNI. This procedure was adopted for the two variables being studied.

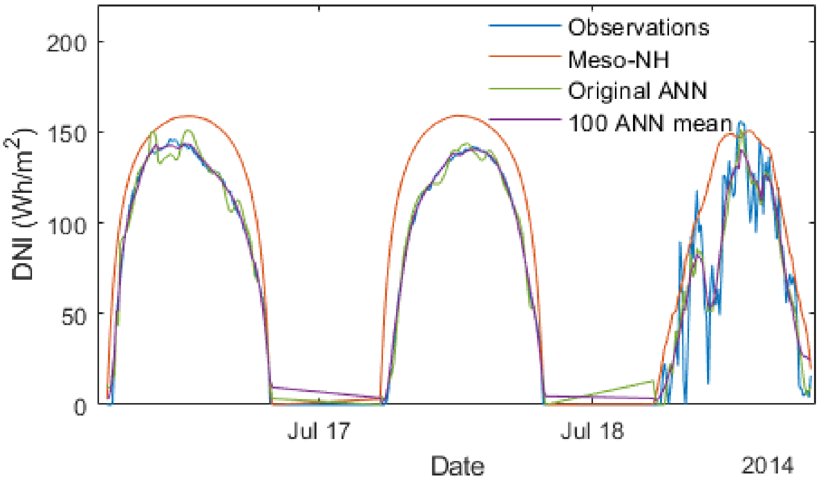
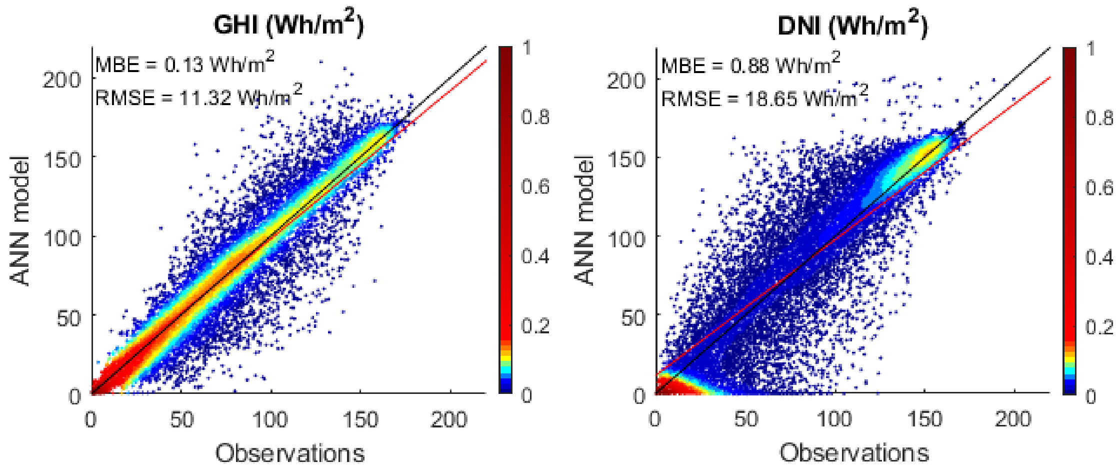


Fig. 2.9 - 10-minute DNI observed, simulated through Meso-NH, corrected using the original ANN and computed through the mean of the results of the trained 100 ANNs for Évora-Verney from 16 to 18 of July 2014.

2.5 Results and discussion

After obtaining the output mean of the 100 ANNs based on the best combination of parameters for each variable (GHI and DNI) for the location of Évora, the developed models showed a MSE_{rd} of 72.36% for GHI and 85.65% for DNI for the TMY showing that the developed method can improve the Meso-NH simulations for the station of Évora – Verney. Moreover, using the mean of 100 ANNs with the same parameters but different initialization values, improves not only the results regarding the oscillation problem, as seen in Fig. 2.10, but also the MSE_{rd} when compared to the use of only one ANN.

Figs. 2.11 and 2.12 show a direct comparison between the obtained values of GHI and DNI, respectively, for the TMY and the station of Évora – Verney and the corresponding observations. The improvement of the ANN on the simulations can be seen when comparing these plots with Figs. 2.6 and 2.7, although some dispersion is still observed. The annual irradiation for the TMY in this location became 1.83 MWh/m^2 for GHI and 2.07 MWh/m^2 for DNI after applying the developed ANN models on the original Meso-NH simulations (1.97 MWh/m^2 for GHI and 2.70 MWh/m^2 for DNI), which is much closer to the observed values (1.82 MWh/m² for GHI and 2.05 MWh/m² for DNI) than the NWP simulations. This represents a reduction in annual relative mean bias error from 8.24 % to 0.55 % for GHI and from 31.71 % to 0.98 % for DNI.



a) *b)*

Fig. 2.10 - Comparison between corrected 10-minute a) GHI and b) DNI simulations using the 100 ANNs and observations for TMY and the station of Évora – Verney (Wh/m²).

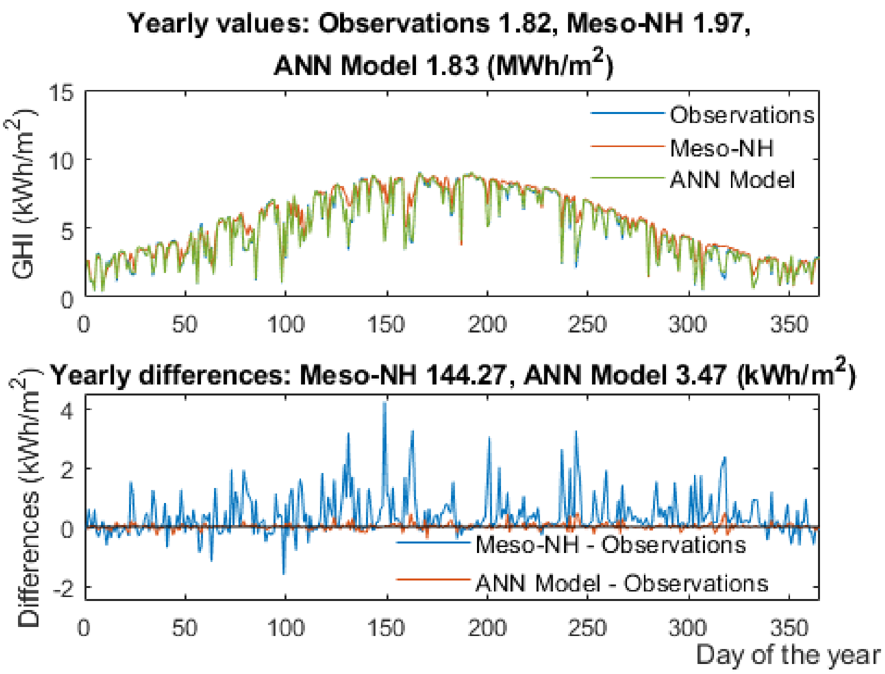


Fig. 2.11 - Daily GHI observed and simulated with Meso-NH and the ANN model (top) and the differences between modelled and observed values (bottom) for the TMY and station of Évora – Verney.

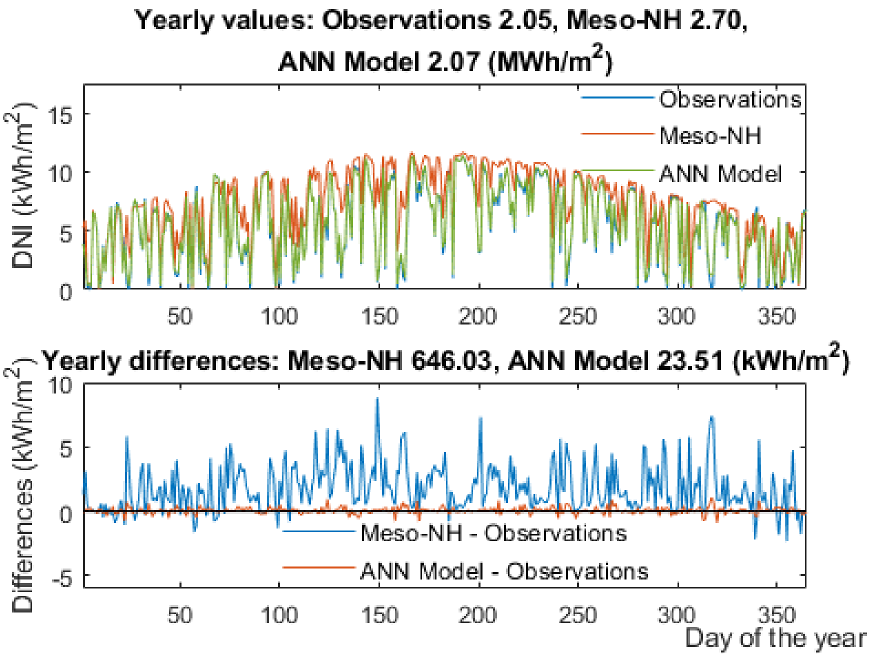


Fig. 2.12 - Daily DNI observed and simulated with Meso-NH and the ANN model (top) and the differences between modelled and observed values (bottom) for the TMY and station of Évora – Verney.

Table 2.5 shows the root mean squared error (RMSE) values for the modelled data for each calendar month and for the TMY. The developed method based on the ANN models showed a significant reduction of annual RMSE from 21.00 Wh/m² for GHI and 49.22 Wh/m² for DNI (obtained with the original Meso-NH simulations) to 12.33 Wh/m² for GHI and 18.65 Wh/m² for DNI.

Figs. 2.11 and 2.12 show the daily values of GHI and DNI, respectively, as simulated by the Meso-NH model, the ANN model and observed, as well as the daily differences between the modelled and observed data, showing the improvement of the simulated daily values when the developed method with the ANN models is applied.

The developed ANN models are then applied to the NWP simulations for the whole region of study and the generated TMY with two main purposes: i) validate the developed method; and ii) to evaluate the usage of the developed method for solar resource assessment and its accuracy.

Table 2.5 - RMSE of 10-minute GHI and DNI simulated by the Meso-NH and ANN models for the generated TMY, in Wh/m².

Month	GHI		DNI	
	Meso-NH	**ANN**	**Meso-NH**	**ANN**
1	14.44	8.32	43.72	20.20
2	17.91	10.54	49.75	21.88
3	24.56	12.63	51.50	20.01
4	25.86	15.20	50.72	21.38
5	30.09	14.96	60.02	20.90
6	23.19	13.07	45.77	16.96
7	20.00	9.84	42.71	15.22
8	16.15	7.95	38.42	13.79
9	22.32	12.51	50.40	19.83
10	19.23	8.98	49.07	16.68
11	20.20	8.46	59.68	19.98
12	12.54	6.35	46.09	16.60
Year	21.52	12.33	49.22	18.65

For the validation of the developed method, Figs. 2.13 and 2.14 were generated where the monthly values of GHI and DNI from the Meso-NH and ANN models were compared against the available monthly values of observed data at other stations from the DNI-A network shown in Fig. 2.1.

The application of the method based on the ANN models resulted in monthly values much closer to observations than the original Meso-NH simulations for all stations. The monthly mean bias errors (MBE) resulting from the use of this method and including the available data from all measuring stations is of 2.67 kWh/m² for GHI and 5.97 kWh/m² for DNI, much lower than the MBEs for the Meso-NH simulations which are of 9.63 kWh/m² for GHI and 43.67 kWh/m² for DNI. This results in a decrease in relative mean bias error from 8.50 % to 2.34 % for GHI and from 29.54 % to 3.41 % for DNI.

The monthly RMSEs of the ANN are approximately 4.26 kWh/m² for GHI and 11.57 kWh/m² for DNI, which are also much lower than those of the Meso-NH (11.49 kWh/m² for GHI and 48.00 kWh/m² for DNI). This means that the use of an ANN model developed for a specific site centered in a limited region of interest, which was trained with typical meteorological data, can still improve

NWP simulations for that region since an ANN model has the capability of capturing and learning the relations between the different input variables and the measured values.

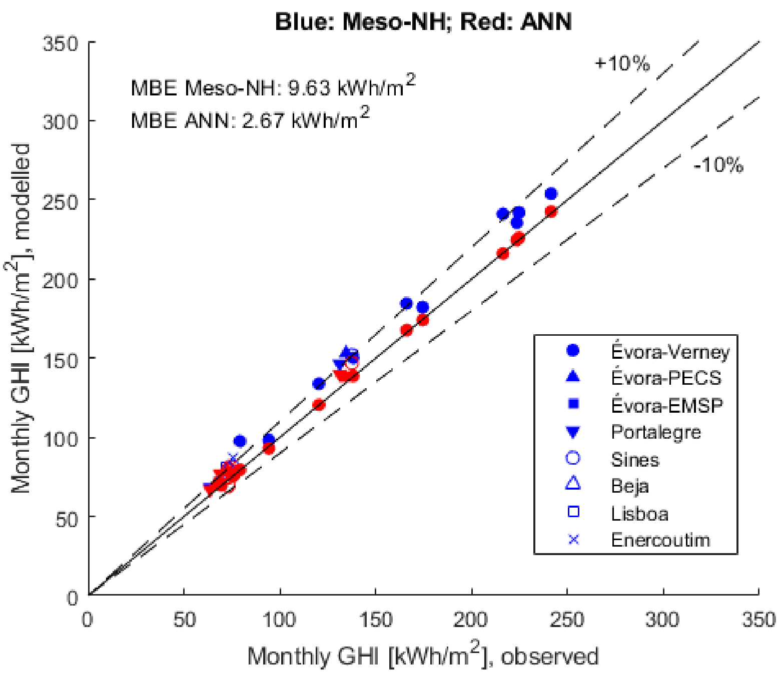


Fig. 2.13 - Comparison between monthly modelled (Meso-NH in blue and ANN in red) and observed GHI values at different locations of the south of Portugal.

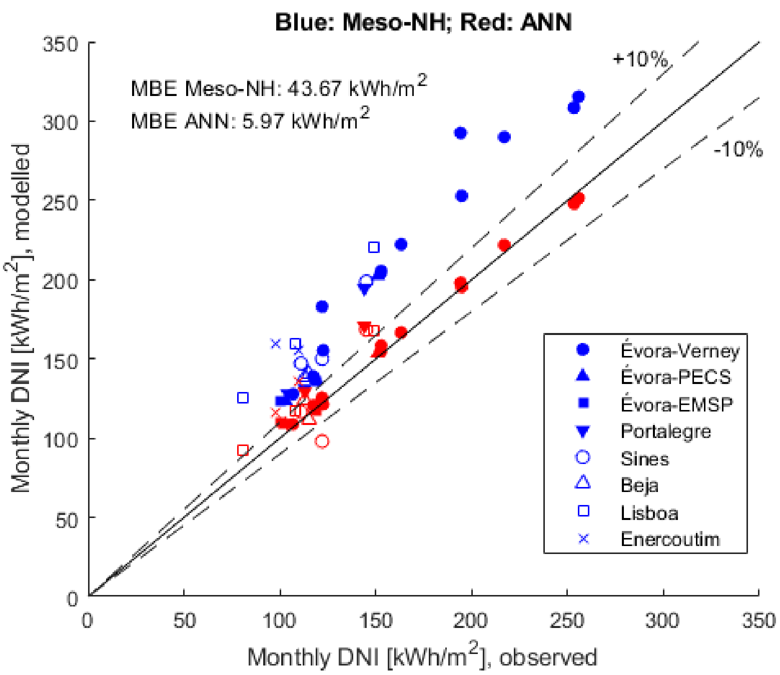


Fig. 2.14 - Comparison between monthly modelled (Meso-NH in blue and ANN in red) and observed DNI values at different locations of the south of Portugal.

Thus, besides approximating the solar irradiation simulated by the Meso-NH model to observations for these various locations and for typical meteorological months in Évora, it was also expected that the developed model would approximate these results to the typical values of each location. To verify that, the simulated GHI was compared with the monthly-mean values obtained from long-term observations, computed using 10 or more years (Cavaco et al, 2016), for two locations in the inland, Portalegre (37.0167, -7.9667) and Beja (38.0249, -7.8672), and other two in the coastal zone, Lisboa (38.7734, -9.1779) and Faro (39.2941, -7.4213), as shown in Fig. 2.15. A very good agreement is observed and, since the goal of a TMY is to reproduce the long-term climatology and the GHI is the meteorological variable with higher statistical weight in the typical meteorological month selection (Section 2.2.2), we may assume that the ANN model can reproduce the typical meteorological data in the entire region based on the NWP simulations.

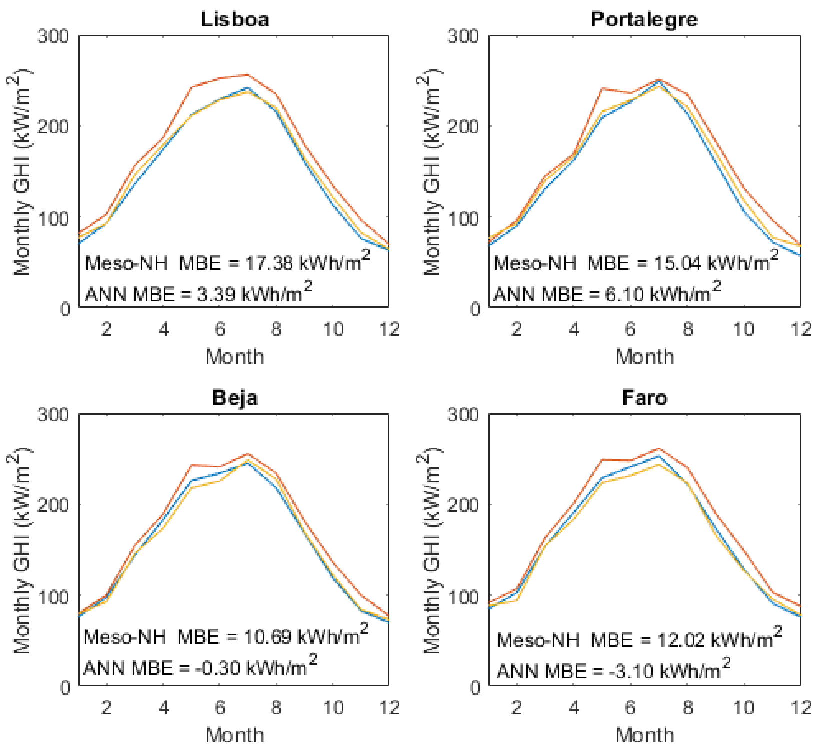


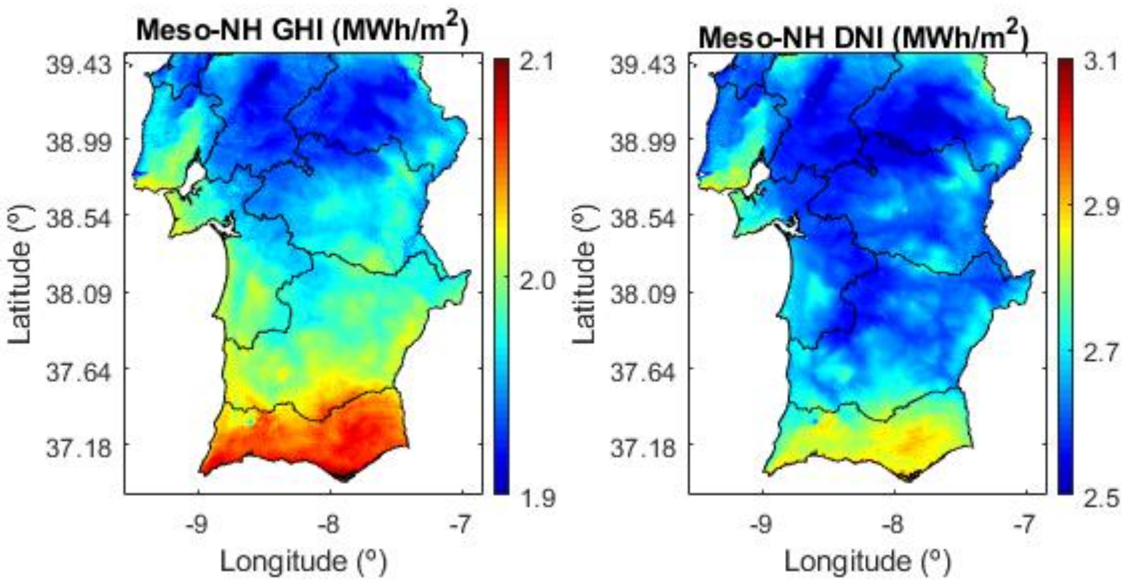
Fig. 2.15 - Comparison between monthly GHI results of the Meso-NH and ANN models for four different locations in the south of Portugal obtained for the TMY of Évora with their long-term monthly means. Blue: Long-term GHI monthly means; Red: Meso-NH monthly GHI for the determined TMY; Yellow: ANN monthly GHI for the determined TMY.

Thus, the developed method successfully results in an accurate assessment of the solar resource by incorporating NWP models, allowing the physical simulation of the weather for a specified region, and ANN models, which work as a site adaptation model and incorporate aerosol data, thus improving the simulations previously done for the TMY data representative of the typical meteorological conditions. With the resulting data, one can generate solar resource maps where the spatial distribution of the resource is easily visible. Fig. 2.16 shows the GHI and DNI resource maps for the south of Portugal from the original Meso-NH simulations for the generated TMY. Fig. 2.17 shows the GHI and DNI resource maps for the same region after applying the method described in this work.

The annual results of both the Meso-NH and the developed method show a positive difference when compared to observations as mentioned before, being the Meso-NH difference much higher than the one of the ANN models. Due to the difference in magnitude of the Meso-NH and ANN simulated variables and in order to be able to observe the distribution of the resource throughout the region, the color scales of these maps are not identical. For a better comparison between the original Meso-NH simulations and the solar resource modelled through the method based on ANNs, Fig. 2.18 was generated showing the differences between them.

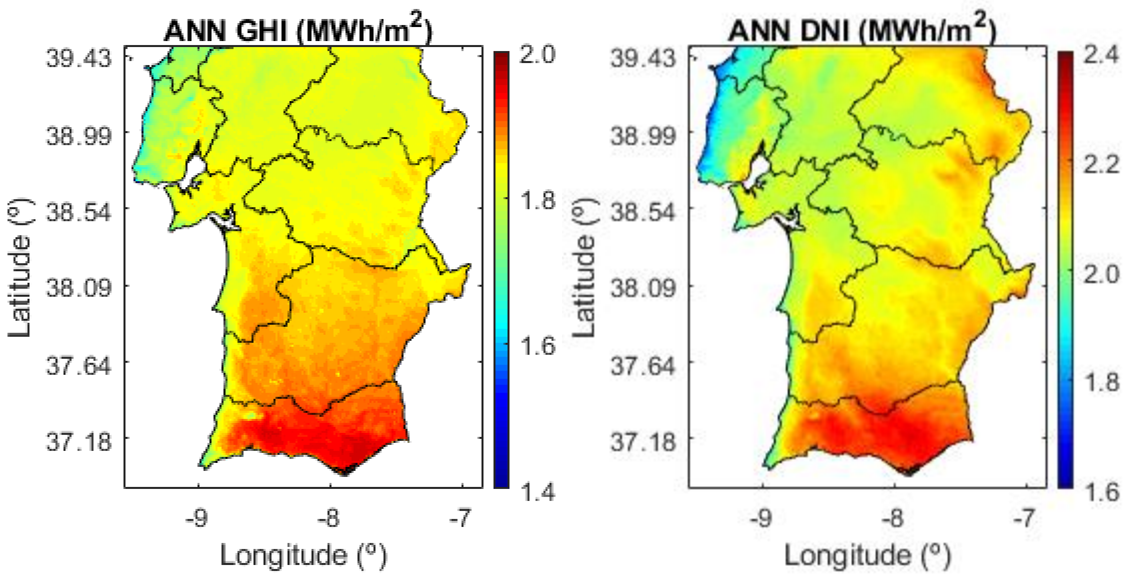
The differences observed are more accentuated for the coastal region of Portugal (mainly the Atlantic one) but are also present at a smaller scale for regions with greater altitudes which can be interpreted as a greater overestimation by the Meso-NH model. The cause of this difference can be explained by the higher frequency of clouds in those regions, which are more difficult to accurately simulate, as well as the usage of monthly aerosol climatologies, more specifically at the coastal region where these climatologies don’t adequately consider the effect of the Atlantic Ocean on the climate of Portugal.

According to the obtained spatial distribution of annual solar irradiation, the highest values of DNI are observed in the south coastal region of Portugal (coastal strip of Algarve) as well as in the Algarve region with highest altitudes. Higher values of DNI are also present close to the east border of Portugal.



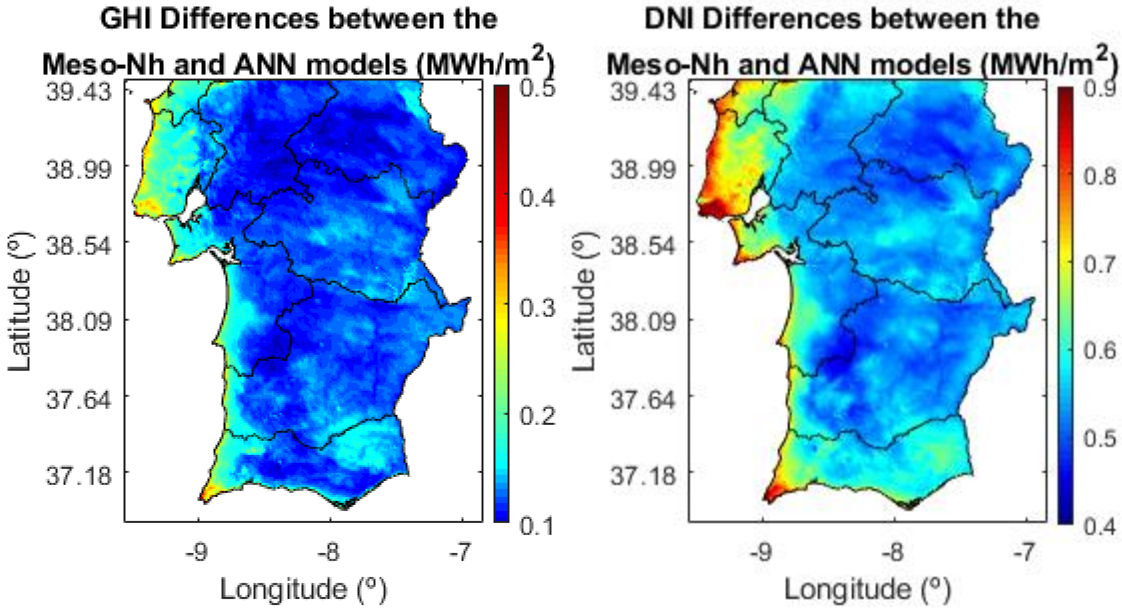
a) *b)*

Fig. 2.16 - Spatial distribution of a) GHI and b) DNI, for the south of Portugal and TMY obtained from Meso-NH simulations.



a) *b)*

Fig. 2.17 - Spatial distribution of a) GHI and b) DNI, for the south of Portugal and TMY obtained from the ANN model.



a) *b)*

Fig. 2.18 - Difference between the original Meso-NH GHI simulations of GHI and the corrected through the ANN model.

2.6 Conclusion

This work presented a method for the assessment of the solar resource with 1.25 km of horizontal resolution obtained by developing a site adaptation model based on artificial neural networks and incorporating aerosol data from CAMS that is applied to simulations made by a numerical weather prediction model for typical meteorological data.

A typical meteorological year was generated for Évora, Portugal, using sixteen years of experimental data, which selected typical meteorological months can also be considered representative of the typical meteorological months of the surrounding region. The weather simulations for this TMY were done using the NWP model Meso-NH coupled with the ecRad radiation scheme which generates atmospheric data for the desired region (south of Portugal) with a resolution of 1.25 km.

The comparison between the NWP simulated and the observed GHI and DNI showed an overestimation of these variables (higher for DNI) that can be caused due to difficulties in correctly simulating clouds and the usage of monthly-mean aerosol climatologies. It was shown that this overestimation is higher in the coastal strip region of Portugal, showing that the NWP model does not correctly account for the effect of maritime aerosols and clouds on the simulated solar radiation in this region.

The developed procedure for site adaptation using an optimized ANN model is based on the testing of different parameters, including the addition of analysis aerosol data from CAMS as inputs, and choosing the variables that result in higher improvements in terms of MSE.

The ANN models are obtained for the specific location used for the generation of the TMY, in this case Évora, and then applied to the whole region simulated by the Meso-NH model (south of Portugal). The monthly values obtained using the Meso-NH model and using the developed procedure were compared against observed values at stations located in this region, showing a reduction in relative mean bias error from 8.50 % to 2.34 % for GHI and from 29.54 % to 3.41 % for DNI. Thus, validating the method proposed by showing that even if the ANN models were designed to improve GHI or DNI simulations at Évora – Verney station, they can improve simulations for other locations in the same region.

This is a general method and can be applied to any region of the world, however, it should be noted that the area simulated in this work is of approximately 46 000 km^2 and the generation of larger maps and/or for locations with very different climates will probably decrease the improvement brought by the application of an ANN model developed for a specific site in that region. The understanding of these implications can be a topic for future research. In the future, this procedure can also be used for other meteorological variables, such as air temperature, which greatly affects the operating efficiency of photovoltaic systems and to generate solar energy potential maps for this specific technology that may consider photovoltaic cells temperature and other loss factors.

The high resolution GHI and DNI maps generated are extremely important for solar radiation related applications like planning and design of solar energy systems and while the usage of an atmospheric model allows for the assessment of the resource in a desired region, these tend to show large errors when solar radiation is concerned. Thus, the usage of post-processing tools like artificial neural networks incorporating aerosol data allows for a more accurate assessment of the solar resource.

Appendix A – Details on the procedure for the development of the ANN model

For the development of the ANN models proposed in this work, firstly the 1-minute Meso-NH atmospheric variable simulations are processed obtaining 10-minute data for the location of Évora, Portugal through bi-linear interpolation of the values of the 4 grid points surrounding the observation site. The data obtained from CAMS for the TMY is also processed in the same way and, additionally, is downscaled from 3-hour data to 10-minute data through piecewise cubic hermite interpolating polynomials which allows for the preservation of the shape of the data. The correlations of these variables with the GHI and DNI observations for the TMY are determined resulting in Fig 2.19.

The simulated variables are ordered according to their Pearson's linear correlation coefficients with the observed variables so they can be added to the input matrix of each ANN model (GHI or DNI) according to what was described in Section 4.

Table 2.6 shows not only the selected ANN models as shown in Table 4 but the five best combinations of parameters for the ANN models developed to improve GHI and DNI simulations for Évora – Verney station. All these models use the training function *trainbr* which is a training function that updates the weight and bias values according to Levenberg-Marquardt optimization and uses Bayesian regularization that minimizes a combination of squared errors and weights, determining the correct combination to produce a network that generalizes well. The best models tend to use higher number of inputs as well as neurons. Again, all models developed for GHI have lower MSE_{rd} than the ones developed for the correction of DNI since the NWP model is already better at simulating GHI.

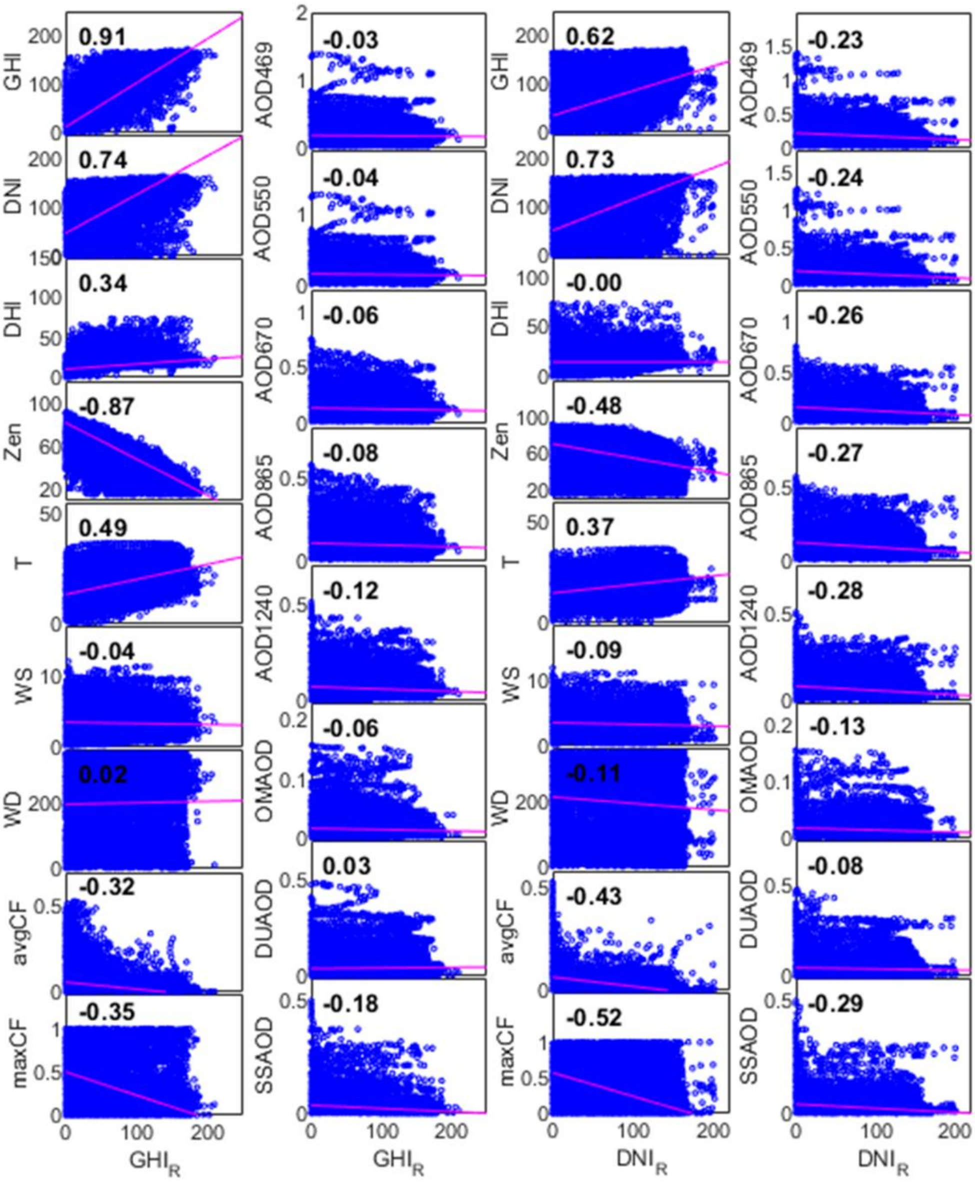


Fig. 2.19 - Correlations and Pearson's linear correlation coefficients between modelled variables and observed global horizontal (GHI_R) and direct normal irradiation (DNI_R). Irradiation variables, GHI, DNI, DHI, GHI_R, DNI_R, in Wh/m^2, temperature, T, in °C, wind speed, WS, in m/s, zenith angle, Zen, and wind direction, WD, in degrees.

Table 2.6 – Five best combinations of ANN parameters for the improvement of GHI and DNI. For all models, the best training function was shown to be trainbr.

Variable of interest		**GHI**					**DNI**				
ANN model rank		**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**
MSE_{rd} (%)		67.60	67.07	66.95	66.92	66.83	82.41	82.34	82.28	82.24	81.89
Number of neurons		100	99	95	97	98	97	99	100	98	96
Inputs	**GHI**	×	×	×	×	×	×	×	×	×	×
	DNI	×	×	×	×	×	×	×	×	×	×
	DHI	×	×	×	×	×	×	×	×	×	×
	Zen	×	×	×	×	×	×	×	×	×	×
	T	×	×	×	×	×	×	×	×	×	×
	WS	×	×	×	×	×	×	×	×	×	×
	WD	×	×	×	×	×	×	×	×	×	×
	avgCF	×	×	×	×	×	×	×	×	×	×
	maxCF	×	×	×	×	×	×	×	×	×	×
	AOD469	×	×	×			×	×			×
	AOD550	×	×	×	×	×	×	×	×	×	×
	AOD670	×	×	×	×	×	×	×	×	×	×
	AOD865	×	×	×	×	×	×	×	×	×	×
	AOD1240	×	×	×	×	×	×	×	×	×	×
	OMAOD	×	×	×	×	×	×	×	×	×	×
	DUAOD	×	×	×	×	×	×	×	×	×	×
	SSAOD	×	×	×	×	×	×	×	×	×	×

To validate the use of 100 ANNs with the same parameters but random initializations as opposed to the use of the ANN trained and defined as the best combination, Fig. 2.20 was generated. This figure shows the MSE of the mean DNI obtained from the results of N trained ANNs, being these ordered from lowest to highest value of MSE, i.e., for N=1, the MSE shown in the figure is the MSE resulting from the ANN with lowest MSE, for N=2, the MSE shown in the figure is the MSE of the mean results of the two ANNs with lowest MSE, and so on. The MSE tends to decrease with the higher number of ANNs used in the average result.

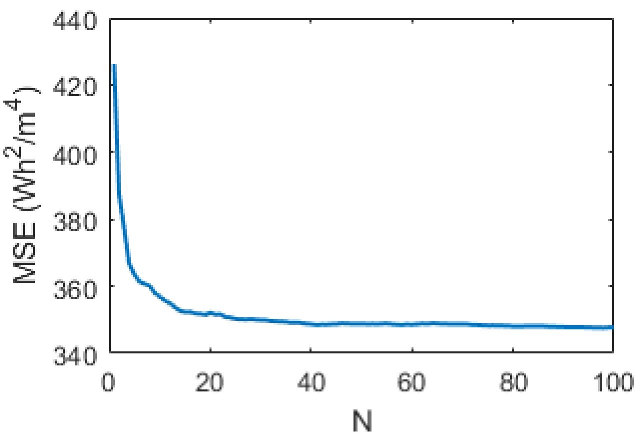


Fig. 2.20 - MSE of the mean DNI of the first N trained ANNs, these being ordered from lowest to highest values of MSE.

References

1. International Renewable Energy Agency (IRENA) *Renewable Energy Capacity Statistics 2021*; 2021; ISBN 9789295111905.
2. Scharmer, K.; Greif, J. *The European Solar Radiation Atlas: Fundamentals and Maps*; The European Solar Radiation Atlas; Les Presses de l'Ecole des Mines, 2000; ISBN 9782911762215.
3. Feng, Y.; Zhang, X.; Jia, Y.; Cui, N.; Hao, W.; Li, H.; Gong, D. High-Resolution Assessment of Solar Radiation and Energy Potential in China. *Energy Convers Manag* **2021**, *240*, 114265, doi:10.1016/J.ENCONMAN.2021.114265.
4. Lhendup, T.; Lhundup, S. Comparison of Methodologies for Generating a Typical Meteorological Year (TMY). *Energy for Sustainable Development* **2007**, *11*, 5–10, doi:10.1016/S0973-0826(08)60571-2.
5. Hall, I.J.; Praire, R.; Anderson, H.; Boes, E. *Generation of Typical Meteorological Years for 26 SOLMET Stations*; Laboratories, S., Ed.; Sandia Laboratories energy report; Sandia Laboratories, 1978;
6. Marion, W.; Urban, K. *User's Manual for TMY2s - Typical Meteorological Years*; United States. Department of Energy. Office of Energy Efficiency and Renewable Energy, 1995;
7. Wilcox, S.; Marion, W. *Users Manual for TMY3 Data Sets*; 2008;

8. Ohunakin, O.S.; Adaramola, M.S.; Oyewola, O.M.; Fagbenle, R.O.; Adelekan, D.S.; Gill, J.; Abam, F.I. Photovoltaic Performance Prediction in Northern Nigeria Using Generated Typical Meteorological Year Dataset. *African Journal of Science, Technology, Innovation and Development* **2018**, *10*, 579–591, doi:10.1080/20421338.2018.1511280.
9. Huld, T.; Paietta, E.; Zangheri, P.; Pascua, I.P. Assembling Typical Meteorological Year Data Sets for Building Energy Performance Using Reanalysis and Satellite-Based Data. *Atmosphere* **2018**, *9*, doi:10.3390/atmos9020053.
10. Al-Azri, N.A.; Al-Saadi, S. Variant Developments of Typical Meteorological Years (TMYs) for Seeb, Oman and Their Impact on Energy Simulation of Residential Buildings. *Journal of Engineering Research* **2018**, *15*, 29–41, doi:10.24200/tjer.vol15iss2p129-141.
11. Ernst, M.; Gooday, J. Methodology for Generating High Time Resolution Typical Meteorological Year Data for Accurate Photovoltaic Energy Yield Modeling. *Solar Energy* **2019**, *189*, 299–306, doi:10.1016/j.solener.2019.07.069.
12. Yang, D.; Gueymard, C.A. Producing High-Quality Solar Resource Maps by Integrating High- and Low-Accuracy Measurements Using Gaussian Processes. *Renewable and Sustainable Energy Reviews* **2019**, *113*, 109260, doi:10.1016/j.rser.2019.109260.
13. Polo, J.; Perez, R. Solar Radiation Modeling from Satellite Imagery. In *Solar Resources Mapping: Fundamentals and Applications*; Polo, J., Martín-Pomares, L., Sanfilippo, A., Eds.; Springer International Publishing: Cham, 2019; pp. 183–197 ISBN 978-3-319-97484-2.
14. Clifton, A.; Hodge, B.M.; Draxl, C.; Badger, J.; Habte, A. Wind and Solar Resource Data Sets. *Wiley Interdiscip Rev Energy Environ* **2018**, *7*, doi:10.1002/wene.276.
15. Jiménez, P.A.; Lee, J.A.; Kosovic, B.; Haupt, S.E. Solar Resource Evaluation with Numerical Weather Prediction Models. In *Solar Resources Mapping: Fundamentals and Applications*; Polo, J., Martín-Pomares, L., Sanfilippo, A.,

Eds.; Springer International Publishing: Cham, 2019; pp. 199–219 ISBN 978-3-319-97484-2.

16. Perdigão, J.; Salgado, R.; Magarreiro, C.; Soares, P.M.M.; Costa, M.J.; Dasari, H.P. An Iberian Climatology of Solar Radiation Obtained from WRF Regional Climate Simulations for 1950–2010 Period. *Atmos Res* **2017**, *198*, 151–162, doi:10.1016/J.ATMOSRES.2017.08.016.
17. Araujo, J.M.S. de Improvement of Coding for Solar Radiation Forecasting in Dili Timor Leste—A WRF Case Study. *Journal of Power and Energy Engineering* **2021**, *09*, 7–20, doi:10.4236/jpee.2021.92002.
18. Soukissian, T.H.; Karathanasi, F.E.; Zaragkas, D.K. Exploiting Offshore Wind and Solar Resources in the Mediterranean Using ERA5 Reanalysis Data. *Energy Convers Manag* **2021**, *237*, 114092, doi:10.1016/J.ENCONMAN.2021.114092.
19. Obiso, V.; Jorba, O. Aerosol-Radiation Interaction in Atmospheric Models: Idealized Sensitivity Study of Simulated Short-Wave Direct Radiative Effects to Particle Microphysical Properties. *J Aerosol Sci* **2018**, *115*, 46–61, doi:10.1016/j.jaerosci.2017.10.004.
20. Zhang, M.; Ma, Y.; Gong, W.; Liu, B.; Shi, Y.; Chen, Z. Aerosol Optical Properties and Radiative Effects: Assessment of Urban Aerosols in Central China Using 10-Year Observations. *Atmos Environ* **2018**, doi:10.1016/J.ATMOSENV.2018.03.040.
21. Bran, S.H.; Jose, S.; Srivastava, R. Investigation of Optical and Radiative Properties of Aerosols during an Intense Dust Storm: A Regional Climate Modeling Approach. *J Atmos Sol Terr Phys* **2018**, *168*, 21–31, doi:10.1016/J.JASTP.2018.01.003.
22. Breitkreuz, H.; Schroedter-Homscheidt, M.; Holzer-Popp, T.; Dech, S. Short-Range Direct and Diffuse Irradiance Forecasts for Solar Energy Applications Based on Aerosol Chemical Transport and Numerical Weather Modeling. *J Appl Meteorol Climatol* **2009**, *48*, 1766–1779, doi:10.1175/2009JAMC2090.1.
23. Polo, J.; Fernández-Peruchena, C.; Salamalikis, V.; Mazorra-Aguiar, L.; Turpin, M.; Martín-Pomares, L.; Kazantzidis, A.; Blanc, P.; Remund, J. Benchmarking on Improvement and Site-Adaptation Techniques for Modeled

Solar Radiation Datasets. *Solar Energy* **2020**, *201*, 469–479, doi:10.1016/J.SOLENER.2020.03.040.

24. Alfadda, A.; Rahman, S.; Pipattanasomporn, M. Solar Irradiance Forecast Using Aerosols Measurements: A Data Driven Approach. *Solar Energy* **2018**, *170*, 924–939, doi:10.1016/J.SOLENER.2018.05.089.
25. Fonseca, J.G.D.S.; Uno, F.; Ohtake, H.; Oozeki, T.; Ogimoto, K. Enhancements in Day-Ahead Forecasts of Solar Irradiation with Machine Learning: A Novel Analysis with the Japanese Mesoscale Model. *J Appl Meteorol Climatol* **2020**, *59*, 1011–1028, doi:10.1175/JAMC-D-19-0240.1.
26. Marzo, A.; Trigo, M.; Alonso-Montesinos, J.; Martínez-Durbán, M.; López, G.; Ferrada, P.; Fuentealba, E.; Cortés, M.; Batlles, F.J. Daily Global Solar Radiation Estimation in Desert Areas Using Daily Extreme Temperatures and Extraterrestrial Radiation. *Renew Energy* **2017**, *113*, 303–311, doi:10.1016/j.renene.2017.01.061.
27. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting : A Review. *Renew Energy* **2017**, *105*, 569–582, doi:10.1016/j.renene.2016.12.095.
28. Pereira, S.; Canhoto, P.; Salgado, R.; Costa, M.J. Development of an ANN Based Corrective Algorithm of the Operational ECMWF Global Horizontal Irradiation Forecasts. *Solar Energy* **2019**, *185*, 387–405, doi:10.1016/j.solener.2019.04.070.
29. Liu, L.; Zhao, Y.; Chang, D.; Xie, J.; Ma, Z.; Sun, Q.; Yin, H.; Wennersten, R. Prediction of Short-Term PV Power Output and Uncertainty Analysis. *Appl Energy* **2018**, *228*, 700–711, doi:10.1016/j.apenergy.2018.06.112.
30. Chu, Y.; Urquhart, B.; Gohari, S.M.I.; Pedro, H.T.C.; Kleissl, J.; Coimbra, C.F.M. Short-Term Reforecasting of Power Output from a 48 MWe Solar PV Plant. *Solar Energy* **2015**, *112*, 68–77, doi:10.1016/j.solener.2014.11.017.
31. Scolari, E.; Sossan, F.; Paolone, M. Photovoltaic Model-Based Solar Irradiance Estimators: Performance Comparison and Application to Maximum Power Forecasting. *IEEE Trans Sustain Energy* **2017**, 1–10, doi:10.1109/TSTE.2017.2714690.

32. Gigoni, L.; Betti, A.; Crisostomi, E.; Franco, A.; Tucci, M.; Bizzarri, F.; Mucci, D. Day-Ahead Hourly Forecasting of Power Generation from Photovoltaic Plants. *IEEE Trans Sustain Energy* **2018**, *9*, 831–842, doi:10.1109/TSTE.2017.2762435.
33. Anderson, W.W.; Yakimenko, O.A. Using Neural Networks to Model and Forecast Solar PV Power Generation at Isle of Eigg. In Proceedings of the 2018 IEEE 12th International Conference on Compatibility, Power Electronics and Power Engineering (CPE-POWERENG 2018); IEEE, April 2018; pp. 1–8.
34. Hachicha, A.A.; Rodríguez, I.; Capdevila, R.; Oliva, A. Heat Transfer Analysis and Numerical Simulation of a Parabolic Trough Solar Collector. *Appl Energy* **2013**, *111*, 581–592, doi:10.1016/j.apenergy.2013.04.067.
35. Son, J.; Park, Y.; Lee, J.; Kim, H. Sensorless PV Power Forecasting in Grid-Connected Buildings through Deep Learning. *Sensors* **2018**, *18*, doi:10.3390/s18082529.
36. Rosiek, S.; Alonso-Montesinos, J.; Batlles, F.J. Online 3-h Forecasting of the Power Output from a BIPV System Using Satellite Observations and ANN. *International Journal of Electrical Power and Energy Systems* **2018**, *99*, 261–272, doi:10.1016/j.ijepes.2018.01.025.
37. Fleetwood, A.; Galarza, A. Predicting Solar Energy Generation through Arti Fi Cial Neural Networks Using Weather Forecasts for Microgrid Control. **2018**, *126*, 855–864, doi:10.1016/j.renene.2018.03.070.
38. Ghimire, S.; Deo, R.C.; Downs, N.J.; Raj, N. Global Solar Radiation Prediction by ANN Integrated with European Centre for Medium Range Weather Forecast Fields in Solar Rich Cities of Queensland Australia. *J Clean Prod* **2019**, *216*, 288–310, doi:10.1016/J.JCLEPRO.2019.01.158.
39. Barreto, G.; Canhoto, P. Estudo e Avaliação Do Recurso Eólico Na Região de Évora a Partir de Dados de Uma Estação Meteorológica. In Proceedings of the III Congresso Jovens Investigadores em Geociências LEG 2013 & 6th Meeting of Post-Graduation in Physics and Earth Sciences of UÉ; 2013.
40. Canhoto, P. *Preliminary Analysis of Irradiance Data for the Generation of a Typical Solar Radiation Year for Évora, Portugal*; 2016;

41. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2.0 Available online: https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 28 May 2024).
42. Abreu, E.F.M.; Canhoto, P.; Prior, V.; Melicio, R. Solar Resource Assessment through Long-Term Statistical Analysis and Typical Data Generation with Different Time Resolutions Using GHI Measurements. *Renew Energy* **2018**, *127*, 398–411, doi:10.1016/J.RENENE.2018.04.068.
43. Cavaco, A.; Canhoto, P.; Collares Pereira, M. Procedures for Solar Radiation Data Gathering and Processing and Their Application to DNI Assessment in Southern Portugal. *Renew Energy* **2021**, *163*, 2208–2219, doi:10.1016/J.RENENE.2020.10.075.
44. Finkelstein, J.M.; Schafer, R.E. Improved Goodness-of-Fit Tests. *Biometrika* **1971**, *58*, 641–645, doi:doi.org/10.1093/biomet/58.3.641.
45. Pissimanis, D.; Karras, G.; Notaridou, V.; Gavra, K. The Generation of a "Typical Meteorological Year" for the City of Athens. *Solar Energy* **1988**, *40*, 405–411, doi:10.1016/0038-092X(88)90095-3.
46. Lac, C.; Chaboureau, J.-P.; Masson, V.; Pinty, J.-P.; Tulet, P.; Escobar, J.; Leriche, M.; Barthe, C.; Aouizerats, B.; Augros, C.; et al. Overview of the Meso-NH Model Version 5.4 and Its Applications. *Geosci Model Dev* **2018**, *11*, 1929–1969, doi:10.5194/gmd-11-1929-2018.
47. Meso-NH The Mesonh User's Guide. **2016**.
48. Masson, V.; Le Moigne, P.; Martin, E.; Faroux, S.; Alias, A.; Alkama, R.; Belamari, S.; Barbu, A.; Boone, A.; Bouyssel, F.; et al. The SURFEXv7.2 Land and Ocean Surface Platform for Coupled or Offline Simulation of Earth Surface Variables and Fluxes. *Geosci Model Dev* **2013**, *6*, 929–960, doi:10.5194/gmd-6-929-2013.
49. Faroux, S.; Kaptué Tchuenté, A.T.; Roujean, J.-L.; Masson, V.; Martin, E.; Le Moigne, P. ECOCLIMAP-II/Europe: A Twofold Database of Ecosystems and Surface Parameters at 1 Km Resolution Based on Satellite Information for Use in Land Surface, Meteorological and Climate Models. *Geosci Model Dev* **2013**, *6*, 563–582, doi:10.5194/gmd-6-563-2013.

50. Capecchi, V. Reforecasting Two Heavy-Precipitation Events with Three Convection-Permitting Ensembles. *Weather Forecast* **2021**, *36*, 769–790, doi:10.1175/WAF-D-20-0130.1.
51. Costa, M.J.; Salgado, R.; Santos, D.; Levizzani, V.; Bortoli, D.; Silva, A.M.; Pinto, P. Modeling of Orographic Precipitation over Iberia: A Springtime Case Study. *Advances in Geosciences* **2010**, *25*, 103–110, doi:10.5194/adgeo-25-103-2010.
52. Couto, F.; Cardoso, E.; Costa, M.J.; Salgado, R.; Guerrero-Rascado, J.; Salgueiro, V. How a Mesoscale Cyclonic Vortex over Sahara Leads to a Dust Outbreak in South-Western Iberia. *Atmos Res* **2021**, *249*, 105302, doi:10.1016/j.atmosres.2020.105302.
53. Policarpo, C.; Salgado, R.; Costa, M.J. Numerical Simulations of Fog Events in Southern Portugal. *Advances in Meteorology* **2017**, *2017*, doi:10.1155/2017/1276784.
54. Santos, D.; Costa, M.J.; Silva, A.M.; Salgado, R. Modeling Saharan Desert Dust Radiative Effects on Clouds. *Atmos Res* **2013**, *127*, 178–194, doi:10.1016/j.atmosres.2012.09.024.
55. Hogan, R.J.; Bozzo, A. ECRAD: A New Radiation Scheme for the IFS. *ECMWF TechnicalMemorandum* **2016**, *787*, 1–33.
56. Hogan, R.J.; Bozzo, A. A Flexible and Efficient Radiation Scheme for the ECMWF Model. *J Adv Model Earth Syst* **2018**, *10*, 1990–2008, doi:10.1029/2018MS001364.
57. Iakunin, M.; Abreu, E.F.M.; Canhoto, P.; Pereira, S.; Salgado, R. Impact of a Large Artificial Lake on Regional Climate: A Typical Meteorological Year Meso-NH Simulation Results. *International Journal of Climatology* **2021**, 1–22, doi:10.1002/joc.7299.
58. Morcrette, J.-J.; Boucher, O.; Jones, L.; Salmond, D.; Bechtold, P.; Beljaars, A.; Benedetti, A.; Bonet, A.; Kaiser, J.W.; Razinger, M.; et al. Aerosol Analysis and Forecast in the European Centre for Medium-Range Weather Forecasts Integrated Forecast System: Forward Modeling. *Journal of Geophysical Research: Atmospheres* **2009**, *114*, doi:10.1029/2008JD011235.

59. Courtier, P.; Thépaut, J.-N.; Hollingsworth, A. A Strategy for Operational Implementation of 4D-Var, Using an Incremental Approach. *Quarterly Journal of the Royal Meteorological Society* **1994**, *120*, 1367–1387, doi:10.1002/qj.49712051912.

Nomenclature

ALBD	Direct albedo (–)
ALBS	Scattered albedo (–)
AOD469	Aerosol optical depth at 469 nm (–)
AOD550	Aerosol optical depth at 550 nm (–)
AOD670	Aerosol optical depth at 670 nm (–)
AOD865	Aerosol optical depth at 865 nm (–)
AOD1240	Aerosol optical depth at 1240 nm (–)
avgCF	Vertical average cloud fraction (–)
DHI	Diffuse horizontal irradiation (Wh/m^2)
DNI	Direct normal irradiation (Wh/m^2)
DUAOD	Dust aerosol optical depth at 550 nm (–)
GHI	Global horizontal irradiation (Wh/m^2)
maxCF	Vertical maximum cloud fraction (–)
MSE_{rd}	Relative mean squared error difference (%)
OMAOD	Organic matter aerosol optical depth at 550 nm (–)
SSAOD	Sea salt aerosol optical depth at 550 nm (–)
T	Air temperature (℃)
WD	Wind direction (°)
WS	Wind speed (m/s)

Acronyms

ANN	Artificial Neural Network
CAMS	Copernicus Atmosphere Monitoring Service
CDF	Cumulative Distribution Function
ECMWF	European Centre for Medium-range Weather Forecasts

FS	Finkelstein-Schaffer
IFS	Integrated Forecasting System
NWP	Numerical Weather Prediction
MBE	Mean Bias Error
MSE	Mean Squared Error
RMSE	Root Mean Squared Error
RRTM	Rapid Radiation Transfer Model
TMM	Typical Meteorological Month
TMY	Typical Meteorological Year

Chapter 3

Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data†

Abstract

Accurate operational solar irradiance forecasts are crucial for better decision making by solar energy system operators due to the variability of resource and energy demand. Although numerical weather prediction (NWP) models can forecast solar radiation variables, they often have significant errors, particularly in the direct normal irradiance (DNI), which is especially affected by the type and concentration of aerosols and clouds. This paper presents a method based on artificial neural networks (ANN) for generating operational DNI forecasts using weather and aerosol forecasts from the European Center for Medium-range Weather Forecasts (ECMWF) and the Copernicus Atmospheric Monitoring Service (CAMS), respectively. Two ANN models were designed: one uses as input the predicted weather and aerosol variables for a given instant, while the other uses a period of the improved DNI forecasts before the forecasted instant. The models were developed using observations for the location of Évora, Portugal, resulting in 10 min DNI forecasts that for day 1 of forecast horizon showed an improvement over the downscaled original forecasts regarding R^2, MAE and RMSE of 0.0646, 21.1

---

†Sara Pereira[a], Paulo Canhoto[a,b] and Rui Salgado[a,c] Development and assessment of artificial neural network models for direct normal solar irradiance forecasting using operational numerical weather prediction data, Energy and AI, Volume 15, 2024, 100314, ISSN 2666-5468, https://doi.org/10.1016/j.egyai.2023.100314.

[a] Institute of Earth Sciences, University of Évora, Évora, Portugal
[b] Department of Mechatronics Engineering, University of Évora, Évora, Portugal
[c] Physics Department, University of Évora, Évora, Portugal

W/m^2 and 27.9 W/m^2, respectively. The model was also evaluated for different timesteps and locations in southern Portugal, providing good agreement with experimental data.

Keywords
Solar radiation; Solar energy; Numerical weather prediction; Artificial neural network; Operational forecasting

3.1 Introduction

Economic growth and an increase in energy demand are usually co-dependent, yet they can be partially decoupled by improving energy efficiency, electrification and wise use of energy. Energy generation, distribution and consumption are becoming more electrified, efficient, interconnected and clean with the increase of the use of renewable energy sources such as solar photovoltaic and wind power and the increasing production and use of electric vehicles [1]. However, renewable resources show a strong spatial and temporal variability that affect the power generation, which makes finding an optimum balance between electric generation and consumption at any moment challenging since direct and reliable large scale storage systems of electric energy are not available.

The solar photovoltaic energy, for example, is mainly dependent on the solar irradiance, namely direct beam irradiance (usually measured in the normal plane, known as Direct Normal Irradiance, DNI) and diffuse irradiance (usually measured in the horizontal plane, known as Diffuse Horizontal Irradiance, DHI), which means that having accurate forecasts of these variables allows for a more accurate estimation of electric power generation. Forecasting solar irradiance can be done through various models and methods, from the physical based methods such as numerical weather prediction (NWP) models, models based on sky/shadow imagery or satellite imagery to the more data-driven and machine-learning based methods such as regression, artificial neural networks (ANN), support vector machines

(SVM) and Kalman filtering, or even a combination of these, which are usually known as hybrid methods.

NWP models provide the evolution of the atmosphere by integrating constitutive and state equations describing physical phenomena in the atmosphere which are subject to boundary and initial conditions. Its main goal is weather forecast with a focus on meteorologic variables such as air temperature, humidity, wind or precipitation. An accurate prediction of the partition between the two components of solar global radiation is not so relevant in that case, i.e., DNI and DHI, with global horizontal irradiance (GHI) being used for the closure of the energy balance at the Earth surface. The Integrated Forecasting System (IFS) developed at the European Center for Medium-range Weather Forecast (ECMWF) is the most widely used global NWP model in Europe being its performance attested by various studies such as in [2], where 24-hour forecasts of global solar irradiation from IFS and the American Global Forecasting System (GFS) were compared with observations made at four stations in Morocco. The authors show that the IFS/ECMWF model performs better than the GFS model for all-sky conditions based on the mean bias error and correlation coefficient. In [3], hourly GHI forecasts made by the IFS/ECMWF global model and GFS-driven Weather Research Forecasting (WRF) mesoscale model (run with different configurations by various forecast providers) were compared with observations made in the US, Canada and Europe showing that the model from the ECMWF performs significantly better for all locations and different climatic conditions. Perdigão et al. [4] analyzed one year of hourly and daily direct normal irradiation forecasts of the ECMWF against observations made at Évora, Portugal, for different temporal forecast horizons (0 to 3 days ahead) showing that the model reproduces hourly and daily experimental values with a RMSE of 210.6 W/m^2 and 68.5 W/m^2, respectively, for the first day ahead and that the performance of the model tends to decrease with higher forecast horizon.

The research on improving accuracy of solar radiation variables in NWP models' output has been recently brought to light due to the need for development of solar energy systems [5,6]. DNI is especially difficult to forecast due to its strong dependency on the presence of clouds and aerosol type and concentration in the atmosphere. In the case of aerosols, NWP models such as the IFS/ECMWF use monthly-mean aerosol climatologies instead of more detailed aerosol forecasts to reduce computation time. In the comprehensive review made by Yang et al. [7], the impact of aerosols in the solar resource is discussed. The importance of including aerosols in the NWP forecasts is mentioned where the global models Copernicus Atmospheric Monitoring Service (CAMS) and Goddard Earth Observing System Version 5 (GEOS-5) are shown to be of particular interest. Breitkreuz et al. [8] used libRadtran [9] and achieved a decrease in the relative mean squared error (rMSE) of 4.3% in the case of hourly GHI from IFS/ECMWF forecasts using aerosol predictions and experimental data. Other studies on aerosol modeling and prediction and its use with NWP models can be found in [10], here the WRF model coupled with Chemistry (WRF-Chem) was used to model aerosol and radiation data which were compared against observations showing 2 to 5 times higher shortwave radiative forcing during a dust storm relative to values during non-dust days, and in [11], where, using the RRTM_SW (Rapid Radiation Transfer Model for Shortwave radiation) widely used in atmospheric models, the propagation of small uncertainties on the aerosol microphysical parameterization on the simulated direct radiative effects is demonstrated.

The knowledge of all the complex phenomena that occur in the atmosphere of the Earth, including the interaction between solar radiation and the atmosphere, and between these and the surface, is still challenging and so is its representation, even in the most complex and detailed NWP models, which are also constrained by data availability [12] and computational limitations. A trade-off between increasing complexity of deterministic physical models, with the consequent higher computational effort, and the development of additional tools that, based on actual physical models output and experimental data, allow for better solar radiation forecasts without such effort, must be considered. Thus, several techniques have been developed and evaluated to further improve solar radiation estimations made by deterministic NWP models. These are divided into classic statistical methods and Machine Learning (ML) methods. Classic statistical methods can be as simple as interpolation to improve temporal and/or spatial resolution or more complex methods based on stepwise linear regression to select the variables that best represent the errors, so they can be incorporated in a multi-variate regression model that gives an estimate of the forecast error as a linear function of the variables that have been selected in the process [13]. ML methods have been reported extensively in the literature for solar forecasts [14], such as k-nearest neighbor, SVMs, random forest (RF) or ANNs, and tend to focus on the prediction of solar global irradiation. Alkhayat et al. [15] developed a novel deep learning-based auto-selective tool that allows for the determination of the best GHI forecasting model from four different ML approaches with 81% accuracy. These models can capture the relation between systematic errors of the outputs of NWP models and relevant variables by comparing historical databases of forecasts and observations and thus provide improved values in a fraction of the time it would take with more detailed physical model approaches [16]. In the work by Alfadda et al. [17], hourly GHI, DNI and DHI forecasts made by various machine learning models and using as inputs solar irradiance and aerosol observations and wind and aerosol forecasts in Saudi Arabia were evaluated resulting in a better performance of the ANN model. Pang et al. [18], compared the use of feedforward artificial neural networks with recurrent neural networks for global solar irradiance prediction using onsite measurements showing that the later moderately improved the prediction performance but with additional computational cost. For a short-range forecast of solar irradiance (15 to 180 min), McCandless et al. [19] developed a regime-dependent artificial neural network forecasting model that showed improvements over a global ANN and the persistence. Fonseca et al. [20] obtained day-ahead solar irradiation forecasts through SVMs using NWP data from the mesoscale model of the Japan Meteorological Agency as input, which resulted in a reduction of the root mean squared error (RMSE) of 16 % in a regional scale. Lima et al. [21] used artificial neural networks for the post-processing of NWP solar irradiation forecasts from the

WRF model for the Brazilian Northeastern region and obtained a reduction of the model bias, MSE and RMSE. The same approach can be used to improve solar resource assessment since, once machine learning models were trained, they can generate more accurate spatial distribution of solar irradiation in a given location or region based on historical data of NWP forecasts. For example, Pereira et al. [22] developed and optimized an ANN model in order to create a method for solar resource assessment from the meso-scale Meso-NH NWP model outputs for a typical meteorological year. In this case, the method was extended and validated for the South of Portugal, showing important improvements regarding the direct normal and global horizontal irradiation mapping with a horizontal resolution of 1.25 km. Comprehensive reviews on machine learning methods for solar irradiance forecasting can be found in [23–29]. Although various ML models have been developed for solar irradiation forecasting, most of them focus only on GHI or cannot be readily used to obtain useful power output forecasts of solar energy systems since the temporal resolution and forecast time horizon are not synchronized with the real-time market forecasting requirements [30,31].

This work proposes a method for operational DNI forecasts which includes: i) the spatial and temporal downscaling of the IFS/ECMWF and aerosol forecasts for a specific location and time step; ii) an ANN model for generating improved DNI forecasts using only the predicted weather and aerosol data for a given instant of a specific time step; iii) a second ANN model that takes as inputs a period of improved DNI forecasts immediately before the time step being forecasted as well as the season and time of day of that time step. The model was developed considering the location of Évora, Portugal (38.567811, -7.911459), a temporal resolution of 10 min and a forecast horizon of 25 to 48 h (day 1 of forecast). After development, evaluation against observations show that this method can be generalized and applied to different locations, temporal resolutions and forecast time horizons up to 72 h.

The paper is organized as follows: in Section 3.2 the different data sets used in this work are presented and the treatment of this data is explained including the methods for spatial and temporal downscaling of the forecast data; in Section 3.3 an evaluation of the downscaled DNI forecasts made by the ECMWF is performed; Section 3.4 presents the development of the ANN models used for the improvement of the DNI forecasts obtained from the ECMWF; in Section 3.5 is performed an evaluation of the models for different time steps while in Section 3.6 this is done for different sites located in the region surrounding the one used for model development; Section 3.7 presents an analysis of the ANN models in an operational forecast setting and finally, in Section 3.8, the conclusions of this work are presented.

3.2. Forecast and experimental data

The method was developed using forecast data from the IFS/ECMWF and CAMS models and observed solar radiation data from a network of measuring stations scattered in the South of Portugal. These models are operational and issued everyday which means that the developed model can also be used operationally. The period and area of the data used was from December 2016 to the end of May 2021 and between latitudes 37.0° and 39.3° and longitudes -7.4° and -9.2°, respectively. A description of the data, quality check, filtering and pre-processing is presented in the following subsections.

3.2.1 Weather forecast data

In this work, various weather forecast variables from the IFS/ECMWF model [32] were used. This NWP model includes the radiative scheme ecRad [33] which solves the 1D radiative transfer equation both for small and long wavelength ranges, considering the vertical profiles of temperature, moisture, cloud droplet and ice cloud effective radius, average monthly climatologies of aerosols, carbon dioxide, ozone and trace gases, and also ground surface temperature, albedo and emissivity for different spectral bands and solar zenith angle. The code is based on the RRTM model (Rapid Radiative Transfer Model) using the McICA method (Monte Carlo Independent Column Approximation), which allows for the parameterization of the interactions between radiation and clouds. The ECMWF versions RRTMLW and

RRTMSW describe the radiative transfer in the long wavelength range, with 16 spectral bands, and in the short wavelength range, with 14 spectral bands, respectively [33].

The operational deterministic model of the ECMWF is run every day at 00UTC and 12UTC, providing hourly forecast values up to 90 h ahead (and then 3 and 6-hourly values up to 144 and 240 h ahead, respectively) at discrete points of a global grid with horizontal spatial resolution of 0.125° × 0.125°. The variables retrieved from the ECMWF database are presented in Table 3.1.

Table 3.1 - Forecast variables retrieved from the IFS/ECMWF database.

Variable	**Symbol**	**Units**	**Range / Comments**
Longitude	*Lon*	° East	0° to 360°
Latitude	*Lat*	° North	-90° to 90°
Time step	*Step*	h	1 to 240 h
Date	*Date*	Days	Days since 1900-01-01 00:00:00
Low cloud cover	*LCC*	0-1	–
Medium cloud cover	*MCC*	0-1	–
High cloud cover	*HCC*	0-1	–
Total cloud cover	*TCC*	0-1	–
10 metre *U* wind component	*u10*	m/s	–
10 metre *V* wind component	*v10*	m/s	–
2 metre temperature	*T*	K	Air temperature at 2 meters
Solar zenith angle	*Zen*	°	–
Surface solar radiation downwards	*SSRD*	J/m^2	Irradiation since forecast issuance
Direct solar radiation	*DSRP*	J/m^2	Irradiation since forecast issuance

Data were retrieved for the maximum grid point density and with a temporal forecast horizon up to 72 h. The retrieved GRIB files were converted to netCDF format and processed with a MATLAB® routine to obtain hourly mean values of global horizontal and direct normal irradiance in W/m^2, converting air temperature (T) to °C and computing wind speed (WS) in m/s and direction (WD) in degrees from North using the 10-meter U and V wind speed components.

3.2.2 Aerosol forecast data

CAMS developed a global atmospheric composition forecast based on the IFS model but with additional modules enabled for aerosols, reactive gases and greenhouse gases taking into consideration phenomena such as the emissions and transport of trace gases and aerosols, uptake and release by vegetation and land and sea surface, removal by dry deposition at the surface and scavenging in precipitation, chemical conversion and aerosol microphysics. It generates atmospheric composition variables, including aerosol optical depth at different wavelengths, in a three-dimensional grid with approximately 40 km of horizontal spatial resolution and 1 h time step [34].

Table 3.2 - Forecast variables retrieved from the CAMS database.

Variable	**Symbol**	**Units**	**Range / Comments**
Longitude	*Lon*	° East	0° to 360°
Latitude	*Lat*	° North	-90° to 90°
Time step	*Step*	h	1 to 120
Date	*Date*	Days	Days since 1900-01-01 00:00:00
Total aerosol optical depth at 469 nm	*AOD469*	–	–
Total aerosol optical depth at 550 nm	*AOD550*	–	–
Total aerosol optical depth at 670 nm	*AOD670*	–	–
Total aerosol optical depth at 865 nm	*AOD865*	–	–
Total aerosol optical depth at 1240 nm	*AOD1240*	–	–
Sea salt aerosol optical depth at 550nm	*SSAOD*	–	–
Organic matter aerosol optical depth at 550nm	*OMAOD*	–	–
Dust aerosol optical depth at 550nm	*DUAOD*	–	–

Hourly mean total aerosol optical depth forecasts at surface level for various discrete wavelengths are computed every day at 00UTC and 12UTC with a temporal forecast horizon of 5 days. The variables retrieved from the CAMS database are shown in Table 3.2, for the same area and period as the variables retrieved from the IFS/ECMWF.

3.2.3 Spatial and temporal downscaling of forecast data

To obtain forecast values for a specific location and with different time steps, spatial and temporal downscaling techniques were employed to all simulated variables. The spatial downscaling is conducted using bilinear interpolation of the values of the four grid points surrounding the desired location. This allows for the development and validation of the presented method against observations made at various solar radiation measuring stations which are not exactly located at a grid point, but also, in an operational setting, the inclusion of this technique allows for the forecasting of solar irradiance at any point of the domain. The temporal downscaling is computed using piecewise cubic hermite interpolation of the hourly mean irradiance. This method might not allow for the conservation of energy in each hour (forecast values) however, the goal is simply to obtain data in shorter timesteps which can be used as input for the more complex machine learning models that will perform the improvement of DNI forecasts. This is a compromise between developing a more elaborated physical downscaling method that preserves the hourly energy predictions (but preserves the error associated with those forecasts) and feeding the machine learning models directly with the original hourly irradiation values, and then having to develop a model for each desired timestep, with the consequent loss of model generalization. In this way, the deviation introduced by this downscaling method in the input values (which, at different instants, can either increase or decrease the error of the energy predictions, taking the ground-based measurements as reference) will also be assimilated by the machine learning model.

3.2.4 Solar radiation measurements and data quality check

The experimental data used for the model development were 1-min DNI and GHI ground-based observations made at Évora – Verney (38.567811°, -7.911459°) with a Kipp & Zonen CHP1 pyrheliometer in the case of DNI and a Kipp & Zonen CMP11 Pyranometer in the case of GHI. Experimental data from this location have been widely validated and used in other works in this field, e.g. [22,35–37].

For the model validation, 1-min DNI and GHI ground-based observations obtained from the network of radiometric stations of the DNI-ALENTEJO project [38] were also used. This is a solar radiation measurement network in the south of Portugal which comprised 13 stations scattered in the region, each station being typically equipped with a Kipp & Zonen Solys2 sun tracker, one CHP1 pyrheliometer and two CMP11 pyranometers. All instruments are periodically calibrated in accordance with ISO 9059:1990 and ISO 9847:1992 and the observations are corrected regarding the zero offset of sensors, filtered according to the Baseline Surface Radiation Network (BSRN) quality control procedure [39] and gaps filled according to the method developed in [38].

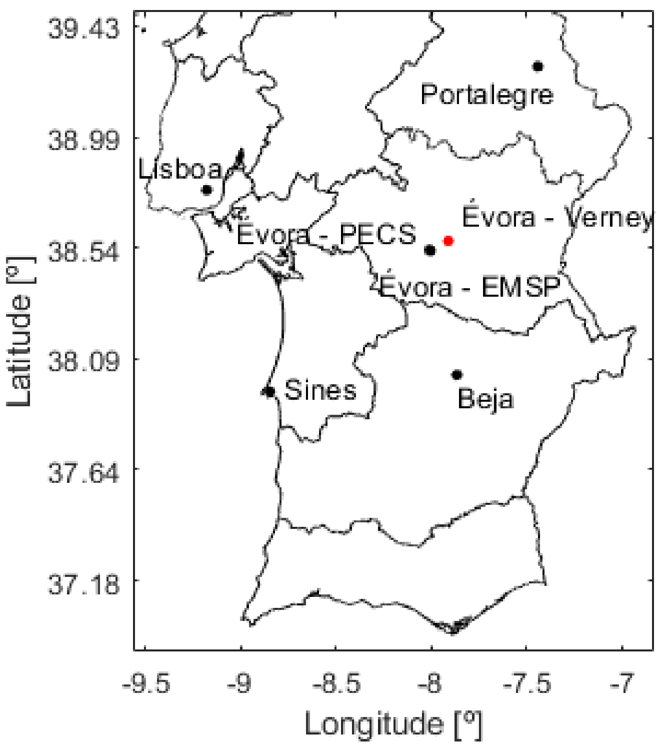


Fig. 3.1 - Location of solar radiation measuring stations from the DNI-ALENTEJO network [22].

The experimental data used for the model validation were retrieved from the following stations of the network: Évora – PECS (38.5306°, -8.0112°), Évora

– EMSP (38.5289°, -8.0053°), Portalegre (39.2692°, -7.4428°), Beja (38.0249°, -7.8672°), Lisboa (38.7734°, -9.1779°) and Sines (37.9576°, -8.8473°). The location of these stations is shown in Fig. 3.1 including the Évora – Verney station which data were used for the model development.

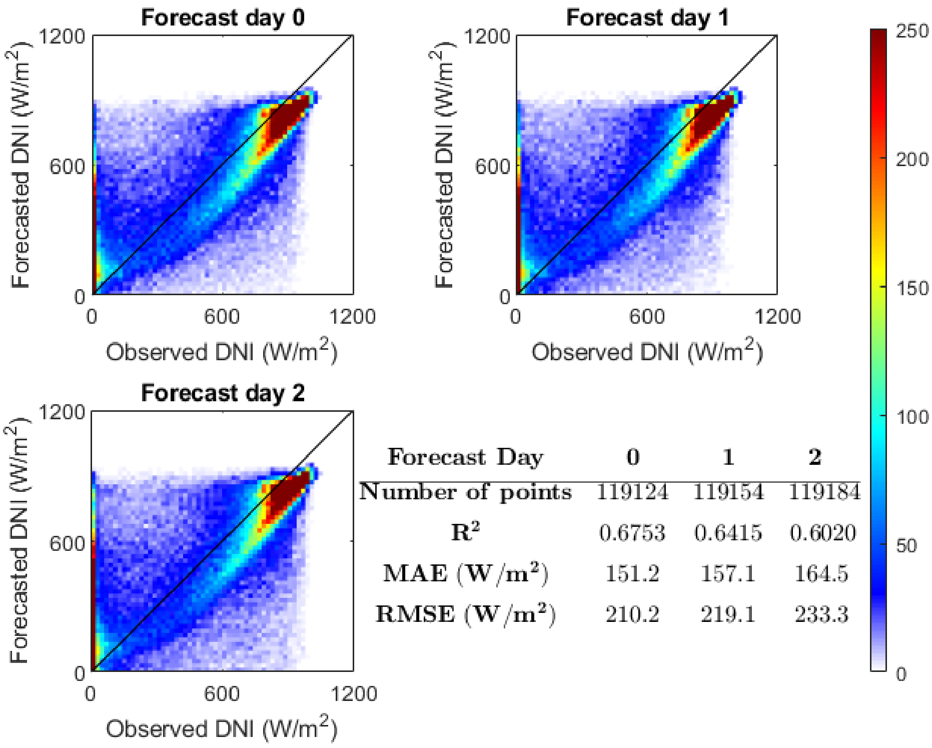
Original solar irradiance records from the radiometric stations are 1 min average, maximum, minimum and standard deviation from which the mean irradiance values for the different temporal resolutions used in this work were calculated (10 min for the development stage). The same period defined for the forecast data was used.

3.3 Analysis of NWP direct normal irradiance forecasts

To understand the accuracy of the original solar irradiance predictions from the IFS/ECMWF model, the forecast data of DNI and other atmospheric variables issued at 00UTC from 1st of December 2016 to the 31st of May 2021 were analyzed and compared against observations made at Évora – Verney station for each of the first three days of forecast time horizon. All forecast data was spatially downscaled to the location of the radiometric station and 10 min mean values were determined as described in Section 3.2.3. Experimental data were also averaged for the same 10 min time step as described in Section 3.2.4.

3.3.1 Comparison of NWP direct normal irradiance forecasts with experimental data

Firstly, a direct comparison between the 10 min DNI forecasts after spatial and temporal downscaling and observations at Évora – Verney station was conducted for each of the 3 days of forecast horizon as presented in Fig. 3.2.



Forecast Day	0	1	2
Number of points	119124	119154	119184
$\mathbf{R^2}$	0.6753	0.6415	0.6020
MAE ($\mathbf{W/m^2}$)	151.2	157.1	164.5
RMSE ($\mathbf{W/m^2}$)	210.2	219.1	233.3

Fig. 3.2 - Comparison between downscaled IFS/ECMWF forecasts and DNI observations (10 min) at Évora – Verney for forecast days 0, 1 and 2 (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m²).

These results show a better performance (lower mean absolute error, MAE, and root mean squared error, RMSE) of the model for day 0 of forecast which decreases for higher forecast time horizons which was also verified by other works in the literature [4]. The correction of these errors can be done with many different tools, from a simple bias correction to more complex machine learning models, as mentioned in Section 3.1. Simple artificial neural networks offer a middle ground where the computational effort is not too high, but the influence of the different variables used as inputs is taken into consideration. Day 1 of forecast (second day of forecast time horizon, or day-ahead) was used for the model development since, even though forecast issuance corresponds to the 00 UTC, data can take until 6:55 UTC to be available, which may turn the forecast unhelpful for that day depending on the location of interest.

3.3.2 Correlations between forecast variables and DNI observations

The various meteorological variables are to some extent all correlated with each other and, in a system so complex as the atmosphere, a way to identify the degree of correlation between two variables is to compute their linear correlation coefficients. Fig. 3.3 was generated with this purpose, comparing forecast values of each variable after spatial and temporal downscaling (10 min) with observed DNI at Évora – Verney station for the three forecast days. The Pearson's linear correlation coefficients (the values shown in each graph of Fig. 3.3) for each forecast variable are similar but tend to show a decrease in correlation across the three days of forecast time horizon. As expected, the highest absolute values of linear correlation occur with the forecasted DNI followed by GHI, cloud cover variables and solar zenith angle. While the aerosol variables show lower values of linear correlation, this does not mean that they are necessarily more independent since they might have a nonlinear relationship.

It is difficult to find any physical world phenomenon which follows linearity straightforwardly. Thus, a non-linear model that can approximate the non-linear phenomenon is needed. Representing these kinds of relationships using classical methods is known to be difficult. In that sense, machine learning models like artificial neural networks with non-linear activation functions will allow the model to create complex mappings between the inputs and outputs of the network. The non-linear layers enable ANNs to learn making conditional decisions for controlling the computational flow which makes them better suited to provide accurate data fitting and forecast.

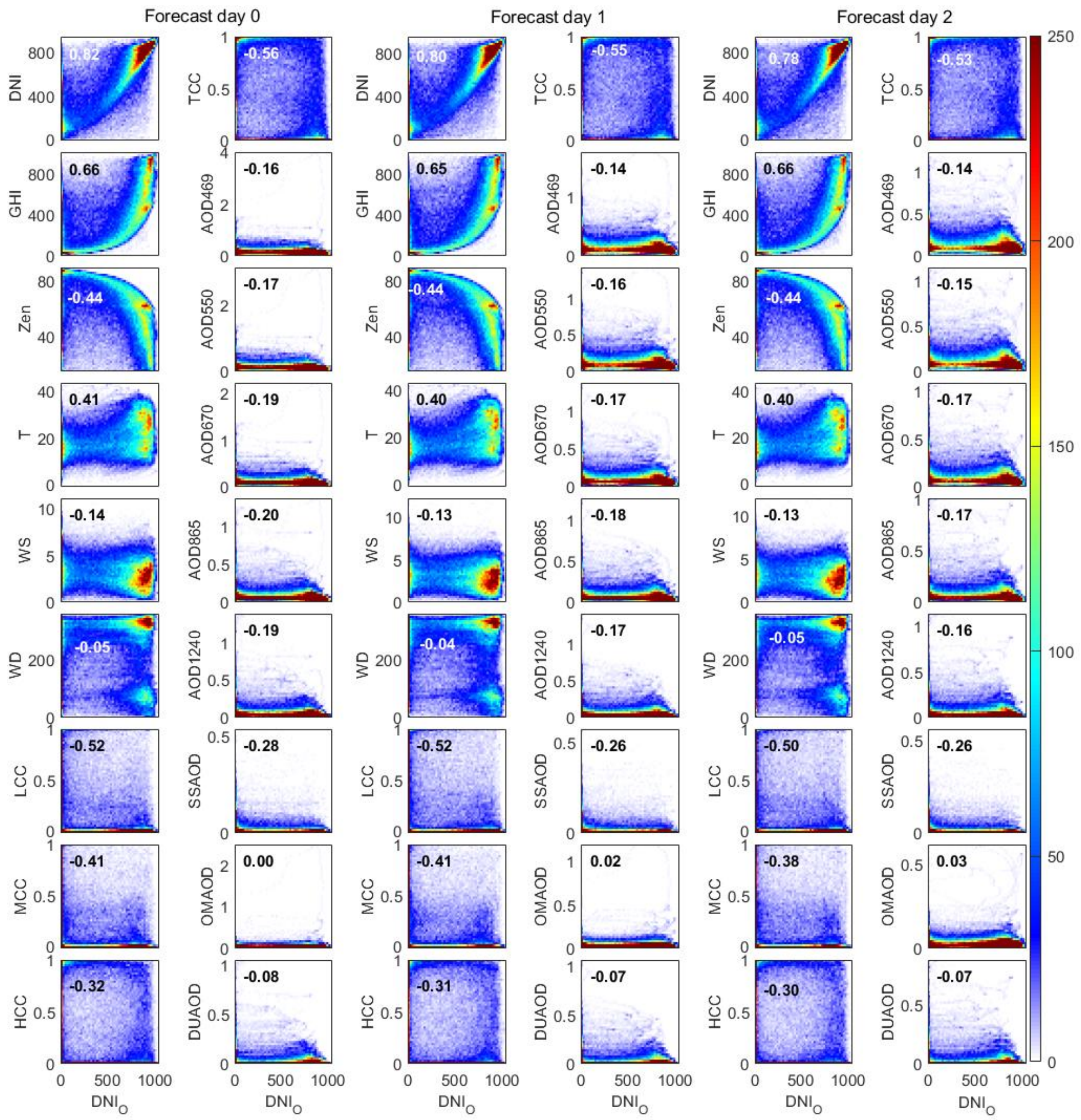


Fig. 3.3 - Comparison and Pearson's linear correlation coefficients of the forecast values of each meteorological variable (y axis) with the 10 min observed DNI at Évora – Verney station (x axis) for each of the forecast days. The colormap represents the number of points (total number of data points for each graph: 236592); Bins grid (50x50 bins).

3.4 Artificial neural network model development

Artificial neural networks (ANN) are constituted of connected artificial neurons forming a network. Since the relevance of each input given to the neuron is not equal, different weights are assigned to each of the inputs, then a linear net function is used to aggregate a bias and the weighted inputs after which a transfer function is applied to obtain the output of the neuron that will then be passed on to the next neuron.

The ANNs developed in this work are multi-layer feed-forward networks with back propagation learning, which are some of the most established ANN architectures due to their ability to perform arbitrary non-linear mappings. These are usually composed of an input and an output layer and one or more hidden layers of neurons. When training, for a given input dataset, the information flows forward through the network until it reaches the output layer and errors are calculated using the given desired output dataset (targets). These errors are then propagated backwards through the network and the weights of each neuron are adjusted so that the next iteration results in outputs with smaller error. This allows solving complex problems that can be stochastic, poorly defined, non-linear, non-analytic and/or non-stationary with low or no intervention in the program itself.

After the spatial and temporal downscaling of the forecast data shown in Section 3.2.3, two ANN models coupled in series are applied in this work according to Fig. 3.4 for the generation of improved DNI forecasts from NWP data. Observations of DNI made at Évora - Verney described in Section 3.2.4 were used for the development of the two ANNs: (i) development of an ANN using only the predicted weather and aerosol variables for a given time step in the temporal horizon of forecasts (Section 3.4.1); (ii) development of an ANN using the predicted DNI data over a period of time before the forecasted instant (Section 3.4.2).

For the evaluation and comparison of the different models, the coefficient of determination (R^2), mean absolute error (MAE) and root mean squared error (RMSE) were used as metrics along with the forecast skill (FS), which represents the improvement in terms of MAE over the original ECMWF forecasts, and a global performance index (GPI) based on the three statistical indicators, for model configurations comparison. The coefficient of determination is intuitively informative as it provides a measure of how well observed outcomes are replicated by the model, based on the proportion of total variation of outcomes explained by the model. According to the literature, the MAE and RMSE metrics are suitable [40] and the most commonly used indicators to assess the performance of machine learning regression algorithms [41]. Each error contributes to MAE in proportion to the absolute value of the error while the RMSE involves squaring the differences, so that a few major differences will increase the RMSE to a greater degree than the MAE.

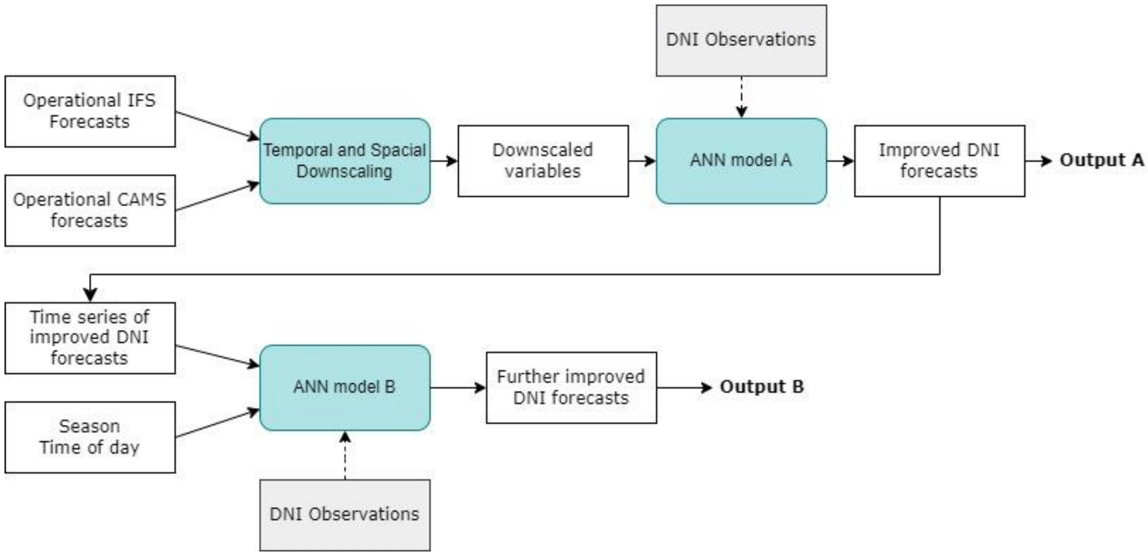


Fig. 3.4 - Flowchart of the developed model. DNI observations are only used for the development and evaluation of the ANN models they are not required as inputs in an operational setting.

The used definition of GPI results in a statistical tool that combines various metrics allowing for the performance comparison of different models and has been widely used in many works in several fields of study, for example in [42–44]. To compute the GPI, the n metrics need to be normalized into values ranging from 0 to 1 and then Eq. (3.1) is used, where the GPI value for the i^{th} model configuration is determined using the median of the normalized values, $\tilde{y}_j$, of the indicator j, the normalized value of indicator j for model configuration i, y_{ij}, and a factor α_j that has a value of 1 for all indicators except the coefficient of determination for which it is -1.

$$GPI_i = \sum_{j=1}^{n} \alpha_j (\tilde{y}_j - y_{ij}) \tag{3.1}$$

In this work, the metrics used for the computation of GPI are the previously mentioned R^2, MAE and RMSE. Since all the metrics used in this definition take the ground-based measurements as reference, a higher value of GPI represents a better performance of the respective model configuration.

3.4.1 ANN model with weather and aerosol forecasts as inputs (ANN model A)

The forecast data included as inputs in the development of this ANN model are the various weather and aerosol variables from the IFS/ECMWF and CAMS global NWP models (see Table 3.1 and Table 3.2) for forecast day 1 (25 to 48 h ahead), after temporal and spatial downscaling as described in Section 3.2.3. The 10 min ground-based observations of DNI at Évora – Verney station as described in Section 3.2.4, are used as targets.

To obtain an ANN configuration that generates improved DNI forecasts while having a good generalization capability, datasets were divided in three subsets: the first subset for training and validation of various internal configurations of the tested ANNs (78,854 data points from 1st December 2016 to 30th November 2019), a second subset for testing and selection of the final ANN configuration (23,654 data points from 1st December 2019 to 30th November 2020) and finally a third subset of data for a blind test so the performance and generalization capability of the selected ANN configuration can be demonstrated (12,470 data points from 1st December 2020 to 31st May 2021). For training and validation, the input data is divided randomly using 80% of the available data points in the first subset for training and the other 20% for validation.

An ANN is defined by numerous parameters and specifications, some of which are pre-established such as using a feedforward ANN with one hidden layer and a linear layer output (fitnet) with an initialization function that initializes the weights and biases of the layers according to the Nguyen-Widrow initialization algorithm (initnw), the hyperbolic tangent sigmoid transfer function (tansig) and the mean squared error as performance function (mse). The input and output data are treated by removing rows with constant values (removeconstantrows) and scaling the mean of each row to 0 and deviations to 1 (mapstd).

Other parameters and configurations were specifically evaluated in this work, namely the training function, the number of neurons and the input variables. The training functions assessed were the Levenberg-Marquardt backpropagation (trainlm) and the Bayesian regularization backpropagation (trainbr) and, for each of these, the number of neurons was varied from 1 to 25 and the input forecasted variables were added one by one according to the sequence DNI, GHI, TCC, Zen, T, SS, AOD1240, WS, LCC, MCC, HCC, AOD865, AOD670, AOD550, AOD469, DU, WD, OM, for each combination of training function and number of neurons.

Additionally, for better results, ten randomly initialized ANNs were trained and validated for each configuration combo mentioned above, being the average output of those ten ANNs considered the result of the corresponding ANN configuration, as already tested in other works in this field [22].

The values of R^2, MAE, RMSE were computed, and the performance of the different ANN configurations is compared using the FS and the GPI. The five ANN configurations with highest GPI at the training and validation stage (among the 900 cases evaluated with GPI values ranging between 1.159 and -1.840) are presented in Table 3.3.

The configurations using the training function trainbr with all input variables and high number of neurons show more accurate DNI forecasts when considering observations as reference. For the best performing configuration, there is an improvement over the original ECMWF forecasts revealed by a forecast skill of 22 % and an increase of 0.1118 in the R^2 and a decrease of 34.2 W/m^2 and 38.4 W/m^2 in the MAE and RMSE, respectively.

Table 3.3 - ANN model A training and validation results (five best configurations in descending order).

Training function	Number of neurons	Number of inputs	R^2	MAE (W/m^2)	RMSE (W/m^2)	FS (%)	GPI
trainbr	25	18	0.7535	122.9	180.7	22.0	1.159
trainbr	24	18	0.7528	123.0	181.0	21.9	1.144
trainbr	23	18	0.7525	122.9	181.1	22.0	1.141
trainbr	25	17	0.7511	123.6	181.6	21.5	1.099
trainbr	24	17	0.7508	123.5	181.7	21.6	1.097

However, regarding the final purpose of the developed model, that is, an algorithm that can generalize well and improve newly generated forecasts, the 900 combos were run for the test data set to select the best configuration. The five ANN configurations with highest GPI are presented in Table 3.4.

Table 3.4 - ANN model A testing results (five best configurations in descending order).

Training function	Number of neurons	Number of inputs	R^2	MAE (W/m^2)	RMSE (W/m^2)	FS (%)	GPI
trainbr	7	13	0.7127	136.7	197.8	10.2	0.876
trainlm	7	13	0.7132	137.1	197.7	9.9	0.867
trainlm	9	13	0.7121	136.8	198.0	10.1	0.844
trainlm	5	13	0.7112	136.5	198.3	10.3	0.832
trainbr	5	13	0.7115	136.9	198.2	10.0	0.818

Thus, the ten ANNs using the trainbr training function, 7 neurons and 13 inputs (DNI, GHI, TCC, Zen, T, SS, AOD1240, WS, LCC, MCC, HCC, AOD865 and AOD670) were selected as the base configuration for this ANN model (ANN model A). This configuration showed an improvement over the original ECMWF forecasts (values in Fig. 3.2) with a forecast skill of 10.2 % and an increase of 0.0712 in the R^2 and a decrease of 20.4 W/m^2 and 21.3 W/m^2 in the MAE and RMSE, respectively. The performance and good generalization of the selected ANN configuration was verified by using a new set of data (blind test), which resulted in values of R^2, MAE and RMSE of

0.6874, 147.2 W/m² and 207.6 W/m², respectively, which is better than all statistical indicators of the original DNI forecasts (Fig. 3.2).

3.4.2 ANN model with forecasted DNI time series and date/time as inputs (ANN model B)

A different approach for an ANN configuration that considers the temporal variation of solar irradiance as input was also addressed and evaluated. This configuration uses the same internal parameters as the one described above, being the evaluated parameters the training function (trainlm, trainbr) and the number of neurons (from 1 to 25), while the inputs are the season and time of day of forecast and the existing time series of DNI values for a given period before the forecast computation. This time series can be either the original IFS/ECMWF predictions or experimental values but, to further improve the solar irradiance forecasts and to make the model independent of ground-based measurements, the corrected DNI forecast data from the ANN model A was used instead in this case.

Table 3.5 - ANN model B training and validation results using a period of 2 h of predicted DNI from ANN model A as input (five best configurations in descending order).

Training function	Number of neurons	R^2	MAE (W/m²)	RMSE (W/m²)	FS (%)	GPI
trainbr	25	0.7332	126.1	188.0	20.0	1.126
trainbr	24	0.7337	126.3	188.1	19.8	1.019
trainbr	23	0.7335	126.4	188.2	19.8	0.977
trainbr	22	0.7321	126.4	188.3	19.8	0.917
trainbr	21	0.7319	126.6	188.4	19.6	0.852

Regarding the length of the time series, two cases were tested: (i) a period corresponding to the previous 2 h (12 time steps of 10 min data); and (ii) a period of 24 h (144 time steps of 10 min data). The same data sets of experimental data were used for training and validation, testing and blind testing as in the Section 3.4.1, as well as the procedure of averaging the results of ten randomly initialized ANNs for each configuration. The resulting metrics for the five configurations with highest GPI are shown in Tables 3.5 and 3.6 for these two cases.

Table 3.6 - ANN model B training and validation results using a period of 24 h of predicted DNI from ANN model A as input (five best configurations in descending order).

Training function	**Number of neurons**	R^2	**MAE (W/m²)**	**RMSE (W/m²)**	**FS (%)**	**GPI**
trainbr	25	0.7518	123.0	181.2	21.9	1.990
trainbr	24	0.7496	123.5	182.0	21.6	1.788
trainbr	23	0.7485	123.9	182.5	21.4	1.672
trainbr	22	0.7479	123.8	182. 7	21.4	1.642
trainbr	21	0.7456	124.4	183.5	21.1	1.418

Similar to the case of ANN model A, the configurations that show better performance tend to have more neurons and the more complex training function *trainbr*. However, the best models that arise from the training process might not be the best at generalization, thus all configurations were also evaluated using the test data set to select the configuration that achieves good performance on new data. Tables 3.7 and 3.8 show the five best configurations using the previous 2 and 24 h of predicted DNI from ANN model A as inputs, respectively.

Selecting the best configurations for each case and applying these models to the blind test data set, values of R^2, MAE and RMSE of 0.6934, 142.4 W/m^2 and 205.1 W/m^2 for the model using a period of 2h and 0.6916, 142.6 W/m^2 and 205.5 W/m^2 for the model using a period of 24h were obtained, respectively. These metrics result in a forecast skill of 12.4 % for both models and an improvement over the predicted values from the ANN model A with an increase in R^2 of 0.0060 and a decrease of MAE and RMSE of 4.8 W/m^2 and 2.5 W/m^2 for the 2h model and an increase in R^2 of 0.0042 and a decrease of MAE and RMSE of 4.6 W/m^2 and 2.1 W/m^2 for the 24 h model, respectively.

Table 3.7 - ANN model B test results using a period of 2 h of predicted DNI from ANN model A as input (five best configurations in descending order).

Training function	Number of neurons	R^2	MAE (W/m^2)	RMSE (W/m^2)	FS (%)	GPI
trainlm	8	0.7197	133.3	195.3	12.4	0.709
trainlm	7	0.7194	133.3	195.4	12.4	0.630
trainbr	7	0.7191	133.2	195.5	12.5	0.572
trainlm	10	0.7191	133.3	195.5	12.4	0.563
trainbr	5	0.7190	133.3	195.5	12.4	0.512

Table 3.8 - ANN model B test results using a period of 24 h of predicted DNI from ANN model A as input (five best configurations in descending order).

Training function	Number of neurons	R^2	MAE (W/m^2)	RMSE (W/m^2)	FS (%)	GPI
trainbr	3	0.7184	133.1	195.7	12.4	0.451
trainlm	12	0.7190	133.4	195.5	12.2	0.439
trainlm	11	0.7183	133.2	195.7	12.4	0.429
trainlm	5	0.7191	133.5	195.4	12.2	0.429
trainbr	4	0.7180	133.3	195.8	12.3	0.299

Although the differences between the two cases are small, the selected configuration of the ANN model B was the one using as input a period of 2 h of DNI predictions due to its slightly better performance and simplicity. The resulting composite ANN model (ANN model B) is thus the combination of an ANN model based on the input of a complete set of meteorological and aerosol forecast data for a given instant (ANN model A) and an ANN model based on the input of a time series of DNI for a given period before the forecast computation (ANN model B configuration), provided that the input values of the second model are the predicted DNI output values of the first model.

Using two separate ANN models can result in better forecasts than a single model which considers both the actual and temporal variations of weather variables on DNI forecasts since it allows for better model specialization. This means that each ANN can be specialized for a specific task. The first ANN was developed for improving DNI forecasts from NWP data at a specific time step being designed and optimized for representing the relationship between the different atmospheric variables and DNI, being optimized to extract relevant features and relationships specific to the current conditions. The second ANN was developed to model the temporal tendencies and patterns in DNI which can be complex and non-linear. By using a separate ANN model to capture these tendencies, it allows the model to focus solely on learning the temporal dependencies and trends, which might be distinct from the relationships governing the immediate forecast.

3.5 Assessment of the developed ANN models using different temporal resolutions

The proposed model was developed using 10 min mean values, but it could easily be adapted to different time steps as required by solar system operators. Also, since mean irradiance values are used instead of irradiation data in the case of DNI, inputs with different time steps can also be directly used with the already developed model.

In order to evaluate the performance of this feature, input data with temporal resolutions of 5 and 15 min and the original hourly (60 min) forecasts were fed to the model, generating the results shown in Table 3.9. It is important to note that, as described in Section 3.4.2, the ANN model B was developed using as inputs 12 records of 10 min DNI predictions, corresponding to the period of 2 h before the forecast time step, while for different temporal resolutions, the number of inputs used is still 12 records but instead of corresponding to a period of 2 h they are equivalent to periods of 1, 3 and 12 h for 5, 15 and 60 min temporal resolutions, respectively.

*Table 3.9 - Metrics for the original downscaled ECMWF predictions, ANN model A and ANN model B with different temporal resolution and for the different data sets used in the development of the models (for all data and each statistical indicator, the best performing model is represented in bold for each time step, the best performing time step is underlined for each model, and the best combination of time step and model is marked with *).*

Time step (min)	**Model**	**R^2**				**MAE (W/m^2)**				**RMSE (W/m^2)**			
		Train	*Test*	*Blind test*	*All*	*Train*	*Test*	*Blind test*	*All*	*Train*	*Test*	*Blind test*	*All*
5	ECMWF	0.6290	0.6489	0.6168	0.6341	160.7	155.9	168.7	160.4	225.2	222.0	236.3	225.7
	A	0.7088	0.7017	0.6776	0.7058	135.9	140.9	151.3	138.7	199.4	204.5	213.7	202.1
	B	0.7129	0.7053	0.6802	**0.7096**	131.1	135.2	146.3	**133.6**	198.2	203.2	213.4	**201.0**
10	ECMWF	0.6353	0.6579	0.6273	0.6415	157.5	152.1	165.1	157.1	220.0	216.1	230.5	219.1
	A	0.7176	0.7130	0.6874	0.7148	132.3	136.7	147.2	135.2	193.4	197.8	207.6	196.1
	B	0.7265	0.7197	0.6934	**0.7234**	128.3	133.3	142.4	**130.9**	190.3	195.3	205.1	**193.0**
15	ECMWF	0.6515	0.6739	0.6422	<u>0.6574</u>	153.0	147.3	159.7	<u>152.5</u>	214.5	210.0	224.3	<u>214.6</u>
	A	0.7339	0.7284	0.7053	<u>0.7314</u>	127.8	131.9	141.7	<u>130.2</u>	187.3	191.5	200.4	<u>189.7</u>
	B	0.7411	0.7355	0.7117	**<u>0.7386</u>***	127.4	130.4	138.4	**<u>129.3</u>***	185.1	189.2	197.1	**<u>187.3</u>***
60	ECMWF	0.5527	0.5877	0.5519	0.5624	176.3	169.0	179.3	175.0	239.9	231.8	245.0	238.7
	A	0.6677	0.6705	0.6505	0.6684	153.3	154.1	157.7	**153.9**	203.5	204.2	209.0	**204.3**
	B	0.6946	0.6904	0.6685	**0.6929**	174.4	171.5	164.3	172.7	217.7	217.8	213.3	217.3

Comparing the results for the different temporal resolutions, the model seems to perform better for a time step of 15 min, followed by 10, 5 and finally 60 min. However, these results are also affected by the number of data points available which differs with temporal resolution (more data for shorter temporal resolution). Thus, to take this aspect into consideration, the results of the models for each temporal resolution were converted to hourly means for comparison and the same metrics were computed, as shown in Table 3.10.

Table 3.10 - Results of hourly mean DNI forecasts with different temporal resolutions and from different models and data sets. Color comparison within each metric and data set where darker color means better performance.

Data set	Time step (min)	R^2			MAE (W/m²)			RMSE (W/m²)			GPI		
		ECMWF	*A*	*B*	*ECMWF*	*A*	*B*	*ECMWF*	*A*	*B*	*ECMWF*	*A*	*B*
Train	5	0.5678	0.6429	0.6493	173.5	156.4	154.5	234.0	211.8	211.1	-1.177	0.140	0.240
	10	0.5280	0.5915	0.6738	183.2	170.2	147.5	247.5	230.3	202.2	-1.926	-0.880	0.740
	15	0.6029	0.6798	0.6916	164.3	146.2	144.2	222.7	199.3	195.6	-0.515	0.864	1.057
	60	0.5527	0.6677	0.6946	176.3	153.3	174.4	239.9	203.5	217.7	-1.456	0.528	-0.126
Test	5	0.5994	0.6505	0.6560	166.6	155.3	153.0	226.9	210.6	210.2	-1.449	-0.111	0.035
	10	0.6164	0.6629	0.6777	162.1	152.6	147.7	221.3	206.5	202.2	-0.974	0.219	0.652
	15	0.6330	0.6848	0.6952	157.6	145.9	143.9	216.0	199.1	196.1	-0.507	0.870	1.125
	60	0.5877	0.6705	0.6904	169.0	154.1	171.5	231.8	204.2	217.8	-1.787	0.298	-0.528
Blind test	5	0.5883	0.6424	0.6507	170.4	158.2	154.8	232.8	212.7	212.0	-1.272	-0.076	0.103
	10	0.6027	0.6533	0.6695	167.0	155.4	149.1	228.4	209.2	204.2	-0.971	0.163	0.577
	15	0.6166	0.6720	0.6854	163.6	151.1	145.6	224.2	203.3	198.1	-0.674	0.555	0.932
	60	0.5519	0.6505	0.6685	179.3	157.7	164.3	245.0	209.0	213.3	-2.068	0.078	-0.076
All	5	0.5790	0.6467	0.6532	171.6	156.3	154.2	232.3	211.6	211.0	-1.136	0.122	0.233
	10	0.5384	0.5962	0.6764	181.6	169.9	147.7	245.9	229.5	202.4	-1.939	-0.925	0.729
	15	0.6130	0.6821	0.6938	162.8	146.6	144.3	221.4	199.7	196.0	-0.463	0.848	1.061
	60	0.5624	0.6684	0.6929	175.0	153.9	172.7	238.7	204.3	217.3	-1.463	0.473	-0.132

Regarding the temporal downscaling, although the metrics for the ECMWF forecasts can, for some instances, deteriorate with downscaling (metrics for hourly values better than for the remaining temporal resolutions) due to a smoothing effect that does not capture the rapid and nonlinear variations of real data, these metrics are improved when using the developed ANN models to the downscaled data when compared with the original hourly ECMWF forecasts. This means that the developed models can not only counteract the possible deviations induced by the temporal downscaling but also further improve the DNI forecasts. As previously discussed, the ANN model B tends to perform better than the ANN model A and ECMWF models except for hourly temporal resolutions.

Finally, the performance (measured by the GPI) of the developed model with a temporal resolution of 15 min is better than the one for 10, 5 and 60 min for the data used in this work which is in accordance with the results shown in Table 3.9.

3.6 Application of the ANN models for different locations

The ANN models described in the previous sections were developed based on data for a specific location (Évora - Verney). In this section, the hypothesis of using the same models in other locations in the region surrounding the reference station of Évora-Verney (south of Portugal) is analyzed. The ECMWF and CAMS forecast data obtained for six different sites where solar radiation measuring stations are installed (DNI-A network, see Fig. 3.1) were used as inputs. The original 10 min downscaled ECMWF forecasts and the DNI forecasts obtained from the ANN model B were compared with the available observed DNI data for each station. The results and the number of data points (N) for each station are shown in Table 3.11, where the data used for evaluating Verney station is comprised of the test and blind test datasets. The results show that the developed model can improve the overall metrics for all stations being best at generating improved DNI forecasts for the station of Verney which data was used in the development of the algorithm, as expected. The stations for which the improvement of DNI forecasts is lower are Lisboa and Sines, both located near the coast of Portugal where the climate is more disparate from that observed at the station of Évora - Verney. Considering all stations and data available, the use of the developed models result in a forecast skill of 14.4 % and improve the original downscaled DNI forecats by 0.0707 in R^2, 23.1 W/m^2 in MAE and 24.1 W/m^2 in RMSE, taking experimental data as reference showing that the application of this model for locations in the surrounding area of the location for which it was developed, can generate improved DNI forecasts.

Table 3.11 - Metrics for the original downscaled ECMWF DNI forecasts and ANN models B predictions (10 min) for the different stations located in the south of Portugal.

Station	R^2		MAE (W/m^2)		RMSE (W/m^2)		FS	N
	ECMWF	**ANN**	**ECMWF**	**ANN**	**ECMWF**	**ANN**	**(%)**	
Verney	0.6630	0.7237	151.3	131.8	215.1	194.4	12.9	28101
EMSP	0.6528	0.7222	153.2	130.5	218.1	194.2	14.8	96323
PECS	0.6348	0.7128	161.7	135.5	226.9	200.4	16.2	113300
Portalegre	0.6329	0.7083	157.5	134.8	225.9	200.8	14.4	115997
Beja	0.6393	0.6996	156.8	135.0	219.8	199.7	13.9	97109
Lisboa	0.5970	0.6659	170.9	147.6	238.3	212.2	13.6	61942
Sines	0.5459	0.6184	175.8	153.3	244.8	220.9	12.8	53192
All	0.6269	0.6976	160.4	137.2	226.5	202.4	14.4	478574

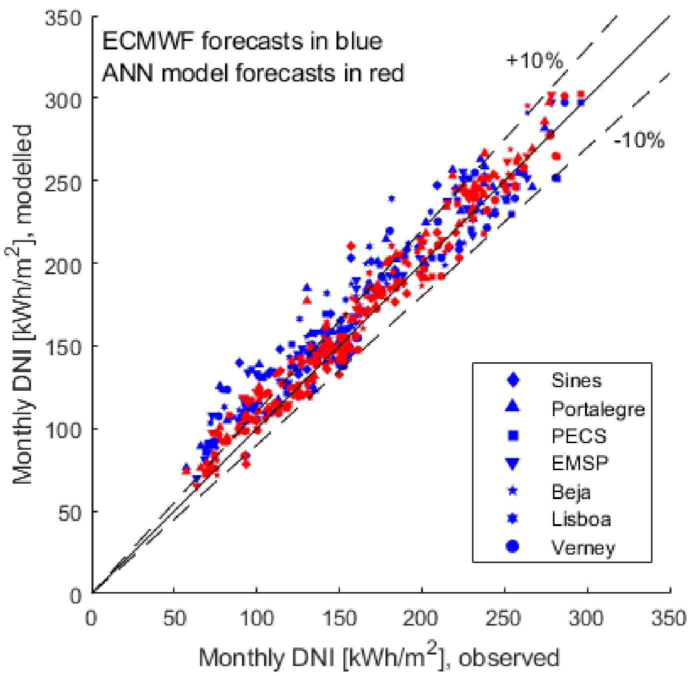


Fig. 3.5 - Comparison between monthly DNI forecasts (irradiation) from the ECMWF (blue) and from ANN model (red) against the available experimental data for each station in the South of Portugal.

Fig. 3.5 shows a direct comparison between the monthly forecasted DNI (irradiation) by the ECMWF and the developed ANN models (based on 10 min data) and the available monthly observed DNI for the different stations, while Table 3.12 presents the metrics resulting from this data. For the Verney station only the data from test and blind test datasets was considered.

Considering the monthly irradiation values for all stations, an improvement over the original DNI forecasts from the ECMWF of 0.0331, 5.9 kWh/m^2 and 6.6 kWh/m^2 for R^2, MAE and RMSE, respectively, is achieved by the developed algorithm with a forecast skill of 38.1 %.

Table 3.12 - Metrics for the monthly original spatially downscaled ECMWF DNI (irradiation) forecasts and ANN model B predictions for the different stations located in the south of Portugal.

Station	**R^2**		**MAE (kWh/m^2)**		**RMSE (kWh/m^2)**		**FS (%)**
	ECMWF	**ANN**	**ECMWF**	**ANN**	**ECMWF**	**ANN**	
Verney	0.9431	0.9771	13.0	7.4	16.3	9.4	43.0
EMSP	0.9483	0.9758	15.0	9.8	17.6	12.1	34.9
PECS	0.9518	0.9803	13.6	7.8	16.8	9.7	43.1
Portalegre	0.9402	0.9717	15.9	11.9	19.7	14.9	25.5
Beja	0.9472	0.9689	14.0	9.6	16.2	12.3	31.0
Lisboa	0.8948	0.9598	22.2	11.4	25.7	13.6	48.6
Sines	0.8244	0.8742	18.3	9.6	22.1	13.9	47.6
All	0.9313	0.9644	15.5	9.6	18.9	12.3	38.1

3.7 Application of the developed model for operational DNI forecasts

The model developed in this work can be used operationally to obtain more accurate DNI forecasts which can help solar energy systems and power grid operators to estimate the energy generation and make better informed decisions. After retrieving the operational forecasts of ECMWF of the 00UTC run, the model is applied resulting in 10 min DNI forecasts for the present day and the next 2 days. As an example, Fig. 3.6 was generated, where the model was used to generate DNI forecasts for 3 consecutive days (forecast issue time at 00:00 UTC of 9, 10 and 11 of January 2020, as example) at the location of Évora - Verney station. Here, the model results show very good agreement with observations for clear sky conditions, with larger errors for overcast or partially cloudy skies, as expected. The day January 11 is forecasted for all presented runs of the model and, in order to assess the variation of model performance along the successive forecast issues, the variation of statistical indicators R^2, MAE and RMSE for this day is shown in Fig. 3.7.

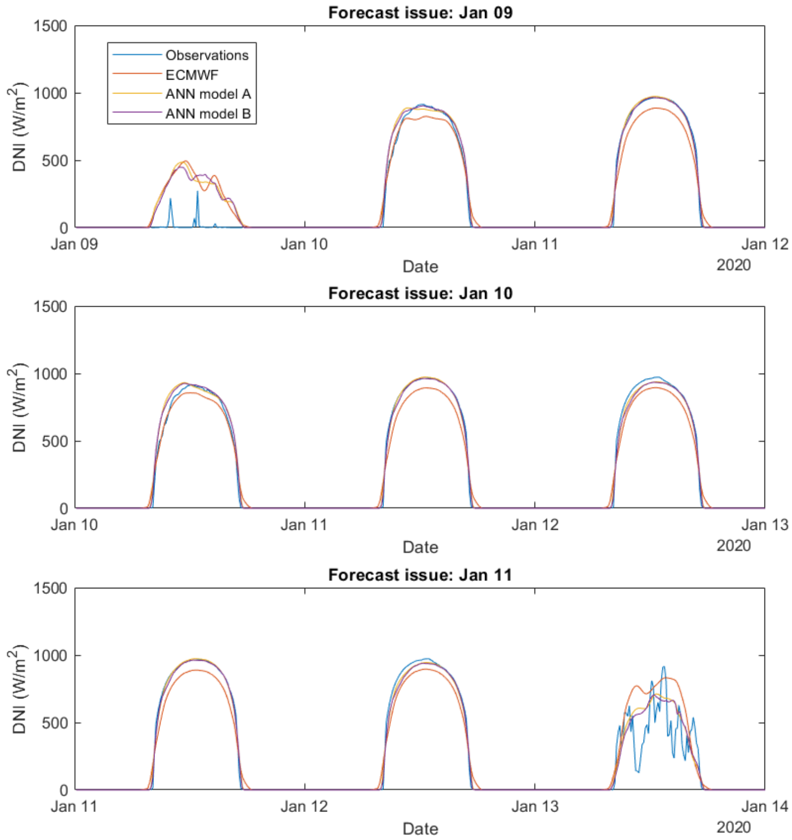
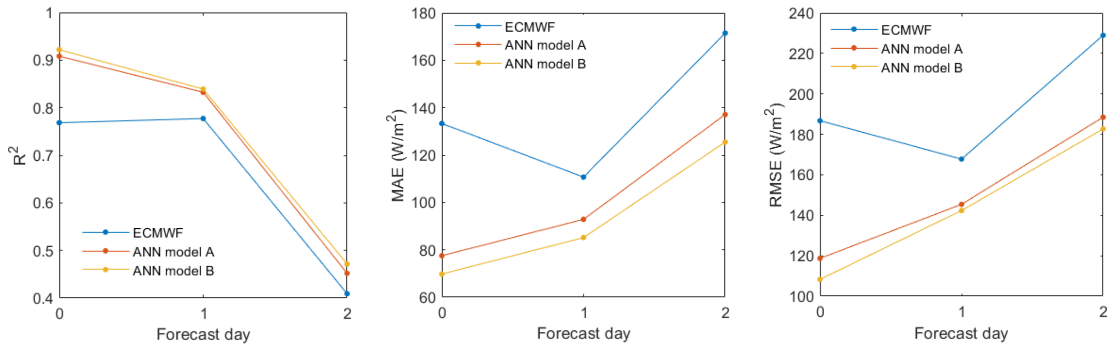


Fig. 3.6 - Example of operational use of the developed model for 3 consecutive days.

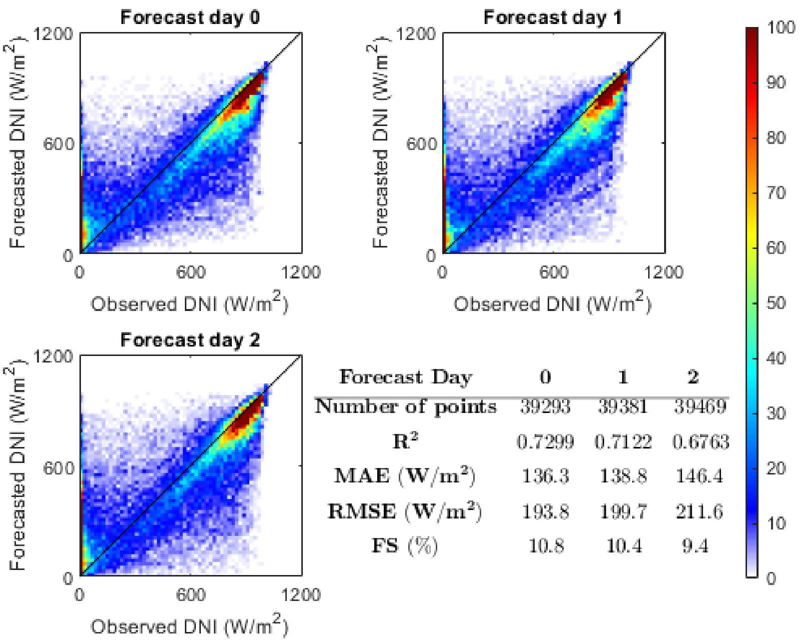


(a) *(b)* *(c)*

Fig. 3.7 - Variation of statistical indicators (a) R^2, (b) MAE and (c) RMSE for Jan. 11, 2020 using 00:00 UTC forecasts at day 0 (forecast issue: Jan. 11), 1 day ahead (forecast issue: Jan. 10) and 2 days ahead (forecast issue: Jan. 9), based on 10 min data.

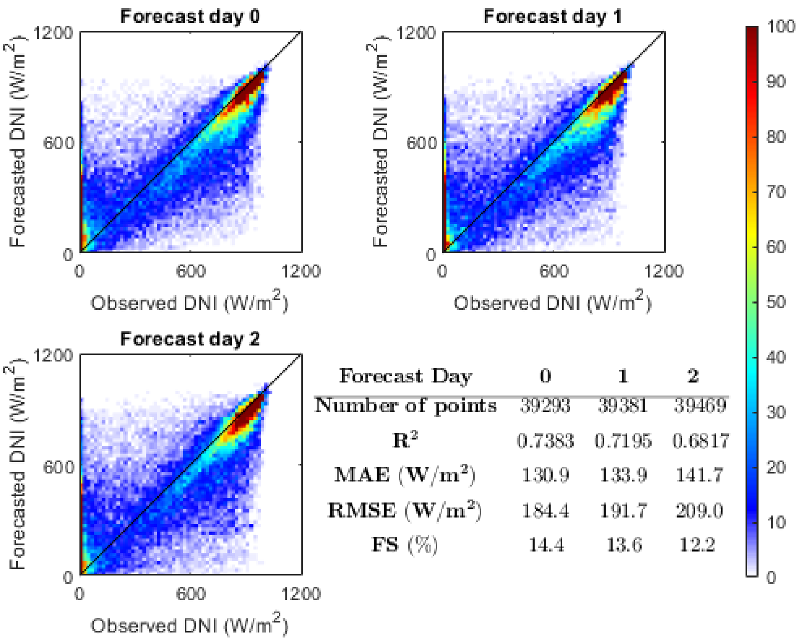
In this case, the performance of both ANN models is better for smaller forecast time horizons where the ANN model B achieved the best results followed by the ANN model A and finally the ECMWF model. It should be noted that this is not a typical behavior. As previously indicated in Figs. 3.2 and 3.3, the irradiation forecast tends to worsen with the horizon. However, although, in general, the predictability of solar irradiance decreases with the time horizon, in particular cases the forecast for two days ahead can be better than the forecast for the next day. A recent and good discussion about the predictability of the numerical prediction of solar irradiation and its variation with the horizon can be found in [45].

As this is just an example for a specific day, the same operational test was carried out for the test and blind test datasets of Évora – Verney station and the results can be seen in Figs. 3.8 and 3.9 for the ANN model A and ANN model B, respectively. Comparing these figures with the one obtained for the original downscaled ECMWF forecasts (Fig. 3.2), a higher density of data points near the bisection line is found for the forecasts of the ANN models.



Forecast Day	0	1	2
Number of points	39293	39381	39469
R^2	0.7299	0.7122	0.6763
MAE (W/m^2)	136.3	138.8	146.4
RMSE (W/m^2)	193.8	199.7	211.6
FS (%)	10.8	10.4	9.4

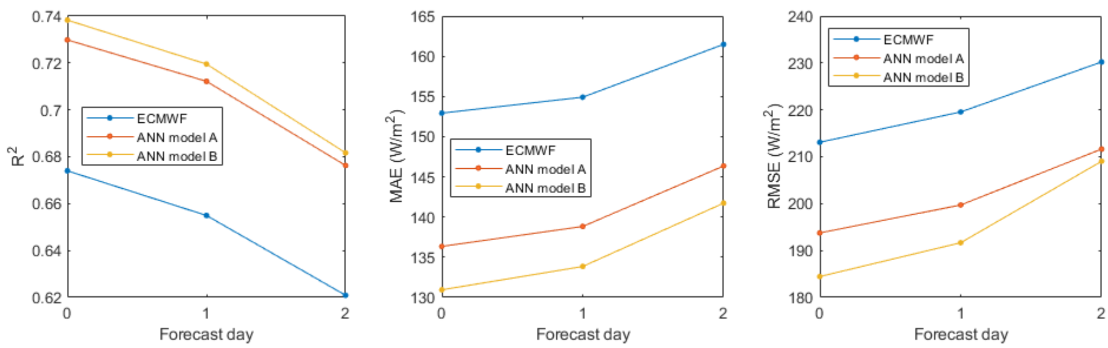
Fig. 3.8 - Results of ANN model A for test and blind test 10 min data of Évora-Verney station (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m^2).



Forecast Day	0	1	2
Number of points	39293	39381	39469
R^2	0.7383	0.7195	0.6817
MAE (W/m²)	130.9	133.9	141.7
RMSE (W/m²)	184.4	191.7	209.0
FS (%)	14.4	13.6	12.2

Fig. 3.9 - Results of ANN model B for test and blind test 10 min data of Évora-Verney station (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m²).

A similar way to visualize these results to that of Fig. 3.7 is shown in Fig. 3.10 for each forecasting model and each of the forecast days, now for all data and not for a specific day. Again, all models show an overall increase in performance for shorter forecast time horizons and all metrics are improved by the models developed in this work for all days of the forecast time horizon.



(a) *(b)* *(c)*

Fig. 3.10 - Variation of statistical indicators (a) R^2, (b) MAE and (c) RMSE results for Évora – Verney station for each forecast day ahead, based on 10 min data.

3.8 Conclusions

This work presents the development of a model that can successfully generate improved DNI forecasts based on data from NWP models in an operational setting. This model uses forecast data from the IFS/ECMWF and CAMS of several atmospheric variables including aerosol data which affects the transport of solar radiation through the atmosphere. These forecasts are spatially and temporally downscaled to the location and time resolution desired using bi-linear interpolation of the values of the four surrounding grid points and piecewise cubic interpolation of the hourly mean variables, respectively. Then, two different models based on artificial neural networks were designed and optimized to generate improved DNI forecasts with the desired temporal resolution and for a forecast time horizon of 72 h. Different configurations of these models were tested, and the selected configuration uses two feedforward ANNs in series, that is, the second ANN uses the output of the first as input. In an operational setting, the model is run every day when the ECMWF and CAMS operational forecasts are available for retrieval (maximum dissemination time 06:55 UTC) and results for the next two days can be used by solar energy producers and power grid operators to estimate the energy production and make better informed decisions.

Comparing the final model results using only test and blind test datasets with a temporal resolution of 10 min and for the location of Évora – Verney (the location used for model development) with the ground-based observations at the same location, values of R^2, MAE and RMSE of 0.7195, 133.9 W/m^2 and 191.7 W/m^2 for forecast day 1 and of 0.6812, 141.7 W/m^2 and 209.0 W/m^2 for forecast day 2 were achieved, respectively. This results in an improvement over the original downscaled DNI forecast from ECMWF of 0.0646, 21.1 W/m^2 and 27.9 W/m^2 for forecast day 1 and of 0.0608, 19.8 W/m^2 and 21.2 W/m^2 for forecast day 2 for R^2, MAE and RMSE, respectively.

The model was also applied to other locations scattered in the region surrounding the site for which it was developed showing improvements of the DNI forecasts of 0.0713, 23.3 W/m^2 and 24.2 W/m^2 for day 1 of forecast and for R^2, MAE and RMSE respectively, when compared to the ECMWF forecasts and taking the ground-based measurements in each location as reference.

It was also shown that the already developed algorithm can be applied to similar temporal resolutions, such as 5 or 15 min, achieving good results which can be extremely helpful when forecasts with different time steps are needed so that they are in accordance with the real-time market forecasting requirements.

With these improved forecasts, more accurate estimations of energy generation in solar energy systems can be achieved which is extremely important for solar power plant and energy market operators.

References

1. IEA World Energy Outlook 2021 - Revised Version October 2021. **2021**.
2. El Alani, O.; Ghennioui, A.; Ghennioui, H.; Saint-Drenan, Y.-M.; Blanc, P. Evaluation of 24-Hours Forecasts of Global Solar Irradiation from IFS, GFS and McClear Models. In Proceedings of the AIP Conference Proceedings; 2020.
3. Perez, R.; Lorenz, E.; Pelland, S.; Beauharnois, M.; Van Knowe, G.; Hemker, K.; Heinemann, D.; Remund, J.; Müller, S.C.; Traunmüller, W.; et al. Comparison of Numerical Weather Prediction Solar Irradiance Forecasts in the US, Canada and Europe. *Solar Energy* **2013**, *94*, 305–326, doi:10.1016/j.solener.2013.05.005.
4. Perdigão, J.; Canhoto, P.; Salgado, R.; Costa, M.J. Assessment of Direct Normal Irradiance Forecasts Based on IFS/ECMWF Data and Observations in the South of Portugal. *Forecasting* **2020**, *2*, 130–150, doi:10.3390/forecast2020007.
5. Araujo, J.M.S. de Improvement of Coding for Solar Radiation Forecasting in Dili Timor Leste—A WRF Case Study. *Journal of Power and Energy Engineering* **2021**, *09*, 7–20, doi:10.4236/jpee.2021.92002.
6. Hogan, R. Radiation in the next Generation of Weather Forecast Models : Workshop Report. **2018**, 1–3.

7. Yang, D.; Wang, W.; Gueymard, C.A.; Hong, T.; Kleissl, J.; Huang, J.; Perez, M.J.; Perez, R.; Bright, J.M.; Xia, X.; et al. A Review of Solar Forecasting, Its Dependence on Atmospheric Sciences and Implications for Grid Integration: Towards Carbon Neutrality. *Renewable and Sustainable Energy Reviews* **2022**, *161*, 112348, doi:10.1016/j.rser.2022.112348.
8. Breitkreuz, H.; Schroedter-Homscheidt, M.; Holzer-Popp, T.; Dech, S. Short-Range Direct and Diffuse Irradiance Forecasts for Solar Energy Applications Based on Aerosol Chemical Transport and Numerical Weather Modeling. *J Appl Meteorol Climatol* **2009**, *48*, 1766–1779, doi:10.1175/2009JAMC2090.1.
9. Mayer, B.; Kylling, A.; Emde, C. LibRadtran User ' s Guide. **2012**.
10. Bran, S.H.; Jose, S.; Srivastava, R. Investigation of Optical and Radiative Properties of Aerosols during an Intense Dust Storm: A Regional Climate Modeling Approach. *J Atmos Sol Terr Phys* **2018**, *168*, 21–31, doi:10.1016/J.JASTP.2018.01.003.
11. Obiso, V.; Jorba, O. Aerosol-Radiation Interaction in Atmospheric Models: Idealized Sensitivity Study of Simulated Short-Wave Direct Radiative Effects to Particle Microphysical Properties. *J Aerosol Sci* **2018**, *115*, 46–61, doi:10.1016/j.jaerosci.2017.10.004.
12. Larson, V.E. Forecasting Solar Irradiance with Numerical Weather Prediction Models. *Solar Energy Forecasting and Resource Assessment* **2013**, 299–318, doi:10.1016/B978-0-12-397177-7.00012-7.
13. Verzijlbergh, R.A.; Heijnen, P.W.; de Roode, S.R.; Los, A.; Jonker, H.J.J. Improved Model Output Statistics of Numerical Weather Prediction Based Irradiance Forecasts for Solar Power Applications. *Solar Energy* **2015**, *118*, 634–645, doi:10.1016/J.SOLENER.2015.06.005.
14. Alkhayat, G.; Mehmood, R. A Review and Taxonomy of Wind and Solar Energy Forecasting Methods Based on Deep Learning. *Energy and AI* **2021**, *4*, 100060, doi:10.1016/J.EGYAI.2021.100060.
15. Alkhayat, G.; Hasan, S.H.; Mehmood, R. SENERGY: A Novel Deep Learning-Based Auto-Selective Approach and Tool for Solar Energy Forecasting. *Energies* **2022**, *15*, doi:10.3390/en15186659.

16. Delle Monache, L.; Eckel, F.A.; Rife, D.L.; Nagarajan, B.; Searight, K. Probabilistic Weather Prediction with an Analog Ensemble. *Mon Weather Rev* **2013**, *141*, 3498–3516, doi:10.1175/MWR-D-12-00281.1.
17. Alfadda, A.; Rahman, S.; Pipattanasomporn, M. Solar Irradiance Forecast Using Aerosols Measurements: A Data Driven Approach. *Solar Energy* **2018**, *170*, 924–939, doi:10.1016/J.SOLENER.2018.05.089.
18. Pang, Z.; Niu, F.; O'Neill, Z. Solar Radiation Prediction Using Recurrent Neural Network and Artificial Neural Network: A Case Study with Comparisons. *Renew Energy* **2020**, *156*, 279–289, doi:10.1016/J.RENENE.2020.04.042.
19. McCandless, T.C.; Young, G.S.; Haupt, S.E.; Hinkelman, L.M. Regime-Dependent Short-Range Solar Irradiance Forecasting. *J Appl Meteorol Climatol* **2016**, *55*, 1599–1613, doi:10.1175/JAMC-D-15-0354.1.
20. Fonseca, J.G.D.S.; Uno, F.; Ohtake, H.; Oozeki, T.; Ogimoto, K. Enhancements in Day-Ahead Forecasts of Solar Irradiation with Machine Learning: A Novel Analysis with the Japanese Mesoscale Model. *J Appl Meteorol Climatol* **2020**, *59*, 1011–1028, doi:10.1175/JAMC-D-19-0240.1.
21. Lima, F.J.L.L.; Martins, F.R.; Pereira, E.B.; Lorenz, E.; Heinemann, D. Forecast for Surface Solar Irradiance at the Brazilian Northeastern Region Using NWP Model and Artificial Neural Networks. *Renew Energy* **2016**, *87*, 807–818, doi:10.1016/j.renene.2015.11.005.
22. Pereira, S.; Abreu, E.F.M.; Iakunin, M.; Cavaco, A.; Salgado, R.; Canhoto, P. 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. *Solar Energy* **2022**, *236*, 225–238, doi:10.1016/j.solener.2022.03.003.
23. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting: A Review. *Renew Energy* **2017**, *105*, 569–582, doi:10.1016/j.renene.2016.12.095.
24. Zhou, Y.; Liu, Y.; Wang, D.; Liu, X.; Wang, Y. A Review on Global Solar Radiation Prediction with Machine Learning Models in a Comprehensive

Perspective. *Energy Convers Manag* **2021**, *235*, 113960, doi:10.1016/j.enconman.2021.113960.

25. Yang, B.; Zhu, T.; Cao, P.; Guo, Z.; Zeng, C.; Li, D.; Chen, Y.; Ye, H.; Shao, R.; Shu, H.; et al. Classification and Summarization of Solar Irradiance and Power Forecasting Methods: A Thorough Review. *CSEE Journal of Power and Energy Systems* **2021**, 1–19, doi:10.17775/CSEEJPES.2020.04930.
26. Liu, W.; Liu, Y.; Zhou, X.; Xie, Y.; Han, Y.; Yoo, S.; Sengupta, M. Use of Physics to Improve Solar Forecast: Physics-Informed Persistence Models for Simultaneously Forecasting GHI, DNI, and DHI. *Solar Energy* **2021**, *215*, 252–265, doi:10.1016/j.solener.2020.12.045.
27. Liu, W.; Liu, Y.; Zhang, T.; Han, Y.; Zhou, X.; Xie, Y.; Yoo, S. Use of Physics to Improve Solar Forecast: Part II, Machine Learning and Model Interpretability. *Solar Energy* **2022**, *244*, 362–378, doi:10.1016/j.solener.2022.08.040.
28. Voyant, C.; Notton, G.; Duchaud, J.; Gutiérrez, L.A.G.; Bright, J.M.; Yang, D. Benchmarks for Solar Radiation Time Series Forecasting. *Renew Energy* **2022**, *191*, 747–762, doi:10.1016/j.renene.2022.04.065.
29. Visser, L.; Alskaif, T.; Hu, J.; Louwen, A.; Sark, W. Van On the Value of Expert Knowledge in Estimation and Forecasting of Solar Photovoltaic Power Generation. *Solar Energy* **2023**, *251*, 86–105, doi:10.1016/j.solener.2023.01.019.
30. Jung, A. hyun; Lee, D.-H.; Kim, J.-Y.; Kim, C.K.; Kim, H.-G.; Lee, Y.-S. Spatial and Temporal Downscaling of Solar Radiation Using Statistical Techniques. *Journal of the Korean Solar Energy Society* **2020**, *40*, 89–94, doi:10.7836/kses.2020.40.6.089.
31. Yang, D.; Wu, E.; Kleissl, J. Operational Solar Forecasting for the Real-Time Market. *Int J Forecast* **2019**, *35*, 1499–1519, doi:10.1016/j.ijforecast.2019.03.009.
32. ECMWF IFS Documentation Available online: https://www.ecmwf.int/en/publications/ifs-documentation (accessed on 10 December 2021).

33. Hogan, R.J.; Bozzo, A. A Flexible and Efficient Radiation Scheme for the ECMWF Model. *J Adv Model Earth Syst* **2018**, *10*, 1990–2008, doi:10.1029/2018MS001364.
34. ECMWF CAMS: Global Atmospheric Composition Forecast Data Documentation Available online: https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation (accessed on 10 December 2021).
35. Abreu, E.F.M.; Gueymard, C.A.; Canhoto, P.; Costa, M.J. Performance Assessment of Clear-Sky Solar Irradiance Predictions Using State-of-the-Art Radiation Models and Input Atmospheric Data from Reanalysis or Ground Measurements. *Solar Energy* **2023**, *252*, 309–321, doi:10.1016/J.SOLENER.2023.01.051.
36. Iakunin, M.; Abreu, E.F.M.; Canhoto, P.; Pereira, S.; Salgado, R. Impact of a Large Artificial Lake on Regional Climate: A Typical Meteorological Year Meso-NH Simulation Results. *International Journal of Climatology* **2021**, 1–22, doi:10.1002/joc.7299.
37. Lopes, F.M.; Conceição, R.; Silva, H.G.; Fasquelle, T.; Salgado, R.; Canhoto, P.; Collares-Pereira, M. Short-Term Forecasts of DNI from an Integrated Forecasting System (ECMWF) for Optimized Operational Strategies of a Central Receiver System. *Energies* **2019**, *12*, doi:10.3390/en12071368.
38. Cavaco, A.; Canhoto, P.; Collares Pereira, M. Procedures for Solar Radiation Data Gathering and Processing and Their Application to DNI Assessment in Southern Portugal. *Renew Energy* **2021**, *163*, 2208–2219, doi:10.1016/J.RENENE.2020.10.075.
39. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2.0 Available online: https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 28 May 2024).
40. Shcherbakov, M.; Brebels, A.; Tyukov, A. A Survey of Forecast Error Measures. *World Appl Sci J* **2013**, *24*, 171–176, doi:10.5829/idosi.wasj.2013.24.itmies.80032.

41. Botchkarev, A. A New Typology Design of Performance Metrics to Measure Errors in Machine Learning Regression Algorithms. *Interdisciplinary Journal of Information, Knowledge, and Management* **2019**, *14*, 045–076, doi:10.28945/4184.
42. Jamil, B.; Akhtar, N. Comparative Analysis of Diffuse Solar Radiation Models Based on Sky-Clearness Index and Sunshine Period for Humid-Subtropical Climatic Region of India: A Case Study. *Renewable and Sustainable Energy Reviews* **2017**, *78*, 329–355, doi:10.1016/j.rser.2017.04.073.
43. Sharifi, S.S.; Rezaverdinejad, V.; Nourani, V. Estimation of Daily Global Solar Radiation Using Wavelet Regression, ANN, GEP and Empirical Models: A Comparative Study of Selected Temperature-Based Approaches. *J Atmos Sol Terr Phys* **2016**, *149*, 131–145, doi:10.1016/j.jastp.2016.10.008.
44. Pereira, S.; Canhoto, P.; Salgado, R.; João, M.; Costa, M.J. Development of an ANN Based Corrective Algorithm of the Operational ECMWF Global Horizontal Irradiation Forecasts. *Solar Energy* **2019**, *185*, 387–405, doi:10.1016/j.solener.2019.04.070.
45. Liu, B.; Wang, J.; Chen, J.; Li, B.; Sun, D.; Zhang, G. A Probabilistic Perspective on Predictability of Solar Irradiance Using Bootstrapped Correlograms and Ensemble Predictability Error Growth. *Solar Energy* **2023**, *260*, 17–24, doi:10.1016/J.SOLENER.2023.05.053.

Nomenclature

AOD469	Aerosol optical depth at 469 nm (–)
AOD550	Aerosol optical depth at 550 nm (–)
AOD670	Aerosol optical depth at 670 nm (–)
AOD865	Aerosol optical depth at 865 nm (–)
AOD1240	Aerosol optical depth at 1240 nm (–)
avgCF	Vertical average cloud fraction (–)
DHI	Diffuse horizontal irradiance (W/m^2)
DNI	Direct normal irradiance (W/m^2)
DUAOD	Dust aerosol optical depth at 550 nm (–)

GHI	Global horizontal irradiance (W/m^2)
maxCF	Vertical maximum cloud fraction (–)
OMAOD	Organic matter aerosol optical depth at 550 nm (–)
SSAOD	Sea salt aerosol optical depth at 550 nm (–)
T	Air temperature (℃)
WD	Wind direction (°)
WS	Wind speed (m/s)
Zen	Solar zenith angle (°)

Acronyms

ANN	Artificial Neural Network
BSRN	Baseline Surface Radiation Network
CAMS	Copernicus Atmosphere Monitoring Service
ECMWF	European Centre for Medium-range Weather Forecasts
FS	Forecast Skill
GFS	Global Forecasting System
GPI	Global Performance Index
NWP	Numerical Weather Prediction
MAE	Mean Absolute Error
ML	Machine Learning
MSE	Mean Squared Error
RF	Random Forest
rMSE	Relative Mean Squared Error
RMSE	Root Mean Squared Error
RRTM	Rapid Radiation Transfer Model
SVM	Support Vector Machines
WRF	Weather Research Forecasting

Chapter 4

Prediction of global solar irradiance on parallel rows of tilted surfaces including the effect of direct and anisotropic diffuse shading†

Abstract

Solar photovoltaic power plants typically consist of rows of solar panels, where the accurate estimation of solar irradiance on inclined surfaces significantly impacts energy generation. Existing practices often only account for the first row, neglecting shading from subsequent rows. In this work, ten transposition models were assessed against experimental data and a transposition model for inner rows was developed and validated. The developed model incorporates view factors and direct and circumsolar irradiances shading from adjacent rows, significantly improving global tilted irradiance (GTI) estimates. This model was validated against one-minute observations recorded between 14 April and 1 June 2022, at Évora, Portugal (38.5306, −8.0112) resulting in values of mean bias error (MBE) and root-mean-squared error (RMSE) of −12.9 W/m^2 and 76.8 W/m^2, respectively, which represent an improvement of 368.3 W/m^2 in the MBE of GTI estimations compared to the best-performing transposition model for the first row. The proposed model was also evaluated in an operational forecast setting where corrected forecasts of direct and diffuse irradiance (0 to 72 h ahead) were used as inputs, resulting in an MBE and RMSE of −33.6 W/m^2 and 169.7

†Sara Pereira[a], Paulo Canhoto[a,b] and Rui Salgado[a,c] 2024. Prediction of Global Solar Irradiance on Parallel Rows of Tilted Surfaces Including the Effect of Direct and Anisotropic Diffuse Shading. Energies 2024, 17, 3444. https://doi.org/10.3390/en17143444.

[a] Institute of Earth Sciences, University of Évora, Évora, Portugal
[b] Department of Mechatronics Engineering, University of Évora, Évora, Portugal
[c] Physics Department, University of Évora, Évora, Portugal

W/m^2, respectively. These findings underscore the potential of the developed model to enhance solar energy forecasting accuracy and operational algorithms' efficiency and robustness.

Keywords
Solar radiation; Solar energy; Transposition model; Solar power plant; Forecast

4.1 Introduction

In recent decades, there has been a significant surge in the installed capacity of PV systems. In 2022 alone, solar photovoltaics comprised two-thirds of the new renewable energy capacity added to the grid, totaling 239 GW [1]. This growth is driven by global efforts to decarbonize the economy and achieve net-zero greenhouse gas emissions by 2050. The International Energy Agency (IEA) estimates that PV systems will continue to expand, with new capacity projected to exceed 700 GW by 2028 [2]. With the rapid increase in the installed capacity of solar energy systems worldwide and the characteristic variability of solar radiation, accurate estimation and forecasting of power output is becoming increasingly important. The factor that most affects this output is the irradiance incident on the solar energy collectors. However, weather stations, satellites, and numerical weather prediction (NWP) models usually provide global (GHI) and diffuse (DIF) irradiance data on the horizontal plane and direct irradiance on a plane normal to the sun's rays (DNI). Therefore, transposition models that can compute the irradiance incident on a tilted surface from the available variables are essential.

Various transposition models have been developed and studied over the years. They can be divided into analytical, semi-empirical, and empirical models [3]. Analytical models are based on the laws of physics and only require the geometric characteristics and position data of the surface, while semi-empirical and empirical models also rely on observation data.

Physical models are usually based on the sum of the direct, diffuse, and reflected irradiances on the tilted surface. The computation of the direct and reflected components is often the same for all models, with the computation of the diffuse irradiance being the distinct factor [4]. Among these models, some assume a uniform sky dome radiance with the same intensity in all directions, termed isotropic. Anisotropic models, on the other hand, define indices representing the irradiance intensity from different regions of the sky dome, such as the circumsolar region (an annular region surrounding the sun disk), the horizon-brightening region (a band along the horizon), and the background or remaining region often called the isotropic region.

Transposition models also tend to show varying degrees of performance depending on sky conditions, whether clear, cloudy, or overcast [5]. The complexity and variability of the shape and position of clouds make the accurate instantaneous evaluation of the diffuse component extremely difficult without knowing the distribution of sky radiance. A model considering this would be more complex and require not only the simple measurement of the irradiance on the tilted plane but many more variables that are not commonly measured, such as the spectral radiance. Thus, physical models must rely on assumptions that can result in high transient errors [6]. Some of the most referenced analytical physical models in the literature are reviewed in this work, such as the Klucher [7], Hay–Davies [8], and Modified Bugler [9]models. On the other hand, abundant research on the comparison of the different models at different locations and positions of the surfaces has been published [3,4,10,11]. From these, no single model emerges as the best for all studied locations/positions, but it has been widely demonstrated that anisotropic models tend to outperform isotropic models when compared to measured data because they consider the angular dependence of sky radiance, thus reflecting real-world conditions more closely compared to isotropic models.

Semi-empirical and empirical models, including machine learning approaches which are data-driven, tend to be site-dependent, which means that they are biased towards specific locations and/or tilt and azimuth angles

[12–16], and will not be considered here. While these models are gaining relevance and can provide substantial benefits in some situations, their applicability is often limited because extensive local experimental data are required for the model calibration. This increases the challenge and difficulty of generalizing them to different geographic locations or environmental conditions. Given the focus of this study on developing a transposition model that is applicable across various conditions without the need for site-specific parameter adjustments, these models were excluded to maintain a broader applicability and robustness of proposed approach.

Most transposition models have been developed for a tilted surface in an open field. However, a solar energy power plant, such as photovoltaic power plants, comprises several rows of modules, where rows other than the first have a different view of the sky dome and the ground compared to the first row. Therefore, when computing the global tilted irradiance (GTI) on a panel not located in the front row, the transposition model should be adjusted to account for the fraction of the sky dome obscured by the rows at the front, as well as the possible blocking of the direct beam and circumsolar irradiance from the sun disk and the reflected irradiance coming from the front row and the ground between them. Since in large solar power plants the number of panels in the first row can be much smaller than the number of panels in the remaining rows, having a more accurate transposition model for these panels can improve the estimation of power output. Few authors have focused on this aspect, with the most notable and recent works being those by Applebaum et al. [17], Varga and Mayer [18], and Tschopp et al. [19].

As mentioned above, one aspect of higher complexity in transposition models is the parametrization of diffuse irradiance. This component of the GTI is strongly related to the diffuse horizontal irradiance (DIF) and the global horizontal irradiance (GHI), which are typically measured in radiometric stations. However, if only GHI is measured, a separation method is needed to estimate DNI and DIF. If there are no measurements at the location of interest, estimations or forecast values need to be used instead. This increases the uncertainty associated with the estimation of GTI and,

consequently, of power output. Of particular interest in this work is the usage of forecast values of solar radiation components, with the goal of predicting power output in a time horizon up to three days with an acceptable accuracy. In this regard, when utilizing forecast DNI and DIF as inputs, a transposition model allows for the forecasting of GTI and power output of photovoltaic systems if coupled with thermal and electric models of photovoltaic systems.
These forecasts can be obtained from a global operational numerical weather prediction model (NWP) such as the Integrated Forecasting System (IFS) of the European Centre for Medium-range Forecasts (ECMWF) [20]. By incorporating constitutive and state equations describing physical phenomena in the atmosphere, subject to boundary and initial conditions, NWP models provide the evolution of the atmospheric state, which include surface-level GHI. In the case of IFS/ECMWFS, DNI is also included, allowing for the study and assessment of solar resources and operational forecasts [21,22].
This work presents a comprehensive review of transposition models, categorizing models that do not include shading or irradiance masking as "first-row models" and those that include these aspects as "inner-row models". The development of a transposition model for inner rows is also presented, which can be applied to a desired first-row model. The proposed model is validated using experimental values, which are also used for the assessment and comparison of all models. The use of operational forecasts of solar irradiance is assessed to estimate the impact of the accuracy of the transposition model on GTI prediction at different time horizons. The primary aim is to develop a transposition model for inner rows of solar power plants, validate the proposed model against experimental measurements, and evaluate its performance using operational solar radiation forecasts. This study addresses gaps in the existing literature, which mainly focuses on the first row of panels, neglecting the impact of direct and anisotropic diffuse shading on inner rows. The hypotheses focus on the expected improvement in the accuracy of the developed model's results over existing models for the first

row, as well as its robustness when used with forecasted irradiance data as input.

This paper is organized as follows: An initial review of the most used analytical models and recent models developed for rows that are not the first is presented in Section 4.2. Section 4.3 details the development of a transposition model based on the works of Tschopp et al. and Varga and Mayer [18,19] for surfaces not located in the first row. Section 4.4 presents the experimental setup and procedure for testing the different models, and Section 4.5 discusses the evaluation and results. Section 4.6 establishes a connection with solar irradiance forecasting by applying the developed model to forecasted values of direct normal and diffuse horizontal irradiance, integrating it as an essential part of an operational algorithm for forecasting solar irradiance on a tilted surface. Finally, Section 4.7 presents the conclusions of this work.

4.2. Solar irradiance transposition models for tilted surfaces

This section provides a review of analytical transposition models, ranging from the most widely used and commonly referenced transposition models that do not consider shading and obscuration to the few and most recently proposed models designed for inner rows in solar power plants.

4.2.1 Transposition models for the first row of collectors in solar power plants

For modeling solar irradiance on a tilted surface (*GTI*), nine of the most commonly used analytical transposition models are reviewed, where one is isotropic and the remaining are anisotropic models. In all the following equations, α is the solar height, Φ is the solar zenith angle, γ_s is the solar azimuth, θ is the incidence angle on the surface, β is the tilt angle of the surface and γ_p is the azimuth of the surface, taking the horizontal plane and the local meridian as references.

Analytical transposition models include the computation of tilt factors for determining the beam (R_b), diffuse (R_d), and reflected (R_r) irradiances on the tilted surface. These factors are used in Eq. (4.1) for determining GTI, where DHI is the direct horizontal irradiance as defined in Eq. (4.2):

$$GTI = R_b DHI + R_d DIF + R_r GHI \tag{4.1}$$

$$DHI = DNI \times \cos(\Phi) \tag{4.2}$$

For the first row of photovoltaic panels, assuming no obstruction of the sun's rays, the direct irradiance on the tilted plane can be calculated at any given instant by multiplying DHI with the beam radiation tilt factor described in Eq. (4.3):

$$R_b = \frac{\cos(\theta)}{\cos(\Phi)} \tag{4.3}$$

The irradiance reflected onto the tilted surface by the surrounding surfaces is assumed isotropic and is modeled by multiplying GHI with the reflected irradiance tilt factor R_r (Eq. (4.4)) which consists of the product of the albedo of the surrounding ground (ρ_g) and the view factor F_g (Eq. (4.5)).

$$R_r = \rho_g F_g, \tag{4.4}$$

$$F_g = \frac{1 - \cos(\beta)}{2} \tag{4.5}$$

Numerous models have been proposed for determining the diffuse tilt factor R_d since the diffuse radiation is highly variable due to its dependency on atmospheric conditions. Models that consider the diffuse irradiance on a tilted surface coming from the entire sky-dome to have the same intensity are isotropic, while others that divide the sky-dome into regions with different intensities of diffuse radiance are called anisotropic models. These regions are usually the circumsolar region, which constitutes an annular region surrounding the Sun, the horizon brightening region, which is the region of the sky-dome near the horizon, and the remaining sky-dome region is termed the isotropic region.

In the following, several models are presented, including the widely used isotropic model that was developed by Liu and Jordan [23]. It assumes that the diffuse solar radiation is uniformly spread across the sky, which simplifies the calculation of solar radiation on inclined surfaces by considering the sky as a homogenous light source. This model only considers the tilt angle of the surface as input, as presented in Eqs. (4.6) and (4.7), where the diffuse tilt factor is simply a view factor between the surface and the sky. Its advantages lie in its simplicity and widespread adoption for preliminary solar energy designs and studies. However, the model's accuracy can be compromised due to its isotropic assumption, leading to potential inaccuracies in non-homogenous sky conditions and environments with significant obstructions. Despite these limitations, it remains a significant tool in solar energy research and engineering, often serving as a benchmark for more complex anisotropic models that account for the directional distribution of diffuse radiation.

Liu and Jordan model (LJ):

$$R_d = F_s \tag{4.6}$$

$$F_s = \frac{1 + \cos(\beta)}{2} \tag{4.7}$$

The remaining models analyzed are anisotropic models, which are often more representative of real sky conditions.

Bugler model (B):

The Bugler model [24] includes a circumsolar region in which its contribution to the diffuse irradiance is assumed to be 5 % of DNI and can be computed using Eq. (4.8), where F_s is the sky view factor given by Eq. (4.7). Bugler derived this model by analyzing solar radiation data and developing equations that incorporate these anisotropic factors, resulting in more accurate predictions of solar irradiance on tilted surfaces. However, this model overlooks the contribution of the circumsolar region to the isotropic irradiance and was later modified.

$$R_d = F_s + 0.05 \frac{DHI \times R_b}{DIF} \tag{4.8}$$

Temps and Coulson model (TC):

The Temps and Coulson [25] model, defined by Eq. (4.9), considers the circumsolar region through the factor $[1 + cos^2(\theta)sin^3(\Phi)]$ and the horizon brightening through $\left[1 + sin^3\left(\frac{\beta}{2}\right)\right]$. This model was obtained through measurements of direct and diffuse solar flux incident on slopes of various orientations for clear sky conditions through a pyranometer.

$$R_d = F_s[1 + cos^2(\theta)sin^3(\Phi)]\left[1 + sin^3\left(\frac{\beta}{2}\right)\right] \tag{4.9}$$

Klucher model (K):

The Klucher model [7], given by Eq. (4.10), is similar to the Temps and Coulson model, except for the inclusion of the clearness index f, defined by Eq. (4.11). Under overcast conditions, this model reduces to the isotropic model (where f tends to 0). It was derived from a 6-month dataset of measured hourly diffuse and total solar radiation on a horizontal plane and total radiation on surfaces with 0° azimuth and tilts of 37° and 60°.

$$R_d = F_s\left[1 + f sin^3\left(\frac{\beta}{2}\right)\right][1 + f cos^2(\theta)sin^3(\Phi)] \tag{4.10}$$

$$f = 1 - \left(\frac{DIF}{DHI + DIF}\right)^2 \tag{4.11}$$

Hay-Davies model (HD):

The Hay-Davies model [8], which does not consider horizon brightening, is given by Eq. (4.12). Here, A is named anisotropy index, also termed direct or beam irradiance index, and is defined in Eq. (4.13), where ENI is the extraterrestrial normal irradiance. This index represents the total transmittance of the atmosphere for beam radiation and is used to define the portion of circumsolar ($A \times R_b$) and isotropic ($F_s(1 - A)$) diffuse irradiance on the tilted surface.

$$R_d = A \times R_b + F_s(1 - A) \tag{4.12}$$

$$A = \frac{DNI}{ENI} \tag{4.13}$$

Ma Iqbal model (MI):

The Ma Iqbal model [26], described through Eq. (4.14), is similar to the Hay-Davies model but uses the clearness index k_T (Eq. (4.15)) to define the circumsolar portion of diffuse irradiance instead of the direct irradiance index. The clearness index is computed as the ratio of Global Horizontal Irradiance (GHI) to extraterrestrial horizontal irradiance (EHI). This formulation allows the Ma Iqbal model to account for varying sky conditions and solar geometry, enhancing its applicability in solar energy studies and irradiance forecasting.

$$R_d = k_T R_b + F_s (1 - k_T) \tag{4.14}$$

$$k_T = \frac{GHI}{EHI} \tag{4.15}$$

Modified Bugler model (MB):

The Modified Bugler model [9] improves the Bugler model by considering the contribution of the circumsolar region to the background isotropic diffuse radiance, computed through Eq. (4.16). This model was validated using extensive datasets from various locations around the world. These datasets included measured solar radiation data collected under different sky conditions, such as clear sky, cloudy sky, and overcast sky conditions. Additionally, the model was validated against experimental data that accounted for various surface orientations, including horizontal, tilted, and vertical surfaces.

$$R_d = \left(1 - 0.05 \frac{DHI}{DIF}\right) F_s + 0.05 \frac{DHI \times R_b}{DIF} \tag{4.16}$$

Modified Ma Iqbal model (MMI):

The modified Ma Iqbal model (Eq. (4.17)) [3] is similar to the original model proposed by this author but uses a clearness index (Eq. (4.18)) adjusted with an optical air mass as defined in Eq. (4.19) [27].

$$R_d = k'_T R_b + (1 - k'_T) F_s, \tag{4.17}$$

$$k'_T = \frac{k_T}{1.031exp\left(\frac{-1.4}{0.9+\frac{9.4}{M}}\right)+0.1} \tag{4.18}$$

$$M = \frac{1}{\cos(\Phi) + 0.15(93.885 - \Phi)^{-1.253}} \tag{4.19}$$

Reindl model (R):

The Reindl model [28], based on the Hay-Davies model, uses the same anisotropic index, A (Eq. (4.13)), but includes an additional term to account for horizon brightening (Eq. (4.20)). This model was developed based on datasets from locations across the USA.

$$R_d = A \times R_b + F_s(1-A)\left[1+\sqrt{\frac{DHI}{DHI+DIF}}sin^3\left(\frac{\beta}{2}\right)\right] \tag{4.20}$$

Perez model (P):

The Perez model, one of the most widely used transposition models in the literature, is included in this work, even though it cannot be classified as an analytical model since it is built upon measured data and is subject to continuous revisions for this reason. The version of the Perez model [29] used in this work, often referred to as Perez3, is considered the most widely accepted version of this model [4]. This model categorizes the sky in three zones: the circumsolar zone, horizon band, and isotropic zone. The diffuse irradiance tilt factor is computed using Eq. (4.21) through (4.27), with the assumption that the circumsolar irradiance originates from a point source and irradiance from the horizon from an infinitesimally thin region. The zenith angle is exceptionally used in radians for Eq. (4.26). The coefficients F_{ij} were determined using data from 10 American and 3 European locations and can be consulted in [4,29].

$$R_d = F_1\frac{a}{b} + (1-F_1)F_s + F_2\sin\beta \tag{4.21}$$

$$a = \max\,(0, \cos\theta) \tag{4.22}$$

$$b = \max[\cos 85°, \sin(90° - \theta_z)] \tag{4.23}$$

$$F_1 = max[0, F_{11}(\varepsilon) + F_{12}(\varepsilon)\Delta + F_{13}(\varepsilon)\theta_z] \tag{4.24}$$

$$F_2 = F_{21}(\varepsilon) + F_{22}(\varepsilon)\Delta + F_{23}(\varepsilon)\theta_z \tag{4.25}$$

$$\varepsilon = \frac{GHI/DIF + 1.041\theta_z^3}{1 + 1.041\theta_z^3} \tag{4.26}$$

$$\Delta = \frac{DIF}{ENI\cos\theta_z} \tag{4.27}$$

4.2.2 Transposition models for surfaces not located in the front row of a solar power plant

The models presented in the Section 4.2.1 were developed for a single row. In the case of photovoltaic power plants, for instance, which are composed of various parallel rows, there is a partial obscuring of the sky radiance and, eventually, the blocking of direct sun rays by the front rows over the second and subsequent rows, thus affecting the global tilted irradiance on these surfaces. An obscuring of the ground between rows can also occur, which affects the reflected irradiance.

Some authors have considered this, namely Appelbaum et al. [17], who adjusted the Klucher transposition model [7] by computing the sky-view factor for the second row and adjusting the indices of the circumsolar and horizon brightening diffuse regions. For this, the authors considered the circumsolar irradiance coming from a point source (the sun) which would be completely obscured when the obscuring angle from the sun to the base of the second row caused by the front row is higher than the sun elevation angle. The correction for the horizon brightening region was considered near sunrise and sunset. This adjusted model was validated with data obtained in Tel Aviv, showing that for rows other than the first the effect of the anisotropic region defined as horizon brightening is negligible.

Varga and Mayer [18] modified the Hay-Davies model [9], which does not include a horizon brightening region, to calculate the distribution of solar irradiance along a tilted surface for rows behind the first. Besides the adjustment of the view-factors, the circumsolar irradiance was corrected through the modeling of the fraction of the region obscured by the front row by considering a circumsolar region with an apparent angular radius of 15°, and the ground was divided into two sections: a shaded section that reflects only isotropic diffuse irradiance and an unshaded section that reflects global irradiance. The model was validated against power generation data from Hungary showing a clear improvement when compared to a model that only considers shading of the direct irradiance.

Tschopp et al. [19] also adjusted the Hay-Davies model for surfaces that are not on the front row by dividing the length of the surface of the row being evaluated, the back surface of the row at its front, and the ground between them in segments and then computing diffuse and direct irradiance in each segment. For this, the view-factors between all segments and between each segment and the sky are computed, while the obscuring of the direct and circumsolar irradiances are determined based on a simple formula for the shadow height on the surface. The model showed a significant improvement in the estimation of GTI for rows other than the first compared to the original transposition model.

Considering this, it is expected that combining the higher detail in the modeling of shadow and circumsolar irradiance obscuration presented in [18] and the inclusion of the back surface of the row at the front and the ground between rows presented in [19] will provide better estimations of GTI for rows that are not at the front of a solar power plant.

4.3 Development of transposition model for surfaces not located in the front row of a solar power plant

The model proposed in this work for determining solar irradiance on tilted rows adjacent to the first row builds upon the models presented in [18,19].

Similar to these models, it assumes rows of panels with lengths much greater than their heights, resulting in a 2D representation as commonly used in the literature [17–19,30,31].

While the aforementioned models only consider cases when the sun is positioned at the front of the rows, here, the modeling of the direct and circumsolar irradiance shading for all involved surfaces is also included, considering any position of the sun, resulting in a more realistic model. Additionally, the developed model was made generic and can be applied to any first-row transposition model as long as it clearly considers direct, circumsolar and isotropic irradiance, instead of relying on a predefined first-row model. Extra detail was also included in the modeling of ground shading by considering rows beyond the two main rows being modeled.

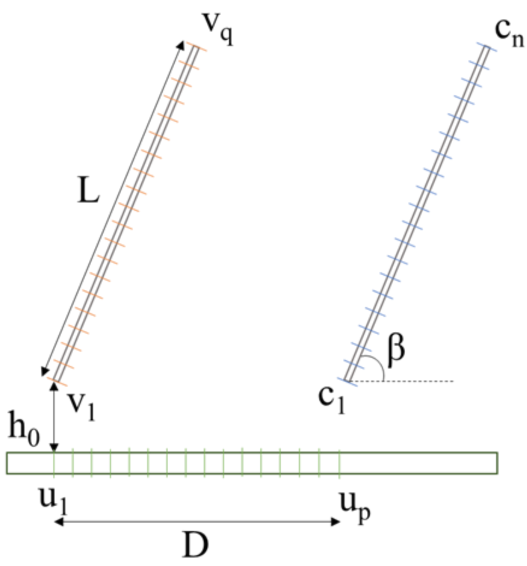


Fig. 4.1 - Schematic for modeling GTI in rows that are not the front row.

The three surfaces considered in this model, namely the front of the panel being evaluated, the back of the panel at its front, and the ground between rows, are divided into segments, as shown in Fig. 4.1, where L, D, h_0 and β are the length of the panels, the horizontal distance between rows, the vertical distance between the ground and the base of the rows, and the tilt angle, respectively. The value of GTI is computed for each segment c of the panel being evaluated and considers the reflected solar irradiance from each segment of the ground u and the back of the front panel v. The number of segments into which these surfaces are divided is defined by the user and can be adjusted, namely the number of segments of the panel being evaluated $i = 1{:}n$, the number of segments of the ground $j = 1{:}p$, and the number of segments of the back of the front panel $k = 1{:}q$.

For each segment of both panels and ground the GTI is computed as in [19] using Eq. (4.28), where $\overrightarrow{GTI}$, $\vec{I}_r$ and $\vec{D}$ are vectors with the values of the global tilted irradiance, the direct normal irradiance, and the diffuse horizontal irradiance, respectively, on each segment of the panel, ground and front panel. $\boldsymbol{I}$ is the identity matrix while $\boldsymbol{F}$ is the view-factor matrix between all segments and $\boldsymbol{R}$ is the reflectivity matrix for each segment. The view-factors between all segments are computed using the Hottel crossed-string rule [32], and the reflectivity of the front of the panel is assumed to be 0 as in [19].

$$\overrightarrow{GTI} = (\boldsymbol{I} - \boldsymbol{FR})^{-1}(\vec{I}_r + \vec{D}), \tag{4.28}$$

$$\vec{I}_r = DHI \times \vec{R}_b \tag{4.29}$$

$$\vec{D} = DIF \times \vec{R}_d \tag{4.30}$$

$$\vec{R}_b = max\left(0, \frac{cos\vec{\theta}}{cos\vec{\Phi}}\right)(1 - \vec{S}) \tag{4.31}$$

$$\vec{R}_d = \vec{F}_s X_i + X_{cs}(1 - \vec{S}_{cs}) \tag{4.32}$$

For the computation of the transposed direct normal (Eq. (4.29)) and diffuse (Eq. (4.30)) irradiance on each segment, the vectors $\vec{R}_b$ and $\vec{R}_d$ are obtained from Eqs. (4.31) and (4.32) which consist of the tilt factors for the direct (beam) and diffuse irradiances. The shading of the direct irradiance in each segment is taken into consideration in the direct tilt factor through the vector $\vec{S}$ whose values are either 0, when direct irradiance is not obscured, or 1, when the direct irradiance is obscured, depending on the geometrical characteristics of the installation and the apparent position of the sun in the sky. Regarding the diffuse tilt factor, an isotropic component, X_i, and a circumsolar component, X_{cs}, are modeled as in a first-row transposition model, which, as mentioned above, can be any first-row model as long as it includes direct, circumsolar and isotropic irradiance components. In the present work, a set of analytical transposition models for first rows are firstly assessed against experimental values and then one is selected, which will then be used in this model evaluation, as reported in the following sections. The sky view factors are computed for each segment considering the summation and reciprocity rules. The fraction of circumsolar irradiance that is obscured is modeled by the vector $\vec{S}_{cs}$ ranging from 0, when there is no obscuration, to 1, when all irradiance from the considered circumsolar region is obscured.

The vectors $\vec{S}$ and $\vec{S}_{cs}$ (Eqs. (4.33) and (4.34), respectively) are determined based on the model proposed by Varga and Mayer [18] with various modifications for the inclusion of cases in which the sun is positioned at the back of the row and the computations for the different segments of the back surface of the row in front of the one being evaluated and the ground. Depending on the surface of each segment, namely the surface of the panel being evaluated, c, the back surface of the panel of the row at its front, v, or the surface of the ground between them, u, the way the shading vectors are modeled differs. Since this is a two-dimensional model, firstly, the projection of the solar elevation angle to the azimuth of the panels, α', is needed, which is obtained through Eq. (4.35), where γ_s is the solar azimuth and γ_p is the azimuth of the surfaces.

$$\vec{S} = \begin{bmatrix} \vec{S}_c \\ \vec{S}_u \\ \vec{S}_v \end{bmatrix} \tag{4.33}$$

$$\vec{S}_{cs} = \begin{bmatrix} \vec{S}_{cs,c} \\ \vec{S}_{cs,u} \\ \vec{S}_{cs,v} \end{bmatrix} \tag{4.34}$$

$$\alpha' = \tan^{-1}\left(\frac{\tan\alpha}{\cos(\gamma_s - \gamma_p)}\right) \tag{4.35}$$

The shading of direct beam and the fraction of obscured circumsolar irradiance are determined for each segment of the panel, ground and back of the front panel according to the projected solar elevation angle relative to each of the angles shown in Fig. 4.2. These angles are computed using Eqs. (4.36) through (4.43), where i, j and k are the indices of the segments in the panel being evaluated, at the back of the front panel, and on the ground between the rows of panels, respectively. The angles with subscript c are obtained for each segment of the panel being evaluated, those with subscript v are obtained for each segment of the back of the front panel, and those with subscript u are obtained for each segment of ground between the two rows. The angles δ and ε result from the view of the top of an adjacent panel from a segment, the angles ζ and ξ result from the view of the bottom of an adjacent panel from a ground segment, and the angles λ and σ result from the view of the top of a subsequent panel from a ground segment.

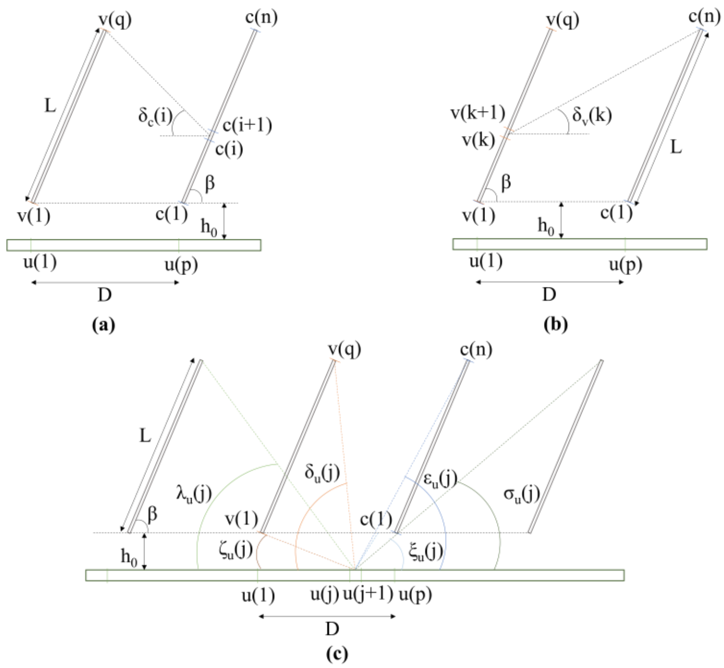


Fig. 4.2 - Schematic of the various angles for the computation of shadows and obscuring of circumsolar radiation for (a) a segment of the panel being evaluated, (b) the back of the front panel and (c) the ground between the rows of panels.

$$\delta_c(i) = \tan^{-1}\left[\frac{\left(L-\frac{c(i+1)-c(i)}{2}\right)\sin\beta}{D+\left(\frac{c(i+1)-c(i)}{2}-L\right)\cos\beta}\right]a = 1, \tag{4.36}$$

$$\delta_v(k) = \tan^{-1}\left[\left(\frac{\left(\mathrm{L}-\frac{v(k+1)-v(k)}{2}\right)\sin\beta}{\mathrm{D}}\right)\right] \tag{4.37}$$

$$\delta_u(j) = \tan^{-1}\left[\frac{\mathrm{L}\sin\beta + h_0}{\frac{u(j+1)-u(j)}{2} - \mathrm{L}\cos\beta}\right] \tag{4.38}$$

$$\varepsilon_u(j) = \tan^{-1}\left[\frac{\mathrm{L}\sin\beta + h_0}{D - \frac{u(j+1)-u(j)}{2} + \mathrm{L}\cos\beta}\right] \tag{4.39}$$

$$\zeta_u(j) = \tan^{-1}\left[\frac{h_0}{\frac{u(j+1)-u(j)}{2}}\right] \tag{4.40}$$

$$\lambda_u(j) = \tan^{-1}\left[\frac{\mathrm{L}\sin\beta + h_0}{D + \frac{u(j+1)-u(j)}{2} - \mathrm{L}\cos\beta}\right] \tag{4.41}$$

$$\xi_u(j) = \tan^{-1}\left[\frac{h_0}{D - \frac{u(j+1)-u(j)}{2}}\right] \tag{4.42}$$

$$\sigma_u(j) = \tan^{-1}\left[\frac{\mathrm{L}\sin\beta + h_0}{2D - \frac{u(j+1)-u(j)}{2} + \mathrm{L}\cos\beta}\right] \tag{4.43}$$

The shading of direct irradiance for each segment of the panel surface, i, is given by Eq. (4.44). Shading can occur ($S_c(i) = 1$) when the sun is positioned either in front of or behind the rows. Specifically, shading occurs when the projected solar elevation angle is positive and lower than the angle from the middle of the segment to the top of the front row (indicating that the sun is behind the front row), or when the projected solar elevation angle is negative and lower than the tilt angle of the surfaces (indicating that the sun is behind the panel). For the remaining cases, the segments are not shaded.

$$S_c(i) = \begin{cases} 1, & \left[\,\delta_c(i) > \alpha' \wedge \left|\gamma_s - \gamma_p\right| < 90°\right] \vee \left(\,\beta > |\alpha'| \wedge \left|\gamma_s - \gamma_p\right| \geq 90°\right) \\ 0, & \left[\,\delta_c(i) \leq \alpha' \wedge \left|\gamma_s - \gamma_p\right| < 90°\right] \vee \left(\,\beta \leq |\alpha'| \wedge \left|\gamma_s - \gamma_p\right| \geq 90°\right) \end{cases} \tag{4.44}$$

The shading of direct irradiance for each segment of the back surface of the front row, denoted by index k, is computed through Eq. (4.45). Each segment is always shaded if the sun is positioned in front of the rows. Additionally, if the sun is not in front of the rows, shading occurs when the projected solar elevation angle exceeds the tilt angle of the surfaces or when it is lower than the angle from the middle of the segment to the top of the panel being evaluated.

$$S_v(k) = \begin{cases} 1, & \left|\gamma_s - \gamma_p\right| < 90° \vee \left(\left|\gamma_s - \gamma_p\right| \geq 90° \wedge [\delta_v(k) > |\alpha'| \vee \beta < |\alpha'|]\right) \\ 0, & \left|\gamma_s - \gamma_p\right| \geq 90° \wedge \delta_v(k) \leq |\alpha'| \wedge \beta \geq |\alpha'| \end{cases} \tag{4.45}$$

The computation of the shading of direct irradiance on the ground between the rows is performed through Eq. (46), which involves comparing the projected solar elevation angle with six other angles as depicted in Fig. 4.2 and calculated through Eqs. (4.38) to (4.43).

$$S_u(j) =$$

$$\begin{cases} 1, & \begin{array}{l} \left|\gamma_s - \gamma_p\right| < 90° \wedge ([\zeta_u(j) < \alpha' < \delta_u(j) \wedge \delta_u(j) > 0] \vee \lambda_u(j) > \alpha' \vee [\delta_u(j) \leq 0 \wedge \alpha' > \zeta_u(j)]) \vee \\ \left|\gamma_s - \gamma_p\right| \geq 90° \wedge \left(\begin{array}{l} [\varepsilon_u(j) > \xi_u(j) \wedge \varepsilon_u(j) > |\alpha'| > \xi_u(j)] \vee [\varepsilon_u(j) \leq \xi_u(j) \wedge \varepsilon_u(j) < |\alpha'| < \xi_u(j)] \vee \\ [\delta_u(j) < 0 \wedge |\delta_u(j)| < |\alpha'|] \vee |\alpha'| \leq \sigma_u(j) \end{array} \right) \end{array} \\ 0, & \begin{array}{l} \left|\gamma_s - \gamma_p\right| < 90° \wedge ([\delta_u(j) \geq 0 \wedge \delta_u(j) \leq \alpha'] \vee [\lambda_u(j) \leq \alpha' \leq \zeta_u(j)]) \vee \\ \left|\gamma_s - \gamma_p\right| \geq 90° \wedge \left(\begin{array}{l} [\delta_u(j) \geq 0 \wedge \varepsilon_u(j) \leq |\alpha'| \wedge \xi_u(j) \leq |\alpha'|] \vee [\sigma_u(j) \leq |\alpha'| \leq \xi_u(j) \wedge |\alpha'| \leq \varepsilon_u(j)] \vee \\ [\delta_u(j) < 0 \wedge \varepsilon_u(j) \leq |\alpha'| \wedge \xi_u(j) \leq |\alpha'| \wedge |\delta_u(j)| \geq |\alpha'|] \end{array} \right) \end{array} \end{cases} \tag{4.46}$$

Each ground segment, denoted by index j, is shaded under the following conditions: when the sun is positioned in front of the rows and behind the row in front, or when its projected solar elevation angle is lower than the angle of the middle of the segment to the top of the subsequent row in front. Additionally, ground segments are shaded if the sun is behind the row being evaluated or when the projected solar elevation angle is lower than the angle from the middle of the segment to the top of the subsequent row behind.

Moreover, for ground segments positioned below the front row, shading occurs when the projected solar elevation angle exceeds the tilt of the surfaces.

The modeling of the circumsolar irradiance obscuring follows a similar approach, albeit with the utilization of the vector $\vec{S}_{cs}$ to represent the fraction of circumsolar irradiance that is obscured. Circumsolar irradiance is assumed as being uniformly distributed within the annular region surrounding the sun disk, with an apparent external angular radius, r, of 15°. The obscured area of the circumsolar region is given by Eq. (4.47) as in [18].

$$C(x) = \frac{sin^{-1}\frac{\sqrt{r^2-(|\alpha'|-x)^2}}{r}}{180} - \frac{\sqrt{r^2-(|\alpha'|-x)^2}}{2\pi r} \tag{4.47}$$

In the case of the segments of the panel being evaluated (Eq. (4.48)), several conditions dictate the obscuring of circumsolar irradiance.

$$S_{cs,c}(i) = \begin{cases} 1, & [|\gamma_s - \gamma_p| < 90° \wedge \alpha' - \delta_c(k) \le -r] \vee (|\gamma_s - \gamma_p| \ge 90° \wedge |\alpha'| - \beta \le -r) \\ |1 - C(\delta_c(i))|, & |\gamma_s - \gamma_p| < 90° \wedge -r < \alpha' - \delta_c(i) < 0 \\ |1 - C(\beta)|, & |\gamma_s - \gamma_p| \ge 90° \wedge -r < |\alpha'| - \beta < 0 \\ C(\delta_c(i)) & |\gamma_s - \gamma_p| < 90° \wedge 0 \le \alpha' - \delta_c(i) < r \\ C(\beta), & |\gamma_s - \gamma_p| \ge 90° \wedge 0 \le |\alpha'| - \beta < r \\ 0, & [|\gamma_s - \gamma_p| < 90° \wedge \alpha' - \delta_c(i) \ge r] \vee (|\gamma_s - \gamma_p| \ge 90° \wedge |\alpha'| - \beta \ge r) \end{cases} \tag{4.48}$$

When the sun is positioned in front of the rows, total obscuration ($S_{cs,c}(i) = 1$) occurs if the entire circumsolar region is below the angle defined by the middle of the segment to the top of the front row, denoted as $\delta_c(i)$. Substantial obscuration (more than 50%, $S_{cs,c}(i) = |1 - C(\delta_c(i))|$) occurs if this angle is higher than the projected solar elevation angle and the difference between these two angles is smaller than the angular radius. It is less obscured (less or equal to 50%, $S_{cs,c}(i) = C(\delta_c(i))$) if the projected solar elevation angle is higher than $\delta_c(i)$ and the difference between these two angles is smaller than the angular radius. Finally, no obscuration ($S_{cs,c}(i) = 0$) occurs if the projected solar elevation angle exceeds $\delta_c(i)$ and the difference between these two angles is higher than the angular radius. A similar principle applies when the

sun is behind the rows, with the difference that instead of using the angle $\delta_c(i)$, the comparison is made with the tilt angle β.
The fraction of circumsolar irradiance obscured for each segment of the back surface of the front panels is given by Eq. (4.49). It is assumed that all circumsolar irradiance is obscured when the sun is positioned in front of the rows. When the sun is positioned behind the rows, the projected solar elevation angles plus or minus the circumsolar angular radius are compared in a similar manner to Eq. (4.48), but with reference to the tilt angle of the surface and the angle of the middle of the back surface segment to the top of the panel being evaluated.

$$S_{cs,v}(k)=\begin{cases}1, & \left(|\gamma_s-\gamma_p|<90°\right)\vee\left(|\gamma_s-\gamma_p|\geq 90°\wedge\left[|\alpha'|-\delta_v(k)\leq -r\vee|\alpha'|-\beta\geq r\right]\right)\\ \left|1-C\left(\delta_v(k)\right)\right|, & |\gamma_s-\gamma_p|\geq 90°\wedge -r<|\alpha'|-\delta_v(k)<0\\ |1-C(\beta)|, & \left(|\gamma_s-\gamma_p|\geq 90°\wedge\ 0<|\alpha'|-\beta<r\right)\\ C\left(\delta_v(k)\right) & |\gamma_s-\gamma_p|\geq 90°\wedge 0\leq|\alpha'|-\delta_v(k)<r\\ C(\beta), & \left(|\gamma_s-\gamma_p|\geq 90°\wedge\ -r<|\alpha'|-\beta\leq 0\right)\\ 0, & \left[|\gamma_s-\gamma_p|\geq 90°\wedge|\alpha'|-\delta_v(k)\geq r\wedge\ |\alpha'|-\beta\leq -r\right]\end{cases}$$

(4.49)

The computation of circumsolar irradiance obscuration for each segment of the ground between rows is more complex (Eq. (4.50)). Complete obscuration is assumed when the entire circumsolar region is positioned either behind the front row panel or the panel being evaluated, considering the middle of the ground segment. Additionally, complete obscuration occurs when the angle of the top of the circumsolar region is lower than the angle of the top of the subsequent rows of panels, whether in front or behind. Similarly to the equations above (Eqs. (4.48) and (4.49)), the projected solar elevation angles plus or minus the circumsolar angular radius are compared with the six different angles shown in the scheme at the bottom of Fig. 4.2 for the computation of the fraction of circumsolar irradiance that is obscured.

$$S_{cs,u}(j) = \begin{cases} 1, & \left[|\gamma_s - \gamma_p| < 90° \wedge \begin{pmatrix} [\delta_u(j) \geq 0 \wedge \alpha' - \zeta_u(j) \geq r \wedge \alpha' - \delta_u(j) \leq -r] \vee \\ \alpha' - \lambda_u(j) \leq -r \vee \\ [\delta_u(j) < 0 \wedge \alpha' - \zeta_u(j) \geq r] \vee \\ [\alpha' - r \leq \lambda_u(j) \wedge \alpha' + r \geq \zeta_u(j)] \end{pmatrix} \right] \vee \\ & \left[|\gamma_s - \gamma_p| \geq 90° \wedge \begin{pmatrix} [|\alpha'| - \xi_u(j) \geq r \wedge |\alpha'| - \varepsilon_u(j) \leq -r] \vee \\ |\alpha'| - \sigma_u(j) \leq -r \vee \\ [\delta_u(j) \leq 0 \wedge |\alpha'| - |\delta_u(j)| \geq r] \vee \\ [|\alpha'| - r \leq \sigma_u(j) \wedge |\alpha'| + r \geq \varepsilon_u(j)] \vee \\ [|\alpha'| - r \leq \sigma_u(j) \wedge |\alpha'| + r \geq \xi_u(j)] \end{pmatrix} \right] \\ |1 - C(|\delta_u(j)|)|, & \left[|\gamma_s - \gamma_p| < 90° \wedge \delta_u(j) \geq 0 \wedge -r < \alpha' - \delta_u(j) < 0 \right] \vee \\ & \left[|\gamma_s - \gamma_p| \geq 90° \wedge \delta_u(j) < 0 \wedge 0 \leq |\alpha'| - |\delta_u(j)| < r \right] \\ |1 - C(\varepsilon_u(j))|, & |\gamma_s - \gamma_p| \geq 90° \wedge -r < |\alpha'| - \varepsilon_u(j) < 0 \\ |1 - C(\zeta_u(j))|, & |\gamma_s - \gamma_p| < 90° \wedge \alpha' - r > \lambda_u(j) \wedge 0 < \alpha' - \zeta_u(j) < r \\ |1 - C(\lambda_u(j))|, & |\gamma_s - \gamma_p| < 90° \wedge \alpha' + r < \zeta_u(j) \wedge -r < \alpha' - \lambda_u(j) < 0 \\ |1 - C(\xi_u(j))| & |\gamma_s - \gamma_p| \geq 90° \wedge |\alpha'| - r > \sigma_u(j) \wedge 0 < |\alpha'| - \xi_u(j) < r \\ |1 - C(\sigma_u(j))|, & |\gamma_s - \gamma_p| \geq 90° \wedge |\alpha'| + r < \xi_u(j) \wedge -r < |\alpha'| - \sigma_u(j) < 0 \\ C(|\delta_u(j)|) & \left[|\gamma_s - \gamma_p| < 90° \wedge \delta_u(j) \geq 0 \wedge 0 \leq \alpha' - \delta_u(j) < r \right] \vee \\ & \left[|\gamma_s - \gamma_p| \geq 90° \wedge \delta_u(j) < 0 \wedge -r \leq |\alpha'| - |\delta_u(j)| < 0 \right] \\ C(\varepsilon_u(j)), & |\gamma_s - \gamma_p| \geq 90° \wedge 0 \leq |\alpha'| - \varepsilon_u(j) < r \\ C(\zeta_u(j)), & |\gamma_s - \gamma_p| < 90° \wedge \alpha' - r > \lambda_u(j) \wedge -r < \alpha' - \zeta_u(j) < 0 \\ C(\lambda_u(j)), & |\gamma_s - \gamma_p| < 90° \wedge \alpha' + r < \zeta_u(j) \wedge 0 < \alpha' - \lambda_u(j) < r \\ C(\xi_u(j)), & |\gamma_s - \gamma_p| \geq 90° \wedge |\alpha'| - r > \sigma_u(j) \wedge -r < |\alpha'| - \xi_u(j) < 0 \\ C(\sigma_u(j)), & |\gamma_s - \gamma_p| \geq 90° \wedge |\alpha'| + r < \xi_u(j) \wedge 0 < |\alpha'| - \sigma_u(j) < r \\ 0, & \left[|\gamma_s - \gamma_p| < 90° \wedge \begin{pmatrix} [\delta_u(j) \geq 0 \wedge \alpha' - \delta_u(j) \geq r] \vee \\ [\alpha' - \zeta_u(j) \leq -r \wedge \alpha' - \lambda_u(j) \geq r] \end{pmatrix} \right] \vee \\ & \left[|\gamma_s - \gamma_p| \geq 90° \wedge \begin{pmatrix} |\alpha'| - \varepsilon_u(j) \geq r \vee \\ [|\alpha'| - \xi_u(j) \leq -r \wedge |\alpha'| - \sigma_u(j) \geq r \wedge |\alpha'| - \varepsilon_u(j) \leq -r] \end{pmatrix} \right] \end{cases} \tag{4.50}$$

4.4 Experimental setup and procedure

In order to obtain observational data for both a first row and subsequent rows of panels with varying tilt angles and inter-row distances, a structure featuring a pyranometer for measuring GTI was constructed in an open field near Évora, Portugal (38.5306°, -8.0112°), as shown in Figs. 4.3 to 4.5.

The experimental setup consists of three frames: a base, a front frame, and a rear frame, with a pyranometer installed on the rear frame. The base was leveled, and two transversal bars on the sides ensured that both front and rear frames maintained the same tilt angle. The apparatus allows for adjustment with three degrees of freedom: tilt angle of the front and rear frames (β, from 20° to 90°) through the solidary adjustment of the inclination of both frames; distance between frames (D, from 0.80 m to 1.10 m) through three positions of where the front frame can be fixed to the base; and position of the pyranometer along the length of the rear frame (c_s, from 0.08 m to 1.08 m) by sliding the instrument along its supporting bar. The uncertainty on the measurements of each of these variables is 1 cm in the cases of distances and 1° in the case of the tilt angles.

To represent the adjacent row, three Alveopan bilaminate white polypropylene boards, with a total width W of 3.03 m and length L of 1.08 m, were installed in the front frame. The reflectivity of these boards was measured using a FieldSpec HandHeld 2 spectroradiometer (ASD, Inc., Boulder, CO, USA) [33], yielding an average reflectivity of 0.921. Although potential edge effects were acknowledged due to board sizes, these were not factored into the general model. For data collection purposes relevant to the assessment of transposition models applied to the first rows, these boards were removed.

Global tilted irradiance was measured using a Kipp & Zonen CMP11 pyranometer (Kipp & Zonen, Delft, The Netherlands), while global horizontal irradiance and reflected irradiances (for the computation of ground albedo) were measured using a Kipp & Zonen CM7B albedometer (Kipp & Zonen, Delft, The Netherlands). Both sensors were connected to a CR300 datalogger from Campbell Scientific (Shepshed, Loughborough, UK). Additionally, DNI, DIF, and GHI observation data were obtained from the Évora–PECS station of the DNI-ALENTEJO project network [34], located 5 m from the experimental setup.

The internal clock of the CR300 data logger used in the apparatus was synchronized with the data logger of the Évora–PECS station, both set to

UTC time. Sensor outputs were sampled at 1 Hz and mean, maximum, minimum, and standard deviation values were recorded every minute. Observations were corrected following the best practices in the field, namely the WMO recommendations and the BSRN (Baseline Solar Radiation Network) guide, including corrections for sensor zero offset and filters according to the BSRN quality control procedure, considering the extremely rare limits [35] and removing measurements for zenith angles equal or above 85°.

Fig. 4.3 - Experimental setup for measuring global tilt irradiance for different positions.

Fig. 4.4 - Overview of the experimental setup including a) the Évora – PECS station and b) the pyranometers used for albedo computations.

Prior to the field measurements, a calibration procedure was conducted specifically for the CMP11 pyranometer and CM7B albedometer using a reference CMP21 pyranometer (Kipp & Zonen, Delft, The Netherlands), according to the ISO 9847:1992 standard [36].

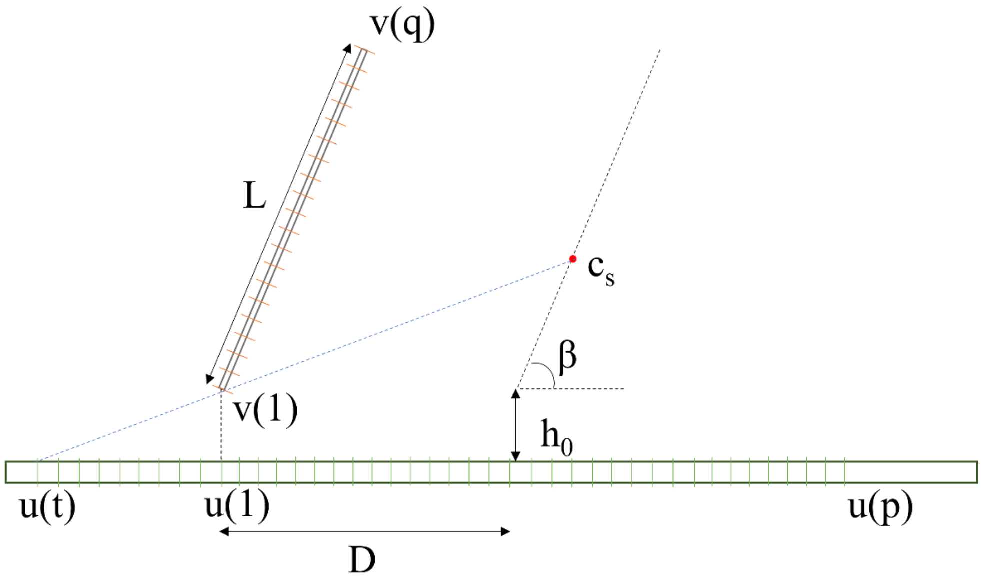


Fig. 4.5 - Schematic for modeling GTI on the sensor in the experimental apparatus.

For the accurate application of transposition models, the ground albedo value is needed. While a standard value of 0.2 is often used, this study estimated the ground albedo using Eq. (4.51) [37], where BSA and WSA represent the black-sky albedo and white-sky albedo, respectively. BSA is defined as the albedo in the absence of the diffuse component and is a function of solar zenith angle, while the WSA is the albedo considering the diffuse component as isotropic and in the absence of the direct component. The mean values of BSA and WSA were obtained by fitting Eq. (4.44) to experimental data, where the albedo, ρ, was computed by applying the ratio of reflected (GRI) to global horizontal (GHI) irradiance observations from the albedometer. Following data treatment, including filtering and removal of the records for solar zenith angles higher than 70°, BSA and WSA were found to be 0.206 and 0.208, respectively, across all recorded data periods.

$$\rho = BSA + (WSA - BSA)\frac{DIF}{GHI} \tag{4.51}$$

Observations were conducted between 14 April and 1 June, 2022. For first-row tests, 5 datasets or periods were generated, each corresponding to a specific tilt angle, as shown in Table 4.1, with the pyranometer positioned at c_s=0.50 m.

Testing of the developed model for other rows resulted in 19 periods with various tilt angles, distance between rows, and pyranometer positions, as shown in Table 4.2. Periods 4, 9, and 19 include instances in which the pyranometer is shaded.

It should be noted that measurements were conducted during a period of relatively high solar elevation, minimizing shading effects. To better capture shading effects, the inter-row distance during experimental tests were shorter than typical photovoltaic power plant configurations. Nonetheless, the developed model is designed to encompass diverse real-world conditions in the field.

Table 4.1 - Structure position and number of 1-min data points for each testing period of the first-row tests.

Period	**β (°)**	**Number of data points**
1	20	551
2	38	582
3	50	620
4	70	545
5	90	508
All	-	2806

Table 4.2 - Structure position and number of 1-minute data points for each testing period of the proposed model for rows other than the first.

Period	D (m)	β (°)	c_s (m)	Number of data points
1	1.10	38	0.08	585
2	1.10	38	0.58	604
3	1.10	38	1.08	1528
4	0.81	38	0.18	2345
5	0.81	38	0.38	605
6	0.81	38	0.58	615
7	0.81	38	0.78	608
8	0.81	38	0.98	585
9	0.80	50	0.15	581
10	0.80	50	0.35	2412
11	0.80	50	0.55	617
12	0.80	50	0.75	632
13	0.80	50	0.95	651
14	0.84	20	0.26	1807
15	0.84	20	0.46	365
16	0.84	20	0.66	648
17	0.84	20	0.86	612
18	0.84	20	1.06	633
19	0.81	38	0.18	718
All	-	-	-	17151
Shaded	-	-	-	2163
Unshaded	-	-	-	14988

4.5 Results and discussion

The different transposition models, including the developed model for rows that are not the first, were applied to the observations of DNI and DIF from the Évora – PECS station. Subsequently, the model outputs were compared with GTI observations from the experimental setup using multiple evaluation metrics developed in the software MATLAB R2018b. These metrics comprised the coefficient of determination (R^2), mean bias error (MBE), and root-mean-squared-error (RMSE) along with a global performance index (GPI) based on R^2, MBE, and RMSE, where a higher value represents the better accuracy of the model [22].

Given the significant influence of atmospheric conditions, particularly cloud cover, on global irradiance, the mean and standard deviation of the clearness index were computed for each period based on 1 min observations. The clearness index, typically ranging from 0 to 1, represents the ratio of global horizontal irradiance measured at ground level to its counterpart estimated at the top of the atmosphere [34] Eq. (4.15)). It serves as an indicator of the total transmittance of the atmosphere, reflecting higher values under clear-sky conditions and lower values under overcast conditions. The subsequent subsections provide detailed tables presenting these clearness index values for each period.

Some clearness index values exceeded unity, with a maximum value of 1.079, attributed to cloud enhancement events. These phenomena occur when partly cloudy skies lead to a temporary increase in local GHI above the extraterrestrial irradiance, facilitated by multiple scatterings and reflections by clouds [35]. These values were kept in the analysis, as they capture the transient nature of atmospheric conditions during the observation periods.

4.5.1 Results for the first row

The transposition models presented in Section 2.1 were computed for the periods shown in Table 4.1, with the resulting mean and standard deviation of the clearness index and GTI, alongside various evaluation metrics, summarized in Table 4.3.

Table 4.3 - Mean and standard deviation of the clearness index and GTI and metrics of GTI from the transposition models analyzed for each and all measuring periods of the first-row tests. The best performing model is represented in bold for each indicator and period and the clearness index data are repeated for each metric for better readability.

Metric	Period	k_t		GTI		Liu Jordan	Klucher	Hay Davies	Reindl	Ma Iqbal	Modified Ma Iqbal	Bugler	Modified Bugler	Temps Coulson	Perez
		Mean	Std	Mean (W/m^2)	Std (W/m^2)										
R^2	1	0.697	0.100	700.0	322.0	0.9995	0.9992	0.9995	0.9995	**0.9995**	0.9995	0.9994	0.9995	0.9987	0.9994
	2	0.720	0.081	702.1	330.0	0.9991	0.9988	0.9991	**0.9992**	0.9991	0.9991	0.9990	0.9991	0.9985	0.9991
	3	0.674	0.087	585.5	297.9	0.9955	0.9941	0.9965	0.9967	0.9959	0.9959	0.9946	0.9957	0.8341	**0.9975**
	4	0.686	0.141	474.8	288.4	**0.9946**	0.9936	0.9945	0.9938	0.9944	0.9944	0.9942	0.9944	0.9317	0.9934
	5	0.676	0.102	327.7	202.8	0.9982	0.9963	0.9978	0.9953	0.9979	0.9978	0.9991	0.9987	0.9837	**0.9993**
	All	0.690	0.105	558.1	323.6	0.9959	0.9969	0.9974	0.9975	0.9970	0.9970	0.9938	0.9947	0.9216	**0.9978**
MBE (W/m^2)	1	0.697	0.100	700.0	322.0	-44.4	**-28.8**	-36.6	-36.4	-38.6	-38.4	58.1	29.8	-43.7	-31.5
	2	0.720	0.081	702.1	330.0	-53.7	-40.7	-44.6	-43.9	-46.9	-46.7	51.2	**23.6**	-34.6	-36.5
	3	0.674	0.087	585.5	297.9	-66.5	-32.4	-48.6	-42.4	-53.9	-53.5	23.0	**5.3**	-157.4	-24.1
	4	0.686	0.141	474.8	288.4	-51.4	-28.4	-45.1	-38.4	-46.5	-46.5	37.2	**19.7**	-75.0	-24.0
	5	0.676	0.102	327.7	202.8	-23.2	**-2.0**	-27.3	-18.3	-25.8	-26.2	57.9	44.6	-19.6	-10.3
	All	0.690	0.105	558.1	323.6	-48.7	-27.1	-40.9	-36.4	-42.9	-42.8	44.8	**23.8**	-68.7	-25.6
RMSE (W/m^2)	1	0.697	0.100	700.0	322.0	49.7	36.1	41.3	41.2	43.5	43.3	59.1	**30.4**	45.8	37.5
	2	0.720	0.081	702.1	330.0	56.1	43.5	46.3	45.6	48.8	48.5	53.9	**25.8**	36.1	38.9
	3	0.674	0.087	585.5	297.9	71.9	38.0	51.7	45.1	57.6	57.2	27.4	**15.2**	181.8	26.9
	4	0.686	0.141	474.8	288.4	56.5	34.8	48.7	42.7	50.4	50.3	42.9	**26.7**	94.8	30.8
	5	0.676	0.102	327.7	202.8	24.9	11.8	28.3	21.7	27.0	27.3	62.6	48.2	41.9	11.3
	All	0.690	0.105	558.1	323.6	54.9	35.0	44.5	40.7	47.1	47.0	50.2	**30.5**	100.2	30.9
GPI	All	0.690	0.105	558.1	323.6	1.068	1.848	1.412	1.569	1.324	1.329	1.195	**1.959**	-1.000	1.953

Across all models, GTI values were generally underestimated, with the exception of the Bugler and Modified Bugler models, where GTI was overestimated. This discrepancy could stem from the treatment of the direct normal component within the factor R_b used for modeling the diffuse component. Nevertheless, the Modified Bugler model showed better results compared with the other models. The Modified Bugler model tends to perform best except for vertical surfaces (period 5), where the Klucher model seems to show better results. Given that the global performance index (GPI) for the overall data presented the Modified Bugler model as the best performing model, it was chosen as the primary model for first-row applications in the proposed model. Despite its tendency to overestimate GTI, with an overall MBE of 23.8 W/m^2 and RMSE of 30.5 W/m^2, it delivered optimal results for period 3, characterized by a tilt angle of 50° and small variation in sky conditions (standard deviation of clearness index of 0.087).

4.5.2 Results for other rows

For the evaluation of the developed transposition model for rows other than the first, some adjustments were implemented to accommodate the experimental setup (refer to Fig. 4.5). Given the absence of panels in the second row, this row was not considered in the modeling. Instead, GTI was computed for each segment of the back of the front panel and ground, followed by the calculation of GTI at a designated point which represents the sensor. In this case, the angles ε_u, λ_u, ξ_u, σ_u, and δ_v used for shadow computation and circumsolar irradiance obscuration were not applicable. Another modification involved the length of ground considered in the model. Since the setup comprised only one panel, reflections from the ground beyond the modeled rows could significantly impact the measured GTI and were thus incorporated into the model validation process (depicted in Fig. 4.5).

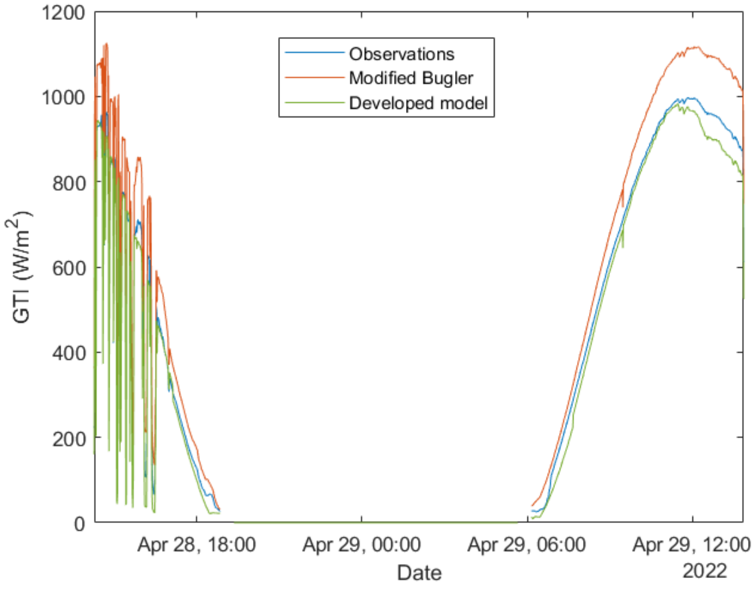


Fig. 4.6 - Global tilted irradiance observed and modeled by the Modified Bugler model (first-row model for reference and comparison) and the developed model for period 1 (1-minute timestep).

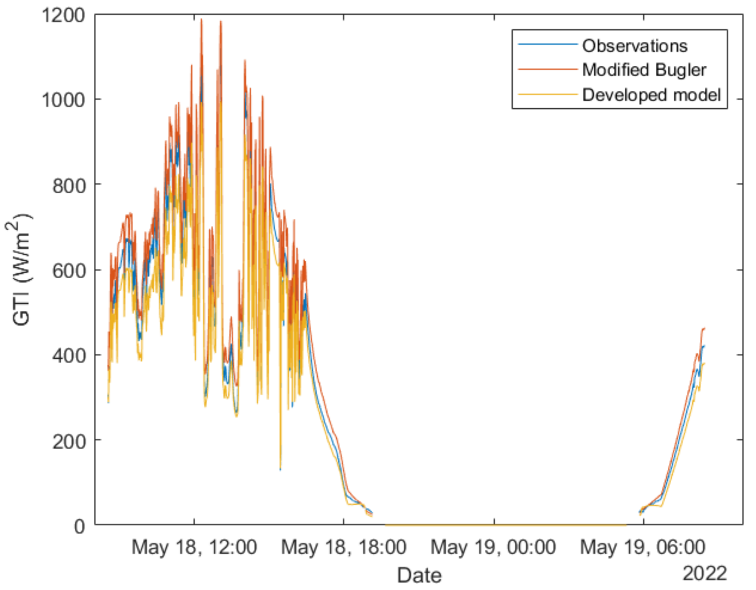


Fig. 4.7 - Global tilted irradiance observed and modeled by the Modified Bugler model (first-row model for reference and comparison) and the developed model for period 12 (1-minute timestep).

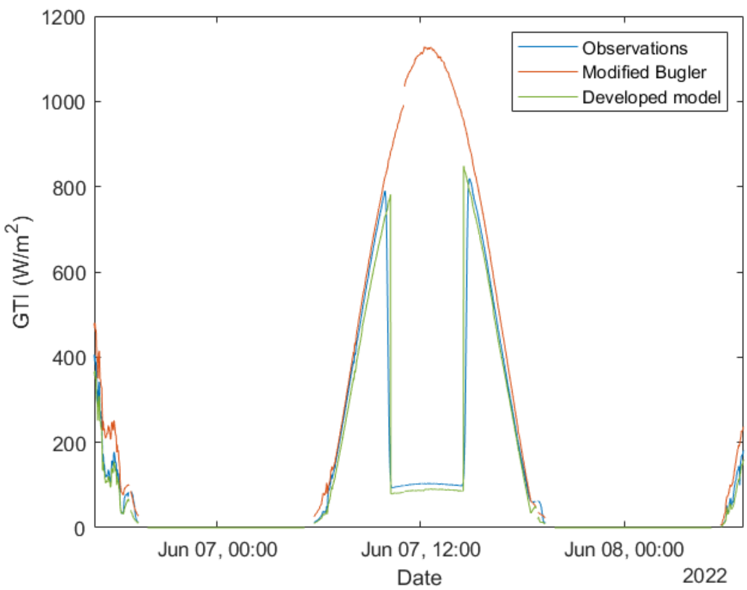


Fig. 4.8 - Global tilted irradiance observed and modeled by the Modified Bugler model (first-row model for reference and comparison) and the developed model for period 19 (1-minute timestep).

GTI estimation was performed using both the developed model and the Modified Bugler model for reference and comparison, which is a common practice in the absence of a specific model for other rows. It is important to note that the configurations used in periods 4, 9, and 19 result in direct shading of the pyranometer for a certain time span of the day. As example, periods 1 and 12, when there was no shading, and period 19, when there was shading and obscuration of the pyranometer by the front row, are shown in Fig. 4.6 through Fig. 4.8 (the small data gaps during the day are a result of the filtering procedure mentioned in Section 4.4). Despite the fact that slightly lesser improvements are observed for period 12 (Figure 4.6), which is attributed to partially cloudy conditions, the effectiveness of the developed model over the Modified Bugler is evident across the evaluated periods.

The results for each period, all periods, and for the data when the pyranometer is shaded or unshaded are presented in Table 4.4. When compared with the original Modified Bugler model, which overestimates the GTI, the proposed model improves the MBE for most periods, albeit with a slight underestimation. Typically, the Modified Bugler performs better in periods characterized by higher pyranometer positioning and greater frame-to-frame distances, resembling first-row irradiance conditions. During periods of direct irradiance shading (periods 4, 9, and 19), the Modified Bugler model, which does not consider shading, shows significantly higher errors. Another aspect to highlight is the impact of clouds in the performance of the model. In periods 9 and 14, for example, when the mean clearness index is lower and its standard deviation is higher (indicating cloudier skies) the metrics show lower performance of both models.

Due to the variation in sky conditions along the different periods, the comparison between the different positioning of the setup proved challenging and thus, more importance is given to the overall results instead of each period. In this regard, the developed model for rows affected by the presence of rows in front showed an MBE of -12.9 W/m^2 and a root-mean-squared-error of 76.8 W/m^2. As expected, this model outperforms the first-row model when the pyranometer is shaded. Even under unshaded conditions, the developed model is better than the model for the first row, showing the impact of the obscuring of the sky dome due to the other rows, considering the sky radiance anisotropy, namely the circumsolar region, and of the reflections from the front row and ground on the GTI.

To quantify the impact of the proposed model on reducing the error for each irradiance component on a tilted surface compared with the Modified Bugler model, a weight (w_i) for each component i of GTI was computed through Eq. (4.52).

$$w_i = \frac{1}{n}\sum \frac{i_D - i_{MB}}{GTI_D - GTI_{MB}} \tag{4.52}$$

Here, D stands for developed model and MB for the Modified Bugler model.

Table 4.4 - Mean and standard deviation of the clearness index and metrics of GTI from the Modified Bugler model and proposed model for other rows.

Period	**k_t**		**GTI**		**Modified Bugler model**			**Developed model for other rows**		
	Mean	Std	Mean (W/m²)	Std (W/m²)	R^2	MBE (W/m²)	RMSE (W/m²)	R^2	MBE (W/m²)	RMSE (W/m²)
1	0.687	0.077	566.1	342.1	0.9905	100.0	106.1	0.9921	-31.1	39.5
2	0.675	0.108	646.2	324.4	0.9988	75.6	76.6	0.9987	-38.5	44.8
3	0.584	0.204	456.2	338.0	0.9413	57.9	96.2	0.9410	-24.7	82.0
4	0.641	0.191	176.2	126.3	0.0873	521.5	651.9	0.9844	-15.7	22.7
5	0.704	0.095	627.8	335.9	0.9984	76.4	78.5	0.9990	-38.6	42.3
6	0.690	0.103	599.6	356.2	0.9995	68.4	70.8	0.9991	-39.4	42.3
7	0.672	0.098	582.9	345.2	0.9996	69.3	71.0	0.9996	-39.4	43.8
8	0.684	0.104	634.2	335.2	0.9990	36.6	37.6	0.9988	-69.5	77.3
9	0.540	0.206	248.4	187.0	0.1385	290.2	376.0	0.3225	-81.1	161.5
10	0.590	0.190	420.5	307.1	0.9769	103.4	112.0	0.9944	-42.3	50.6
11	0.674	0.145	548.9	307.6	0.9986	61.2	62.1	0.9962	-51.6	57.8
12	0.622	0.157	432.0	270.4	0.9949	52.4	55.2	0.9946	-54.0	62.1
13	0.561	0.134	436.4	272.7	0.9962	37.9	40.6	0.9961	-44.7	58.7
14	0.351	0.192	124.5	79.55	0.4669	250.6	304.9	0.4610	87.0	141.2
15	0.636	0.154	440.7	300.0	0.9980	28.3	30.3	0.9987	-62.7	69.6
16	0.524	0.275	481.7	370.5	0.9792	35.1	61.7	0.9793	-50.3	78.2
17	0.695	0.172	679.9	360.3	0.9907	41.5	50.6	0.9909	-65.9	78.0
18	0.677	0.147	643.9	334.1	0.9966	37.4	42.4	0.9964	-63.4	69.0
19	0.644	0.177	248.6	232.2	0.0372	373.2	570.2	0.9063	-12.4	75.5
All	0.638	0.188	323.8	304.6	0.4698	381.2	301.2	0.9543	-12.9	76.8
Shaded	0.682	0.197	203.7	169.2	0.0028	717.3	775.0	0.2585	-28.9	84.2
Unshaded	0.632	0.186	421.6	336.7	0.9047	88.9	131.0	0.9552	-26.5	75.7

The results are presented in Table 4.5 for the direct, I_b, diffuse circumsolar D_{cs}, diffuse isotropic, D_{iso}, and reflected, I_{refl}, components and for both unshaded and shaded conditions. For the proposed model, the reflected component includes reflections in the ground and in the back side of the front panel. The mean bias error of each model is also included in the table for reference.

Overall, the masking of the isotropic diffuse irradiance has the highest weight in the difference between the Modified Bugler and the developed model followed by the modeling of reflected irradiance. As expected, when the direct irradiance is shaded, this becomes the most impactful component, while for unshaded conditions, it has no impact.

Table 4.5 - MBE and weight of each component of GTI on the reduction of bias error when comparing the developed model with the Modified Bugler model.

Period	**MBE Modified Bugler (W/m²)**	**MBE Developed model (W/m²)**	w_{I_b} **(%)**	$w_{D_{cs}}$ **(%)**	$w_{D_{iso}}$ **(%)**	$w_{I_{refl}}$ **(%)**
Unshaded	88.9	-26.5	0.0	0.5	56.6	42.9
Shaded	717.3	-28.9	66.5	2.6	28.7	2.2
All	381.2	-12.9	8.4	0.8	53.1	37.7

4.6 Operational algorithm for forecasting of solar irradiance on tilted surfaces

This section presents the application of the developed model to predict irradiance on tilted surfaces using operational direct normal and diffuse irradiance forecasts as inputs.

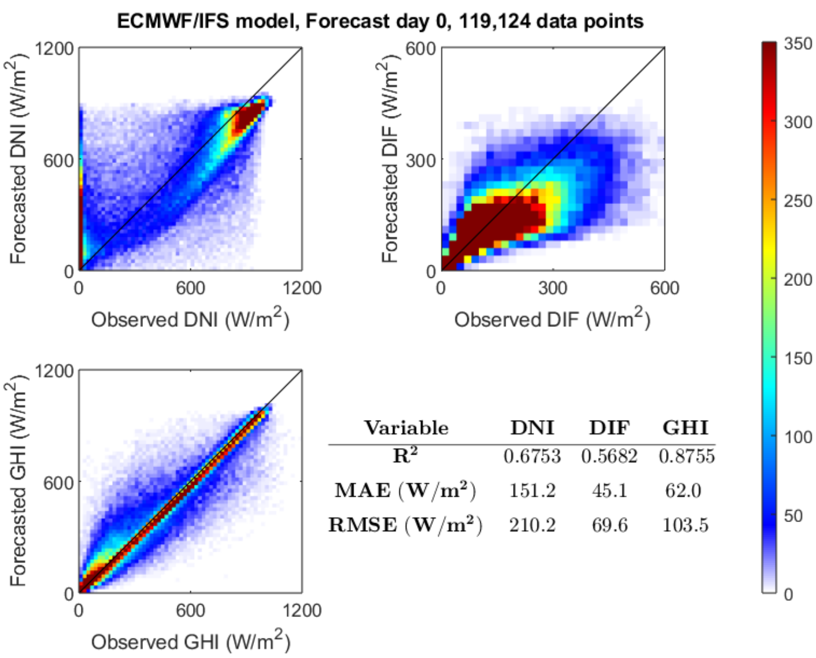
4.6.1 Forecast input data

The transposition models presented in this work, in particular the developed model for rows that are not the first, require several geometrical details on the solar panels of the power plant as input. However, only two meteorological variables, specifically DNI and DIF, are necessary for the determination of the GTI. The European Centre for Medium-range Weather Forecasts

(ECMWF) developed the Integrated Forecasting System (IFS), which issues global hourly forecasts every day at 00 UTC up to 10 days ahead. These forecasts include the GHI and DNI variables, while DIF can be computed through the closure function $GHI = BHI + DIF$ (where BHI is given by Eq. (4.2)).

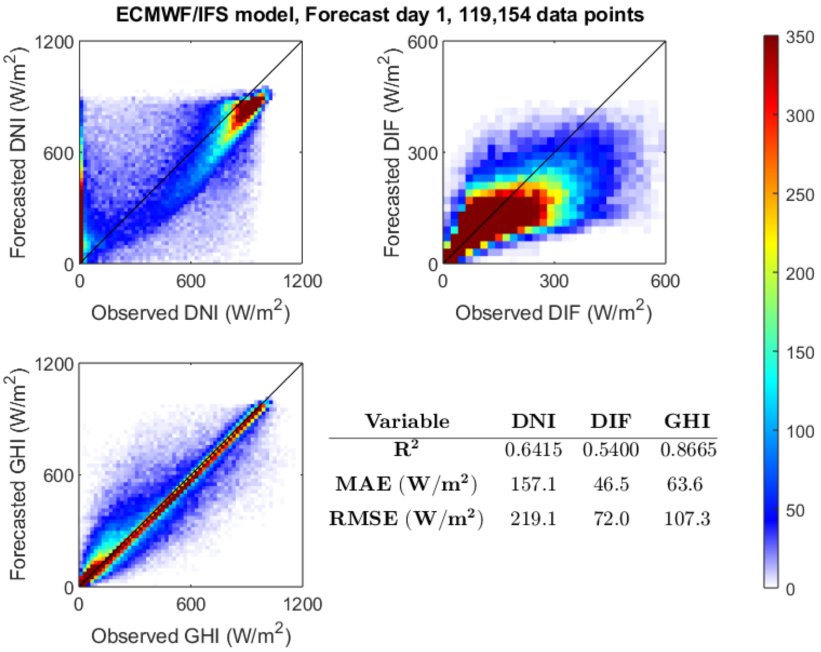
In the work by Pereira et al. [22], a method based on artificial neural networks (ANNs) was developed to generate improved spatial and temporal downscaled DNI forecasts using operational forecast outputs from the ECMWF/IFS and the Copernicus Atmospheric Monitoring Service (CAMS).

This work employs the same models and procedure, while also including the analysis and usage of GHI as compared with previous work. Fig. 4.9 to Fig. 4.11 provide a direct comparison between ECMWF/IFS forecasts and observations made through calibrated sensors at Évora – Verney station in Portugal (38.5678, -7.9114).



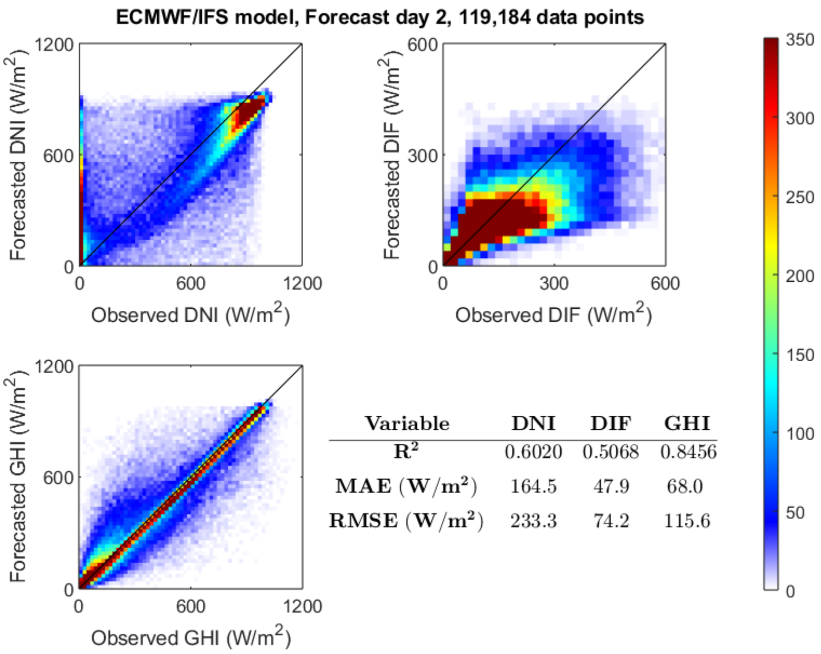
Variable	DNI	DIF	GHI
R^2	0.6753	0.5682	0.8755
MAE (W/m^2)	151.2	45.1	62.0
RMSE (W/m^2)	210.2	69.6	103.5

Fig. 4.9 - Comparison between downscaled 10-minute forecasts of the ECMWF/IFS and observations made at Évora Verney of DNI, DIF and GHI for forecast day 0 (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m^2).



Variable	DNI	DIF	GHI
R^2	0.6415	0.5400	0.8665
MAE (W/m^2)	157.1	46.5	63.6
RMSE (W/m^2)	219.1	72.0	107.3

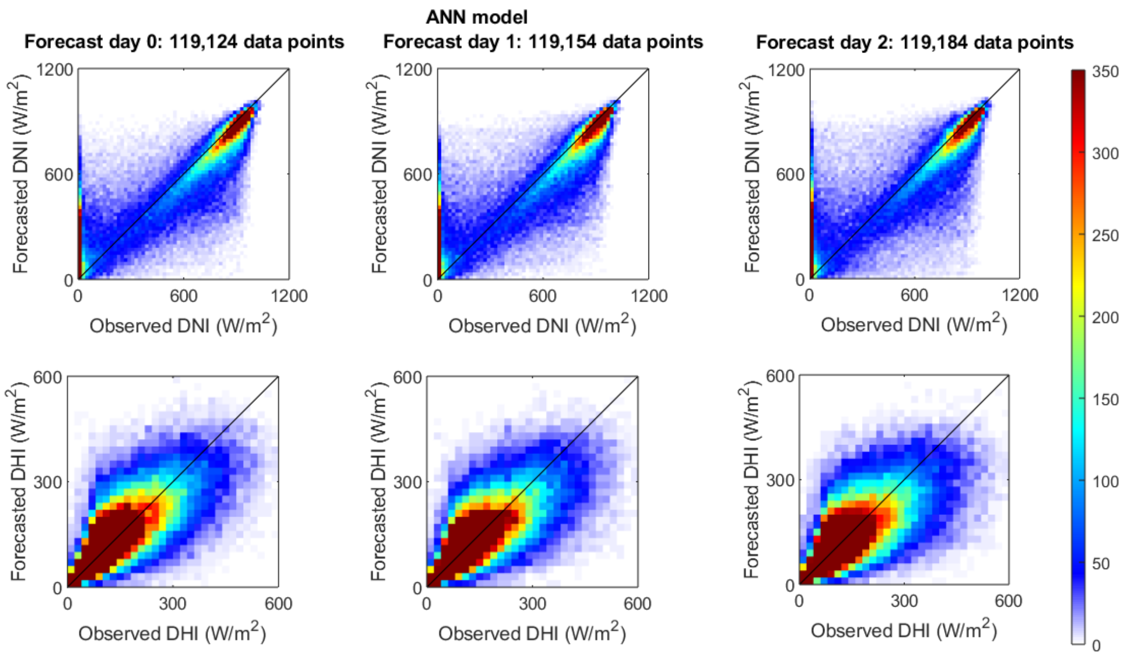
Fig. 4.10 - Comparison between 10-minute downscaled forecasts of the ECMWF/IFS and observations made at Évora Verney of DNI, DIF and GHI for forecast day 1 (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m²).



Variable	DNI	DIF	GHI
R^2	0.6020	0.5068	0.8456
MAE (W/m^2)	164.5	47.9	68.0
RMSE (W/m^2)	233.3	74.2	115.6

Fig. 4.11 - Comparison between10-minute downscaled forecasts of the ECMWF/IFS and observations made at Évora Verney of DNI, DIF and GHI for forecast day 2 (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m²).

These figures show data for DNI, GHI, and DIF for each forecast day of the forecast horizon (day 0: 1 h to 24 h; day 1: 25 h to 48 h; day 2: 49 h to 72 h) over the period from 1 December 2016 to the 31 May 2021, with a timestep of 10 min. As expected, shorter forecast horizons correspond to smaller errors in the forecasts. Improved DNI forecasts were obtained using artificial neural network models, accounting for the nonlinear relationships between the DNI and various meteorological forecasted variables, such as GHI, cloud cover, and aerosol optical depth, as well as the variation in DNI over a given period before the forecasted instant. A brief explanation of this model and procedure can be found in Appendix A. The application of this procedure results in the metrics shown in Fig. 4.12, where the diffuse component was computed using the GHI forecasted by the ECMWF/IFS and the improved DNI forecasts generated by the ANN models.



Variable	DNI	DHI
R^2	0.7498	0.6087
MAE (W/m^2)	125.5	42.2
RMSE (W/m^2)	184.4	66.6

Variable	DNI	DHI
R^2	0.7234	0.5883
MAE (W/m^2)	130.9	43.0
RMSE (W/m^2)	193	68.1

Variable	DNI	DHI
R^2	0.6789	0.5475
MAE (W/m^2)	139.2	45.4
RMSE (W/m^2)	209.0	71.8

Fig. 4.12 - Comparison between improved 10-minute forecasts and observations made at Évora Verney of DNI and DIF for forecast day 0, 1 and 2 (the colormap represents the number of data points in each bin. Bin size: 20x20 W/m²).

The application of the ANN model shows improvements in the forecasting of both DNI and DIF over a dataset that encompasses different sky conditions. The same procedure was applied by retrieving weather forecasts for the measurement periods used in this work to evaluate the different transposition models (Section 4.4). The resulting 10 min improved DNI and computed DIF forecasts were then evaluated against the 10 min mean observations obtained from the experimental data presented in Section 4.4.
Table 4.6 shows the results for first-row and inner-row tests. Detailed results for each testing period can be found in Appendix B. Similar to the larger dataset, the longer the forecast horizon, the higher the errors. However, the errors obtained tend to be greater than the ones aforementioned. This discrepancy is due to the smaller dataset used for this analysis, where even minor differences between forecasts and observations can lead to significant errors. Forecasts of DNI and DIF tend to have larger errors compared to global variables such as GHI or GTI, since accurate forecasting of clouds, specifically their location, is critical for this temporal resolution. The GTI results from the transposition models, based on the output of these forecast models, are presented in the following.

Table 4.6 - DNI and DIF forecast results for all data of first-row tests and for all, shaded and unshaded data of the developed model testing (MBE, MAE and RMSE in W/m²).

Data and Variable		Day 0				Day 1				Day 2			
		R²	**MBE**	**MAE**	**RMSE**	**R²**	**MBE**	**MAE**	**RMSE**	**R²**	**MBE**	**MAE**	**RMSE**
First-row tests													
All	DNI	0.4859	-38.5	87.2	132.1	0.2568	-41.5	107.4	169.2	0.2373	-99.8	151.2	225.7
	DIF	0.5258	-1.6	44.2	62.1	0.5540	4.0	41.3	60.3	0.5382	7.7	48.6	62.4
Inner-row tests													
All	DNI	0.5386	34.0	146.3	208.7	0.5458	39.4	145.1	208.6	0.4930	31.8	156.0	224.6
	DIF	0.4252	66.4	66.5	94.6	0.4720	64.0	64.0	91.5	0.4316	68.3	68.3	95.2
Shaded	DNI	0.6166	126.4	185.6	267.4	0.6488	134.8	183.6	263.8	0.5840	121.6	196.2	272.8
	DIF	0.4670	-49.7	92.8	126.8	0.5025	-53.1	97.1	126.9	0.4255	-44.4	101.0	131.0
Unshaded	DNI	0.5351	16.9	139.0	196.0	0.5389	21.8	138.0	196.7	0.4876	15.2	148.5	214.6
	DIF	0.3884	-8.9	61.6	87.4	0.4437	-17.6	57.8	83.3	0.4038	-15.6	62.3	87.1

4.6.2 Operational analysis of transposition models

To understand how the selected transposition models perform in an operational forecast context, the transposition models were applied using the resulting forecasts of DNI and DIF for the testing location with a 10-minute timestep. The results were then compared with the observations.

Table 4.7 presents the results for the first row using the Modified Bugler model and for inner rows using the developed model for each day of the forecast horizon.

Regarding first-row results, when compared with the results using DNI and DIF measurements instead of forecasts (Table 4.*3*), it is evident that for the analyzed periods, the use of forecasts shows better results for days 0 and 1 regarding MBE. However, there is a deterioration for day 2 and across all forecast horizons regarding the MAE and RMSE.
As for the inner-row test results, when compared with GTI based on DNI and DIF measurements instead of forecasts (Table 4.4), there is a general deterioration in the results, except for the MBE in unshaded conditions for forecast days 0 and 1. Detailed results for each testing period can be found in Appendix B. The use of the developed transposition model resulted in overall MBE and RMSE values of 33.6 W/m^2 and 169.7 W/m^2, respectively, for the entire forecast horizon. These values show that the model is beneficial for linking irradiance forecast models with energy generation modeling in solar power plants, aiding in the production of power output forecasts. These forecasts are essential for better decision-making by operators of such energy systems due to the variability of resource and energy demand.

Table 4.7 - GTI results of the developed model for rows that are not the front row for each day of forecast horizon using DNI and DIF forecasts (MAE and RMSE in W/m^2, MAPE and RMSPE in %).

Data	**Day 0**				**Day 1**				**Day 2**			
	R^2	**MBE**	**MAE**	**RMSE**	**R^2**	**MBE**	**MAE**	**RMSE**	**R^2**	**MBE**	**MAE**	**RMSE**
First-row tests												
All	0.9646	6.7	34.5	48.7	0.9607	15.5	39.4	55.0	0.8001	71.5	67.3	128.5
Inner-row tests												
All	0.7197	-31.5	100.6	164.0	0.7162	-32.1	101.1	165.6	0.6727	-37.2	106.8	179.4
Shaded	0.0074	-115.3	152.4	285.1	0.0242	-116.7	152.4	281.8	0.0078	-113.9	151.3	284.5
Unshaded	0.8031	-16.0	91.0	129.8	0.7907	-16.5	91.6	133.6	0.7318	-23.0	98.6	152.2

To understand how the mean bias error of DNI and DIF forecasts affects the bias of the computed GTI values, Fig. 4.13 was created. This figure shows the difference between the MBE of GTI from the developed model using forecasts and experimental data as input as a function of the MBE of DNI and DIF predictions for the forecast at day 0. Positive values of this difference are represented with blue circles, while negative values are represented with white circles. For better readability, only periods with an MBE difference of less than 100 W/m² are represented, which resulted in excluding only period 1.

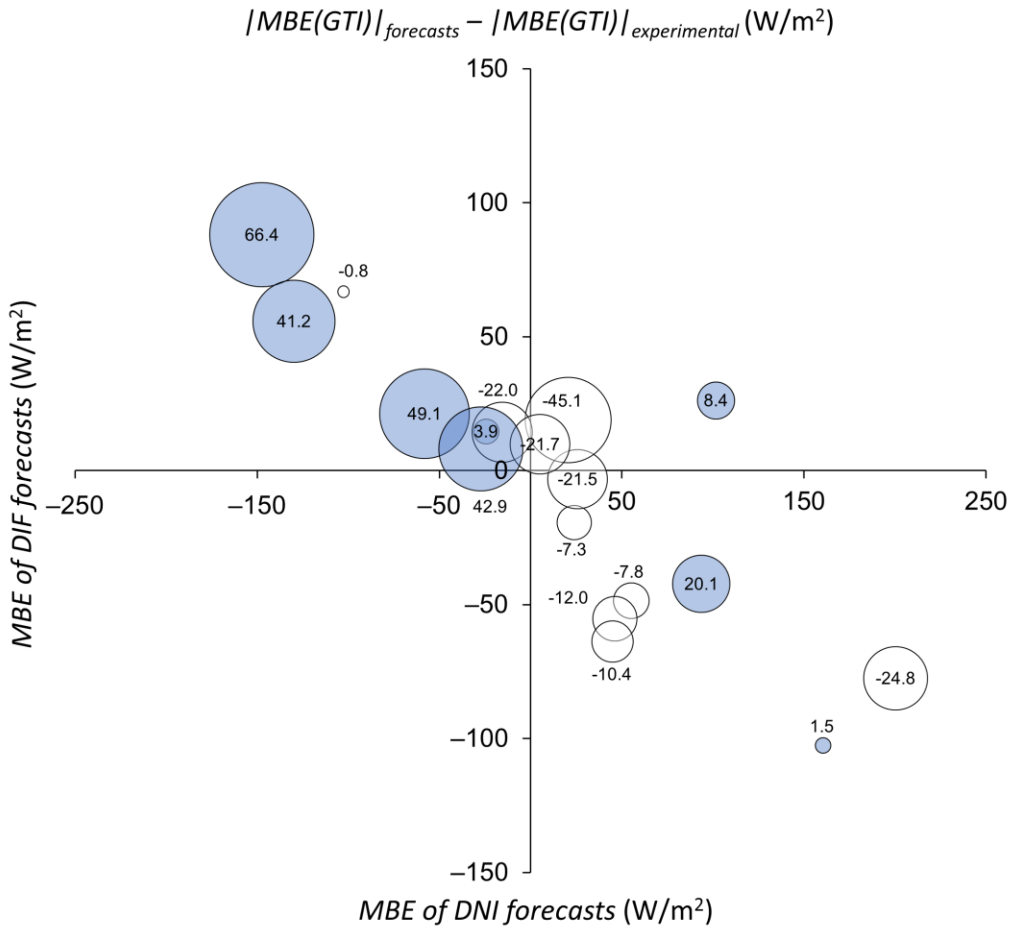


Fig. 4.13 - Difference between mean bias errors of GTI from the developed model results using forecasted or experimental data as input for day 0 (negative differences in white and positive differences in blue).

Since the MBE of the proposed model is typically negative for the tests carried out (Table 4.7 and 4.11), and considering the relationship between the MBE values for the forecasted DNI and DIF components (when one tends to higher positive values, the other tends to more negative values, Table 4.6 and 4.10, as a consequence of the closure equation and knowing that the MBE of the GHI forecasts is lower), the bias of the GTI using prediction values may decrease as shown in Fig. 4.13 due to a favorable combination of forecast and model errors. This is the reason why the MBE for unshaded conditions using forecast values (Table 4.7) is lower than the bias error for the same conditions using experimental values (Table 4.4). However, this trend is not observed for the shaded conditions, where the mean bias error is still lower when using ground-based measurements. A more detailed analysis of these aspects is needed in the future, which extends beyond the scope of the present work. Also, due to spatial resolution limitations, the forecasted variables from ECMWF (current operational horizontal resolution: ~9 km) are identical for the first and second rows in this experiment. In future work, this algorithm should be validated for a larger power plant that encompasses multiple grid points of a high-resolution forecast model at a hectometric scale.

4.7 Conclusions

This work presented a comprehensive analysis of nine analytical transposition models based on physics alongside the Perez transposition model to compute the global tilted irradiance on photovoltaic module surfaces. Additionally, it presented a model for the computation of this variable in rows of modules other than the first, which usually comprises most rows of solar power plants. The developed model can be applied to any first-row transposition model, provided it considers direct, circumsolar, and isotropic diffuse irradiance. This model computes the GTI for different longitudinal segments of the surfaces of the row of modules, the back of the row in front, and the ground between the rows. It takes into consideration the different view factors and the obscuring of direct and circumsolar irradiance for each of the segments for any apparent solar position and includes the shading effect of the succeeding rows on the ground segments.

The evaluation of these models utilized data collected in Évora, Portugal, for different tilt angles for first-row tests and also for different inter-row distances, including shading conditions, to assess the performance of the developed model. The clearness index helps address potential confounding variables by providing a baseline for sunshine conditions. However, we acknowledge that factors such as wind, relative humidity, precipitation, and aerosols also affect the experiment to some extent. In future research, these factors should be considered. Also, our dataset is limited to our experimental setup. While expanding to diverse locales would enhance generalizability, the focus of this study was on addressing challenges specific to inner rows of solar panels, rather than aiming for global applicability. Thus, while our findings offer valuable insights, they may not directly apply to all regions.

Results showed that the best analytical transposition model for the first row is the Modified Bugler model, showing an overall MBE of 23.8 W/m^2 and RMSE of 30.5 W/m^2. Conversely, for other rows, the developed model showed an MBE of -12.9 W/m^2 and RMSE of 76.8 W/m^2, resulting in an improvement of 368.3 W/m^2 and 224.4 W/m^2, respectively, compared to using the selected reference transposition model for first rows. This shows the importance of considering the direct shading and obscuring of the sky dome when computing GTI for surfaces in rows that are not the first.

Furthermore, the operational performance of transposition models was evaluated for GTI forecasting, using improved irradiance forecast values instead of measurements of DNI and DIF. These forecast values were obtained from artificial neural network models using numerical weather prediction and aerosol forecast data. Results of the first-row tests showed a MBE and RMSE for all data of 6.7 W/m^2 and 48.7 W/m^2 for forecast day 0, 15.5 W/m^2 and 55.0 W/m^2 for forecast day 1, and 71.5 W/m^2 and 128.5 W/m^2 for forecast day 2. This shows an increased error compared to results using observations which are a mean MBE and RMSE increase across the three days of forecast of 7.4 W/m^2 and 46.9 W/m^2. It also shows how the forecast performance tends to deteriorate with time. The same is visible for the tests performed for other rows, which show an overall MBE and RMSE of -31.5 W/m^2 and 164.0 W/m^2 for forecast day 0, -32.1 W/m^2 and 165.6 W/m^2 for forecast day 1 and -37.2 W/m^2 and 179.4 W/m^2 for forecast day 2.

This work demonstrated that transposition models that neglect shading and irradiance obscuration are not suitable for the accurate estimation of GTI in surfaces that are not in the front row of a solar power plant. The use of a dedicated model for these conditions, such as the one presented in this work, is of great importance, given that GTI is the main factor influencing the energy generation of solar photovoltaic systems.

Appendix A

The flowchart of the model used to generate forecast data of Direct Normal Irradiance (DNI) [22] used in this work is shown in Fig. 4.14.

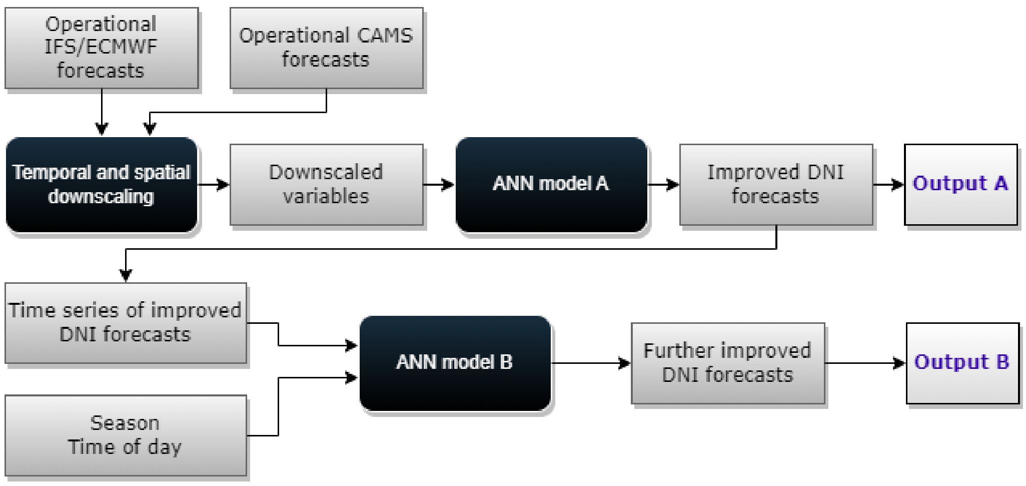


Fig. 4.14 - Flowchart of the model used to generate DNI forecasts [22].

Table 4.8 - Input variables obtained from numerical weather prediction models.

Variables obtained from IFS/ECMWF	Variables obtained from CAMS
Direct normal irradiance	Total aerosol optical depth at 670 nm
Global horizontal irradiance	Total aerosol optical depth at 865 nm
Low cloud cover	Total aerosol optical depth at 1240 nm
Medium cloud cover	Sea salt aerosol optical depth at 550nm
High cloud cover	
Total cloud cover	
Wind speed	
Air temperature	
Solar zenith angle	

The model takes as input data the variables shown in Table 4.8 from the operational Integrated Forecasting System (IFS) of the European Centre for Medium-range Weather Forecasts (ECMWF) and the Copernicus Atmospheric Monitoring Service (CAMS). These are run every day at 00 UTC, providing hourly forecast values up to 90 h ahead at discrete points of a global grid with a horizontal spatial resolution of 0.125° × 0.125°.

A temporal and spatial downscaling is performed on these variables, which results in forecasts for a specific location and higher temporal resolution. This downscaling is obtained through a bi-linear interpolation of the values in the four surrounding grid points of the desired location for the spatial downscaling and a piecewise cubic hermite interpolation of the hourly values into smaller timesteps (in this work, it was 10 min values) for the temporal downscaling.

The downscaled variables are then fed to an artificial neural network (ANN model A), which is a feed-forward network with one hidden layer comprising of 7 neurons and uses: (i) a backpropagation learning function, namely Bayesian regularization backpropagation; (ii) a linear layer output with an initialization function that initializes the weights and biases of the layers according to the Nguyen-Widrow initialization algorithm; (iii) the hyperbolic tangent sigmoid transfer function; and (iv) the mean-squared error as a performance function. The input and output data are processed by removing rows with constant values and scaling the mean of each row to 0 and deviations to 1. This ANN takes into consideration the nonlinear relationships between the different atmospheric and aerosol variables and DNI resulting in improved DNI forecasts for the specified location and temporal resolution.

A second artificial neural network (ANN model B) is similar to ANN model A, but uses the Levenberg-Marquardt backpropagation algorithm and has eight neurons in the hidden layer, taking as input a time series of 12 timesteps of the improved DNI forecasts prior to the forecast moment (from the output of the ANN model A) along with the season and time of day. This model takes into consideration the temporal variation in DNI and further improves the DNI forecasts from ANN model A.

Appendix B

Table 4.9 - Metrics of DNI and DIF forecasts for the measuring periods used for first-row tests (MBE, MAE and RMSE in W/m²).

Period and Variable		Day 0				Day 1				Day 2			
		R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE
1	DNI	0.5114	-33.1	57.0	87.6	0.6120	-14.9	38.3	67.9	0.4664	-15.3	72.2	99.5
	DIF	0.1228	18.5	26.5	39.1	0.2874	6.0	11.0	13.7	0.0912	-2.8	32.0	37.9
2	DNI	0.9082	-68.7	68.9	72.7	0.7794	-30.4	39.5	58.6	0.6228	-36.1	61.6	85.1
	DIF	0.8119	40.8	41.1	46.5	0.0063	6.8	15.9	21.1	0.0019	7.0	31.7	36.5
3	DNI	0.3542	-50.9	92.0	115.0	0.0803	-17.7	175.5	204.5	0.0014	-13.2	131.0	164.0
	DIF	0.6937	-8.7	59.7	70.4	0.5290	0.6	78.9	86.8	0.6843	-9.4	70.4	77.7
4	DNI	0.2470	47.4	74.0	127.7	0.2273	-33.2	79.1	128.7	0.1022	-322.1	322.1	391.3
	DIF	0.2873	-57.5	58.4	94.0	0.4425	-24.5	40.2	69.2	0.4884	11.7	43.2	69.3
5	DNI	0.4456	-86.6	149.2	220.4	0.1410	-121.6	206.9	283.7	0.2980	-132.1	180.7	256.1
	DIF	0.5380	-3.2	32.5	39.9	0.0513	33.5	57.9	70.8	0.0195	36.9	65.1	76.6

Table 4.10 - DNI and DIF forecast results for the measuring periods used for testing of the developed model (MBE, MAE and RMSE in W/m²).

Period and Variable		Day 0				Day 1				Day 2			
		R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE
1	DNI	0.1957	26.3	101.8	141.4	0.1127	14.1	92.3	146.1	0.0151	-28.4	128.6	162.6
	DIF	0.1687	-59.2	64.2	76.0	0.4663	-36.4	36.9	52.9	0.0006	4.7	59.5	67.6
2	DNI	0.9552	55.4	55.6	66.8	0.4768	28.5	59.8	80.9	0.9153	9.4	21.7	33.9
	DIF	0.5153	-48.5	48.5	52.3	0.0240	-20.4	31.3	40.1	0.2740	-23.8	30.7	39.8
3	DNI	0.6753	24.0	121.2	190.3	0.6811	-16.6	134.1	187.7	0.6930	2.9	134.2	184.3
	DIF	0.6857	-19.4	48.7	72.4	0.6798	-15.5	51.3	71.1	0.5439	-15.4	60.7	85.3
4	DNI	0.0879	46.2	178.6	273.2	0.1934	59.5	162.1	253.6	0.2277	51.8	162.0	251.3
	DIF	0.2560	-55.3	63.1	97.3	0.5052	-60.9	65.2	90.1	0.4056	-55.6	64.6	91.0
5	DNI	0.1872	-58.2	106.8	140.4	0.0197	-45.0	133.4	184.8	0.0065	0.4	83.2	149.6
	DIF	0.1999	21.2	40.3	46.0	0.0189	-11.8	38.5	58.5	0.0143	-4.1	35.6	56.4
6	DNI	0.3289	-147.5	169.6	210.3	0.3673	46.18	47.2	89.7	0.3013	38.0	46.6	89.3
	DIF	0.0152	88.0	96.9	129.0	0.0550	-33.5	33.5	46.3	0.0019	-25.4	28.5	44.2
7	DNI	0.8260	-24.4	36.7	44.5	0.8981	11.9	31.0	37.3	0.8394	37.6	49.0	65.9
	DIF	0.6205	14.6	22.6	28.4	0.4374	-3.9	23.9	26.6	0.0125	-25.9	38.1	49.2
8	DNI	0.8672	25.8	53.5	60.2	0.8169	30.5	74.9	85.3	0.8656	57.5	61.8	69.2
	DIF	0.5586	-3.2	11.8	15.8	0.3760	-13.7	26.8	30.6	0.0822	-26.5	26.5	30.9
9	DNI	0.1415	160.6	213.8	270.0	0.1635	99.9	184.8	223.0	0.0117	190.2	290.4	341.6
	DIF	0.3964	-102.6	115.7	143.9	0.0298	-96.5	126.6	157.6	0.0196	-102.5	131.1	161.4
10	DNI	0.3396	93.7	196.6	266.0	0.2661	33.3	204.1	268.5	0.4108	-20.3	201.3	235.9
	DIF	0.2564	-42.3	90.2	122.9	0.2894	-20.0	85.2	113.8	0.3719	-17.3	87.1	105.8
11	DNI	0.2717	-129.9	170.9	207.2	0.6058	-42.8	83.0	111.2	0.6257	22.0	69.2	103.2
	DIF	0.0832	55.7	70.8	96.5	0.0929	22.9	37.4	45.3	0.3765	-7.9	23.0	30.9
12	DNI	0.0420	45.0	197.2	228.6	0.1069	95.6	258.8	310.0	0.3188	197.5	295.7	363.3
	DIF	0.5633	-63.8	97.1	113.2	0.6490	-114.3	128.1	152.1	0.7377	-108.5	121.3	150.0
13	DNI	0.2652	200.3	217.9	265.3	0.2650	231.0	252.1	300.2	0.2762	262.2	299.9	341.5
	DIF	0.1240	-77.5	104.1	127.2	0.1153	-99.3	119.9	141.8	0.1656	-119.1	135.3	154.5
14	DNI	0.3747	101.8	137.5	163.5	0.4259	156.6	162.3	210.6	0.2201	107.0	145.8	186.7
	DIF	0.4555	26.2	89.6	111.4	0.4369	14.8	85.0	112.2	0.4314	55.2	95.9	122.8
15	DNI	0.1945	-102.7	125.5	161.5	0.2203	-148.3	173.5	190.0	0.2292	-116.1	139.4	170.9
	DIF	0.3872	66.8	68.7	93.0	0.4400	81.7	81.7	97.4	0.4533	73.0	76.7	102.6
16	DNI	0.0805	20.7	230.0	265.3	0.1606	-34.3	209.9	246.7	0.0197	-197.8	326.9	405.4
	DIF	0.5675	18.9	53.0	65.6	0.4792	31.2	64.2	77.2	0.7111	-15.6	46.0	52.8
17	DNI	0.3496	-15.5	92.7	146.0	0.4274	-37.6	119.2	158.3	0.4238	-96.0	151.4	205.5
	DIF	0.1479	14.3	29.4	38.5	0.2745	11.2	31.1	39.7	0.3782	31.2	41.1	50.6
18	DNI	0.0262	5.2	145.8	211.6	0.0983	19.4	139.8	198.7	0.0154	39.0	142.2	210.7
	DIF	0.0994	9.8	76.4	102.3	0.1921	3.7	72.6	95.8	0.0518	-4.0	81.5	111.6
19	DNI	0.7739	-27.3	65.6	99.7	0.8940	0.1	42.2	67.5	0.7421	-33.4	89.0	121.4
	DIF	0.4384	8.1	20.8	27.4	0.8150	-2.1	12.9	18.2	0.0654	34.8	45.4	60.6

Table 4.11- GTI metrics of first-row tests for each day of forecast horizon and for each period using the Modified Bugler transposition model with DNI and DIF forecasts (MBE, MAE and RMSE in W/m²).

Period	Day 0				Day 1				Day 2			
	R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE
1	0.9795	29.2	33.5	46.4	0.9798	29.5	35.6	46.4	0.9719	23.0	41.4	48.7
2	0.9980	8.9	13.8	16.6	0.9941	12.7	22.6	25.6	0.9933	11.0	23.8	27.7
3	0.9224	-29.4	45.4	63.9	0.9511	31.6	45.7	65.9	0.9345	16.3	36.4	56.3
4	0.9177	9.8	44.7	59.4	0.8715	19.8	51.8	74.6	0.0893	-203.9	210.1	277.7
5	0.9633	21.6	34.9	40.6	0.9329	21.8	41.7	48.4	0.9742	47.0	29.9	34.6

Table 4.12 - GTI results of the developed model for rows that are not the front row for each day of forecast horizon and for each period using DNI and DIF forecasts (MAE and RMSE in W/m², MAPE and RMSPE in %).

Period	Day 0				Day 1				Day 2			
	R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE	R^2	MBE	MAE	RMSE
1	0.0043	-163.0	207.1	364.9	0.0059	-163.5	195.8	364.0	0.0048	-179.4	215.5	365.9
2	0.9913	-30.7	34.9	38.8	0.9835	-41.0	42.7	57.2	0.9768	-50.3	53.2	70.5
3	0.6449	-17.4	117.5	175.0	0.6302	-36.0	129.9	181.5	0.5980	-25.8	128.2	187.2
4	0.3640	-3.7	81.8	137.8	0.3764	-3.0	84.0	138.3	0.3630	1.1	85.3	142.8
5	0.4550	-87.7	108.9	214.5	0.4017	-99.6	131.8	234.1	0.4038	-72.3	105.8	217.9
6	0.9196	-105.8	113.3	141.0	0.9651	-29.2	45.2	57.8	0.9636	-30.8	46.8	59.7
7	0.9938	-43.3	43.8	48.8	0.9981	-31.1	31.6	35.0	0.9971	-26.5	27.0	30.9
8	0.9958	-48.0	50.0	52.5	0.9963	-50.2	51.5	53.8	0.9949	-49.7	51.7	54.6
9	0.1442	-82.6	112.5	166.3	0.1479	-93.7	114.7	166.2	0.0815	-84.5	127.3	179.2
10	0.1607	-62.4	142.8	242.5	0.1907	-78.1	143.9	239.9	0.2246	-105.8	147.0	242.8
11	0.8884	-92.8	97.4	125.5	0.9728	-62.1	67.4	78.0	0.9823	-37.5	44.1	51.3
12	0.5099	-43.6	108.8	150.6	0.4232	-36.7	136.7	181.2	0.5488	44.0	111.4	167.7
13	0.9111	19.9	57.1	71.5	0.9154	28.6	54.7	71.8	0.9025	41.8	64.9	87.6
14	0.1142	95.4	105.1	135.2	0.1097	107.8	113.6	155.5	0.0710	111.3	121.3	155.7
15	0.8651	-61.9	74.8	91.6	0.8613	-68.5	87.4	96.9	0.8723	-64.5	78.0	91.8
16	0.4764	5.2	166.3	204.4	0.4233	-30.8	172.6	212.8	0.0174	-227.0	295.8	380.3
17	0.9180	-43.9	70.8	85.8	0.9211	-59.2	76.7	91.8	0.9147	-79.1	90.5	107.8
18	0.8772	-41.7	81.4	99.7	0.8803	-39.9	78.1	97.9	0.8741	-41.9	83.2	100.8
19	0.6672	-55.3	65.9	114.3	0.6605	-51.8	62.1	114.6	0.6384	-39.4	51.7	109.7

References

1. SolarPower Europe. *Global Market Outlook for Solar Power 2023–2027*; SolarPower Europe: Brussels, Belgium, 2023. Available online:

https://www.solarpowereurope.org/insights/market-outlooks/global-market-outlook-for-solar-power-2023-2027-1 (accessed on 5 June 2024).

2. *International Energy Agency Renewables 2023*; Analysis and Forecast to 2028; International Energy Agency: Paris, France, 2023.
3. Nassar, Y.F.; Hafez, A.A.; Alsadi, S.Y. 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.* **2020**, *7*, 1–19. https://doi.org/10.3389/fenrg.2019.00163.
4. Yang, D. Solar Radiation on Inclined Surfaces: Corrections and Benchmarks. *Sol. Energy* **2016**, *136*, 288–302. https://doi.org/10.1016/j.solener.2016.06.062.
5. Gueymard, C.A. Direct and Indirect Uncertainties in the Prediction of Tilted Irradiance for Solar Engineering Applications. *Sol. Energy* **2009**, *83*, 432–444. https://doi.org/10.1016/j.solener.2008.11.004.
6. Gueymard, C. An Anisotropic Solar Irradiance Model for Tilted Surfaces and Its Comparison with Selected Engineering Algorithms. *Sol. Energy* **1987**, *38*, 367–386. https://doi.org/10.1016/0038-092X(87)90009-0.
7. Klucher, T.M. Evaluation of Models to Predict Insolation on Tilted Surfaces. *Sol. Energy* **1979**, *23*, 111–114. https://doi.org/10.1016/0038-092X(79)90110-5.
8. Hay, J.E. Calculating Solar Radiation for Inclined Surfaces: Practical Approaches. *Renew. Energy* **1993**, *3*, 373–380. https://doi.org/10.1016/0960-1481(93)90104-O.
9. Hay, J.; McKay, D. Final Report IEA Task IX-Calculation of Solar Irradiances for Inclined Surfaces: Verification of Models Which Use Hourly and Daily Data; International Energy Agency: Paris, France, 1988.
10. Nassar, Y.F.; Abuhamoud, N.M.; Miskeen, G.M.; El-Khozondar, H.J.; Alsadi, S.Y.; Ahwidi, O.M. Investigating the Applicability of Horizontal to Tilted Sky-Diffuse Solar Irradiation Transposition Models for Key Libyan Cities. In Proceedings of the 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic

Control and Computer Engineering (MI-STA 2022), Sabratha, Libya, 23–25 May 2022; pp. 9–14. https://doi.org/10.1109/MI-STA54861.2022.9837500.

11. Mayer, M.J.; Gróf, G. Extensive Comparison of Physical Models for Photovoltaic Power Forecasting. *Appl. Energy* **2021**, *283*, 116239. https://doi.org/10.1016/j.apenergy.2020.116239.
12. Khan, M.M.; Ahmad, M.J. Estimation of Global Solar Radiation Using Clear Sky Radiation in Yemen. *J. Eng. Sci. Technol. Rev.* **2012**, *5*, 12–19. https://doi.org/10.25103/jestr.052.03.
13. Padovan, A.; Del Col, D. Measurement and Modeling of Solar Irradiance Components on Horizontal and Tilted Planes. *Sol. Energy* **2010**, *84*, 2068–2084. https://doi.org/10.1016/j.solener.2010.09.009.
14. Muneer, T.; Gueymard, C.; Kambezidis, H. Hourly Slope Irradiation and Illuminance. In *Solar Radiation and Daylight Models*; Butterworth-Heinemann: Oxford, UK, 2004; p. 143–I. https://doi.org/10.1016/B978-075065974-1/50012-7.
15. Badescu, V. 3D Isotropic Approximation for Solar Diffuse Irradiance on Tilted Surfaces. *Renew. Energy* **2002**, *26*, 221–233. https://doi.org/10.1016/S0960-1481(01)00123-9.
16. Tian, Y.Q.; Davies-Colley, R.J.; Gong, P.; Thorrold, B.W. Estimating Solar Radiation on Slopes of Arbitrary Aspect. *Agric. Meteorol.* **2001**, *109*, 67–74. https://doi.org/10.1016/S0168-1923(01)00245-3.
17. Appelbaum, J.; Massalha, Y.; Aronescu, A. Corrections to Anisotropic Diffuse Radiation Model. *Sol. Energy* **2019**, *193*, 523–528. https://doi.org/10.1016/j.solener.2019.09.090.
18. Varga, N.; Mayer, M.J. Model-Based Analysis of Shading Losses in Ground-Mounted Photovoltaic Power Plants. *Sol. Energy* **2021**, *216*, 428–438. https://doi.org/10.1016/J.SOLENER.2021.01.047.
19. Tschopp, D.; Jensen, A.R.; Dragsted, J.; Ohnewein, P.; Furbo, S. Measurement and Modeling of Diffuse Irradiance Masking on Tilted Planes for Solar Engineering Applications. *Sol. Energy* **2022**, *231*, 365–378. https://doi.org/10.1016/J.SOLENER.2021.10.083.

20. ECMWF IFS Documentation. Available online: https://www.ecmwf.int/en/publications/ifs-documentation (accessed on 10 December 2021).
21. Pereira, S.; Abreu, E.F.M.; Iakunin, M.; Cavaco, A.; Salgado, R.; Canhoto, P. 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* **2022**, *236*, 225–238. https://doi.org/10.1016/j.solener.2022.03.003.
22. Pereira, S.; Canhoto, P.; Salgado, R. Development and Assessment of Artificial Neural Network Models for Direct Normal Solar Irradiance Forecasting Using Operational Numerical Weather Prediction Data. *Energy AI* **2023**, *15*, 100314. https://doi.org/10.1016/j.egyai.2023.100314.
23. Liu, B.Y.H.; Jordan, R.C. The Long-Term Average Performance of Flat-Plate Solar-Energy Collectors. *Sol. Energy* **1963**, *7*, 53–74. https://doi.org/10.1016/0038-092X(63)90006-9.
24. Bugler, J.W. The Determination of Hourly Insolation on an Inclined Plane Using a Diffuse Irradiance Model Based on Hourly Measured Global Horizontal Insolation. *Sol. Energy* **1977**, *19*, 477–491. https://doi.org/10.1016/0038-092X(77)90103-7.
25. Temps, R.C.; Coulson, K.L. Solar Radiation Incident upon Slopes of Different Orientations. *Sol. Energy* **1977**, *19*, 179–184. https://doi.org/10.1016/0038-092X(77)90056-1.
26. Iqbal, M. *An Introduction to Solar Radiation*; Elsevier: Vancouver, BC, Canada, 1983; ISBN 9780123737502.
27. Kasten, F. A New Table and Approximation Formula for the Relative Optial Air Mass. *Arch. Für. Meteorol. Geophys. Und. Bioklimatol. Ser. B* **1965**, *14*, 206–223. https://doi.org/10.1007/BF02248840.
28. Reindl, D.T.; Beckman, W.A.; Duffie, J.A. Evaluation of Hourly Tilted Surface Radiation Models. *Sol. Energy* **1990**, *45*, 9–17. https://doi.org/10.1016/0038-092X(90)90061-G.
29. Perez, R.; Ineichen, P.; Seals, R.; Michalsky, J.; Stewart, R. Modeling Daylight Availability and Irradiance Components from Direct and Global

Irradiance. *Sol. Energy* **1990**, *44*, 271–289. https://doi.org/10.1016/0038-092X(90)90055-H.

30. Saint-Drenan, Y.-M.; Barbier, T. Data-Analysis and Modeling of the Effect of Inter-Row Shading on the Power Production of Photovoltaic Plants. *Sol. Energy* **2019**, *184*, 127–147. https://doi.org/10.1016/j.solener.2019.03.086.
31. Maor, T.; Appelbaum, J. View Factors of Photovoltaic Collector Systems. *Sol. Energy* **2012**, *86*, 1701–1708. https://doi.org/10.1016/j.solener.2012.03.017.
32. Eckert, E.R.G. Radiative Transfer, H.C. Hottel and A. F. Sarofim, McGraw-Hill Book Company, New York, 1967. 52 Pages. *AIChE J.* **1969**, *15*, 794–796. https://doi.org/10.1002/aic.690150504.
33. ASD FieldSpec® HandHeld 2 User Manual 2010, 1–140. Available online: https://www.geo-informatie.nl/courses/grs60312/material2017/manuals/600860-dHH2Manual.pdf (accessed on 5 June 2024).
34. Cavaco, A.; Canhoto, P.; Collares Pereira, M. Procedures for Solar Radiation Data Gathering and Processing and Their Application to DNI Assessment in Southern Portugal. *Renew. Energy* **2021**, *163*, 2208–2219. https://doi.org/10.1016/J.RENENE.2020.10.075.
35. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2.0. Available online: https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 28 May 2024).
36. *ISO 9847:1992*; Solar energy—Calibration of field pyranometers by comparison to a reference pyranometer. International Standard Organization: Geneva, Switzerland, 1992.
37. Betti, A.; Blanc, P.; David, M.; Saint-Drenan, Y.-M.; Driesse, A.; Freeman, J.; Fritz, R.; Gueymard, C.; Habte, A.; Höller, R.; et al. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications, Report IEA-PVPS 16-04:2021; 3rd ed.; International Energy Agency: Paris, France, 2021; ISBN 978-3-907281-19-2.

Nomenclature

A	Anisotropy index
BHI	Beam horizontal irradiance (W/m^2)
C	Fraction of circumsolar area obscured by the adjacent row
c	Length along the panel (m)
c_s	Position of the pyranometer along the length of the panel (m)
D	Distance between rows in the horizontal plane (m)
$\vec{D}$	Vector of diffuse horizontal irradiance vector (W/m^2)
D_{cs}	Circumsolar diffuse component of GTI (W/m^2)
DIF	Diffuse horizontal irradiance (W/m^2)
D_{iso}	Isotropic diffuse component of GTI (W/m^2)
DNI	Direct normal irradiance (W/m^2)
EHI	Extraterrestrial horizontal irradiance (W/m^2)
ENI	Extraterrestrial normal irradiance (W/m^2)
F	View-factor matrix
F_g	Ground view factor
F_s	Sky view factor
f	Clearness index as defined in the Klucher model
GTI	Global tilted irradiance (W/m^2)
GRI	Global reflected irradiance (W/m^2)
h_0	Vertical distance between the ground and the panel base (m)
I	Identity matrix
I_b	Direct component of GTI (W/m^2)
$\vec{I}_r$	Vector of direct horizontal irradiance vector (W/m^2)
I_{refl}	Reflection component of GTI (W/m^2)
k_T	Clearness index
L	Length of the panel (m)
MAE	Mean absolute error (W/m^2)
MBE	Mean bias error (W/m^2)
MAPE	Mean absolute percentage error (%)
R	Reflectivity matrix

R_b	Beam irradiance tilt factor
R_d	Diffuse irradiance tilt factor
RMSE	Root mean squared error (W/m^2)
RMSPE	Root mean squared percentage error (%)
R^2	Coefficient of determination
R_r	Reflected irradiance tilt factor
r	Apparent angular radius of circumsolar irradiance (°)
S	Shading of direct component coefficient
S_{cs}	Shading of circumsolar component coefficient
u	Length along the ground (m)
v	Length along the back of the front panel (m)
W	Row width (m)
w_i	Weight of irradiance component i to the bias reduction (%)
X_i	Isotropic component of first-row model
X_{cs}	Circumsolar component of first-row model

Greek symbols

Φ	Solar zenith angle (°)
α	Solar elevation angle (°)
α’	Projection of α in the vertical plane of the local meridian (°)
β	Surface tilt angle (°)
γ_p	Surface azimuth (°)
γ_s	Solar azimuth (°)
δ	Angle between the horizontal and the top of the front panel (°)
ε	Angle between the horizontal and the top of the panel being considered (°)
ζ	Angle between the horizontal and the bottom of the front panel (°)
θ	Angle of incidence (°)
λ	Angle between the horizontal and the bottom of the panel being considered (°)

ξ	Angle between the horizontal and the top of the panel in front of the front panel (°)
ρ	Ground albedo
σ	Angle between the horizontal and the top of the panel behind the panel being considered (°)

Acronyms

ANN	Artificial neural networks
CAMS	Copernicus atmospheric monitoring service
ECMWF	European centre for medium-range weather forecasts
IFS	Integrated forecasting system
NWP	Numerical weather prediction

Chapter 5

Assessment of thermal modeling of photovoltaic panels for predicting power generation using only manufacturer data†

Abstract

This study presents an assessment of thermal modeling for photovoltaic modules, focusing on power output prediction using manufacturer-provided data along with irradiance and weather-related variables. Several steady-state thermal models based on empirical correlations were evaluated for computing the temperature of the photovoltaic module. Additionally, a dynamic model was developed based on the energy conservation equation, incorporating the effects of wind speed and direction, using only manufacturer data and other parameters available in the literature. The performance of these models was evaluated against measured temperatures on the backsides of photovoltaic modules. The models were further integrated with the simple estimate with temperature correction and single diode and five-parameter electrical models to assess combined power output prediction performance. Results show that the Mattei steady-state model is the most accurate for temperature estimation, with a mean bias error of −0.4°C and a root mean squared error of 2.7°C. However, for power output estimation, the Kurtz (Sandia1) model combined with the simple estimate with temperature

†Sara Pereira[a], Paulo Canhoto[a,b], Takashi Oozeki[c] and Rui Salgado[a,d] 2024. Assessment of thermal modeling of photovoltaic panels for predicting power generation using only manufacturer data. Energy Reports, 12, 1431-1448. https://doi.org/10.1016/j.egyr.2024.07.039.

[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, Fukushima, Japan
[d] Physics Department, University of Évora, Évora, Portugal

correction outperforms others, showing a mean bias error of 4.6 W and a root mean squared error of 54.5 W. This study systematically evaluates and compares the performance of thermal models for different photovoltaic systems, offering a framework for selecting appropriate models based on their accuracy in temperature estimation and power output prediction. These models can support operational photovoltaic forecasts without the need for production data and facilitate decision-making in the deployment and management of photovoltaic technology.

Keywords

Modeling and simulation; Photovoltaic; Solar energy; Thermal model

5.1 Introduction

A photovoltaic module is a device that converts sunlight into electrical energy through the photovoltaic effect occurring in a junction of two dissimilar semiconductor materials. The generated electricity can vary rapidly due to the variability of the solar resource. Solar irradiance is affected by factors such as the presence of clouds and aerosols in the atmosphere. Moreover, photovoltaic power output is also very impacted by temperature since higher values of module temperature result in a decrease in conversion efficiency [1]. Previous studies have extensively explored the impact of environmental parameters such as irradiation, temperature, and dust on the performance of PV systems, leading to the development of various empirical correlation forms to enhance the accuracy of PV modeling. All these aspects are important because, with the increase in installed capacity of photovoltaic systems throughout the world, the need for modeling such a variable electricity generation has also increased. Accurate models of the conversion from sunlight into electricity can allow for better planning and development of photovoltaic power plants, monitoring system performance, and can also be included in electricity generation forecasting for better control of the electric grid.

Many electrical models of photovoltaic cells have been developed in the literature. Among these models, the level of detail, and therefore the number of input parameters, may vary widely [2], ranging from simple models where the electrical power produced is assumed to be proportional to solar irradiance, to very detailed models such as the equivalent electric circuit models [3,4]. The most widely used models are single diode models, typically involving three to five parameters for optimal performance analysis. The parameter extraction is a key issue due to the number of parameters to determine and the nonlinearity of the optimization problem, usually requiring numerical or analytical methods and curve fitting methods [5]. Precisely adapting the parameters to real conditions is essential for accurate power output estimates. For the crystalline silicon technology, the five parameter single diode model shows a good balance between accuracy and computation time [2], although additional adaptations are needed for other technologies such as microcrystalline, amorphous, and cadmium telluride [6]. Numerous studies review and compare different electric models, presenting advantages and disadvantages derived from simulation or experimental tests [2,3,7–12]. These analyses emphasize the importance of accurately capturing the electrical behavior of photovoltaic modules for reliable power output estimation.

When the temperature of photovoltaic cells/modules is not available, a thermal model is also needed for accurate power output estimation. Most thermal models used in the literature are steady-state and empirical, often biased towards specific technologies, material properties or locations. Tuomiranta et al. [13] obtained different correlations between module temperature (seven different semiconductor materials) and solar irradiance, air temperature and wind speed for Abu Dhabi. Coskun et al. [14] tested seventeen different correlations available in the literature for a photovoltaic powerplant in Turkey, modifying them for a better fit to the experimental data. The modified Chenni correlation [14] was shown to be the most suitable under variable wind speed conditions. A comprehensive list of explicit and implicit correlations for module temperature estimation found in the literature is available in [15]. Most models typically incorporate solar irradiance, air temperature and wind speed, while additional parameters such as air humidity and dust deposition can also impact photovoltaic system performance. Gholami et al. [16] investigated the effect of these variables on photovoltaic panel temperature, finding standardized Beta coefficients of 0.541, 0.645, −0.035, −0.055, and −0.068 for irradiance, ambient temperature, humidity, wind speed, and dust accumulation, respectively. Their proposed semi-empirical correlations achieved coefficients of determination ranging from 0.728 to 0.974. Additionally, studies have shown that dust accumulation impacts the performance of photovoltaic systems, causing both reduced power generation and increased module temperatures [17,18].

Physical models based on the energy conservation equation for dynamic regimes and heat transfer modeling can reproduce the thermal response of photovoltaic modules over time under variable conditions, making them suitable for simulations with smaller timesteps, typically at a second or minute time scale. Most studies in the literature consider either the electrical or the thermal model alone, or when both are considered, they are often uncoupled [2]. Since the photovoltaic cell/module temperature depends on the electrical power generated and vice-versa, coupling thermal and electrical models is advantageous. Some of the few works that include this coupling were developed recently [19,20]. Perovic et al. [19] developed a dynamic model that includes the dependence of electric power on module efficiency and temperature, comparing temperature results with measurements in three different points on the back of a photovoltaic module. Li et al. [20] developed a 3-D thermal model which, when coupled with the single diode 5 parameters electric model, resulted in values of power output, back-surface temperature, and current-voltage curves consistent with observations. In [21], a model for module temperature and power output estimation incorporating a dynamic thermal model and the single diode 5 parameter model, considering parasitic effects and bypass diodes, as well as site-specific adjusting parameters, was developed. In this case, root mean square error of 1.34°C for module temperatures and a normalized root mean square error of 1.98% for power-voltage curves were found. Gu et al. [22] assessed thermal resistance distribution and environmental effects on photovoltaic module performance through a coupled electrical-thermal model, highlighting the significance of radiative and convective thermal resistances. In [23], coupled electrical and thermal models for photovoltaic module temperature estimation was presented using inputs such as electric voltage and current, ambient temperature, wind speed and direction, total irradiance, and relative humidity. Bevilacqua et al. [24] developed and validated a one-dimensional thermal model for photovoltaic modules, using the energy balance of individual layers in dynamic conditions and determining the temperature distribution across the module cross-section through a finite difference method. In [25] a dynamic coupled thermal-electric model considering irradiance, air temperature, wind speed and direction, module parameters and geospatial data was developed for hourly data and assessed against observations of a poly-crystalline photovoltaic module, resulting in an MBE of 3.35 W/m^2.

This work systematically evaluates and compares thermal models for different photovoltaic systems, focusing on the accuracy of temperature estimation and power output prediction. It focuses on models that take only as inputs the characteristic values commonly provided by manufacturers in the datasheets, beside solar irradiance and meteorological variables. Data-based models such as regression or machine learning models were not included in the analysis. Experimental values of environmental variables and power output of a photovoltaic system are used for model assessment. Additionally, a developed electric-thermal coupled model for dynamic regimes, considering the wind direction relative to module rows, is also included in the analysis and validated against the same data. All models are compared to find the most accurate method for estimating power output of photovoltaic systems without needing measurements from the system. The novelty of this work lies in several key aspects. Firstly, a transient thermal model was developed that extends beyond traditional steady-state models, enabling predictions under varying operational scenarios. Secondly, a thorough comparative analysis of multiple thermal models was conducted, evaluating their performance against measured data from several PV technologies. This comprehensive evaluation framework ensures robustness of the proposed methodology and strengthens the practical implications for operational decision-making in PV systems. Ultimately, the approach used supports PV forecasts even in the absence of energy generation data, thereby advancing the field's capability for accurate and reliable performance estimation.

This paper is organized as follows: Section 5.2 presents the methodology of this work specifically the experimental data used for model assessment, including the environmental conditions (solar irradiance, air temperature and wind speed and direction) and photovoltaic system characteristics and power generation data. The various steady-state thermal models of the photovoltaic modules for obtaining module temperature are assessed, and the developed dynamic thermal model is presented. These models are then coupled with the simple estimate with temperature correction and single diode five parameter electric models and validated against observations in Section 5.3. Finally, Section 5.4 presents the conclusions of this work.

5.2. Methodology

The methodology encompasses two main components: data collection and processing, and thermal and electric modeling. Data collection involved retrieving and processing solar resource and meteorological variables, as well as photovoltaic system data. Thermal modeling utilized both steady-state and dynamic models to estimate module temperatures under changing environmental conditions. Coupling the thermal with electric models allows for the estimation of photovoltaic power output. The chapter discusses the formulation of empirical and physics-based models and the rationale behind model selection.

5.2.1 Data

Data from four different photovoltaic systems located at Fukushima Renewable Energy Institute (FREA, AIST), Koriyama, Japan (37.4495° N, 140.3144° E, Fig. 5.1) were utilized in this study. The systems employ different photovoltaic cell technologies: one with polycrystalline modules, two with monocrystalline modules, and one using heterojunction technology with intrinsic thin layer (HIT), which combines monocrystalline and amorphous silicon with dissimilar band gaps. The characteristics of these systems are presented in Table 5.1.

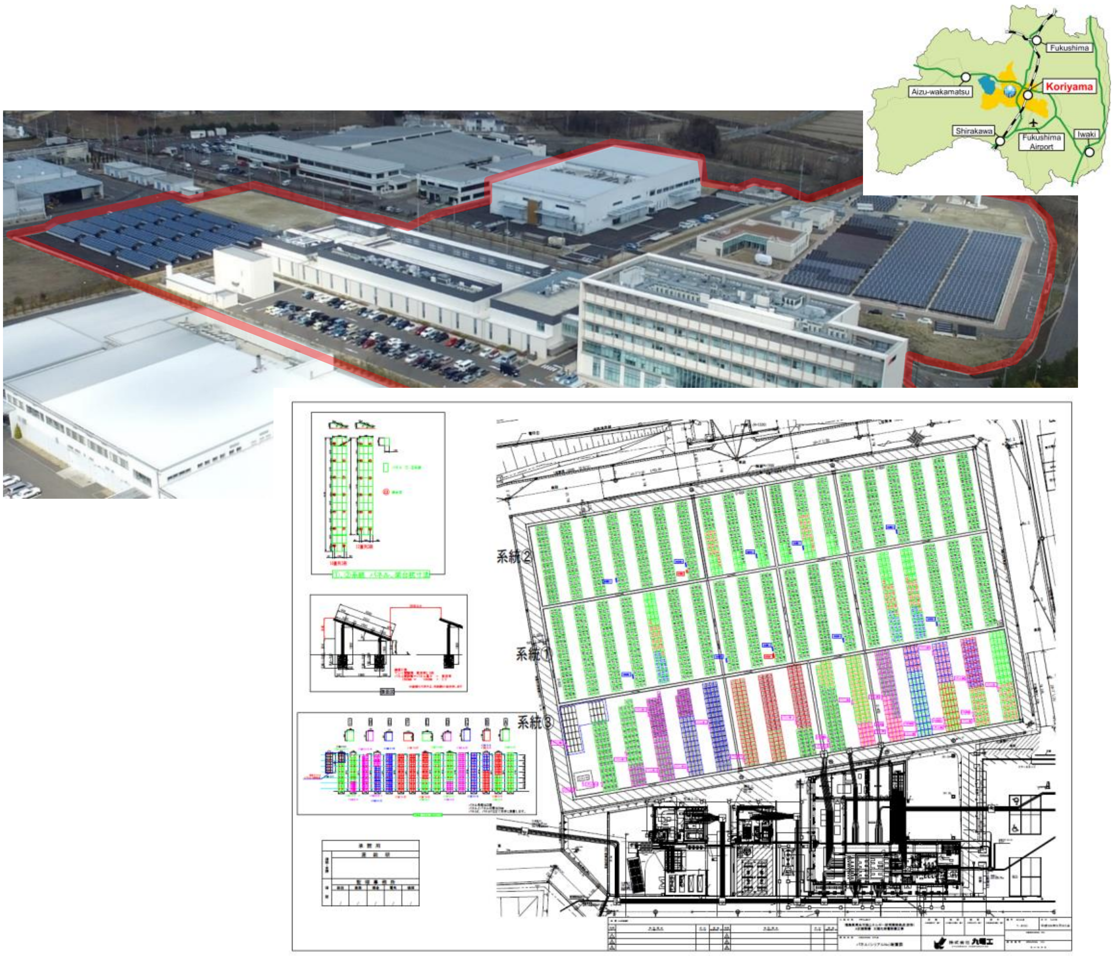


Fig. 5.1 - Location and plan of the photovoltaic systems at FREA, AIST.

The experimental data comprise environmental variables and photovoltaic system variables. Environmental variables include global tilted solar irradiance (*GTI*) measured using a class A pyranometer (Hukseflux CHF-

SR20-JM), direct normal irradiance (DNI) measured using a pyrheliometer, air temperature (T_a), wind speed (WS) and wind direction (WD) at 7.4 m height.

Table 5.1 - Characteristics of the photovoltaic systems for Standard Test Conditions (STC).

System	**1**	**2**	**3**	**4**
Photovoltaic modules				
Manufacturer	Toshiba	Panasonic	Kyocera	Sharp
Model	TA60M250WA	VBHN238SJ23A	KS242P-3CF3CE	NQ-198AC
Technology	Monocrystalline	HIT	Polycrystalline	Monocrystalline
$\boldsymbol{P_{max}}$ [W]	250.0	238.1	242.0	198.0
$\boldsymbol{V_{mp}}$ [V]	30.96	43.4	29.8	22.4
$\boldsymbol{I_{mp}}$ [A]	8.07	5.50	8.13	8.84
$\boldsymbol{V_{oc}}$ [V]	37.92	52.20	36.90	27.50
$\boldsymbol{I_{sc}}$ [A]	8.62	5.85	8.80	9.50
$\boldsymbol{\alpha_{P_{max}}}$ [%/$\boldsymbol{K}$]	−0.440	−0.290	−0.450	−0.392
$\boldsymbol{\alpha_{V_{oc}}}$ [%/K]	−0.300	−0.288	−0.320	−0.357
$\boldsymbol{L}$ [m]	1.650	1.580	1.662	1.165
$\boldsymbol{W}$ [m]	0.991	0.812	0.990	0.990
Thickness [mm]	40	35	46	46
Strings				
No. of modules	8	8	8	10
Position	1st row	2nd row	3rd row	5th row
Peak power [W]	2000.0	1904.8	1936.0	1980.0
Inverters				
Manufacturer	Tabuchi Electric	Tabuchi Electric	Tabuchi Electric	Tabuchi Electric
Model	EPU-A-T100P-SB	EPU-A-T100P-SB	EPU-A-T100P-SB	EPU-A-T100P-SB
Efficiency (%)	93	93	93	93
Capacity [kW]	10	10	10	10

Photovoltaic system variables comprise power output (P) and temperature at the back of the module (T_{mod}), measured with a type T thermocouple attached with Kapton tape directly to the back of one module per system. These variables were recorded at 1-minute timestep and converted to 10-minute mean values. The dataset spans from January 1st 2017 to December 31st 2020. Data preprocessing excluded winter months (December through March) due to snow accumulation on modules, retaining records with valid values across all variables. This resulted in a total of 24543 10-minute data points per system during the daylight hours.

5.2.2 Thermal modeling of photovoltaic modules

When the measured temperature of the modules is unavailable, a model for its estimation is necessary. Specifically, a model that can accurately estimate the temperature using easily available meteorological variables and characteristics of the photovoltaic systems is required. From the models described in the literature, two main categories can be found: steady-state and dynamic models.

Steady-state models are most common and assume that the heat transfer rate between the modules and the environment is quasi-stationary. In contrast, dynamic models consider the variation in thermal energy storage in the module and heat transfer rates over time. Steady-state models are usually correlations based on experimental data, which can make these models biased towards specific climates or technologies. Dynamic models tend to be based on physical principles and are more suitable for estimations made over small timesteps.

5.2.1.1 Steady state models

The steady-state models analyzed in this work are amongst the most commonly used thermal models in the literature and typically require air temperature (T_a), global tilted irradiance (GTI) and wind speed (WS).

Modified Chenni thermal model

The Chenni thermal model [8] was improved by examining the trend in the temperature difference between the model's results and observation data, resulting in the modified Chenni thermal model [14], as shown in Eq. (5.1). The observation data used results from a year of records from a polycrystalline powerplant in Turkey.

$$T_{mod} = T_a - 1.93666 + 0.0138\,GTI(1 + 0.031\,T_a)(1 - 0.042\,WS) + 0.007882GTI - 0.0000134647(GTI)^2 \quad (5.1)$$

Mattei thermal model

The Mattei thermal model [26] was developed using observation data from a grid-connected system and a polycrystalline module with eight temperature sensors integrated into the module, resulting in the correlation shown in Eq. (5.2). The coefficient of thermal losses, U_{pv}, is computed through Eq. (5.3), and the transmittance and absorption product at a normal incidence angle of irradiance, $(\tau\alpha)_n$, assumes a value of 0.81.

$$T_{mod} = \frac{U_{pv}\,T_a + GTI\,((\tau\alpha)_n - \eta_{STC}\left(1 - \alpha_{P_{max}} \times T_{stc}\right)}{U_{pv} + GTI\alpha_{P_{max}}\eta_{STC}} \quad (5.2)$$

$$U_{pv} = 26.6 + 2.3WS \quad (5.3)$$

Sandia thermal model

The Sandia thermal model [27], shown in Eq. (5.4), was developed for different module and mount types through the experimentally determined coefficients a and b, as presented in Table 5.2.

$$T_{mod} = T_a + GTIe^{a+b\,WS} \quad (5.4)$$

Table 5.2 - Parameters of the Sandia models.

Name	**Module Type**	**Mount**	**a**	**b**
Kurtz (Sandia1)	Glass/cell/glass	Open rack	−3.47	−0.0594
Sandia2	Glass/cell/glass	Close roof mount	−2.98	−0.0471
Sandia3	Glass/cell/polymer sheet	Open rack	−3.56	−0.0750
Sandia 4	Glass/cell/polymer sheet	Insulated back	−2.81	−0.0455

Faiman thermal model

The Faiman thermal model [28] is computed using the correlation shown in Eq. (5.5), derived from observations on seven different photovoltaic modules with mono and polycrystalline silicon cells located at Sede Boqer in the Negev Desert. The observations of GTI and T_{mod} were made using a pyranometer on the plane of array and thermocouples attached to the back of the modules using a small bead of quick-drying epoxy putty.

$$T_{mod} = T_a + \frac{GTI}{25 + 6.84\,WS} \tag{5.5}$$

JIS thermal model

The JIS thermal model presented in Eq. (5.6), originates from the Japanese Industrial Standard [29] and was determined for two different mount types, resulting in parameters a and b shown in Table 5.3.

$$T_{mod} = T_a + \left(\frac{a}{b\,WS^{0.8} + 1} + 2\right)\frac{GTI}{G_{STC}} - 2 \tag{5.6}$$

Table 5.3 - Parameters of the JIS models.

Name	**Mount**	**a**	**b**
JIS1	Frame	46	0.41
JIS2	Roof	50	0.38

Migan thermal model

The Migan model [30] is computed using the correlation on module temperature, air temperature, global tilted irradiance, and wind speed, as shown in Eq. (5.7). This equation is derived from several different assumptions regarding a 5-layer module considering conduction and convection losses and disregarding heat losses due to radiation.

$$T_{mod} = T_a + \frac{0.32\ GTI}{8.91 + 2\ WS} \tag{5.7}$$

5.2.1.2 Dynamic model

Dynamic thermal models are based on the energy conservation equation derived from the first law of thermodynamics. For a specific photovoltaic module, an energy balance equation for each material layer can be included in the model [24] or, alternatively, a simpler model can be developed by considering the entire module [19]. In this work, since the focus is on general models that only uses parameters provided by the manufacturers, the later approach is followed, and thus the module is considered as a block with an average temperature T_{mod}. This approach is a compromise between the need for input data and parameters and computational effort. The energy balance can be defined by Eq. (5.8), where C_{mod} is the equivalent heat capacity of the module, Q_{sun} is the net rate of in-plane solar irradiance and Q_{cond}, Q_{conv} and Q_{rad} are the net rates of heat transfer to the environment through conduction, convection and thermal radiation, respectively, and P_e is the electric power output generated by the photovoltaic cells.

$$C_{mod} \frac{dT_{mod}}{dt} = Q_{sun} - Q_{cond} - Q_{conv} - Q_{rad} - P_e \tag{5.8}$$

In [25,31], a sensitivity analysis was performed to evaluate the effect of the heat capacity value on the performance of dynamic thermal models. The results showed that varying this value had very little impact on the modeled temperature and power output of the modules when compared to the

remaining terms of the equation. As such, the recommended value of 22800 J/K is applied to any photovoltaic system [31].
The absorbed solar energy is computed through Eq. (5.9) requiring the area of the module, A, provided by the manufacturer, and where S is the absorbed irradiance, $(\tau\alpha)_n$ is the optical efficiency for normal incidence, which is obtained through the product of transmissivity by the typical absorptivity value of 0.9 [27]. For the equations for the computation of S and $(\tau\alpha)_n$, please check Appendix A. Although each module has both active and inactive areas, in this study, the total area of the photovoltaic module is assumed to be the active area, since the ratio of active to total area is typically close to 1.

$$Q_{sun} = A \times S \times (\tau\alpha)_n \tag{5.9}$$

The heat transfer by conduction is considered negligible since the contact area between the module and the supporting structure is small. The net rate of heat transfer by convection can be computed through Eq. (5.10), where h_f and h_b are the convective heat transfer coefficients at the front and back of the photovoltaic module, respectively.

$$Q_{conv} = A\left(h_f + h_b\right)(T_{mod} - T_a) \tag{5.10}$$

Convection can be defined as free (or natural) convection, driven by buoyancy due to temperature gradients in the fluid, or forced convection, driven by external forces, or a combination of both depending on the flow conditions. Eqs. (5.11) through (5.16) are used to compute the overall convective heat transfer coefficient of each surface [32]. In these equations, L_c represents the length of the module along the airflow direction. The thermodynamic and transport properties of dry air, obtained at 1 atm and film temperature T_f, are used. The film temperature is computed through the formula $T_f = T_a + 0.25(T_{mod} - T_a)$ for natural convection [33–35]. This model considers the wind direction, WD_s, in which 0° is for a wind direction from the South (local reference meridian), through the characteristic length L_c computed by Eq. (5.16) [34], with A_s representing the photovoltaic module azimuth.

$$h_{f/b} = \begin{cases} h_{f/b,\,free}, & if\ \dfrac{Gr}{Re_{f/b}^2} > 100 \\ \sqrt[3]{h_{f,\,free}^3 + h_{f,\,forced}^3}, & if\ 0.1 \le \dfrac{Gr}{Re_f^2} \le 100 \\ \sqrt[3]{h_{b,\,free}^3 + h_{b,\,forced}^3}, & if\ 0.1 \le \dfrac{Gr}{Re_b^2} \le 100 \wedge leeward \\ \sqrt[3]{h_{b,\,free}^3 - h_{b,\,forced}^3}, & if\ 0.1 \le \dfrac{Gr}{Re_b^2} \le 100 \wedge windward \\ h_{f/b,\,forced}, & if\ \dfrac{Gr}{Re_{f/b}^2} < 0.1 \end{cases} \tag{5.11}$$

$$Gr = Ra/Pr \tag{5.12}$$

$$Ra = g\big(1/T_f\big)(T_{mod} - T_a)L^3/(v_{T_f}\alpha_{T_f}) \tag{5.13}$$

$$Pr = c_{p,T_f} v_{T_f} /\ k_{T_f} \tag{5.14}$$

$$Re_{f/b} = WS \times L_{c,f/b}/v_{T_f} \tag{5.15}$$

$$L_c = \begin{cases} W,\ if\ windward\ \wedge\ [45 < (WD_s - As) < 135 \vee 225 < (WD_s - As) < 315] \\ L,\ if\ windward\ \wedge\ [(WD_s - As) \le 45 \vee 135 \le (WD_s - As) \le 225 \vee (WD_s - As) > 315] \\ \frac{4L\times W}{2(L+W)},\ if\ leeward \end{cases} \tag{5.16}$$

The convective heat transfer coefficient for natural convection is computed through Eq. (5.17), where L_{free} is the length of the module along the natural airflow [34]. The Nusselt number, Nu, is computed using the correlations in Eq. (5.18) presented by [35]. The first correlation is used for laminar flow, while the second is used for turbulent flow.

$$h_{f/b,\,free} = \frac{Nu_{f/b} \times k_{T_f}}{L_{free}} \tag{5.17}$$

$$Nu_{f/b} = \begin{cases} \left(0.825 + 0.387\sqrt[6]{Ra\,sin\,(\beta)\left(1 + \left(\frac{0.492}{Pr}\right)^{9/16}\right)^{-16/9}}\right)^2,\ if\ 0.1 < Ra < Ra_{cr} \wedge Pr > 0.001 \\ 0.56(Ra_{cr}cos\,(\theta_r))^{0.25} + 0.13\big(\sqrt[3]{Ra} - \sqrt[3]{Ra_{cr}}\big),\ if\ Ra \ge Ra_{cr} \wedge Pr > 0.001 \end{cases} \tag{5.18}$$

in which the critical Rayleigh value is computed using Eq. (5.19), where θ_v is given by Eq. (5.20).

$$Ra_{cr} = 10^{8.9-0.00178(\theta_v)} \tag{5.19}$$

$$\theta_v = 90 - \beta \tag{5.20}$$

In the case of forced convection, the convective heat transfer coefficients for the front and back surfaces of the module are calculated considering the wind speed and direction, as shown in Eq. (5.21). The first correlation is used for laminar flow, the second for transition flow, and the third for turbulent flow [36]. The critical Reynolds number (Re_{cr}) is 4×10[5], and L_c is computed using Eq. (5.16), which also depends on the wind direction.

$$h_{forced,\,f/b} = \begin{cases} 3.83WS^{0.5}L_c^{-0,5}, \; if \dfrac{x_{cr}}{L_c} \geq 0.95 \\ 5.74WS^{0.8}L_c^{-0,2} - 16.46/L_c, \; if \; 0.05 \leq \dfrac{x_{cr}}{L_c} < 0.95 \\ 5.74WS^{0.8}L_c^{-0,2}, \; if \dfrac{x_{cr}}{L_c} < 0.05 \end{cases} \tag{5.21}$$

in which the critical value x_{cr} is given by Eq. (5.22).

$$x_{cr} = Re_{cr} v_{T_f} / WS \tag{5.22}$$

The net rate of heat transfer with the environment by thermal radiation is computed using Eq. (5.23). The radiative heat transfer coefficients for the front and back surfaces of the module are computed using Eqs. (5.24) and (5.25), respectively. The sky temperature is obtained in Kelvin through the correlation presented in Eq. (5.26) [37] and the ground temperature (T_{ground}) is assumed to be equal to the air temperature. The emissivity coefficients (ε) are assumed to be 0.85 for the front surface and 0.91 for the back surface [32].

$$Q_{rad} = A\left(h_{rad,f}\left(T_{mod} - T_{sky}\right) + h_{rad,b}\left(T_{mod} - T_{ground}\right)\right) \tag{5.23}$$

$$h_{rad,\,f} = \varepsilon_f \sigma_{SB}\left(T_{mod}^2 + T_{sky}^2\right)\left(T_{mod} + T_{sky}\right)\frac{1+cos(\beta)}{2} + \varepsilon_f \sigma_{SB}\left(T_{mod}^2 + T_{ground}^2\right) \times \left(T_{mod} + T_{ground}\right)\frac{\left(T_{mod}-T_{ground}\right)}{\left(T_{mod}-T_{sky}\right)}\frac{1-cos(\beta)}{2} \tag{5.24}$$

$$h_{rad,\,b} = \varepsilon_b \sigma_{SB}\left(T_{mod}^2 + T_{sky}^2\right)\left(T_{mod} + T_{sky}\right)\frac{\left(T_{mod}-T_{sky}\right)}{\left(T_{mod}-T_{ground}\right)}\frac{1+cos(180-\beta)}{2} + \varepsilon_b \sigma_{SB}\left(T_{mod}^2 + T_{ground}^2\right)\left(T_{mod} + T_{ground}\right)\frac{1-cos(180-\beta)}{2} \tag{5.25}$$

$$T_{sky} = 0.0552 T_a^{1.5} \tag{5.26}$$

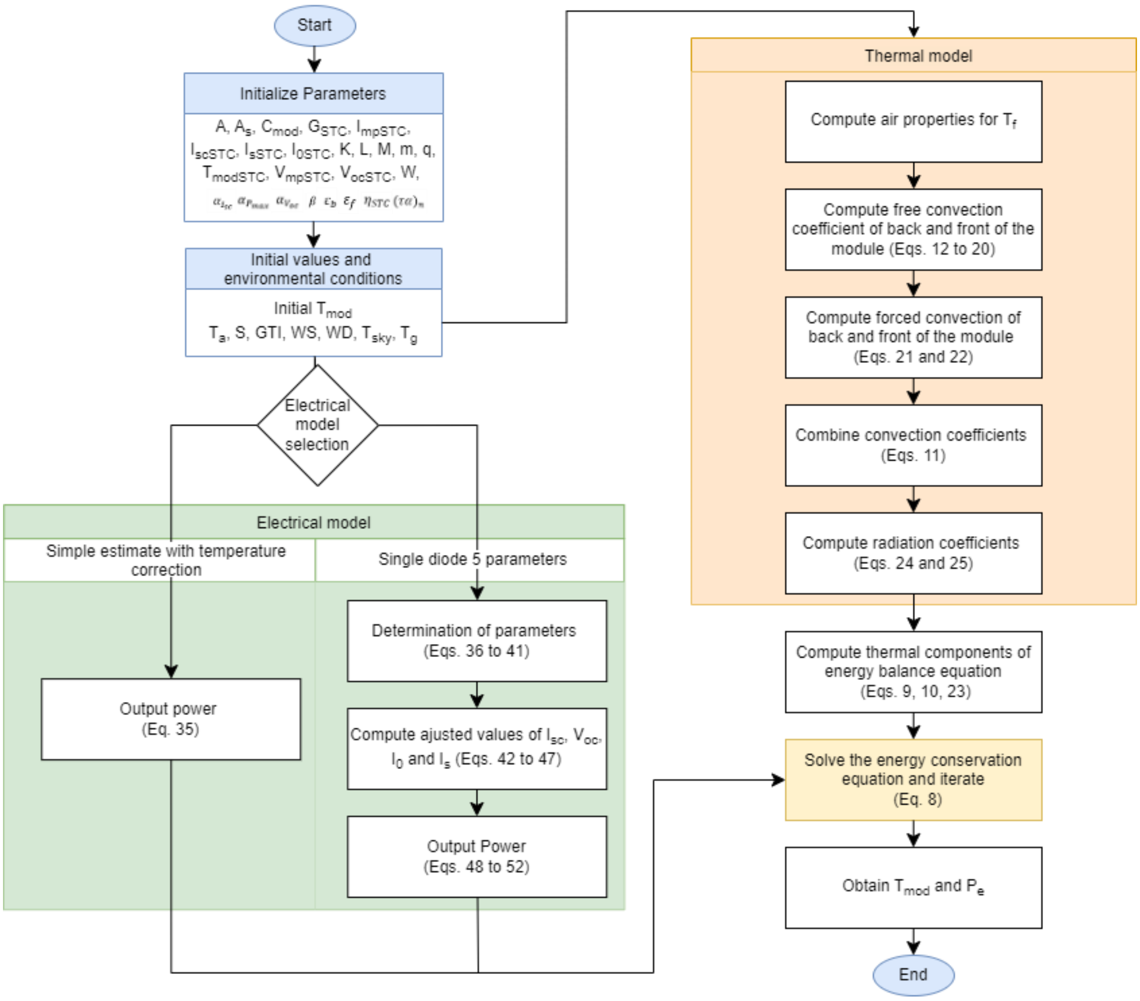


Fig. 5.2 – Dynamic model flowchart.

Finally, the electrical power output of the module, P_e, can be obtained from different models. In this work, the most commonly used electric models–simple estimate with temperature correction, and the single diode and five parameters models (presented in Appendix A)–were coupled to the presented dynamic thermal model. Typical values available in the literature were assumed for different parameters of photovoltaic modules, including surface emissivity [32], glazing transmissivity, module absorptivity [27], and heat capacity [31]. This means the model can be further optimized for improved results by adjusting these parameters.

To solve these equations and simultaneously obtain the module temperature and electrical power output, an iterative method must be applied. In this work, the Dorman and Prince version of the Runge-Kutta formula [38] was used through the Matlab function *ode45*.

Fig. 5.2 illustrates the dynamic model used to estimate the module temperature and electrical power output of photovoltaic systems. This structured approach ensures accurate modeling by integrating thermal and electrical processes.

5.3 Results and discussion

The thermal models discussed in Section 2 were applied to the study's open rack mounted photovoltaic systems. Models originally designed for roof-mounted or insulated modules, namely Sandia2, Sandia4, and JIS2, were excluded from consideration. The results were assessed against observations as shown in Fig. 5.3, which presents a direct comparison between estimated module temperature and the measured temperature on the back of the photovoltaic module.

The dynamic thermal model, which incorporates electric power output, was evaluated when coupled with the simple estimate with temperature correction (SET) and single diode and 5 parameters (1d5p) models. Selection of these electrical models for coupling with the dynamic thermal model was based on preliminary tests demonstrating their suitability and accuracy for predicting the electrical photovoltaic system performance. As detailed in Appendix A, these preliminary evaluations showed that the SET model offers a straightforward yet effective approach for incorporating temperature effects, making it a practical choice for dynamic simulations. On the other hand, the

1d5p model, with its detailed parameterization, provides a more nuanced and precise representation of the electrical characteristics of photovoltaic modules.

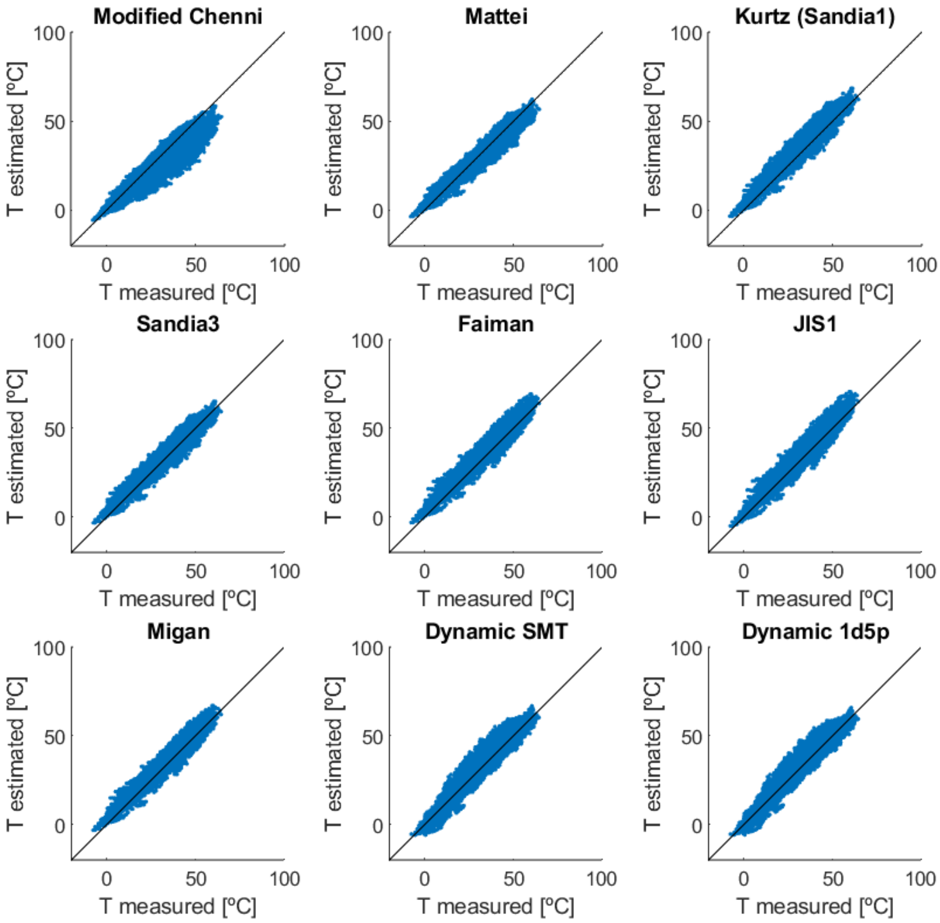


Fig. 5.3 - Comparison between measured and estimated temperature with the various models.

The models used in this study are derived from experimental correlations and theoretical assumptions or empirical relations developed by various researchers using data from different technologies and under different environmental conditions. As a result, each model's performance can be affected by the specific characteristics and assumptions inherent in its development. For instance, models differ in how they account for the different heat transfer mechanisms, material properties, and environmental conditions, which leads to variations in their accuracy. While some models, such as the Mattei and Modified Chenni, show negative bias, others overestimate temperatures due to these inherent differences.

This overestimation could also be due to the fact that temperature measurements are taken outside of the photovoltaic module, usually resulting in values slightly lower than the actual photovoltaic cell temperature, which is the output of the thermal models. Detailed results are provided in Table 5.4, which presents various statistical indicators, namely the correlation coefficient, R^2, mean bias error, MBE, relative mean bias error, rMBE, root mean squared error, RMSE, relative root mean squared error, rRMSE, and a global performance indicator GPI. The GPI allows for the ranking of the models from best to worst (higher to lower value) according to the above metrics (further details can be found in [39]). Fig. 5.4 displays a boxplot of these results, providing a visual comparison of the performance of each model relative to others based on MBE, RMSE and GPI metrics.

Table 5.4 - Statistical indicators of thermal models when compared to the measured temperature.

Model	**R^2**	**MBE (°C)**	**rMBE (%)**	**RMSE (°C)**	**rRMSE (%)**	**GPI**
Mattei	0.975	−0.4	−1.8	2.7	13.7	0.620
Sandia3	0.973	0.9	4.6	2.8	14.0	0.328
Migan	0.975	1.2	6.0	2.7	13.6	0.297
JIS1	0.967	0.8	4.0	3.1	15.7	0.072
Faiman	0.969	1.6	8.2	3.0	15.1	−0.159
Dynamic + 1d5p	0.958	−0.4	−2.0	3.6	18.1	−0.205
Dynamic + SET	0.959	−1.0	−5.0	3.6	17.9	−0.395
Kurtz (Sandia1)	0.955	2.0	10.2	3.6	18.3	−0.932
Modified Chenni	0.926	−2.9	−14.4	4.7	23.4	−2.370

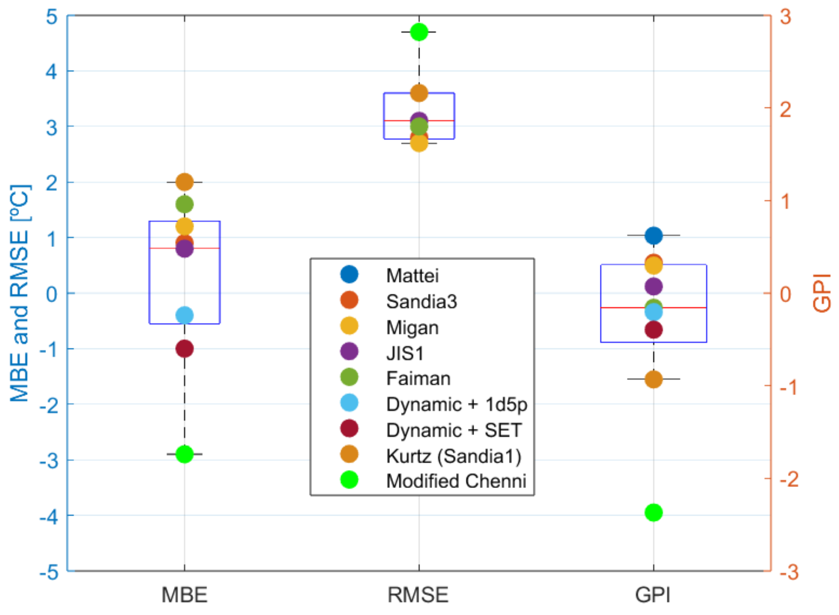


Fig. 5.4 - Boxplot of thermal model results.

The assumptions discussed above are confirmed, and the steady-state Mattei model is identified as the best in estimating module temperature, having the highest GPI along with a mean bias error of −0.4 °C and root mean squared error of 2.7 °C. The best dynamic model is shown to be the one coupled with the single diode and 5 parameters electrical model with a mean bias error of −0.4 °C and root mean squared error of 3.6 °C. In Fig. 5.4, it is visible that the Modified Chenni and Kurtz (Sandia1) models show lower performance, particularly in terms of higher bias and lower accuracy. Detailed results for each model and photovoltaic system can be consulted in Table 5.9 in Appendix B.

However, the measured temperature is not the actual mean temperature of the module and, since the goal of this work is to have an accurate estimation of the photovoltaic power output, the thermal models were evaluated in terms of estimated power when coupled with either the simple estimate with temperature correction or the single diode and 5 parameters electrical models. For steady-state models, this involved using the thermal model to estimate module temperature based on environmental conditions such as irradiance and ambient temperature. This estimated temperature is then used in the electrical model to adjust the electrical power output accordingly.

The results for the SET model are presented in Fig. 5.5, Fig. 5.6 and Table 5.5. Here, the estimated and measured electric power are directly compared, with a remarkably similar overestimation for all thermal models. This is also observed when using the measured temperature with the SET model (upper left plot of Fig. 5.5). The best thermal model (with the highest GPI) coupled with the SET model in terms of power output prediction is shown to be the Kurtz (Sandia1) model with a mean bias error of 4.6 W and root mean squared error of 54.5 W. Using the temperature measured in the back of the module only returns better power estimates when compared with the results for Mattei and Modified Chenni models which show the worst metrics as visible in Fig. 5.6. The results for each model and photovoltaic system are presented in Table 5.10 of Appendix B.

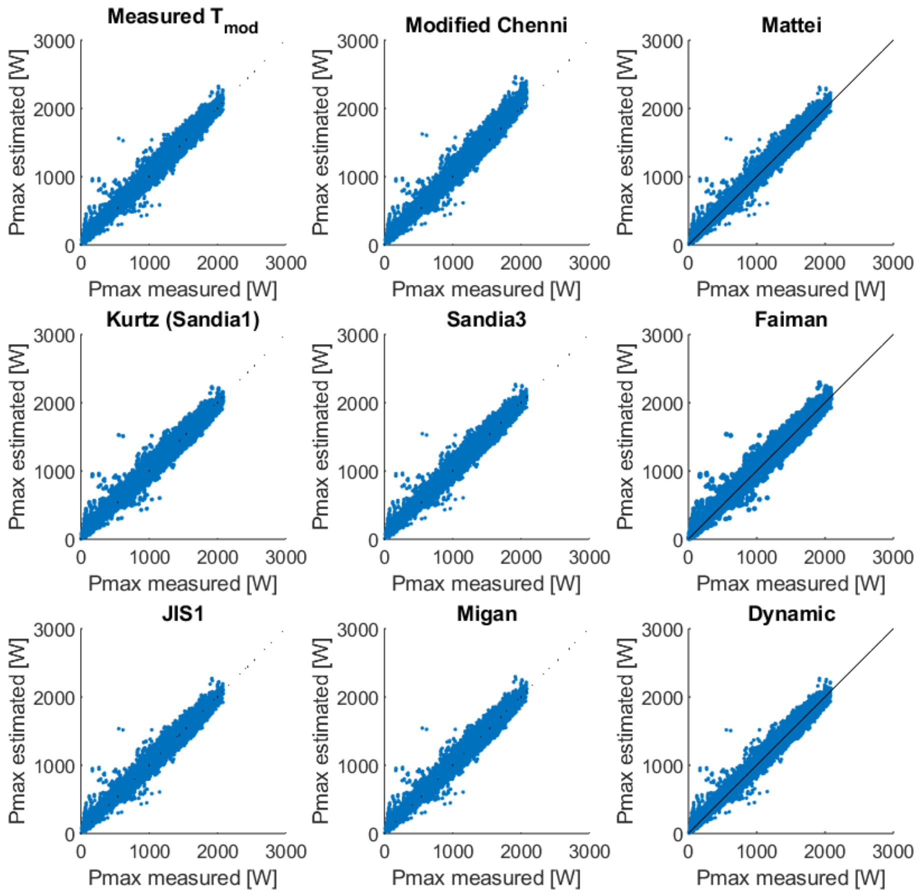


Fig. 5.5 - Comparison between measured and estimated power output with SET model coupled with each thermal model.

Table 5.5 - Statistical indicators of SET model output coupled with each thermal model.

Model	**R^2**	**MBE (W)**	**rMBE (%)**	**RMSE (W)**	**rRMSE (%)**	**GPI**
Kurtz (Sandia1)	0.995	4.6	0.7	54.5	8.5	0.329
Faiman	0.994	7.4	1.2	55.3	8.6	0.088
JIS1	0.994	7.2	1.1	55.5	8.6	0.071
Dynamic	0.995	12.0	1.9	54.4	8.5	0.061
Sandia3	0.994	10.2	1.6	54.9	8.6	0.033
Migan	0.994	9.5	1.5	55.3	8.6	0.008
Measured temperature	0.994	13.7	2.1	57.4	8.9	−0.505
Mattei	0.994	16.3	2.5	57.0	8.9	−0.562
Modified Chenni	0.992	27.6	4.3	66.7	10.4	−2.606

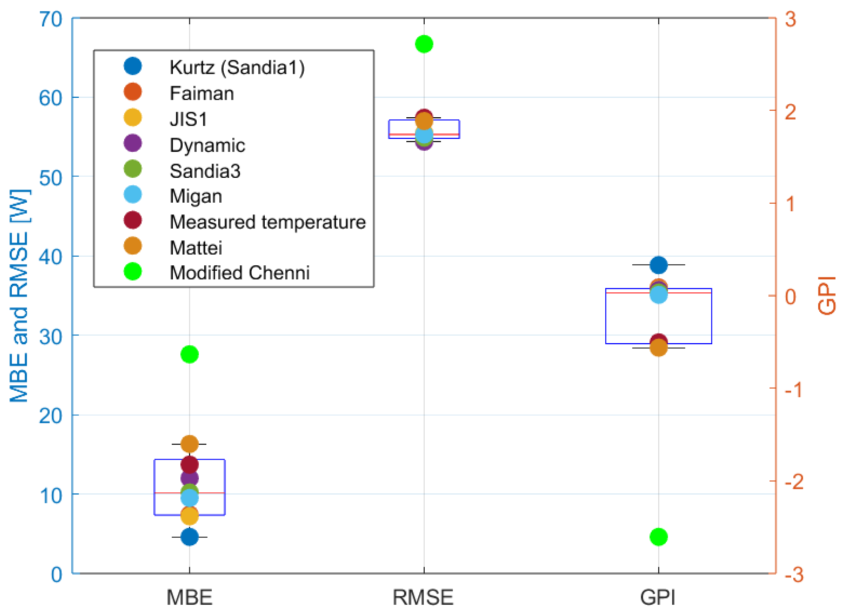


Fig. 5.6 - Boxplot of SET results.

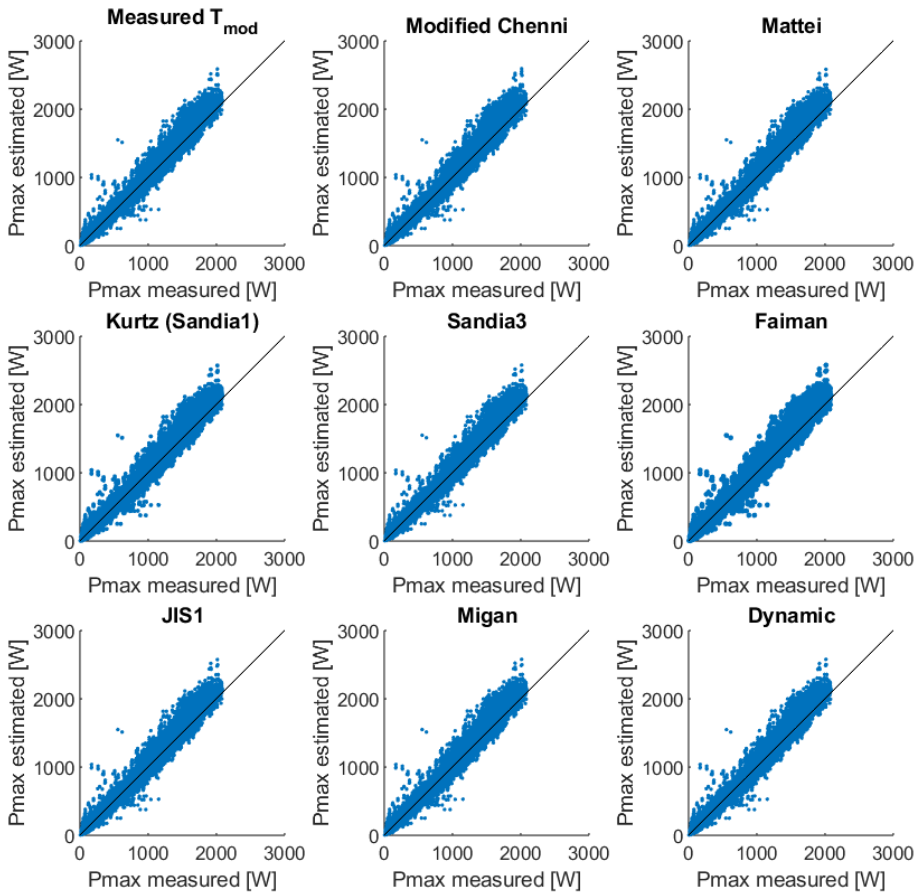


Fig. 5.7 - Comparison between measured and estimated power output with the single diode and 5 parameters model coupled with each thermal model.

The same analysis was carried out by coupling the thermal models with the single diode and 5 parameters model and the corresponding results can be seen in Figs. 5.7, 5.8 and Table 5.6.

With this electrical model, an overestimation of the output power is observed for all steady-state thermal models, being the best model the Kurtz (Sandia1) thermal model with a mean bias error of 15.7 W and root mean squared error of 83.6 W. All models, except the Mattei and Modified Chenni models, show better results than the use of measured module temperature when estimating power output with the 1d5p electrical model. The results for each model and photovoltaic system are presented in Table 5.11 in Appendix B.

Table 5.6 - Statistical indicators of single diode and 5 parameters model output coupled with each thermal model.

Model	**R^2**	**MBE (W)**	**rMBE (%)**	**RMSE (W)**	**rRMSE (%)**	**GPI**
Kurtz (Sandia1)	0.987	15.7	2.5	83.6	13.0	0.250
JIS1	0.987	15.8	2.5	83.7	13.0	0.216
Faiman	0.987	16.0	2.5	84.0	13.1	0.119
Dynamic	0.987	16.0	2.5	84.2	13.1	0.046
Migan	0.987	16.2	2.5	84.4	13.1	0.000
Sandia3	0.987	16.3	2.5	84.5	13.2	−0.059
Measured Temperature	0.987	16.3	2.5	84.8	13.2	−0.140
Mattei	0.987	16.7	2.6	85.4	13.3	−0.348
Modified Chenni	0.986	17.2	2.7	86.5	13.5	−0.750

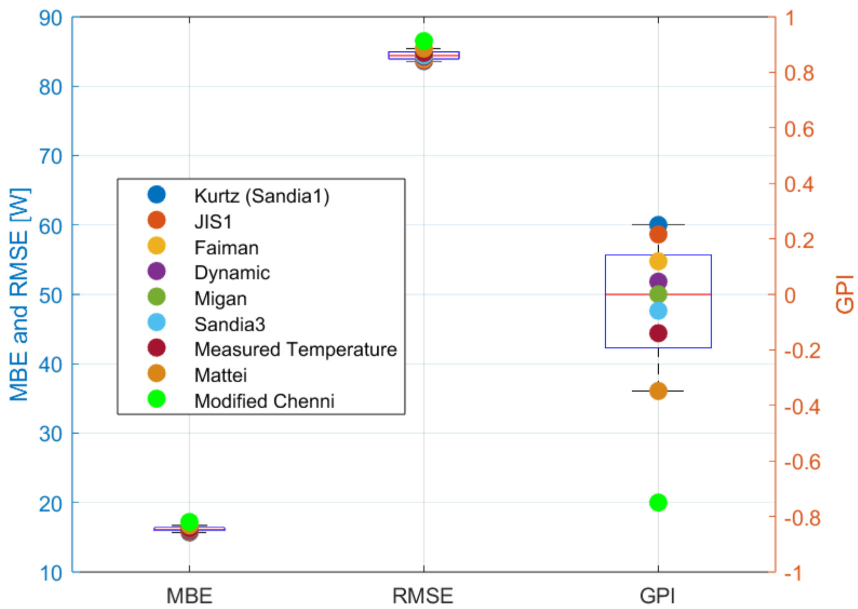


Fig. 5.8 - Boxplot of 1d5p model results.

Comparing the results shown in Table 5.5 and Table 5.6, the best combination of thermal and electrical models for the estimation of power output is shown to be the Kurtz (Sandia1) thermal model with the simple estimate with temperature correction model which shows a mean bias error of 4.6 W and a root mean squared error of 54.5 W. Also, clearly visible when comparing Fig. 5.6 and Fig. 5.8, the use of the simple estimate with temperature correction model for electric power calculation shows greater variability in statistical indicators across different temperature models compared to the single diode 5 parameters model, which has a maximum variation in MBE of 1.5 W between thermal models. This means that when using the 1d5p electrical model, the choice of thermal model is less critical for accurate power output estimation.

While the models considered in this study include the impact of module temperature, irradiance, and some of them wind speed and direction, other meteorological variables such as relative humidity, precipitation in the form of rain or snow, and dust accumulation can also affect the thermal response of photovoltaic modules and deserve further investigation to enhance the robustness of the models.

5.4 Conclusions

In this work, several thermal models of a photovoltaic module, including a developed dynamic thermal model, were evaluated for their ability to estimate temperature and electrical power output using only the characteristic parameters provided by the module manufacturer. Measurements of GTI and temperature on the back of the module from four different technologies were used for assessment. The best model for estimating the measured temperature was the steady-state Mattei model, achieving an MBE of −0.4 °C. In terms of power output, the Kurtz model coupled with the simple estimate with temperature correction electrical model demonstrated best performance, with an MBE of 4.6 W. When coupled with the single diode and 5 parameter electric model, the MBE was 15.7 W. The single diode 5 parameter model showed consistent results across all thermal models, with a maximum MBE variation of only 1.5 W, indicating the choice of thermal model is less critical when combined with a highly detailed electrical model.

This study provides a comprehensive evaluation of both steady-state and dynamic thermal models, highlighting the importance of selecting appropriate coupling of thermal and electrical models to improve the accuracy of module temperature and power output estimations for photovoltaic systems. It is shown that selecting the most accurate model for estimating module temperature does not necessarily yield the best power output estimates, as the accuracy of power output predictions strongly depends on the selected electrical model. It was also shown that using one of the studied thermal models (excluding the Mattei and Modified Chenni models) results in better estimates of photovoltaic power output than using temperature measurements on the back of the modules.

The evaluated models allow researchers and power plant operators to estimate the temperature and power output of PV systems without the need for extensive and costly experimental testing. These models can be employed for forecasting the power output of photovoltaic modules if forecasts of GTI and other relevant weather variables are available. Further refinements in thermal modeling could focus on improving accuracy under varying environmental conditions or specific module technologies, while continuous validation and adaptation of these models with experimental data from power plants could enhance their applicability and reliability in different operational settings.

The findings can aid in the planning and development of PV systems by providing reliable tools for performance estimation, ultimately enhancing the efficiency and feasibility of PV installations. This study highlights the practical benefits and scientific and technological importance of these models, offering valuable insights for optimizing and managing photovoltaic systems, and paving the way for more accurate and efficient renewable energy solutions.

Appendix A – Electric models of photovoltaic modules

Here, a preliminary review and assessment of standard electric models of photovoltaic modules, namely, a simple estimate model and an equivalent circuit model using a data-driven model as reference, is presented.

The models were evaluated against observations using as input the characteristics of the photovoltaic modules, observed GTI and temperature measured at the back of the modules.

For these models, the incident irradiance, S, is obtained through Eq. (5.27) [24], where θ_b is the incidence angle of direct solar radiation, θ_d is the equivalent diffuse incidence angle obtained through Eq. (5.28) [40] which are also used to compute the beam and diffuse incidence angle modifiers, K_{θ_b} and K_{θ_d}, respectively. The incident angle modifier is computed as in Eq. (5.29) [40]. The transmittance values can be obtained through Eqs. (5.30) and (5.31), respectively, where θ_r is the angle of refraction computed through Eq. (5.32). Typical input parameters for photovoltaic modules suggested in [41] were used in this work, namely, the refractive index $n = 1.526$, the extinction coefficient $K_g = 4\,m^{-1}$ and the thickness $L_g = 0.002\,m$ of the glazing.

$$S = DNI \cos\theta_b\, K_{\theta_b} + (GTI - DNI\cos\theta_b)K_{\theta_d} \tag{5.27}$$

$$\theta_d = 59.7 - 0.1388\beta + 0.001497\beta^2 \tag{5.28}$$

$$K_\theta = \frac{\tau(\theta)}{\tau(0)} \tag{5.29}$$

$$\tau(\theta) = e^{-\left(\frac{K_g L_g}{cos(\theta_r)}\right)}\left[1 - \frac{1}{2}\left(\frac{sin^2(\theta_r - \theta\,)}{sin^2(\theta_r + \theta\,)} + \frac{tan^2(\theta_r - \theta\,)}{tan^2(\theta_r + \theta\,)}\right)\right] \tag{5.30}$$

$$\tau(0) = e^{-(K_g L_g)}\left[1 - \left(\frac{1-n}{1+n}\right)^2\right] \tag{5.31}$$

$$\theta_r = sin^{-1}\left[\frac{1}{n} sin\,(\theta)\right] \tag{5.32}$$

where β stands for the tilt angle of the modules. The air mass modifier is computed as in Eqs. (5.33) and (5.34) [40], where the constants denoted as a_i

can be obtained from [41] for different photovoltaic cells. In this case, the coefficients for monocrystalline silicon were employed, since De Soto et al. [41] also showed that selecting a single set of coefficients and applying them for any cell type results in outcomes which are comparable to using different air mass modifier relationships for each cell type.

$$M = \sum_{0}^{4} a_i (AM)^i \tag{5.33}$$

$$AM = \frac{1}{\cos(\theta_z) + 0.5057(96.080 - \theta_z)^{-1.634}} \tag{5.34}$$

The simple estimate model with temperature correction (SET) [4] is described by Eq. (5.35), where the estimated power output is directly proportional to the solar irradiance absorbed by the photovoltaic module (S) and the air mass modifier (M) as well as to some characteristics of the modules, namely its efficiency at STC given by the manufacturer (η_{STC}) and the surface area (A), including a correction due to the module temperature as in [4].

$$P_e = \eta_{STC} \times A \times S \times M \times \left(1 + \alpha_{P_{max}} \times (T_{mod} - T_{modSTC})\right) \tag{5.35}$$

This correction is done through the peak power temperature coefficient, $\alpha_{P_{max}}$, often provided in the datasheet of the photovoltaic modules, being T_{modSTC} the temperature of the photovoltaic cell at STC, namely 25°C.

The necessary parameters for the direct computation of the power output of a module in real conditions using equivalent circuit models are not readily available. Still, the only commonly available electrical characteristics from the manufacturers are the maximum power (P_{max}), current (I_{mp}) and voltage (V_{mp}) at maximum power, short-circuit current (I_{sc}) and open-circuit voltage (V_{oc}), all for standard test conditions (STC). The parameters needed for a given equivalent electric circuit model need to be computed from these values and then adjusted for the real conditions in the field, namely different conditions of incident solar irradiance and cell temperature.

The single diode and 5 parameters model (1d5p) as represented in Fig. 5.9 and with I-V curve equation as given in Eq. (5.36) includes a current source which is caused by the photovoltaic effect and called photovoltaic current, I_{pv}, in parallel with a diode behaving as the p-n junction and a resistance, R_p, which accounts for the leakage current in the p-n junction and a series resistance, R_s, that represents the electric resistances of the silicon, of the electrodes and the contact between these materials. In this model, the five parameters were determined based on the method presented by Castro [3] in which expressions for the parameters I_{0STC} and I_{pvSTC} are obtained as a function of the remaining three parameters, which in turn are obtained from the equations at the three characteristic points in the I-V curve, thus resulting in the system of equations given by Eqs. (5.37) through (5.41).

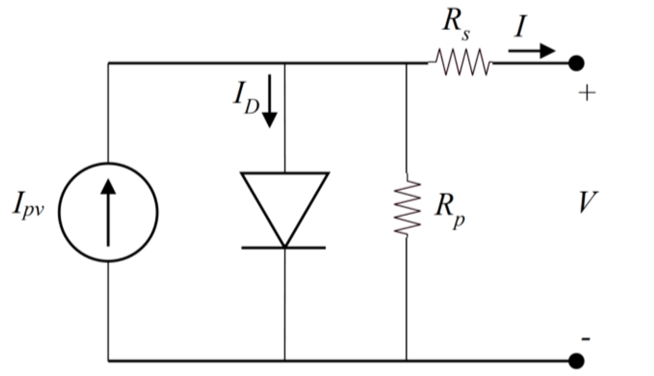


Fig. 5.9 - Single diode and 5 parameters equivalent electric circuit.

$$I = I_{pv} - I_0\left(e^{\frac{V+IR_s}{mV_T}} - 1\right) - \frac{V + IR_s}{R_p} \tag{5.36}$$

$$\begin{aligned} I_{mpSTC} = I_{scSTC} - \frac{V_{mpSTC} + R_s I_{mpSTC} - R_s I_{scSTC}}{R_p} \\ - \left(I_{scSTC} - \frac{V_{ocSTC} - R_s I_{scSTC}}{R_p}\right) e^{\frac{V_{mpSTC}+R_s I_{mpSTC}-V_{ocSTC}}{mV_{TSTC}}} \end{aligned} \tag{5.37}$$

$$0 = I_{mpSTC} + \frac{-\frac{(R_p I_{scSTC} - V_{ocSTC} + R_s I_{scSTC}) e^{\frac{V_{mpSTC}+R_s I_{mpSTC}-V_{ocSTC}}{mV_{TSTC}}}}{mV_{TSTC} R_p} - \frac{1}{R_p}}{1 + \frac{R_s (R_p I_{scSTC} - V_{ocSTC} + R_s I_{scSTC}) e^{\frac{V_{mpSTC}+R_s I_{mpSTC}-V_{ocSTC}}{mV_{TSTC}}}}{mV_{TSTC} R_p} + \frac{R_s}{R_p}} \tag{5.38}$$

$$\frac{-\frac{\left(R_p I_{scSTC} - V_{ocSTC} + R_s I_{scSTC}\right) e^{\frac{R_s I_{scSTC} - V_{ocSTC}}{m V_{TSTC}}}}{m V_{TSTC} R_p} - \frac{1}{R_p}}{1 + \frac{R_s\left(R_p I_{scSTC} - V_{ocSTC} + R_s I_{scSTC}\right) e^{\frac{R_s I_{scSTC} - V_{ocSTC}}{m V_{TSTC}}}}{m V_{TSTC} R_p} + \frac{R_s}{R_p}} = -\frac{1}{R_p} \tag{5.39}$$

$$I_{0STC} = \left(I_{scSTC} - \frac{V_{ocSTC} - R_s I_{scSTC}}{R_p}\right) e^{\frac{-V_{ocSTC}}{m_{STC} V_{TSTC}}} \tag{5.40}$$

$$I_{pvSTC} = I_{0STC} e^{\frac{V_{ocSTC}}{m_{STC} V_{TSTC}}} + \frac{V_{ocSTC}}{R_p} \tag{5.41}$$

The real conditions in the field must be considered, which is done through Eqs. (5.42) to (5.47).

$$I_{sc} = I_{scSTC} \frac{S \times M}{GSTC}\left(1 + \alpha_{Isc}(T_{mod} - T_{modSTC})\right) \tag{5.42}$$

$$V_{oc,g} = m V_{TSTC} ln\left(\frac{\frac{S \times M}{GSTC} I_{pvSTC} R_p - V_{oc,g}}{I_{0STC} R_p}\right) \tag{5.43}$$

$$V_{oc} = V_{oc,g}\left(1 + \alpha_{Voc}(T_{mod} - T_{modSTC})\right) \tag{5.44}$$

$$m = m_{STC} \tag{5.45}$$

$$I_0 = \left(I_{sc} - \frac{V_{oc} - R_s I_{sc}}{R_p}\right) e^{\frac{-V_{oc}}{m V_T}} \tag{5.46}$$

$$I_{pv} = I_0 e^{\frac{V_{oc}}{m V_T}} + \frac{V_{oc}}{R_p} \tag{5.47}$$

Finally, the power output is estimated through Eq. (5.48) and a system of non-linear equations, Eqs. (5.49) and (5.50), with initial values given by Eqs. (5.51) and (5.52).

$$P_e = V_{mp} \times I_{mp} \tag{5.48}$$

$$I_{mp} = I_{sc} - \frac{V_{mp} + R_s I_{mp} - R_s I_{sc}}{R_p} - \left(I_{sc} - \frac{V_{oc} - R_s I_{sc}}{R_p}\right) e^{\frac{V_{mp} + R_s I_{mp} - V_{oc}}{m V_T}} \tag{5.49}$$

$$\frac{-\frac{\left(R_p I_{sc}-V_{oc}+R_s I_{sc}\right)e^{\frac{V_{mp}+R_s I_{mp}-V_{oc}}{mV_T}}}{mV_T R_p}-\frac{1}{R_p}}{1+\frac{R_s\left(R_p I_{sc}-V_{oc}+R_s I_{sc}\right)e^{\frac{V_{mp}+R_s I_{mp}-V_{oc}}{mV_{TSTC}}}}{mV_T R_p}+\frac{R_s}{R_p}}=-\frac{I_{mp}}{V_{mp}} \tag{5.50}$$

$$I_{mp}(0)=I_{mpSTC}\left(1+\alpha_{Isc}(T_{mod}-T_{modSTC})\right) \tag{5.51}$$

$$V_{mp}(0)=V_{mpSTC}\left(1+\alpha_{Voc}(T_{mod}-T_{modSTC})\right) \tag{5.52}$$

The recently published international standard IEC 60891 ED3 [42], specifically the procedure 4.4.4, was used in this work as an empirical model based on measurements for comparison and as a reference for this type of modeling approach. This model requires the current and voltage values at the three characteristic points of the I-V curve for four different conditions of incident irradiance and temperature.

This dataset was obtained for the four systems considered in this work through indoor measurements in the lab for various combinations of irradiance at normal incidence and cell temperature, namely (1000 W/m^2, 25 °C), (1000 W/m^2, 60 °C), (200 W/m^2, 25 °C) and (200 W/m^2, 60 °C) by means of a solar simulator and a temperature chamber at AIST [43]. The solar simulator is comprised of six xenon lamps with elliptical mirrors and ultraviolet, AM 1.5 G and iron net dark filters with an output range of 100 to 1300 W/m^2 while the temperature chamber has a range of 15 °C to 65 °C [43]. The current and voltage for maximum power in real conditions is then obtained through bi-linear interpolation thus allowing for the estimation of the power output.

The electric models presented were used to estimate the power output of the four different photovoltaic systems using the measured GTI, DNI and module temperature data, thus generating 10-minute power output values for each electric model. Results were compared to the power output measurements and with each other as shown in Table 5.7. The same statistical indicators for each system/technology are presented separately in Table 5.8. All models tend to show an overestimation of the power output which can result in part from the fact that the temperature used as input to these models is the one measured at the back of the photovoltaic modules and not the photovoltaic cell temperature. It might be the case, as it often is, that the temperature measured this way is lower than the actual temperature of the cell and consequently the power estimated will be higher than the actual power output of the photovoltaic module.

Table 5.7 - Statistical metrics of the electric model results.

Model	**R^2**	**MBE (W)**	**rMBE (%)**	**RMSE (W)**	**rRMSE (%)**	**GPI**
IEC	0.994	−3.6	−0.6	59.0	9.2	0.511
SET	0.994	13.7	2.1	57.4	8.9	0.108
1d5p	0.987	16.3	2.5	84.8	13.2	−2.020

Table 5.8 - Statistical indicators of the results of electric models using measured temperature for each system.

Model	**System**	**R^2**	**MBE (W)**	**rMBE (%)**	**RMSE (W)**	**rRMSE (%)**
SET	1	0.988	26.2	4.0	61.8	9.3
	2	0.992	4.0	0.6	52.5	7.9
	3	0.991	2.3	0.3	55.4	8.4
	4	0.989	22.4	3.4	59.4	9.0
1d5p	1	0.970	36.6	5.5	99.1	15.0
	2	0.982	4.6	0.7	77.6	11.7
	3	0.981	5.1	0.8	78.9	12.0
	4	0.980	18.8	2.8	81.7	12.4
IEC	1	0.989	18.5	2.8	59.4	9.0
	2	0.987	−32.8	−5.0	66.3	10.0
	3	0.991	−0.9	−0.1	55.6	8.4
	4	0.991	0.9	0.1	54.0	8.2

The IEC model was shown to have better results which was expected since it is based on measurements of the specific systems assessed in this work. The following model is the simple estimate with temperature correction which, although not as detailed, achieved in this case better results than the single diode and 5 parameters model, especially in terms of RMSE. Both of these models perform slightly better for systems 2 and 3 (HIT and poly-crystalline technologies, see Table 5.8).

Appendix B – Temperature model results for each photovoltaic system

Table 5.9 - Statistical indicators of the results of thermal models when compared to the measured temperature for each system.

Model	System	R^2	MBE (°C)	rMBE (%)	RMSE (°C)	rRMSE (%)
Modified Chenni	1	0.857	-2.9	-11.3	4.6	18.0
	2	0.885	-2.1	-8.2	4.2	16.9
	3	0.828	-3.6	-13.5	5.1	19.4
	4	0.857	-2.9	-11.3	4.6	18.0
Mattei	1	0.951	-0.3	-1.0	2.7	10.5
	2	0.952	0.2	0.8	2.7	10.9
	3	0.952	-0.9	-3.3	2.7	10.2
	4	0.949	-0.5	-1.9	2.8	10.8
Kurtz (Sandia1)	1	0.909	2.0	7.7	3.7	14.3
	2	0.896	2.8	11.4	4.0	16.1
	3	0.935	1.3	5.0	3.1	11.9
	4	0.909	2.0	7.7	3.7	14.3
Sandia3	1	0.947	0.9	3.4	2.8	11.0
	2	0.941	1.7	7.0	3.0	12.1
	3	0.959	0.2	0.8	2.5	9.4
	4	0.947	0.9	3.4	2.8	11.0
Faiman	1	0.939	1.6	6.2	3.0	11.7
	2	0.923	2.5	9.9	3.4	13.8
	3	0.958	0.9	3.5	2.5	9.5
	4	0.939	1.6	6.2	3.0	11.7
JIS1	1	0.933	0.8	2.9	3.2	12.3
	2	0.929	1.6	6.5	3.3	13.2
	3	0.948	0.1	0.4	2.8	10.7
	4	0.933	0.8	2.9	3.2	12.3
Migan	1	0.951	1.1	4.4	2.7	10.5
	2	0.939	2.0	8.0	3.1	12.3
	3	0.965	0.5	1.8	2.3	8.7
	4	0.951	1.1	4.4	2.7	10.5
Dynamic + SET	1	0.907	-0.8	-2.9	3.7	14.4
	2	0.935	-0.6	-2.3	3.2	12.6
	3	0.909	-1.4	-5.1	3.7	13.9
	4	0.915	-1.2	-4.8	3.6	13.8
Dynamic + 1d5p	1	0.902	-0.2	-0.8	3.9	14.8
	2	0.932	0.0	0.0	3.3	12.9
	3	0.909	-0.8	-2.8	3.7	13.9
	4	0.915	-0.6	-2.5	3.6	13.7

Table 5.10 - Statistical indicators of SET model output coupled with each thermal model for each system.

Model	System	R^2	MBE (W)	rMBE (%)	RMSE (W)	rRMSE (%)
Measured temperature	1	0.988	26.2	4.0	61.8	9.3
	2	0.992	4.0	0.6	52.5	7.9
	3	0.991	2.3	0.3	55.4	8.4
	4	0.989	22.4	3.4	59.4	9.0
Modified Chenni	1	0.982	41.6	6.3	76.6	11.6
	2	0.991	12.5	1.9	55.9	8.5
	3	0.989	20.3	3.1	60.4	9.2
	4	0.984	36.0	5.4	71.8	10.9
Mattei	1	0.989	28.1	4.2	61.3	9.3
	2	0.992	5.4	0.8	52.8	8.0
	3	0.991	6.5	1.0	53.4	8.1
	4	0.989	25.2	3.8	60.1	9.1
Kurtz (Sandia1)	1	0.991	15.3	2.3	54.6	8.3
	2	0.992	-4.0	-0.6	52.9	8.0
	3	0.990	-5.7	-0.9	56.8	8.6
	4	0.991	12.8	1.9	53.7	8.1
Sandia3	1	0.990	21.7	3.3	57.2	8.6
	2	0.992	0.0	0.0	52.3	7.9
	3	0.991	0.7	0.1	54.1	8.2
	4	0.991	18.5	2.8	55.8	8.4
Faiman	1	0.990	18.6	2.8	57.0	8.6
	2	0.992	-2.0	-0.3	52.7	8.0
	3	0.991	-2.5	-0.4	55.9	8.5
	4	0.991	15.7	2.4	55.5	8.4
JIS1	1	0.990	18.2	2.8	56.8	8.6
	2	0.992	-2.2	-0.3	52.8	8.0
	3	0.990	-2.8	-0.4	56.9	8.6
	4	0.991	15.4	2.3	55.3	8.4
Migan	1	0.990	20.9	3.2	57.6	8.7
	2	0.992	-0.5	-0.1	52.4	7.9
	3	0.991	-0.2	0.0	54.8	8.3
	4	0.991	17.7	2.7	56.1	8.5
Dynamic	1	0.990	22.4	3.4	56.2	8.5
	2	0.992	2.9	0.4	51.5	7.8
	3	0.991	0.7	0.1	54.1	8.2
	4	0.991	21.9	3.3	55.9	8.4

Table 5.11 - Statistical indicators of single diode and 5 parameters model output coupled with each thermal model for each system.

Model	System	R^2	MBE (W)	rMBE (%)	RMSE (W)	rRMSE (%)
Measured temperature	1	0.970	36.6	5.5	99.1	15.0
	2	0.982	4.6	0.7	77.6	11.7
	3	0.981	5.1	0.8	78.9	12.0
	4	0.980	18.8	2.8	81.7	12.4
Modified Chenni	1	0.970	36.7	5.5	99.8	15.1
	2	0.982	3.6	0.5	77.3	11.7
	3	0.980	6.5	1.0	81.1	12.3
	4	0.978	21.8	3.3	86.2	13.0
Mattei	1	0.970	36.8	5.6	99.5	15.0
	2	0.982	4.7	0.7	77.6	11.7
	3	0.981	5.7	0.9	79.6	12.1
	4	0.979	19.7	3.0	83.1	12.6
Kurtz (Sandia1)	1	0.971	36.5	5.5	98.4	14.9
	2	0.982	5.5	0.8	77.6	11.7
	3	0.982	4.4	0.7	77.9	11.8
	4	0.981	16.6	2.5	78.8	11.9
Sandia3	1	0.970	36.7	5.6	99.0	15.0
	2	0.982	5.2	0.8	77.6	11.7
	3	0.981	5.1	0.8	78.8	11.9
	4	0.980	18.1	2.7	80.8	12.2
Faiman	1	0.971	36.6	5.5	98.6	14.9
	2	0.982	5.3	0.8	77.5	11.7
	3	0.981	4.7	0.7	78.3	11.9
	4	0.981	17.3	2.6	79.8	12.1
JIS1	1	0.971	36.4	5.5	98.4	14.9
	2	0.982	5.2	0.8	77.6	11.7
	3	0.982	4.5	0.7	77.9	11.8
	4	0.981	17.0	2.6	79.2	12.0
Migan	1	0.971	36.7	5.5	98.9	15.0
	2	0.982	5.2	0.8	77.6	11.7
	3	0.981	5.0	0.8	78.6	11.9
	4	0.980	17.8	2.7	80.6	12.2
Dynamic	1	0.971	36.5	5.5	98.8	14.9
	2	0.982	5.0	0.8	77.8	11.8
	3	0.982	4.7	0.7	78.2	11.9
	4	0.981	18.0	2.7	80.3	12.1

References

1. Kalogirou, S.A. *Solar Energy Engineering: Processes and Systems*; Academic Press Inc., 2009;
2. Gholami, A.; Ameri, M.; Zandi, M.; Gavagsaz Ghoachani, R. Electrical, Thermal and Optical Modeling of Photovoltaic Systems: Step-by-Step Guide and Comparative Review Study. *Sustainable Energy Technologies and Assessments* **2022**, *49*, 101711, doi:10.1016/j.seta.2021.101711.
3. Castro, R. Data-Driven PV Modules Modeling: Comparison between Equivalent Electric Circuit and Artificial Intelligence Based Models. *Sustainable Energy Technologies and Assessments* **2018**, *30*, 230–238, doi:10.1016/J.SETA.2018.10.011.
4. Evans, D.L.; Florschuetz, L.W. Cost Studies on Terrestrial Photovoltaic Power Systems with Sunlight Concentration. *Solar Energy* **1977**, *19*, 255–262, doi:10.1016/0038-092X(77)90068-8.
5. Abbassi, R.; Abbassi, A.; Jemli, M.; Chebbi, S. Identification of Unknown Parameters of Solar Cell Models: A Comprehensive Overview of Available Approaches. *Renewable and Sustainable Energy Reviews* 2018, *90*.
6. Mermoud, A.; Lejeune, T. Performance Assessment of a Simulation Model for PV Modules of Any Available Technology. In Proceedings of the Proceedings of the 25th European Photovoltaic Solar Energy Conference; 2010.
7. Chin, V.J.; Salam, Z.; Ishaque, K. Cell Modeling and Model Parameters Estimation Techniques for Photovoltaic Simulator Application: A Review. *Appl Energy* **2015**, *154*, 500–519, doi:10.1016/j.apenergy.2015.05.035.
8. Chenni, R.; Makhlouf, M.; Kerbache, T.; Bouzid, A. A Detailed Modeling Method for Photovoltaic Cells. *Energy* **2007**, *32*, 1724–1730, doi:10.1016/J.ENERGY.2006.12.006.
9. Ma, T.; Yang, H.; Lu, L. Solar Photovoltaic System Modeling and Performance Prediction. *Renewable and Sustainable Energy Reviews* **2014**, *36*, 304–315, doi:10.1016/J.RSER.2014.04.057.

10. Hasan, M.A.; Parida, S.K. An Overview of Solar Photovoltaic Panel Modeling Based on Analytical and Experimental Viewpoint. *Renewable and Sustainable Energy Reviews* **2016**, *60*, 75–83, doi:10.1016/j.rser.2016.01.087.
11. Roberts, J.J.; Mendiburu Zevallos, A.A.; Cassula, A.M. Assessment of Photovoltaic Performance Models for System Simulation. *Renewable and Sustainable Energy Reviews* **2017**, *72*, 1104–1123, doi:10.1016/J.RSER.2016.10.022.
12. Fahim, S.R.; Hasanien, H.M.; Turky, R.A.; Aleem, S.H.E.A.; Ćalasan, M. A Comprehensive Review of Photovoltaic Modules Models and Algorithms Used in Parameter Extraction. *Energies* **2022**, *15*, 1–56, doi:10.3390/en15238941.
13. Tuomiranta, A.; Marpu, P.; Munawwar, S.; Ghedira, H. Validation of Thermal Models for Photovoltaic Cells under Hot Desert Climates. *Energy Procedia* **2014**, *57*, 136–143, doi:10.1016/j.egypro.2014.10.017.
14. Coskun, C.; Toygar, U.; Sarpdag, O.; Oktay, Z. Sensitivity Analysis of Implicit Correlations for Photovoltaic Module Temperature: A Review. *J Clean Prod* **2017**, *164*, 1474–1485, doi:10.1016/j.jclepro.2017.07.080.
15. Skoplaki, E.; Palyvos, J.A. Operating Temperature of Photovoltaic Modules: A Survey of Pertinent Correlations. *Renew Energy* **2009**, *34*, 23–29, doi:10.1016/j.renene.2008.04.009.
16. Gholami, A.; Ameri, M.; Zandi, M.; Gavagsaz Ghoachani, R.; Jafarzadegan Gerashi, S.; Kazem, H.A.; Al-Waeli, A.H.A. Impact of Harsh Weather Conditions on Solar Photovoltaic Cell Temperature: Experimental Analysis and Thermal-Optical Modeling. *Solar Energy* **2023**, *252*, 176–194, doi:10.1016/J.SOLENER.2023.01.039.
17. Kazem, H.A.; Chaichan, M.T.; Al-Waeli, A.H.A.; Al-Badi, R.; Fayad, M.A.; Gholami, A. Dust Impact on Photovoltaic/Thermal System in Harsh Weather Conditions. *Solar Energy* **2022**, *245*, 308–321, doi:10.1016/J.SOLENER.2022.09.012.
18. Gholami, A.; Ameri, M.; Zandi, M.; Gavagsaz Ghoachani, R.; Kazem, H.A. Predicting Solar Photovoltaic Electrical Output under Variable Environmental Conditions: Modified Semi-Empirical Correlations for Dust.

Energy for Sustainable Development **2022**, *71*, 389–405, doi:10.1016/J.ESD.2022.10.012.
19. Perovic, B.; Klimenta, D.; Jevtic, M.; Milovanovic, M. A Transient Thermal Model for Flat-Plate Photovoltaic Systems and Its Experimental Validation. *Elektronika ir Elektrotechnika* **2019**, *25*, 40–46, doi:10.5755/j01.eie.25.2.23203.
20. Li, F.; Wu, W. Coupled Electrical-Thermal Performance Estimation of Photovoltaic Devices : A Transient Multiphysics Framework with Robust Parameter Extraction and 3-D Thermal Analysis. *Appl Energy* **2022**, *319*, 119249, doi:10.1016/j.apenergy.2022.119249.
21. Torres-Lobera, D.; Valkealahti, S. Inclusive Dynamic Thermal and Electric Simulation Model of Solar PV Systems under Varying Atmospheric Conditions. *Solar Energy* **2014**, *105*, 632–647, doi:10.1016/j.solener.2014.04.018.
22. Gu, W.; Ma, T.; Shen, L.; Li, M.; Zhang, Y.; Zhang, W. Coupled Electrical-Thermal Modeling of Photovoltaic Modules under Dynamic Conditions. *Energy* **2019**, *188*, 116043, doi:10.1016/j.energy.2019.116043.
23. Tina, G. A Coupled Electrical and Thermal Model for Photovoltaic Modules. *J Sol Energy Eng* **2010**, *132*, doi:10.1115/1.4001149.
24. Bevilacqua, P.; Perrella, S.; Bruno, R.; Arcuri, N. An Accurate Thermal Model for the PV Electric Generation Prediction: Long-Term Validation in Different Climatic Conditions. *Renew Energy* **2021**, *163*, 1092–1112, doi:10.1016/j.renene.2020.07.115.
25. Tuncel, B.; Ozden, T.; Balog, R.S.; Akinoglu, B.G. Dynamic Thermal Modeling of PV Performance and Effect of Heat Capacity on the Module Temperature. *Case Studies in Thermal Engineering* **2020**, *22*, 100754, doi:10.1016/j.csite.2020.100754.
26. Mattei, M.; Notton, G.; Cristofari, C.; Muselli, M.; Poggi, P. Calculation of the Polycrystalline PV Module Temperature Using a Simple Method of Energy Balance. *Renew Energy* **2006**, *31*, 553–567, doi:10.1016/J.RENENE.2005.03.010.

27. Sandia National Laboratories Sandia Module Temperature Model Available online: https://pvpmc.sandia.gov/modeling-steps/2-dc-module-iv/module-temperature/sandia-module-temperature-model/ (accessed on 24 January 2023).
28. Faiman, D. Assessing the Outdoor Operating Temperature of Photovoltaic Modules. *Progress in Photovoltaics: Research and Applications* **2008**, *16*, 307–315, doi:10.1002/pip.813.
29. Japanese Industrial Standard - Japanese Standards Association Estimation Method of Generating Electric Energy by PV Power System 2005.
30. Migan, G.-A. Project Report 2013 MVK160 Heat and Mass Transfer Study of the Operating Temperature of a PV Module. **2013**.
31. Luketa-Hanlin, A.; Stein, J.S. Improvement and Validation of a Transient Model to Predict Photovoltaic Module Temperature. In Proceedings of the World Renewable Energy Forum; 2012.
32. Kaplanis, S.; Kaplani, E. A New Dynamic Model to Predict Transient and Steady State Pv Temperatures Taking into Account the Environmental Conditions. *Energies* **2019**, *12*, doi:10.3390/en12010002.
33. Fujii, T.; Imura, H. Natural-Convection Heat Transfer from a Plate with Arbitrary Inclination. *Int J Heat Mass Transf* **1972**, *15*, 755–767, doi:10.1016/0017-9310(72)90118-4.
34. Kaplani, E.; Kaplanis, S. Thermal Modeling and Experimental Assessment of the Dependence of PV Module Temperature on Wind Velocity and Direction, Module Orientation and Inclination. *Solar Energy* **2014**, *107*, 443–460, doi:https://doi.org/10.1016/j.solener.2014.05.037.
35. Reiter, C.N.; Trinkl, C.; Zörner, W.; Hanby, V.I. A Dynamic Multinode Model for Component-Oriented Thermal Analysis of Flat-Plate Solar Collectors. *Journal of Solar Energy* **2015**, *2015*, 1–16, doi:10.1155/2015/280694.
36. Sartori, E. Convection Coefficient Equations for Forced Air Flow over Flat Surfaces. *Solar Energy* **2006**, *80*, 1063–1071, doi:10.1016/J.SOLENER.2005.11.001.

37. Swinbank, W.C. Long-Wave Radiation from Clear Skies. *Quarterly Journal of the Royal Meteorological Society* **1963**, *89*, 339–348, doi:https://doi.org/10.1002/qj.49708938105.
38. Dormand, J.R.; Prince, P.J. A Reconsideration of Some Embedded Runge-Kutta Formulae. *J Comput Appl Math* **1986**, *15*, 203–211, doi:10.1016/0377-0427(86)90027-0.
39. Pereira, S.; Canhoto, P.; Salgado, R. Development and Assessment of Artificial Neural Network Models for Direct Normal Solar Irradiance Forecasting Using Operational Numerical Weather Prediction Data. *Energy and AI* **2023**, 100314, doi:10.1016/j.egyai.2023.100314.
40. Duffie, J.A.; Beckman, W.A. Solar Engineering of Thermal Processes. **2013**, 1–885.
41. De Soto, W.; Klein, S.A.; Beckman, W.A. Improvement and Validation of a Model for Photovoltaic Array Performance. *Solar Energy* **2006**, *80*, 78–88, doi:10.1016/j.solener.2005.06.010.
42. IEC *IEC 60891:2021 ED3 Photovoltaic Devices - Procedures for Temperature and Irradiance Corrections to Measured I-V Characteristics*; 2021;
43. Ustun, T.S.; Nakamura, Y.; Hashimoto, J.; Otani, K. Performance Analysis of PV Panels Based on Different Technologies after Two Years of Outdoor Exposure in Fukushima, Japan. *Renew Energy* **2019**, *136*, 159–178, doi:10.1016/J.RENENE.2018.12.100.

Nomenclature

A	Area of the photovoltaic module (m^2)
A_s	Photovoltaic module azimuth (°)
C_{mod}	Heat capacity of the module (J/K)
c_{p,T_f}	Specific heat capacity of air film at temperature T_f (J/Kg/K)
DHI	Direct horizontal irradiance (W/m^2)
DIF	Diffuse horizontal irradiance (W/m^2)
DNI	Direct normal irradiance (W/m^2)
g	Gravitational Acceleration (9.8 m/s^2)

Gr	Grashof number
GTI	Global tilted irradiance (W/m^2)
h	Convective heat transfer coefficient (W/m^2/K)
h_{rad}	Radiative heat transfer coefficient (W/m^2/K)
I_0	Reverse saturation current (A)
IAM	Incidence angle modifier
I	Electric current (A)
I_{mp}	Current at maximum power point (A)
I_{pv}	Photovoltaic current (A)
I_{sc}	Short-circuit current (A)
K	Boltzman constant (1.38×10^{-23} J/K)
K_g	Extinction coefficient of the glazing (m^{-1})
k_{T_f}	Thermal conductivity of air film at temperature T_f (W/m/K)
L	Length of the module (m)
L_c	Characteristic length of the module (m)
L_{free}	Length of the module along the natural convection air flow (m)
L_g	Glazing thickness (m)
m	Diode ideality factor
n	Refractive index
N_s	Number of photovoltaic cells in series
Nu	Nusselt number
P	Power output (W)
P_e	Estimated power output (W)
P_{max}	Maximum power output (W)
Pr	Prandtl number
q	Electrical charge of the electron (1.6×10^{-19} C)
Q_{cond}	Net rate of energy loss to the environment by conduction (W)
Q_{conv}	Net rate of energy loss to the environment by convection(W)
Q_{rad}	Net rate of energy loss to the environment by radiation (W)
Q_{sun}	Net in-plane solar irradiance (W)
Re	Reynolds number
Ra	Rayleigh number

R_p	Parallel or shunt resistance (Ω)
R_s	Series resistance (Ω)
S	Incident irradiance (W/m^2)
T	Temperature (°C or K)
T_a	Air temperature (°C)
T_f	Temperature of air film at the surface of the module (°C)
T_{ground}	Temperature of the ground surface(°C)
T_{mod}	Mean temperature of the photovoltaic module (°C)
T_{sky}	Apparent sky temperature (°C)
U_{pv}	Coefficient of thermal losses from Mattei model (W/°C/m^2)
V	Voltage (V)
V_{mp}	Voltage at maximum power point (V)
V_{oc}	Open-circuit voltage (V)
V_T	Thermal voltage (V)
W	Width of the module (m)
WD	Wind direction (North is assumed 0°) (°)
WD_s	Wind direction (South is assumed 0°) (°)
WS	Wind speed (m/s)
x_{cr}	Critical length (m)

Greek symbols

$\alpha_{I_{sc}}$	Thermal coefficient of short-circuit current (%/°C)
$\alpha_{P_{max}}$	Thermal coefficient of maximum power (%/°C)
$\alpha_{V_{oc}}$	Thermal coefficient of open-circuit voltage (%/°C)
α_{T_f}	Thermal diffusivity of air film at temperature T_f (m^2/s)
β	Tilt angle (°)
ε	Silicon bandgap (= 1.12 eV)
ε_b	Emissivity of the back surface of the module
ε_f	Emissivity of the front surface of the module
η	Photovoltaic module efficiency
θ	Angle of incidence (°)

θ_r	Angle of refraction (°)
ν_{T_f}	Kinematic viscosity of air (m^2/s)
σ_{SB}	Stefan-Boltzmann constant (5.670 x 10^{-8} $W/m^2/K^4$)
τ	Transmittance

Acronyms and Abbreviations

1d3p	Single diode and 3 parameters model
1d4p	Single diode and 4 parameters model
$1d4p_m$	Single diode and 4 parameters model with ideality factor adjustment
1d5p	Single diode and 5 parameters model
GPI	Global performance index
HIT	Heterojunction intrinsic thin layer
IEC	International electrotechnical commission
MBE	Mean bias error
rMBE	Relative mean bias error
RMSE	Root mean squared error
R^2	Coefficient of determination
rRMSE	Relative root mean squared error
SE	Simple estimate model
SET	Simple estimate with temperature correction model
STC	Standard test conditions

Chapter 6

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

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-hour power forecasts with customizable temporal resolution, without the need for on-site observations. Validation against 15-minute 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

†Sara Pereira[a], Paulo Canhoto[a,b], Takashi Oozeki[c] and Rui Salgado[a,d] 2025. Comprehensive approach to photovoltaic power forecasting using numerical weather prediction data and physics-based models and data-driven techniques. Renewable Energy (under review).
[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, Fukushima, Japan
[d] Physics Department, University of Évora, Évora, Portugal

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.

Keywords
Forecasting; Grid integration; Photovoltaic power; Solar energy; Solar power plant.

6.1 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 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 [37] a hybrid architecture of recurrent neural networks and shallow neural networks was developed showing improved performance in predicting daily photovoltaic power generation. In [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 [39] the authors improved hourly direct normal irradiance predictions. In [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 six hours 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-hour 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 6.2 while its validation against observations, including results and discussion, is presented in Section 6.3. Finally, the conclusions are drawn in Section 6.4.

6.2. Algorithm for photovoltaic power forecasting

The flowchart of the developed integrated forecasting algorithm is presented in Fig. 6.1, 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).

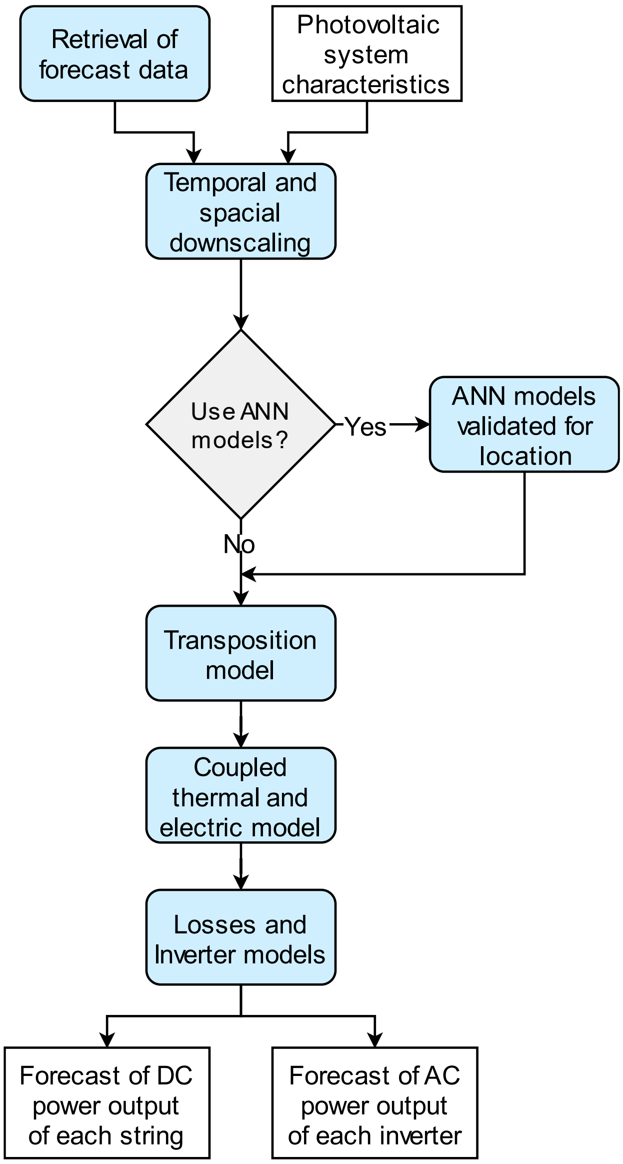


Fig. 6.1 - Flowchart of the developed algorithm for photovoltaic power forecasting using non-observational data.

The IFS/ECMWF is the most widely used global NWP model in Europe being its performance attested by various studies such as in [2–4]. These forecasts are downscaled to the desired temporal resolution and for the location of interest with a temporal horizon of 72 hours. The choice of a 72-hour forecast horizon ensures the algorithm's adaptability to various operational needs. While 24-hour forecasts are standard for grid operators, extending the forecast horizon to 72 hours provides valuable insights for medium-term planning, including resource allocation and maintenance scheduling. Additionally, the algorithm is designed to run and update every 24 hours, 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 as those developed in [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 modelled 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.

6.2.1 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 hours into the future. Beyond this period, the model provides forecasts at 3-hour intervals up to 144 hours and at 6-hour intervals up to 240 hours, all at discrete points across a global grid covering the entire globe with a horizontal spatial resolution of 0.125° × 0.125°.

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 kilometers and a temporal step of 1 hour [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 hours 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 6.1. The ANN model developed in [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 the region of interest through a similar procedure as described in [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.

Table 6.1 - 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 550nm*
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 (°)	

Table 6.2 - 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)		

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 6.2.

6.2.2 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. 6.2.

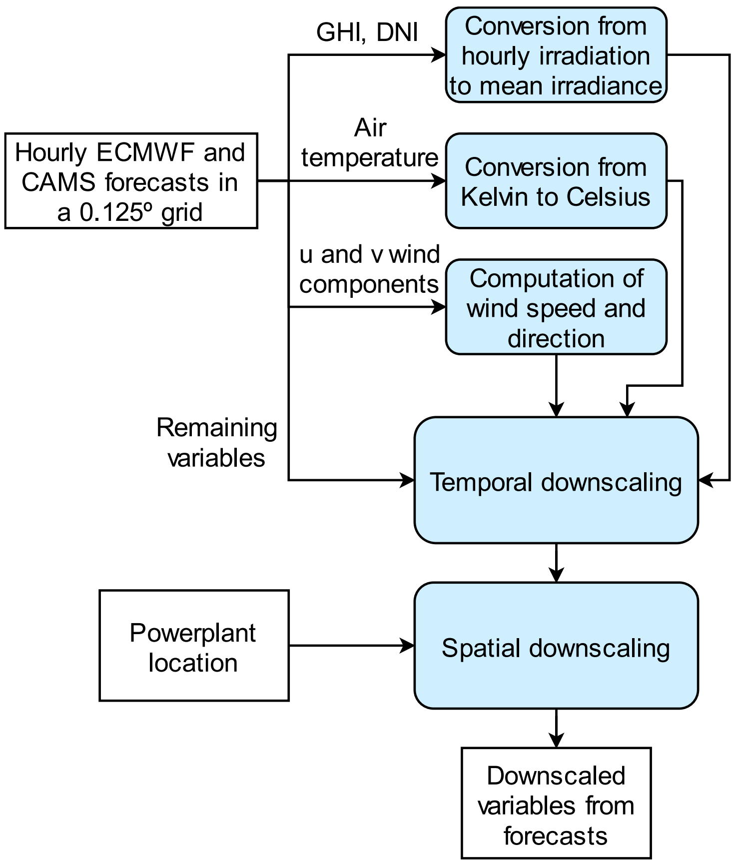


Fig. 6.2 - Flowchart of the temporal and spacial downscaling procedure.

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.

6.2.3 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. 6.3.

The data used as inputs are downscaled forecasts of operational outputs from the ECMWF/IFS and the CAMS models as described in Section 6.2.1 and 6.2.2 of the variables presented in Table 6.1. 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].

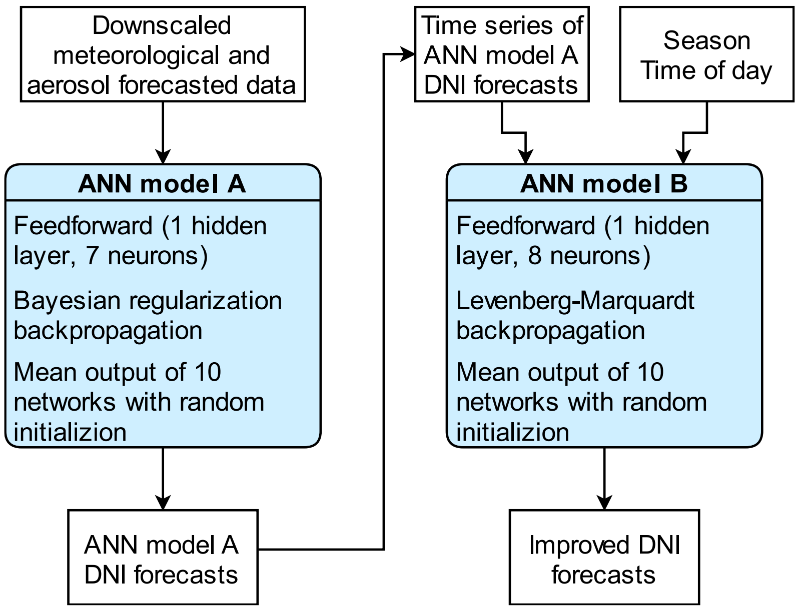


Fig. 6.3 - Flowchart of the ANN models used in the algorithm.

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.

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].

6.2.4 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 componets 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. 6.4 shows this part of the algorithm.

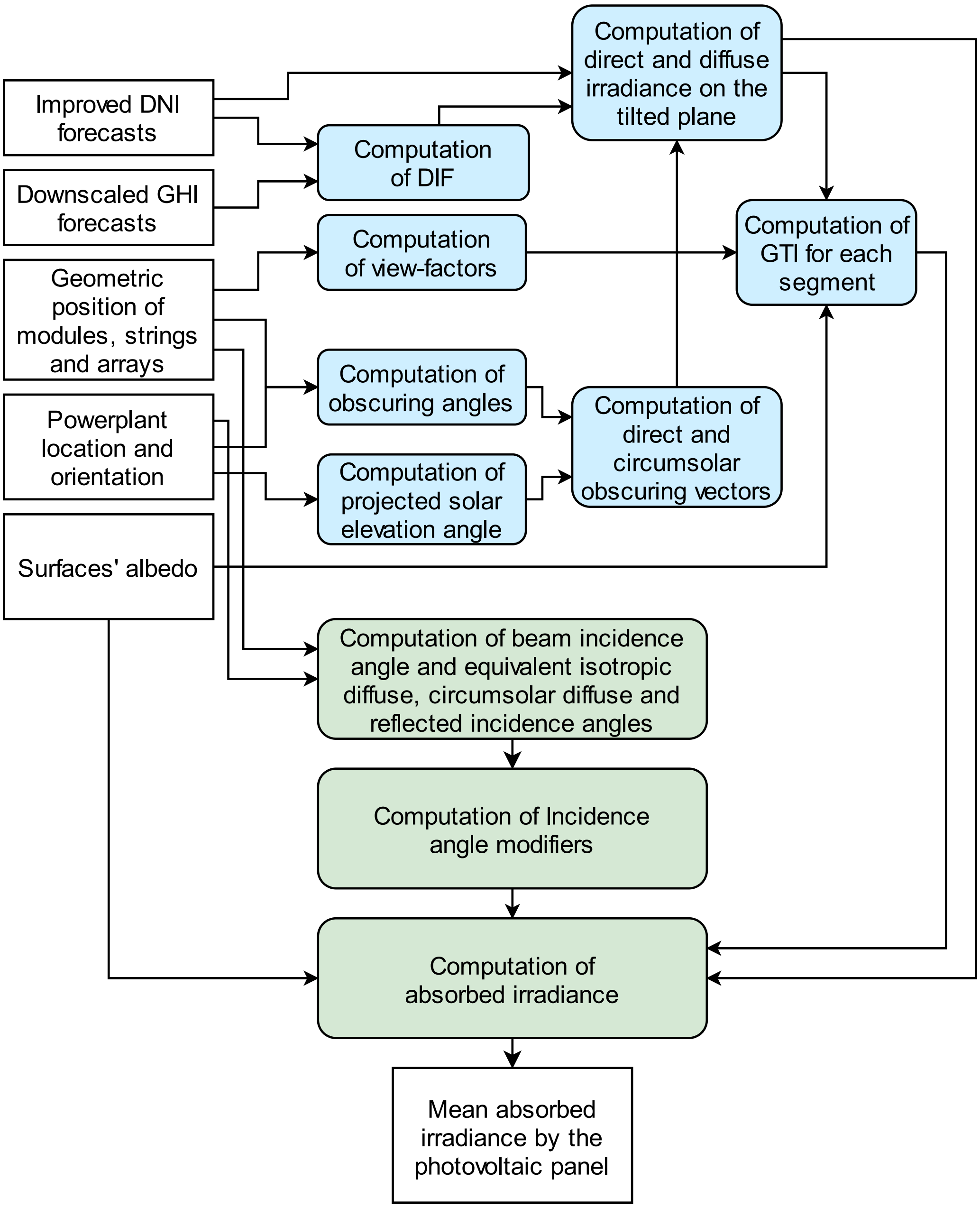


Fig. 6.4 - Flowchart of the transposition model and computation of absorbed irradiance.

The transposition model from reference [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 [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 [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 [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.

6.2.5 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 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 [17] is used, which is a dynamic coupled thermal-electric model designed for crystalline silicon cells using readily available information provided by the manufacturers.

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.

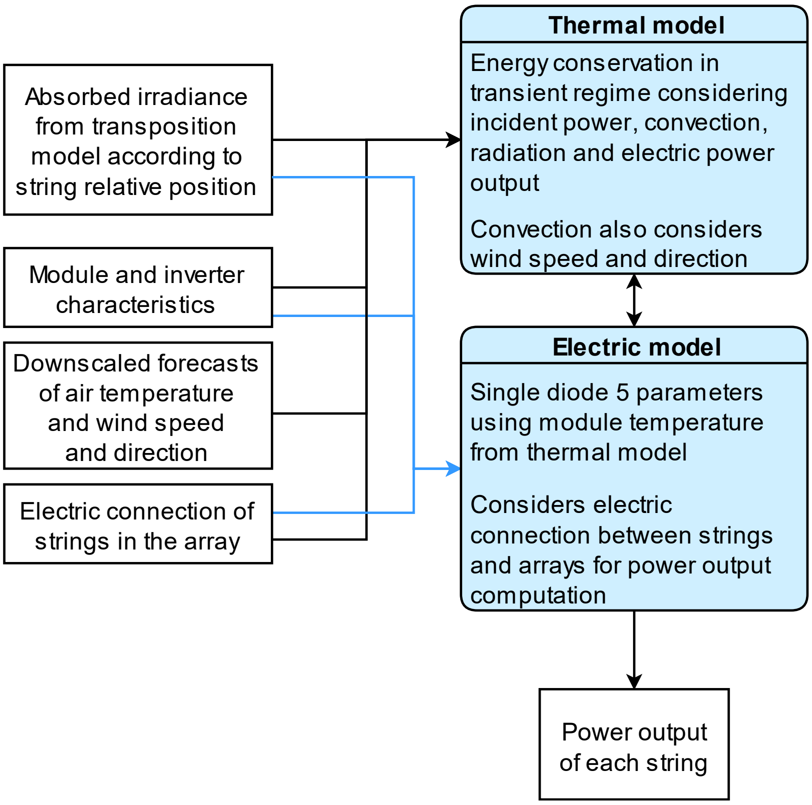


Fig. 6.5 - Flowchart of the coupled thermal-electric model of photovoltaic modules.

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.

6.2.6 Electric losses and inverter model

Fig. 6.6 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 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 6.3. 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.

Table 6.3 - Typical values of the derating factors [58].

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	0.985
Initial light-induced degradation	0.980

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.

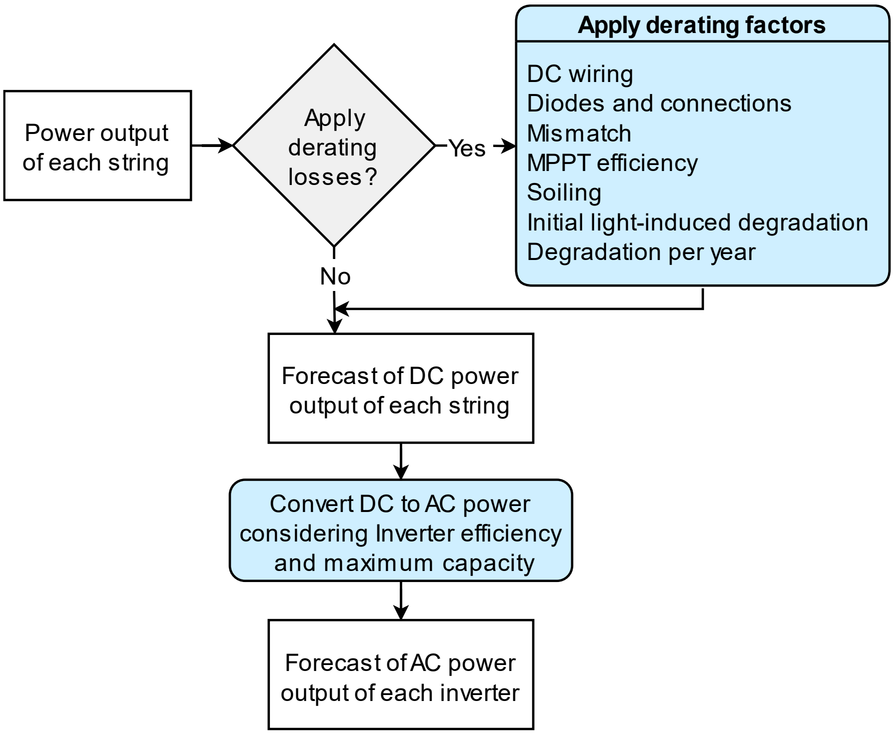


Fig. 6.6 - Flowchart of the electric losses and inverter efficiency models.

6.2.7 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 minutes.

The following is an example of an excerpt of an output file defined with 15-minute 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
(…)

6.3 Algorithm results and validation

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

6.3.1 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. 6.7) 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 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 6.4, 6.5 and 6.6, respectively.

Fig. 6.7 - Aerial view of the photovoltaic powerplant (1-7: strings considered in the work).

Table 6.4 - 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 6.5 - 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 6.6 - 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

This powerplant was commissioned in 2021 and the data used ranges from April 1st, 2022, and April 30th, 2023, with a timestep of 5 minutes. 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 minute 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 1.80GHz and 16 GB of RAM was used.

6.3.2 Results and Discussion

The integrated algorithm was run for the observation site and period described in Section 6.3.1 with a temporal forecast horizon of 72 hours and a timestep of 15 minutes. 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 (FSGTI) 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 6.7 based on all 395 algorithm runs, each providing data over a 72-hour time horizon with 15-minute 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 %.

Table 6.7 - 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

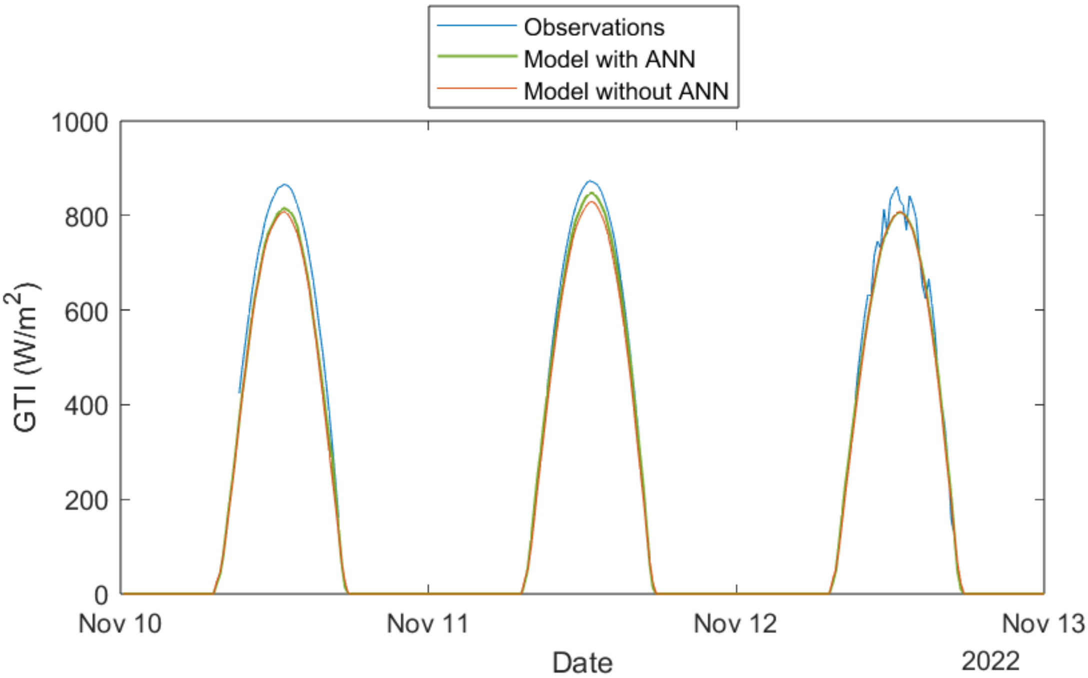


Fig. 6.8 - Observed and predicted values of GTI using the original ECMWF data and the improved DNI and DIF forecasts from the ANN models as input to the transposition model for the operational example.

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 day there are some clouds around midday. Fig. 6.8 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 6.2.5, 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 6.8 for each string and forecast day based on all 395 algorithm runs, each providing data over a 72-hour time horizon with 15-minute timestep. Here, a forecast skill score (FSPOW) 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 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.

Table 6.8 - 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

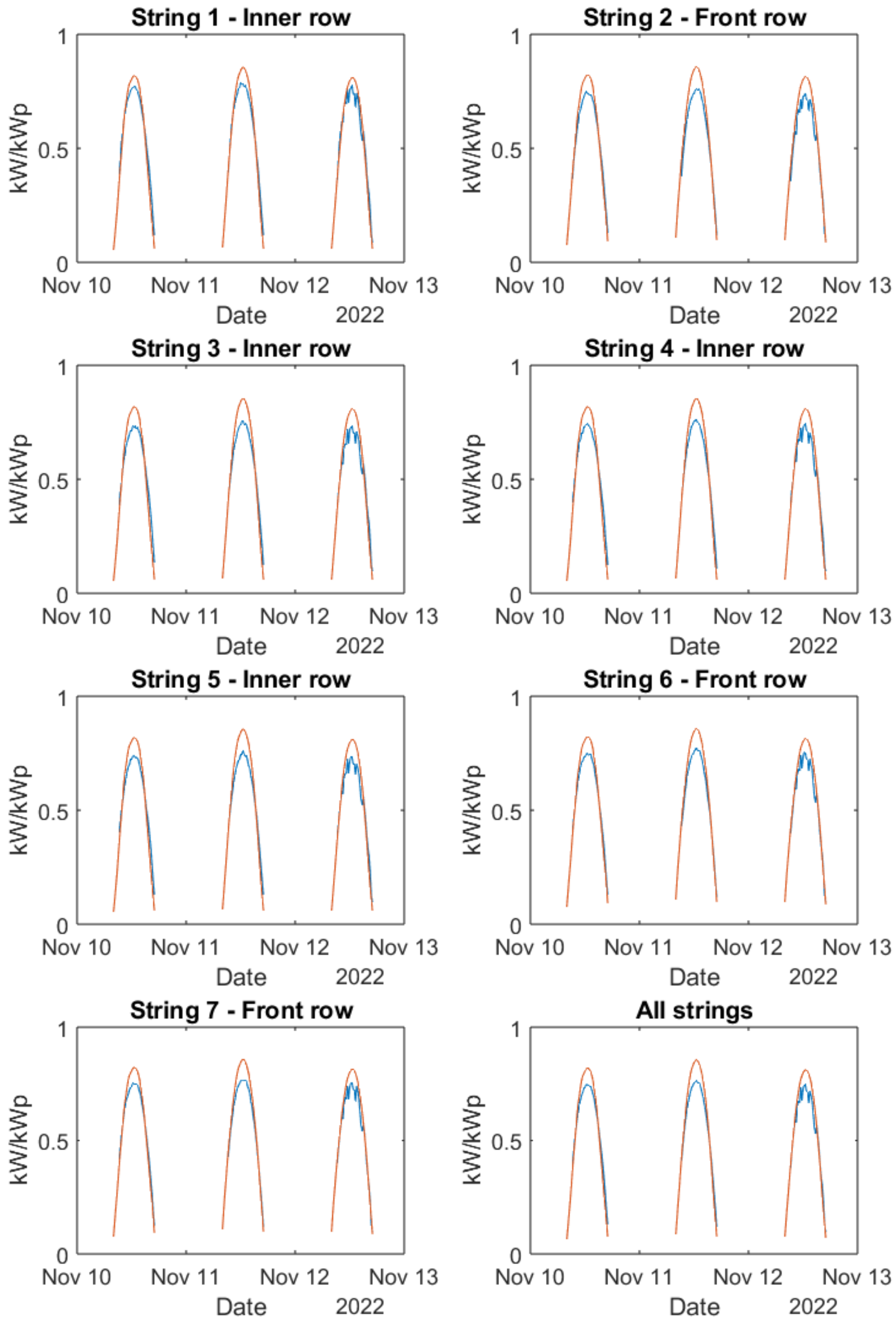


Fig. 6.9 - Power generation (DC) forecasts using the ANN model (orange) and measurements (blue) of each string for the operational example.

Fig. 6.9 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 6.2.6, the results in Table 6.9 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-hour time horizon with a 15-minute 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 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.

Table 6.9 - 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-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

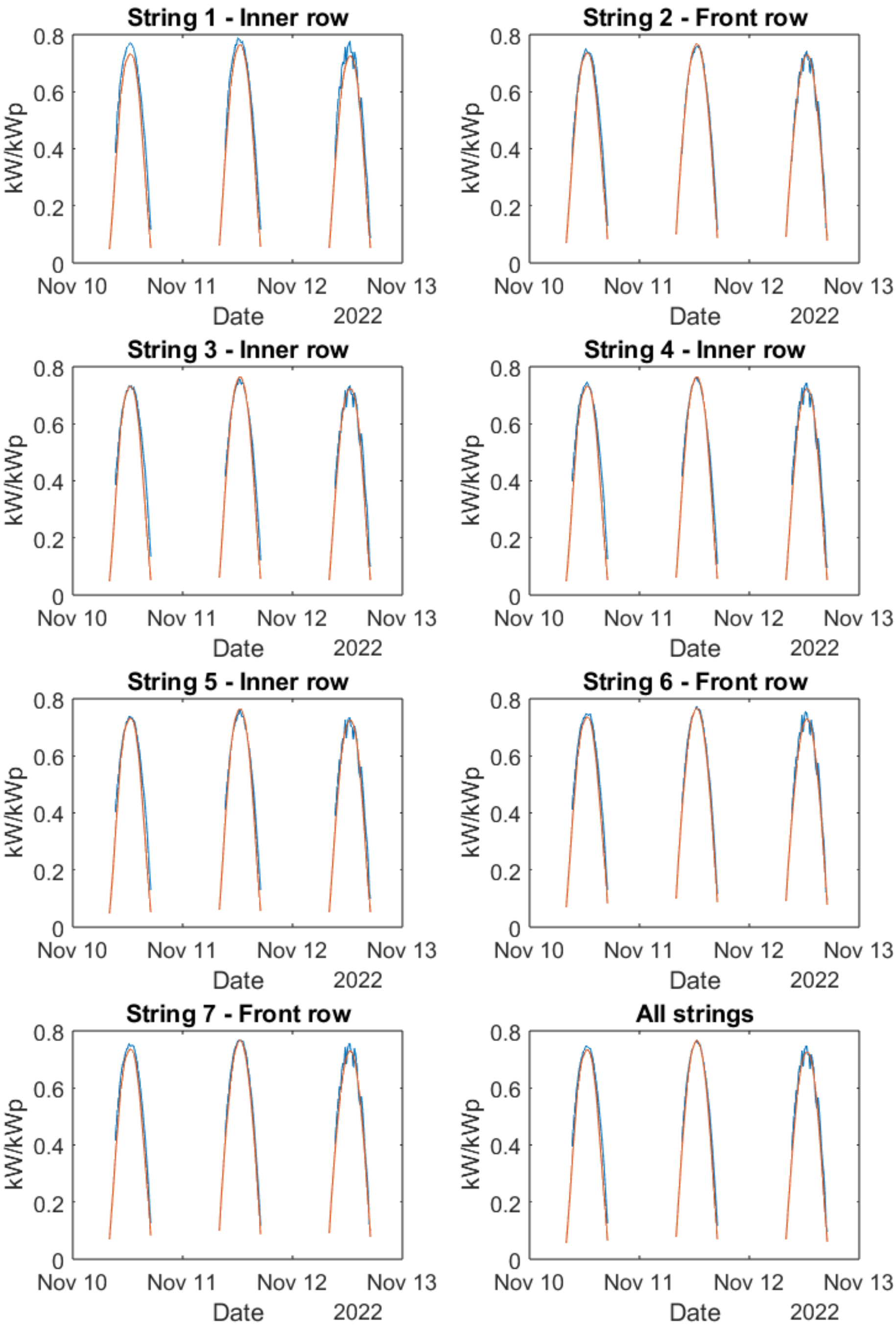


Fig. 6.10 - Power output (DC) forecasts (orange) and measurements (blue) of each string for the operational example including the electric losses.

Fig. 6.10 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 Fig. 6.9 and 6.10 and thus, the following analysis will consider the use of the losses model. There is 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 6.10 are obtained.

Table 6.10 - 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

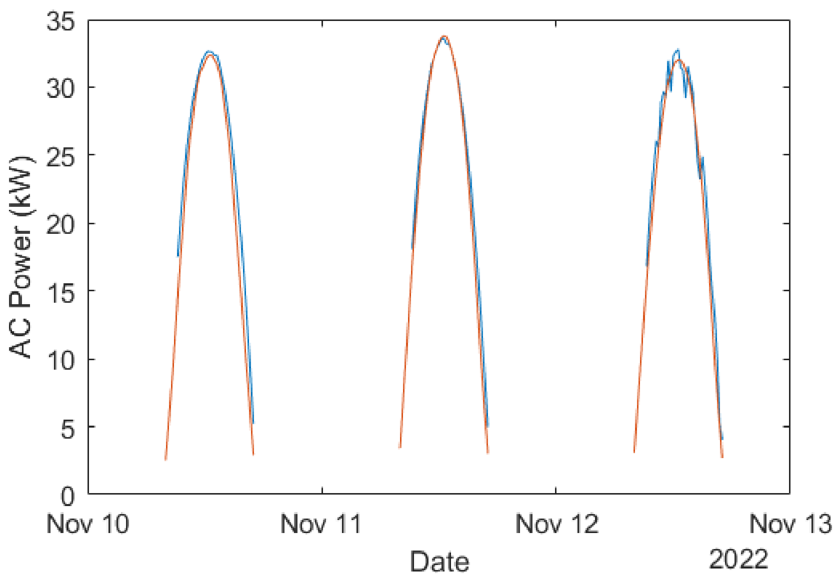


Fig. 6.11 - AC power forecasts (orange) and measurements (blue) for the operational period taken as example.

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. 6.11 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 6.10, 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 [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 6.11). As explained in Section 6.2.1, 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 minutes, thus allowing for the efficient processing and use of the forecast data.

Table 6.11 - 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

6.4 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-hour photovoltaic power forecasts, with user-defined temporal resolution.

The algorithm was validated for a temporal resolution of 15 minutes 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 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 would result in better management of photovoltaic energy generation variability.

Appendix A – Computation of absorbed irradiance

The transposition model presented in [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 [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. (6.1) 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. (6.2) 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. (6.3) to (6.6) based on the principles of Snell's and Bougher's laws, as outlined in [57]. 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 [66].

$$\vec{S} = \vec{\gamma}_b\left(\vec{I} + \vec{D}_{cs}\right) + \vec{\gamma}_d\vec{D}_{iso} + \boldsymbol{R}\overrightarrow{GTI} \tag{6.1}$$

$$\boldsymbol{R} = \boldsymbol{\rho F \gamma_r} \tag{6.2}$$

$$\gamma = \frac{\tau(\theta)}{\tau(0)} \tag{6.3}$$

$$\tau(\theta) = e^{-\left(\frac{K_g L_g}{cos(\theta_r)}\right)}\left[1 - \frac{1}{2}\left(\frac{sin^2(\theta_r - \theta)}{sin^2(\theta_r + \theta)} + \frac{tan^2(\theta_r - \theta)}{tan^2(\theta_r + \theta)}\right)\right] \tag{6.4}$$

$$\tau(0) = e^{-(K_g L_g)}\left[1 - \left(\frac{1-n}{1+n}\right)^2\right] \tag{6.5}$$

$$\theta_r = sin^{-1}\left[\frac{1}{n}sin\,(\theta)\right] \tag{6.6}$$

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. (6.7), 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 [57]. For first rows Eq. (6.8) can be used as presented in [57]. 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. (6.9)), 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 $\boldsymbol{\theta}_{\boldsymbol{ref,v}}$ and from the ground between rows $\boldsymbol{\theta}_{\boldsymbol{ref,u}}$. This is done through Eqs. (6.10) and (6.11) 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.

$$\theta_b = \cos^{-1}\begin{pmatrix}\sin Lat \sin\delta\cos\beta - \cos Lat \sin\delta\sin\beta\cos\varphi + \cos Lat\cos\delta\cos h_s\cos\beta + \\ + \sin Lat\cos\delta\cos h_s\sin\beta\cos\varphi + \cos\delta\sin h_s\sin\beta\sin\varphi\end{pmatrix} \tag{6.7}$$

$$\theta_d = 59.7 - 0.1388\beta + 0.001497\beta^2 \tag{6.8}$$

$$\overrightarrow{\beta'} = \tan^{-1}\left(\frac{rL \times \sin\beta - \vec{c}\sin\beta}{D - rL \times \cos\beta + \vec{c}\cos\beta}\right) \tag{6.9}$$

$$\boldsymbol{\theta_{ref,v}} = \left|90 - \beta - \tan^{-1}\left(\frac{(\vec{v} - \vec{c})\sin\beta}{D + (\vec{c} - \vec{v})\cos\beta}\right)\right| \tag{6.10}$$

$$\boldsymbol{\theta_{ref,u}} = 90 - \beta + \tan^{-1}\left(\frac{\vec{c}\sin\beta + h_0}{D + (\vec{c} - \vec{u})\cos\beta}\right) \tag{6.11}$$

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:

$$C_{mod}\frac{dT_{mod}}{dt} = Q_{sun} - Q_{conv} - Q_{rad} - P_e \tag{6. 12}$$

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. 6.12 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 [17].

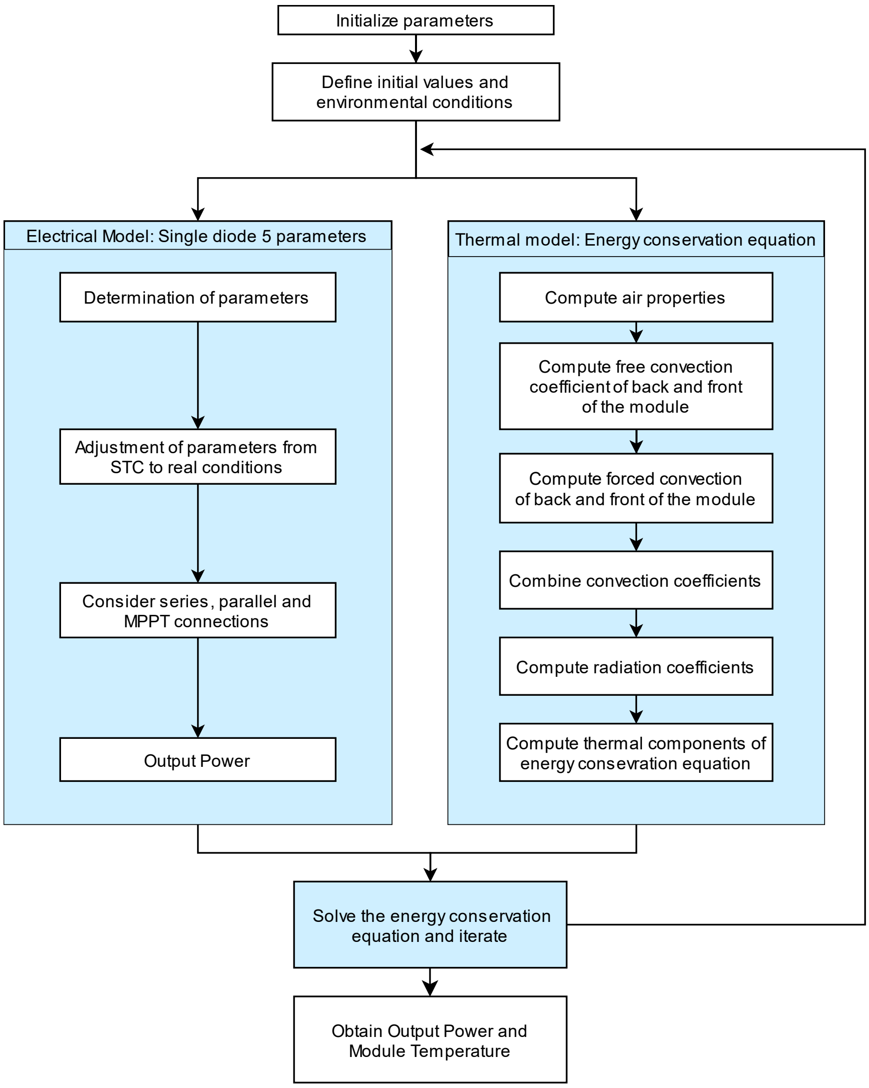


Fig. 6.12 - Flowchart of the thermal and the electric models and its coupling.

References

1. Pereira, S.; Canhoto, P.; Salgado, R. Prediction of Global Solar Irradiance on Parallel Rows of Tilted Surfaces Including the Effect of Direct and Anisotropic Diffuse Shading. *Energies (Basel)* **2024**, *17*, 3444, doi:10.3390/en17143444.
2. El Alani, O.; Ghennioui, A.; Ghennioui, H.; Saint-Drenan, Y.-M.; Blanc, P. Evaluation of 24-Hours Forecasts of Global Solar Irradiation from IFS, GFS and McClear Models. In Proceedings of the AIP Conference Proceedings; 2020.
3. Perdigão, J.; Canhoto, P.; Salgado, R.; Costa, M.J. Assessment of Direct Normal Irradiance Forecasts Based on IFS/ECMWF Data and Observations in the South of Portugal. *Forecasting* **2020**, *2*, 130–150, doi:10.3390/forecast2020007.

4. Perez, R.; Lorenz, E.; Pelland, S.; Beauharnois, M.; Van Knowe, G.; Hemker, K.; Heinemann, D.; Remund, J.; Müller, S.C.; Traunmüller, W.; et al. Comparison of Numerical Weather Prediction Solar Irradiance Forecasts in the US, Canada and Europe. *Solar Energy* **2013**, *94*, 305–326, doi:10.1016/j.solener.2013.05.005.
5. Mayer, M.J.; Yang, D.; Szintai, B. Comparing Global and Regional Downscaled NWP Models for Irradiance and Photovoltaic Power Forecasting: ECMWF versus AROME. *Appl Energy* **2023**, *352*, 121958, doi:10.1016/J.APENERGY.2023.121958.
6. Gueymard, C.A. Direct and Indirect Uncertainties in the Prediction of Tilted Irradiance for Solar Engineering Applications. *Solar Energy* **2009**, *83*, 432–444, doi:10.1016/j.solener.2008.11.004.
7. Gueymard, C.A. 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. *Solar Energy* **2017**, *153*, 766–779, doi:10.1016/j.solener.2017.04.068.
8. Yang, D.; Gueymard, C.A. Ensemble Model Output Statistics for the Separation of Direct and Diffuse Components from 1-Min Global Irradiance. *Solar Energy* **2020**, *208*, 591–603, doi:10.1016/j.solener.2020.05.082.
9. Yang, D. Solar Radiation on Inclined Surfaces: Corrections and Benchmarks. *Solar Energy* **2016**, *136*, 288–302, doi:10.1016/j.solener.2016.06.062.
10. Mayer, M.J.; Gróf, G. Extensive Comparison of Physical Models for Photovoltaic Power Forecasting. *Appl Energy* **2021**, *283*, 116239, doi:10.1016/j.apenergy.2020.116239.
11. Nassar, Y.F.; Abuhamoud, N.M.; Miskeen, G.M.; El-Khozondar, H.J.; Alsadi, S.Y.; Ahwidi, O.M. 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**, 9–14, doi:10.1109/MI-STA54861.2022.9837500.
12. Nassar, Y.F.; Hafez, A.A.; Alsadi, S.Y. 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* **2020**, *7*, 1–19, doi:10.3389/fenrg.2019.00163.
13. Appelbaum, J.; Massalha, Y.; Aronescu, A. Corrections to Anisotropic Diffuse Radiation Model. *Solar Energy* **2019**, *193*, 523–528, doi:10.1016/j.solener.2019.09.090.

14. Tschopp, D.; Jensen, A.R.; Dragsted, J.; Ohnewein, P.; Furbo, S. Measurement and Modeling of Diffuse Irradiance Masking on Tilted Planes for Solar Engineering Applications. *Solar Energy* **2022**, *231*, 365–378, doi:10.1016/J.SOLENER.2021.10.083.
15. Varga, N.; Mayer, M.J. Model-Based Analysis of Shading Losses in Ground-Mounted Photovoltaic Power Plants. *Solar Energy* **2021**, *216*, 428–438, doi:10.1016/J.SOLENER.2021.01.047.
16. Gholami, A.; Ameri, M.; Zandi, M.; Gavagsaz Ghoachani, R. Electrical, Thermal and Optical Modeling of Photovoltaic Systems: Step-by-Step Guide and Comparative Review Study. *Sustainable Energy Technologies and Assessments* **2022**, *49*, 101711, doi:10.1016/j.seta.2021.101711.
17. Pereira, S.; Canhoto, P.; Oozeki, T.; Salgado, R. Assessment of Thermal Modeling of Photovoltaic Panels for Predicting Power Generation Using Only Manufacturer Data. *Energy Reports* **2024**, *12*, 1431–1448, doi:10.1016/j.egyr.2024.07.039.
18. Castro, R. Data-Driven PV Modules Modelling: Comparison between Equivalent Electric Circuit and Artificial Intelligence Based Models. *Sustainable Energy Technologies and Assessments* **2018**, *30*, 230–238, doi:10.1016/J.SETA.2018.10.011.
19. Chenni, R.; Makhlouf, M.; Kerbache, T.; Bouzid, A. A Detailed Modeling Method for Photovoltaic Cells. *Energy* **2007**, *32*, 1724–1730, doi:10.1016/J.ENERGY.2006.12.006.
20. Chin, V.J.; Salam, Z.; Ishaque, K. Cell Modelling and Model Parameters Estimation Techniques for Photovoltaic Simulator Application: A Review. *Appl Energy* **2015**, *154*, 500–519, doi:10.1016/j.apenergy.2015.05.035.
21. Li, F.; Wu, W. Coupled Electrical-Thermal Performance Estimation of Photovoltaic Devices : A Transient Multiphysics Framework with Robust Parameter Extraction and 3-D Thermal Analysis. *Appl Energy* **2022**, *319*, 119249, doi:10.1016/j.apenergy.2022.119249.
22. Perovic, B.; Klimenta, D.; Jevtic, M.; Milovanovic, M. A Transient Thermal Model for Flat-Plate Photovoltaic Systems and Its Experimental Validation. *Elektronika ir Elektrotechnika* **2019**, *25*, 40–46, doi:10.5755/j01.eie.25.2.23203.
23. Cossu, S.; Baccoli, R.; Ghiani, E. Utility Scale Ground Mounted Photovoltaic Plants with Gable Structure and Inverter Oversizing for Land-Use Optimization. *Energies 2021, Vol. 14, Page 3084* **2021**, *14*, 3084, doi:10.3390/EN14113084.
24. González-Peña, D.; García-Ruiz, I.; Díez-Mediavilla, M.; Dieste-Velasco, M.I.; Alonso-Tristán, C. Photovoltaic Prediction Software: Evaluation with Real Data

from Northern Spain. *Applied Sciences (Switzerland)* **2021**, *11*, doi:10.3390/app11115025.

25. Natural Resources Canada RETScreen Available online: https://natural-resources.canada.ca/maps-tools-and-publications/tools/modelling-tools/retscreen/7465 (accessed on 21 January 2025).
26. National Renewable Energy Laboratory; Golden, C. System Avisor Model™ Available online: https://https://sam.nrel.gov (accessed on 21 January 2025).
27. European Commission Joint Research Centre Photovoltaic Geographical Information System Available online: http://re.jrc.ec.europa.eu/pvgis (accessed on 21 January 2025).
28. PVsyst PVsyst Available online: https://www.pvsyst.com/ (accessed on 21 January 2025).
29. Valentin Software PV*SOL Available online: https://pvsol.software/en/ (accessed on 21 January 2025).
30. Bruneau, P.; Fiorelli, D.; Braun, C.; Koster, D. Hybrid-Physical Probabilistic Forecasting for a Set of Photovoltaic Systems Using Recurrent Neural Networks. *Int J Forecast* **2023**, 1–22.
31. Cotfas, D.T.; Marzband, M.; Cotfas, P.A.; Siroux, M.; Sera, D. Editorial: Forecasting Solar Radiation, Photovoltaic Power and Thermal Energy Production Applications. *Front Energy Res* **2022**, *10*, doi:10.3389/fenrg.2022.1115096.
32. Pereira, S.; Canhoto, P.; Salgado, R. Development and Assessment of Artificial Neural Network Models for Direct Normal Solar Irradiance Forecasting Using Operational Numerical Weather Prediction Data. *Energy and AI* **2024**, *15*, 100314, doi:10.1016/j.egyai.2023.100314.
33. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renewable and Sustainable Energy Reviews* **2020**, *124*, 109792, doi:10.1016/J.RSER.2020.109792.
34. Ohtake, H.; Takamatsu, T.; Oozeki, T. A Review on Photovoltaic Power Forecasting Technics. *IEEJ Transactions on Power and Energy* **2022**, *142*, 533–541, doi:10.1541/ieejpes.142.533.
35. Ramirez-Vergara, J.; Bosman, L.B.; Wollega, E.; Leon-Salas, W.D. Review of Forecasting Methods to Support Photovoltaic Predictive Maintenance. *Clean Eng Technol* **2022**, *8*, 100460, doi:https://doi.org/10.1016/j.clet.2022.100460.

36. Mathiesen, P.; Kleissl, J. Evaluation of Numerical Weather Prediction for Intra-Day Solar Forecasting in the Continental United States. *Solar Energy* **2011**, *85*, 967–977, doi:10.1016/j.solener.2011.02.013.
37. Castillo-Rojas, W.; Bekios-Calfa, J.; Hernández, C. Daily Prediction Model of Photovoltaic Power Generation Using a Hybrid Architecture of Recurrent Neural Networks and Shallow Neural Networks. *International Journal of Photoenergy* **2023**, *2023*, 1–19, doi:10.1155/2023/2592405.
38. Bofinger, S.; Heilscher, G. Solar Electricity Forecast: Approaches and First Results. *Proceedings of the 21st European Photovoltaic Solar Energy Conference and Exhibition* **2006**.
39. Lopes, F.M.; Conceição, R.; Silva, H.G.; Salgado, R.; Collares-Pereira, M. Improved ECMWF Forecasts of Direct Normal Irradiance: A Tool for Better Operational Strategies in Concentrating Solar Power Plants. *Renew Energy* **2021**, *163*, 755–771, doi:10.1016/J.RENENE.2020.08.140.
40. Aguiar, L.M.; Pereira, B.; Lauret, P.; Díaz, F.; David, M. Combining Solar Irradiance Measurements, Satellite-Derived Data and a Numerical Weather Prediction Model to Improve Intra-Day Solar Forecasting. *Renew Energy* **2016**, *97*, 599–610, doi:10.1016/j.renene.2016.06.018.
41. Fonseca, J.G.D.S.; Uno, F.; Ohtake, H.; Oozeki, T.; Ogimoto, K. Enhancements in Day-Ahead Forecasts of Solar Irradiation with Machine Learning: A Novel Analysis with the Japanese Mesoscale Model. *J Appl Meteorol Climatol* **2020**, *59*, 1011–1028, doi:10.1175/JAMC-D-19-0240.1.
42. Lima, F.J.L.L.; Martins, F.R.; Pereira, E.B.; Lorenz, E.; Heinemann, D. Forecast for Surface Solar Irradiance at the Brazilian Northeastern Region Using NWP Model and Artificial Neural Networks. *Renew Energy* **2016**, *87*, 807–818, doi:10.1016/j.renene.2015.11.005.
43. Liu, W.; Liu, Y.; Zhou, X.; Xie, Y.; Han, Y.; Yoo, S.; Sengupta, M. Use of Physics to Improve Solar Forecast: Physics-Informed Persistence Models for Simultaneously Forecasting GHI, DNI, and DHI. *Solar Energy* **2021**, *215*, 252–265, doi:10.1016/j.solener.2020.12.045.
44. Pereira, S.; Abreu, E.F.M.; Iakunin, M.; Cavaco, A.; Salgado, R.; Canhoto, P. 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. *Solar Energy* **2022**, *236*, 225–238, doi:10.1016/j.solener.2022.03.003.

45. Visser, L.; Alskaif, T.; Hu, J.; Louwen, A.; Sark, W. Van On the Value of Expert Knowledge in Estimation and Forecasting of Solar Photovoltaic Power Generation. *Solar Energy* **2023**, *251*, 86–105, doi:10.1016/j.solener.2023.01.019.
46. Voyant, C.; Notton, G.; Duchaud, J.; Gutiérrez, L.A.G.; Bright, J.M.; Yang, D. Benchmarks for Solar Radiation Time Series Forecasting. *Renew Energy* **2022**, *191*, 747–762, doi:10.1016/j.renene.2022.04.065.
47. Voyant, C.; Notton, G.; Kalogirou, S.; Nivet, M.L.; Paoli, C.; Motte, F.; Fouilloy, A. Machine Learning Methods for Solar Radiation Forecasting: A Review. *Renew Energy* **2017**, *105*, 569–582, doi:10.1016/j.renene.2016.12.095.
48. Yang, B.; Zhu, T.; Cao, P.; Guo, Z.; Zeng, C.; Li, D.; Chen, Y.; Ye, H.; Shao, R.; Shu, H.; et al. Classification and Summarization of Solar Irradiance and Power Forecasting Methods: A Thorough Review. *CSEE Journal of Power and Energy Systems* **2021**, 1–19, doi:10.17775/CSEEJPES.2020.04930.
49. Zhou, Y.; Liu, Y.; Wang, D.; Liu, X.; Wang, Y. A Review on Global Solar Radiation Prediction with Machine Learning Models in a Comprehensive Perspective. *Energy Convers Manag* **2021**, *235*, 113960, doi:10.1016/j.enconman.2021.113960.
50. Jung, A. hyun; Lee, D.-H.; Kim, J.-Y.; Kim, C.K.; Kim, H.-G.; Lee, Y.-S. Spatial and Temporal Downscaling of Solar Radiation Using Statistical Techniques. *Journal of the Korean Solar Energy Society* **2020**, *40*, 89–94, doi:10.7836/kses.2020.40.6.089.
51. Yang, D.; Wu, E.; Kleissl, J. Operational Solar Forecasting for the Real-Time Market. *Int J Forecast* **2019**, *35*, 1499–1519, doi:10.1016/j.ijforecast.2019.03.009.
52. Jaszczur, M.; Hassan, Q.; Abdulateef, A.M.; Abdulateef, J. Assessing the Temporal Load Resolution Effect on the Photovoltaic Energy Flows and Self-Consumption. *Renew Energy* **2021**, *169*, 1077–1090, doi:10.1016/J.RENENE.2021.01.076.
53. Yang, C.; Thatte, A.A.; Xie, L. Multitime-Scale Data-Driven Spatio-Temporal Forecast of Photovoltaic Generation. *IEEE Trans Sustain Energy* **2015**, *6*, 104–112, doi:10.1109/TSTE.2014.2359974.
54. MathWorks MATLAB (R2024a) [Computer Software]. Available online: https://www.mathworks.com/products/matlab/student.html (accessed on 21 January 2025).
55. ECMWF IFS Documentation Available online: https://www.ecmwf.int/en/publications/ifs-documentation (accessed on 10 December 2021).

56. Schroedter-Homscheidt, M.; Hoyer-klick, C.; Killius Armines, N.; Lefèvre, M.; Transvalor, L.W. Copernicus Atmosphere Monitoring Service User's Guide to the CAMS Radiation Service. **2017**, 1–74.
57. Hogan, R.J.; Bozzo, A. A Flexible and Efficient Radiation Scheme for the ECMWF Model. *J Adv Model Earth Syst* **2018**, *10*, 1990–2008, doi:10.1029/2018MS001364.
58. ECMWF CAMS: Global Atmospheric Composition Forecast Data Documentation Available online: https://confluence.ecmwf.int/display/CKB/CAMS%3A+Global+atmospheric+composition+forecast+data+documentation (accessed on 10 December 2021).
59. Hay, J.; McKay, D. Final Report IEA Task IX-Calculation of Solar Irradiances for Inclined Surfaces: Verification of Models Which Use Hourly and Daily Data. **1988**.
60. Duffie, J.A.; Beckman, W.A. Solar Engineering of Thermal Processes. **2013**, 1–885.
61. Roberts, J.J.; Mendiburu Zevallos, A.A.; Cassula, A.M. Assessment of Photovoltaic Performance Models for System Simulation. *Renewable and Sustainable Energy Reviews* **2017**, *72*, 1104–1123, doi:10.1016/j.rser.2016.10.022.
62. SERAPHIM; SOLARFORCE ENERGY SII Series 390-405 W Available online: https://solarforce-energy.co.za/wp-content/uploads/2021/03/SRP-390-405-BMA-HV_158.75_EN_2020V2.0.pdf (accessed on 26 August 2024).
63. Huawei SUN2000-36KTL Smart String Inverter Available online: https://solar.huawei.com/en/download?p=%2F-%2Fmedia%2FSolar%2Fattachment%2Fpdf%2Fmea%2Fdatasheet%2FSUN2000-36KTL.pdf (accessed on 26 August 2024).
64. meteocontrol IRRADIANCE SENSORS Available online: 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 on 26 August 2024).
65. Long, C.N.; Dutton, E.G. BSRN Global Network Recommended QC Tests, V2.0 Available online: https://epic.awi.de/id/eprint/30083/1/BSRN_recommended_QC_tests_V2.pdf (accessed on 28 May 2024).
66. Abreu, E.F.M.; Canhoto, P.; Prior, V.; Melicio, R. Solar Resource Assessment through Long-Term Statistical Analysis and Typical Data Generation with Different Time Resolutions Using GHI Measurements. *Renew Energy* **2018**, *127*, 398–411, doi:10.1016/J.RENENE.2018.04.068.

67. Cavaco, A.; Canhoto, P.; Collares Pereira, M. Procedures for Solar Radiation Data Gathering and Processing and Their Application to DNI Assessment in Southern Portugal. *Renew Energy* **2021**, *163*, 2208–2219, doi:10.1016/J.RENENE.2020.10.075.
68. Mayer, M.J.; Yang, D. Pairing Ensemble Numerical Weather Prediction with Ensemble Physical Model Chain for Probabilistic Photovoltaic Power Forecasting. *Renewable and Sustainable Energy Reviews* **2023**, *175*, 113171, doi:10.1016/j.rser.2023.113171.
69. De Soto, W.; Klein, S.A.; Beckman, W.A. Improvement and Validation of a Model for Photovoltaic Array Performance. *Solar Energy* **2006**, *80*, 78–88, doi:10.1016/j.solener.2005.06.010.

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

Chapter 7

Conclusions

The main objective of this work was to develop a comprehensive and integrated model for improving accuracy of energy output forecasts of photovoltaic (PV) modules and power plants. This was achieved by combining various models and using data from Numerical Weather Prediction (NWP) and aerosol models. Through the successful realization of several key sub-objectives, the research carried out resulted in the creation of a robust tool designed to improve the accuracy and reliability of PV power predictions, thus supporting the efficient operation and management of solar energy systems.

The core of the developed method combines numerical weather prediction (NWP) models with artificial neural networks (ANNs) to either assess or predict solar resources and photovoltaic power generation at different spatial and time scales. Firstly, this method was used for regional solar resource assessment by improving solar irradiance data from NWP models that effectively integrate predictions from the Meso-NH model and aerosol forecast data from the Copernicus Atmosphere Monitoring Service (CAMS) with artificial neural networks (ANNs). This method was applied to a typical meteorological year and the region of southern Portugal and demonstrated a substantial reduction in the errors associated with solar radiation simulations for eight stations in this region, with 2.34 % and 4.26 kWh/m^2 for global horizontal irradiance (GHI) and 3.41 % and 11.57 kWh/m^2 for direct normal irradiance (DNI) for the relative mean bias error and root mean squared error, respectively. By incorporating aerosol data, the approach addressed a limitation of current NWP models, which often struggle to accurately simulate atmospheric aerosols. This approach has broader applicability, although its performance in different climatic regions and over larger geographical areas requires further investigation. The potential to apply this method to other meteorological variables, such as air temperature, also opens new research topics, particularly in improving the efficiency of PV systems by accurately predicting the environmental conditions and how they will affect their operation.

In addition to the solar resource assessment, this work presents an ANN-based model specifically designed to improve DNI forecasting. By integrating NWP and aerosol forecast data, from the European Centre for Medium-range Weather Forecasts (ECMWF) and CAMS, respectively, the model achieved improvements in DNI predictions, with reductions in the mean absolute error of 21.1 W/m^2 and in the root mean squared error of 24.2 W/m^2 compared to the original forecasts. This model was rigorously validated in an operational setting for a location in Évora using four years of observations, demonstrating its robustness and effectiveness. Furthermore, the model's adaptability was tested by applying it to other six locations within the same region, where it consistently showed enhanced forecasting accuracy. The flexibility of the model to operate with different temporal resolutions makes it highly relevant for energy market forecasting, where accurate predictions are essential.

Another key focus of this work was the accurate estimation of solar irradiance on tilted surfaces, particularly in photovoltaic arrays with multiple rows of modules. State-of-art transposition models often fail to account for direct and anisotropic shading and irradiance obscuration between rows, leading to significant errors in energy generation predictions, particularly from modules located in inner rows. To address this, an improved transposition model was developed, which specifically considered the complex contributions of direct, circumsolar and isotropic diffuse and reflected irradiance components for the global tilted irradiance (GTI) in the case of inter-rows. This model was thoroughly evaluated using experimental data collected in Évora, Portugal, under varying conditions of tilt angles and inter-row distances. The results demonstrated a significant improvement in the accuracy of GTI estimates for rows other than the first, with mean bias error improvements of 368.3 W/m^2 and root mean square error reductions of 224.4 W/m^2. These findings underscore the importance of using dedicated models for non-front-row surfaces in solar power plants, as neglecting these factors can lead to substantial inaccuracies in energy generation forecasts. The developed model is versatile and can be adapted to different configurations, making it a valuable tool not only for optimizing the design and operation of large-scale PV installations, but also for applications such as backtracking optimization in dual-axis tracking systems, where accurate modeling of mutual shading is critical for minimizing losses.

Additionally, in this work an in-depth evaluation of various thermal models for predicting PV module temperature and power output was conducted. The accurate estimation of module temperature is crucial, as it is related and directly affects the efficiency and power output of PV systems. Among the models assessed, the steady-state Mattei model emerged as the most accurate for temperature prediction, achieving a mean bias error of −0.4°C and root mean squared error of 2.7°C. In terms of power output, the Kurtz model, when coupled with a simple electrical model incorporating temperature correction, showed the best performance, with a mean bias error of 4.6 W and root mean squared error of 54.5 W. A coupled thermal-electric model was also presented using single diode 5 parameter electric model and achieving MBE values for cell's temperature and power output of −0.4 °C and 16.0 W and of RMSE values of 3.6 ºC and 84.2 W. The research highlighted the critical interplay between thermal and electrical models, demonstrating that the selection of an appropriate thermal model is essential for accurate power output predictions, particularly under varying environmental conditions. Additionally, the study showed that using these thermal models can result in more accurate power output estimates than relying solely on temperature measurements, which is particularly beneficial in scenarios where extensive experimental data are not available. This finding is significant for both researchers and power plant operators, as it provides a reliable method for forecasting PV system performance without the need for costly and time-consuming experimental testing.

The integration of the developed models into a single forecasting algorithm represents a major contribution of this work. The algorithm was designed to be highly versatile and scalable, capable of providing PV power forecasts for different systems and locations, including those with no on-site observations. This algorithm also incorporates several key components, including the retrieval and processing of forecast data from the IFS/ECMWF NWP and CAMS models, the developed transposition model for GTI estimation, and a dynamic coupled thermal-electric model for temperature and power output predictions. The algorithm also includes models for losses and inverter efficiency, allowing for the computation of both DC and AC power output. The algorithm was validated using data from a real PV power plant in the region of Lisbon, Portugal, with results showing a mean bias error of 9.3 W/kWp and a root mean square error of 121.0 W/kWp for AC power output forecasts over a 72-hour horizon. These results indicate a high level of accuracy, with relative errors as low as 1.8 % for DC power and 2.3 % for AC power. The algorithm’s scalability and adaptability are particularly noteworthy, as it can be easily extended to cover a wide range of PV installations, offering a cost-effective alternative to expensive on-site monitoring through accurate forecasting. Moreover, the algorithm’s ability to provide continuous forecasting makes it an invaluable tool for grid operators, who can use it to manage the variability of PV energy generation more effectively. Its practical strengths, including efficient data acquisition and a relatively fast execution time, further enhance its value, making it a dependable and practical tool for integrating PV power into the grid.

Several areas of research could further enhance the models and algorithms developed in this work. Expanding the application of these models to different climates and larger geographical areas would help to validate their robustness and identify any limitations in their performance. Refining the thermal models to improve accuracy under varying environmental conditions, particularly for specific PV module technologies, would also be a valuable area of study. Finally, ongoing validation and adaptation of the models with real-time data from a diverse range of PV installations will be essential for ensuring their continued reliability and applicability in different operational settings. By addressing these areas of future work, the models and algorithms developed in this work can contribute to the advancement of renewable energy technologies and support the global transition to more sustainable energy systems.

Nomenclature

DNI	Direct normal irradiance (W/m^2)
GHI	Global horizontal irradiance (W/m^2)
GTI	Global tilted irradiance (W/m^2)

Acronyms

AC	Alternate Current
ANN	Artificial Neural Network
CAMS	Copernicus Atmosphere Monitoring Service
DC	Direct Current
ECMWF	European Centre for Medium-range Weather Forecasts
IFS	Integrated Forecasting System
MBE	Mean Bias Error
NWP	Numerical Weather Prediction
PV	Photovoltaic
RMSE	Root Mean Squared Error

Contactos:
Universidade de Évora
Instituto de Investigação e Formação Avançada - IIFA
Palácio do Vimioso | Largo Marquês de Marialva, Apart. 94
7002-554 Évora | Portugal
Tel: (+351) 266 706 581