REPUPLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITÉ SAAD DAHLAB DE BLIDA

Faculté de Téchnologie

Département des Énergies Renouvelables

THÈSE DE DOCTORAT

Filière : Énergies Renouvelables

Spécialité : Énergies Renouvelables en Électronique

Par

Ferial EL ROBRINI

Le data mining et l'intelligence artificielle au service de la gestion et du diagnostic des centrales photovoltaïques connectées au réseau électrique en Algérie

Soutenue publiquement, le 20/11/2025, devant le jury composé de :

Nom et Prénom	Grade	Établissement	Qualité
D. SEMMAR	Professeur	USDB	Président
H. MAZOUZ	Maitre de Conférence "A"	USDB	Examinatrice
A. HADJ ARAB	Directeur de Recherche	CDER	Examinateur
N. BELHAOUAS	Maitre de Recherche "A"	CDER	Examinateur
B. AMROUCHE	Professeur	USDB	Directrice de thèse

DEMOCRATIC AND POPULAR REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

SAAD DAHLAB UNIVERSITY OF BLIDA

Faculty of Technology

Renewable Energies Department

DOCTORAL DISSERTATION

In

Renewable Energies in Electronic Engineering

Authored By

Ferial EL ROBRINI

Data Mining and Artificial Intelligence for the management and Diagnosis of Grid-Connected Photovoltaic Power Plants in Algeria

Defended publicly, on November 20, 2025, before the jury composed of

Name	Rank	Establishement	Members
D. SEMMAR	Full Professor	USDB	President
H. MAZOUZ	Senior Lecturer	USDB	Examiner
A. HADJ ARAB	Research Director	CDER	Examiner
N. BELHAOUAS	Senior researcher	CDER	Examiner
B. AMROUCHE	Full Professor	USDB	Supervisor

To my beloved parents

Acknowledgement

<div dir="rtl">

ملخص

تشكل تقلبات الطاقة الشمسية تحدياً رئيسياً لتوسيع استخدام الأنظمة الكهروضوئية بسبب تأثيرها على استقرار الشبكة الكهربائية. ولمواجهة ذلك، تركز هذه الدراسة على التنبؤ قصير المدى بالطاقة الكهروضوئية في منطقة الجلفة الواقعة بالهضاب العليا بالجزائر، اعتمادا على بيانات حقيقية و تطوير نماذج التعلم العميق لتحسين دقة التوقعات، مع مراعاة التغيرات الشهرية. وتساهم النتائج في تحسين التنبأ بتقلبات الإنتاج الكهروضوئي ، مما يساعد على حسن تزويد الشبكة الكهربائية بالطاقة الشمسية الكهروضوئية ، ما يسهل عملية اتخاذ القرار لمشغلي الشبكات وصانعي السياسات في قطاع الطاقة.

الكلمات المفتاحية: الطاقة الشمسية الكهروضوئية، تنقيب البيانات، الذكاء الاصطناعي، محطات الطاقة الكهروضوئية المتصلة بالشبكة.

</div>

Abstract

The variability of solar energy presents a major challenge for the expansion of photovoltaic systems, as it affects grid stability and overall energy management. To address this issue, advanced forecasting techniques are required to accurately predict production fluctuations. This study focuses on short-term photovoltaic energy forecasting in Algeria's High Plains, using real-world data from the Djelfa power plant. Deep learning models are developed to improve forecasting accuracy while capturing monthly variations. The findings contribute to optimizing photovoltaic energy integration and supporting decision-making for grid operators and energy policymakers.

Keywords: Solar photovoltaic energy, data mining, artificial intelligence, grid-connected PV power plants.

Résumé

La variabilité de l'énergie solaire constitue un défi majeur pour le développement du photovoltaïque, car elle influence la stabilité du réseau et la gestion énergétique. Pour y répondre, des techniques de prévision avancées sont indispensables afin d'anticiper les fluctuations de production. Cette étude porte sur la prévision à court terme de l'énergie photovoltaïque dans les Hautes Plaines d'Algérie, à partir de données réelles issues de la centrale photovoltaïque de Djelfa. Des modèles d'apprentissage profond sont développés pour améliorer la précision des prévisions et intégrer les variations mensuelles. Les résultats obtenus contribuent à une meilleure intégration du photovoltaïque et à l'appui des décisions des opérateurs de réseau.

Mots Clée : énergie solaire photovoltaique,data mining, intelligence artificielle,centrales PV connectées au réseau.

Contents

List of Figures

List of Tables

Nomenclature

Autres Symboles

P_{atm}	Atmospheric pressure (HPa)
PV	Produced photovoltaic power (MW)
R_F	Rainfull amount (mm)
R_H	Relative humidity (%)
$G_{diffused}$	Global solar radiation (W/m^2)
G_{direct}	Direct solar radiation (W/m^2)
G_{global}	Global solar radiation (W/m^2)
G_{tilted}	Tilted solar radiation (W/m^2)
SS_D	Sunshine duration (hours)
T_{amb}	Ambient Temperature (°C)
T_{comp}	Component Temperature (C)
W_d	Wind direction (degrees°)
W_d	Wind speed (m/s)

Chapter 1

GENERAL INTRODUCTION AND RESEARCH CONTEXT

1.1   The global energy transition

Since ancient times, the unwavering drive for self-improvement and societal advancement has been a fundamental aspect of human ambition, and over the decades, human ingenuity has manifested itself in impressive ways to achieve this, ranging from the construction of simple huts on mountainsides and valleys to towering skyscrapers and venturing beyond the Earth's atmosphere. In fact, these accomplishments would not have reached their peak had the energy sector not also undergone transitional phases. Shifting from traditional sources relying on manual labor, the energy sector has embraced more efficient and less physically demanding alternatives, such as petroleum, gas, and coal. This evolution in energy sources reflects not only advancements in technology but also the adaptability and innovation inherent in human endeavors to meet the growing demands of a rapidly changing world.

The growing energy demand, in combination with socio-economic pressures and the severe ecological consequences of conventional energy sources [1], has created an urgent imperative to adopt alternative pathways centered on sustainability. According to the International Renewable Energy Agency, *« a profound and systemic transformation of the global energy system must occur within the next 30 years if the world is to avoid devastating consequences from climate change and a steady erosion of energy security »* [2]. Achieving this transformation requires more than simply decarbonizing energy supply; it calls for the establishment of an energy framework that simultaneously reduces carbon emissions, enhances resilience, and fosters inclusivity within the global economy.

Renewable energies are pioneers in the global energy transition, driving the shift toward a more sustainable and resilient energy system. Among these sources, solar photovoltaic (PV) energy has emerged as a pivotal solution, offering a sustainable and environmentally friendly alternative to fossil fuels. As an abundant and inexhaustible resource, solar energy

addresses the growing global energy demand while significantly reducing greenhouse gas emissions. Continuous advancements in photovoltaic technology, including improvements in efficiency, energy storage, and grid integration, have enhanced its feasibility and economic competitiveness. Additionally, the declining costs of solar PV systems, combined with policy incentives and international commitments to carbon neutrality, have accelerated their widespread adoption across residential, commercial, and industrial sectors. By enabling decentralized energy production and reducing dependence on centralized power grids, solar photovoltaics improves energy security and resilience, particularly in regions with high solar potential. As the global energy landscape undergoes a profound transformation, the integration of solar PV systems plays a crucial role in achieving a sustainable, low-carbon future.

1.2 Algeria's commitment to the global energy transition

Under the vast skies of Algeria, solar energy holds immense promise. With over 3,500 hours of sunshine annually in its desert regions, the country boasts a solar potential of 1,850 to 2,100 kWh per square meter [3]. Recognizing this, Algeria has been steadily expanding its renewable energy sector, particularly in solar PV. By 2021, the nation had installed 438.2 MW of solar capacity, with 24 grid-connected PV plants contributing 366.1 MW to the national grid [4–6].

To accelerate the deployment of solar energy, the Algerian Ministry of Energy Transition and Renewable Energies launched a 1,000 MW grid-connected PV project in 2021, which entered commissioning in 2024. This large-scale initiative was awarded to several companies, including CSCEC China (300 MW), Cosider Canalisation-Fimer Spa (250 MW), Amimer Energie SPA (150 MW), and Ozgun & Bouzida (300 MW) [7]. In parallel, Sonelgaz-EnR initiated another major solar expansion in 2023 with a 2,000 MW tender for the construction of 15 PV plants [8].

These large-scale projects mark a significant step in Algeria's renewable energy roadmap, reinforcing its commitment to reducing reliance on fossil fuels and enhancing energy security. The integration of these PV plants will contribute to stabilizing the national grid, reducing carbon emissions, and supporting economic growth by fostering a more sustainable energy sector. The locations of both operational and planned PV plants are illustrated in Figure. 1.1.
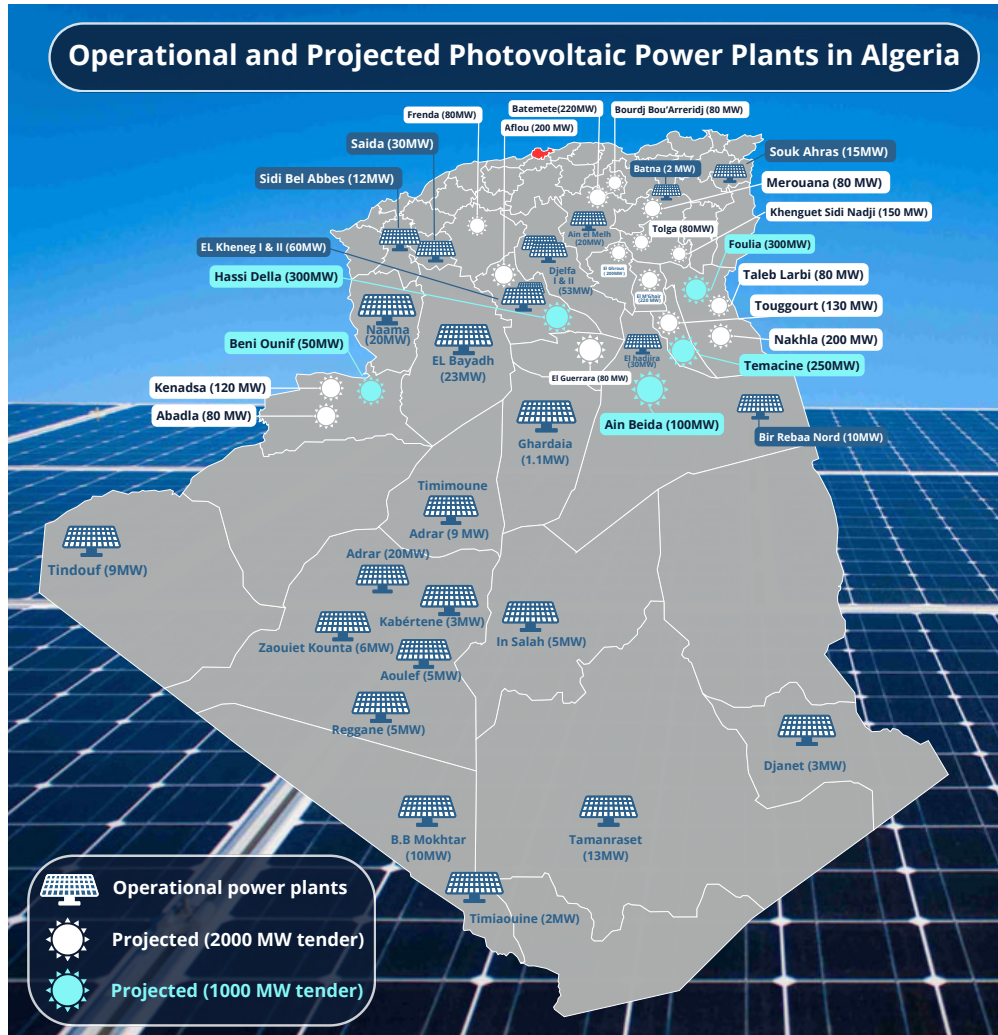
Figure 1.1: Operational and Projected PV power plant in Algeria [5, 7, 8]

1.3 Critical challenges and strategic considerations

Unlike conventional power plants, which can be dispatched to match electricity demand, solar energy generation is inherently variable, depending on weather conditions and natural resource availability. This intermittency poses significant challenges to grid stability, requiring sophisticated mechanisms to balance supply and demand. These challenges are even more pronounced in isolated and islanded energy systems, where the absence of interconnections exacerbates the complexities of integrating variable renewable energy sources, as well as in grids with high renewable energy penetration. Key technical concerns include voltage regulation, frequency stability, harmonic distortions, and protection coordination, all of which can significantly impact overall grid performance and reliability. As the share of renewable energy in the global energy mix continues to grow, these challenges become increasingly critical, necessitating a deeper understanding of their implications on modern power systems [9].

1.3.1 Voltage flickers

A major technical concern arising from the widespread integration of PV systems is voltage instability, with voltage flicker representing a prominent issue. As outlined in IEEE Standard 1453-2022 [10], flicker is defined as *«the subjective impression of fluctuating lighting caused by fluctuations in the supply voltage»*. This phenomenon is primarily triggered by abrupt voltage fluctuations, typically caused by the frequent switching of high-power electrical equipment, including arc furnaces and electric motors. While voltage flicker is often associated with perceptible fluctuations in lighting, its impact extends beyond illumination systems. Sensitive electronic devices, including medical equipment, can be adversely affected by voltage deviations, potentially leading to malfunction or damage. Furthermore, voltage imbalances introduced by PV systems can impair the efficiency of voltage regulation mechanisms, potentially triggering the unintended disconnection of PV installations due to conflicts with the grid's voltage control schemes [11, 12].

1.3.2 Reverse power flow

The large-scale incorporation of photovoltaic systems into distribution networks also gives rise to another significant concern, namely reverse power flow. This phenomenon arises when locally generated solar power surpasses immediate demand, especially at midday when solar irradiance is at its peak and residential consumption is minimal. As a result, any excess generation is supplied back into the grid(Figure. 1.2 (a,b)), leading to a negative net load, as illustrated in Figure. 1.2 (c). Such reverse power flow can pose significant operational challenges, including increased voltage levels, transformer overloading, and potential disruptions to conventional grid protection schemes [11, 13].
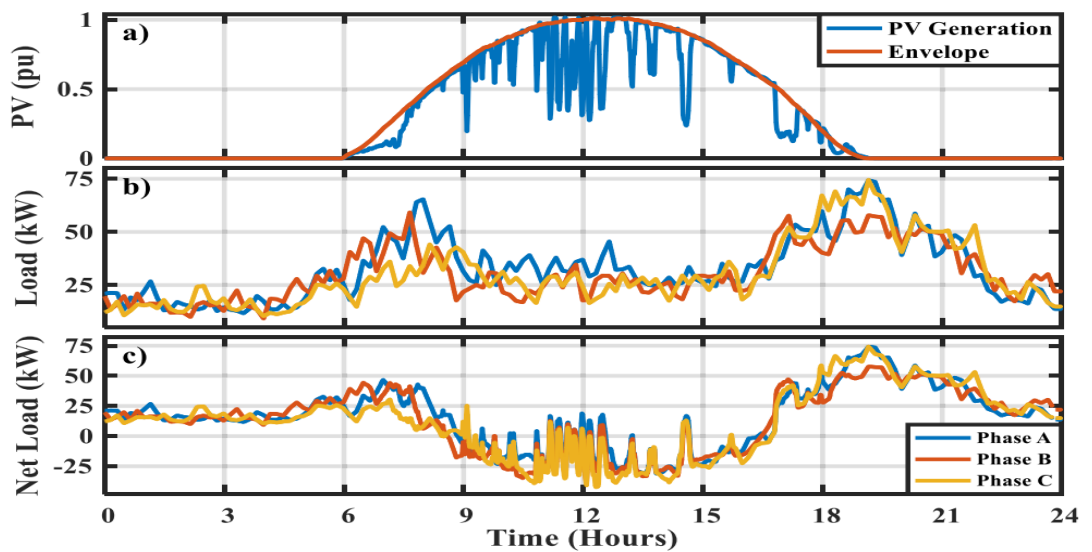
Figure 1.2: (a) Normalized PV power profile and its envelope curve; (b) Load demand; (c) netload with PV power integration in one day [11]

1.3.3 Frequency perturbation

In power networks with substantial penetration of renewable resources, fluctuations in frequency represent a critical operational issue. The concept of frequency stability is defined as the capacity of the system to preserve or re-establish the balance between electricity supply and demand after experiencing a disturbance (see Figure. 1.3) [14]. To ensure reliable operation, the National Grid Electricity System Operator Limited has established permissible frequency limits ranging from 49.5 Hz to 50.5 Hz [15]. Deviations beyond these thresholds can lead to severe consequences, including degraded system performance, malfunctions in sensitive equipment, and, in extreme cases, automatic disconnections or large-scale blackouts.
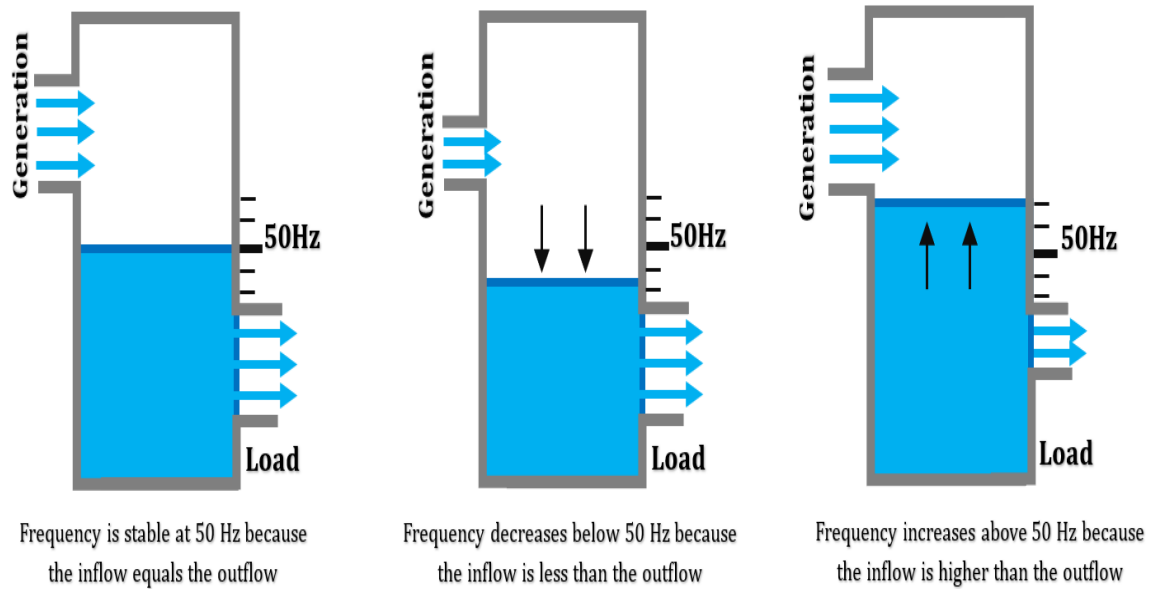
Figure 1.3: Visualization of grid frequency behavior through a container analogy [14].

In conventional power systems, grid frequency is directly governed by the rotational speed of synchronous generators [14, 16, 17]. These generators regulate frequency by adjusting their speed in response to load-generation imbalances, either accelerating or decelerating to restore equilibrium. The kinetic energy stored in their rotating masses provides essential inertial support, mitigating the initial rate of frequency deviations. However, PV systems lack rotating components, eliminating this inherent inertia. This absence, coupled with solar resource variability, results in a higher rate of change of frequency (ROCOF) [14], increased power generation losses [18], and greater complexity in frequency regulation [19]. Consequently, the large-scale integration of PV plants (>10 MW) into power transmission networks necessitates the implementation of additional frequency support mechanisms to maintain grid stability, as discussed in [20–22].

1.3.4 Flexibility factor

Power produced from variable renewable sources is generally configured to be fully delivered to the grid. To adapt the prioritized dispatch of renewable electricity, conventional power plants must accordingly adjust their production by decreasing or increasing output as needed. However, the feasibility of such adjustments is constrained by the technical limitations associated with the permissible rate of change in conventional generation cycles, which are characterized by relatively slow response times. In certain cases, these constraints may impose strict dynamic limitations or even render some traditional power plants incapable of adapting to rapid fluctuations in renewable generation. Consequently, the flexibility of conventional power plants measured by factors such as energy storage capacity, ramping speed, and fast-start capability must be considered as a critical criterion when evaluating the integration of renewable resources into a given electrical grid [9, 23, 24].

In the realm of these technical challenges, energy transition planning, complemented by policy and regulatory enablers, should be grounded in precise technological constraints that align with the requirements of the evolving energy landscape. The shift towards a diversified and interconnected energy system necessitates the modernization and expansion of infrastructure particularly transmission and distribution networks to accommodate the decentralized and intermittent nature of numerous renewable sources. This upgrade is essential to meeting increased capacity demands and facilitating bidirectional power flows. Additionally, the integration of smart grid technologies and advanced analytics can optimize grid operations and enhance overall system reliability [9, 25–27].

Within modern electrical grids, forecasting is a fundamental component of network control, particularly in countries with significant wind energy production [28, 29]. Similarly, PV power forecasting has gained increasing importance, especially in regions where regulatory frameworks actively support the deployment of solar energy. Detailed and accurate forecasts of both energy production and consumption are essential to effectively implement the transition strategies, as they serve as a foundation for grid management and operational planning. Reliable forecasting of intermittent renewable energy generation is particularly crucial for ensuring the efficient integration of large-scale renewable resources into the energy mix. By enhancing predictive accuracy, forecasting methodologies contribute to reducing uncertainties, improving resource allocation, and minimizing grid instability risks [30]. Consequently, substantial research efforts have been devoted to enhancing solar radiation forecasting, which directly improves the accuracy of corresponding PV power predictions.

1.3.5 PV System Faults

The rapid and widespread deployment of PV systems has simultaneously exposed a diverse range of operational faults, which have become the subject of growing academic investigation [31]. Among the earliest and most critical issues identified were inverter failures, frequently

reported throughout the 1990s [32]. Over time, a wider array of electrical anomalies, including line-to-line faults, line-to-ground faults, and arc faults—has been increasingly recognized as a persistent source of disruption within PV operations [33, 34].

In practice, commonly occurring faults such as open- and short-circuited modules, cracked cells, and inverter breakdowns [35], alongside various performance-limiting conditions like partial shading, soiling [36], clipping-induced curtailment, and module degradation [37], including potential-induced degradation (PID) [36], have been shown to compromise energy yield, increase maintenance needs, and raise safety concerns. These factors contribute significantly to the intermittency of PV output, undermining system reliability and grid stability.

1.4   Research objectives and methodological approach

Building upon the essential role of PV power forecasting in modern energy systems, this research focuses on short-term forecasting of PV generation in one of Algeria's High Plains regions. By employing advanced machine learning and deep learning techniques, the study aims to enhance forecasting precision, optimize energy management strategies, and support the large-scale integration of solar energy. Accurate predictive models are crucial for improving grid stability, reducing dependence on conventional backup generation, and fostering a more resilient and sustainable electricity network. To achieve these objectives, this research will:

1. Advance predictive modeling by developing and optimizing machine learning and deep learning frameworks to improve the accuracy of short-term PV power forecasts. This involves fine-tuning model architectures, selecting relevant input features, and employing rigorous validation techniques to ensure robust performance.

2. Generate comprehensive forecasting insights by incorporating meteorological and historical PV production data while addressing seasonal variability. Capturing the impact of seasonal fluctuations will enable more precise predictions, facilitating informed decision-making in energy planning, load scheduling, and resource management.

3. Enhance decision-making in energy management by providing accurate forecasting outputs that can aid in optimizing grid operations, scheduling power generation, and balancing supply and demand. The improved predictive capabilities will help mitigate uncertainties, leading to more efficient utilization of renewable energy resources.

Drawing upon an extensive review and analysis of existing research, along with the utilization of real-world large-scale photovoltaic power data from Algeria and extensive experiments, this study provides advanced forecasting insights to support grid operators and policymakers. By improving the accuracy of short-term PV power predictions, it aims to

enhance decision-making in energy management and contribute to the effective integration of renewable energy into the power grid. Ultimately, this research serves as a foundation for developing more adaptive and resilient energy systems.

1.5   Manuscript organisation

This manuscript is organized into five chapters, each addressing a fundamental aspect related to PV power plants. **Chapter 2** presents a comprehensive review of PV power forecasting methodologies, including physical, statistical, and artificial intelligence-based approaches. **Chapter 3** provides a detailed description of the photovoltaic power plant under investigation, including its geographical location, technical specifications, and the data acquisition process. **Chapter 4** delineates the methodological framework employed in this study, covering data preprocessing techniques, model selection, training strategies, and evaluation metrics. Results and discussion are presented in the last section, offering a comparative analysis of different forecasting models based on key performance indicators. The findings are contextualized within the broader domain of renewable energy forecasting, highlighting their implications for improving solar energy integration into power grids.

Chapter 2

PV POWER FORECASTING: A COMPREHENSIVE REVIEW

2.1 Introduction

The prediction of PV generation is based on forecasting techniques that analyze historical data to estimate the future energy output of a specific PV station. Predictions can be made for various time horizons, such as one hour ahead, one day ahead, or one week ahead, depending on the operational and strategic requirements of grid operators, energy providers, and policymakers. This process relies on a thorough exploration of historical data to identify trends and patterns, establish correlations between key influencing factors, and extrapolate this information to produce reliable forecasts.

Based on a comprehensive review of forecasting methodologies, this chapter examines the critical aspects of PV generation prediction, including temporal and spatial resolution, key forecasting techniques and their utilization across different time horizons, and the primary factors affecting prediction accuracy. Moreover, the discussion addresses the economic aspects of PV forecasting, including its role in optimizing energy markets, minimizing operational costs, and enhancing grid stability. In addition, it provides an overview of key pioneering industries and leading companies in energy forecasting, emphasizing their innovations and contributions to advancing predictive technologies in the renewable energy sector.

2.2 Terminology and key concepts

Before engaging in more advanced analyses, it is crucial to establish a clear understanding of the fundamental aspects related to forecasting. Key elements such as the context in which forecasting is applied, the accuracy measures used to assess precision, and the management of data inputs must be well understood. Furthermore, the parameters that define the forecasting

process, such as the temporal and spatial horizons, are of utmost importance. Clarifying these fundamental aspects early on allows for a more focused and systematic exploration of the forecasting models and contributes to the robustness of the analysis.

2.2.1 Temporal resolution

Temporal resolution denotes the frequency or spacing of time steps at which forecasting is performed. As highlighted in prior studies, photovoltaic output can be predicted over multiple horizons, extending from ultra-short-term intervals to short-, medium-, and long-term ranges. The choice of temporal resolution is determined by the intended application, whether for enhancing energy management efficiency or supporting strategic decisions related to grid integration [38–40].

2.2.1.1 Very-short term horizon

Ultra-short-term forecasting focuses on predicting PV output variations occurring over timescales of seconds to minutes. Such forecasts are particularly important for real-time operational decisions, including load regulation and the management of storage systems, enabling rapid responses to abrupt shifts in demand or weather conditions.

2.2.1.2 Short-term horizon

Spanning horizons from a single hour up to several hours in advance, these forecasts incorporate information on anticipated meteorological conditions, including cloud dynamics and solar irradiance. Such predictions are instrumental for grid operators, as they enhance the ability to match generation with demand and improve the efficiency of solar energy integration within existing power systems.

2.2.1.3 Medium term horizon

Extending over periods of several weeks to months, these forecasts rely on seasonal trends, past meteorological records, and climate projections to estimate solar power output. They play a crucial role in long-term energy planning by informing infrastructure investments, guiding policy development, and supporting capacity expansion strategies for solar energy systems.

2.2.1.4 Long term horizon

Long-range forecasting addresses horizons from several years to multiple decades. Such projections integrate considerations of climate evolution, technological progress, and broader

systemic factors affecting solar energy production. They provide essential insights for strategic decision-making, including energy transition pathways, renewable deployment targets, and long-term investment planning in solar infrastructure.

2.2.2   Spacial resolution

The concept of spatial resolution refers to the geographical scale at which photovoltaic (PV) forecasts are generated. While much of the existing literature emphasizes point-based or localized predictions, the spatial applicability of these forecasts must also be considered. For instance, a model may yield accurate results for a single location within a PV plant but fail to generalize across the entire installation. As an illustration, a forecasting area of 10 km² may be subdivided into cells of 1 km², defining the spatial resolution of the analysis. Accordingly, PV predictions may range from regional-scale evaluations down to localized assessments at the level of a city or even an individual solar system [41].

At the finer end of the spectrum, localized PV forecasts are carried out at the city or system level, where high-resolution data are employed. These predictions incorporate detailed site-specific factors such as panel orientation, shading effects, system architecture, and local environmental conditions. Such granularity supports system operators and designers in enhancing PV performance, improving energy yield, and increasing overall efficiency. The appropriate spatial resolution is therefore contingent on the study's objectives and the extent of data availability.

Conversely, regional-scale PV predictions provide broader insights into the expected electricity generation over large geographic areas. These assessments account for solar resource availability, climatic variations, and regional characteristics, serving as a strategic tool for energy planners. By identifying zones with high PV potential, such projections contribute to informed policy development and the design of effective energy transition strategies.

2.3   Forecasting techniques

The forecasting technique is a crucial factor affecting precision, as the choice of method directly influences the reliability of predictions. these techniques are presented in a dedicated section for a more structured discussion. The primary goal of developing advanced models is to enhance accuracy while reducing computational costs [42]. Over the years, researchers have explored various methodologies, each offering distinct advantages and limitations depending on the application and available data. These techniques are broadly categorized into persistence methods, physical models, statistical approaches, and hybrid frameworks, each playing a significant role in improving forecast performance under different conditions.

2.3.1 Persistence model

The Persistence Model (PM), often referred to as the naïve predictor, represents the most elementary approach to forecasting. Its underlying assumption is that the target variable does not vary with time; hence, the forecasted value $Y_{(t)}$ is directly taken as the previous observation $Y_{(t-1)}$ [43, 44]. Due to its straightforward nature, it serves as a benchmark for assessing the performance of more advanced forecasting techniques [45]. The persistence model is particularly suitable for short-term forecasting, especially for one-hour-ahead predictions of solar and PV power generation [46]. However, its precision decreases significantly as the forecasting horizon extends, primarily due to the inherent variability of weather conditions [43]. Consequently, while PM provides reasonable accuracy for immediate predictions, it becomes inadequate for medium- and long-term forecasting, necessitating the adoption of more sophisticated predictive models.

2.3.2 Physical approaches

A second category of forecasting models comprises physical models, which incorporate meteorological and environmental factors and are particularly effective at intra-day and day-ahead scales [47,48]. Among these, the Numerical Weather Predictor (NWP) model is widely utilized. This model employs a mathematical framework to describe the physical state and dynamic motion of the atmosphere [49], accounting for parameters such as relative humidity, ambient temperature, and wind speed [50]. Another widely adopted approach within physical models involves the integration of satellite imagery and ground-based sky cameras, which take into account the effects of solar position on meteorological factors [51], and neighboring conditions [52]. Such models are capable of delivering detailed insights into atmospheric dynamics, encompassing wind velocity and orientation within the lower, middle, and upper atmospheric layers. Additionally, they generate a variety of cloud analysis images, capturing details such as cloud cover, shape, and thickness [53].

While physical models generally achieve high accuracy under stable weather conditions, their performance can be significantly impacted by sudden fluctuations in meteorological variables, leading to a decline in forecast reliability [54]

2.3.3 Statistical approaches

While physical models capture atmospheric dynamics through meteorological and environmental parameters, statistical approaches offer an alternative perspective by relying on historical data and data-driven techniques to enhance forecast accuracy. Statistical forecasting methods are commonly divided into two categories. The first encompasses linear models, including the Auto-Regressive Moving Average (ARMA), the Auto-Regressive Integrated Moving Average (ARIMA) [55], the Seasonal Autoregressive Moving Average

(SARIMA) [56, 57], and the Coupled Auto-Regressive and Dynamical System (CARDS) model [58]. These approaches rely on the principle of regressing the predicted value against its own historical observations. Formally, the forecasted variable $Y_{(t)}$ is expressed as a function of its lagged values $Y_{(t-i)}$, their associated coefficients $\beta_{(i)}$, and a stochastic error component $e_{(t)}$.

The second category employs artificial intelligence (AI) and deep learning (DL) techniques. Numerous studies highlight the superior predictive capacity of AI-based methods, with neural networks being a widely applied example [59–61]. In particular, the multi-layer perceptron, representing a deeper and more sophisticated extension of conventional neural networks, has shown strong forecasting performance, as demonstrated by Khelifi et al. [62]. Beyond neural models, other AI-based techniques used for forecasting include support vector machines [63–65], genetic algorithms [66], random forests [67], boosted decision trees [68], as well as clustering-based methods such as K-means [69] and c-means [70].

2.3.4 Hybrid approaches

Earlier research predominantly assessed the effectiveness of individual methods; however, recent studies have increasingly focused on enhancing the efficiency and overall performance of predictive models. Consequently, it is now uncommon to encounter studies that rely solely on a single technique. In contemporary research, the limitations of solitary methods have prompted the integration of multiple approaches whether by pairing different models [71–73], or incorporating optimization theorems to develop hybrid methods that capitalize on the strengths of various forecasting techniques, thereby enhancing PV forecasting accuracy [68, 74] . These hybrid approaches benefit from the complementary features of each constituent method, enabling more robust handling of diverse data characteristics and dynamic environmental conditions. As the complexity of PV systems and their operational contexts increases, the synergistic application of multiple techniques offers a promising avenue for achieving greater predictive precision and operational resilience.

2.3.5 Summary of recent appoaches

Table 2.1 synthesizes a range of contributions addressing PV power prediction, employing distinct methodological frameworks and datasets from various geographic regions. A key observation across the literature is the increasing emphasis on hybrid deep learning paradigms. For instance, the work in [75] introduced a forecasting architecture that integrates frequency-domain decomposition with advanced DL models. The procedure begins with ensemble empirical mode decomposition, which partitions the raw PV time series into high- and low-frequency components, followed by a statistical feature extraction step. Prediction is then performed using an optimized long short-term memory (LSTM) network, whose hyperparameters are tuned through Bayesian optimization. To further refine forecast stabil-

ity, support vector regression (SVR) is applied at the initial time node, effectively reducing short-term fluctuations. Validation on multi-scale datasets from two large PV plants (100 MW and 110 MW) located in Arizona, USA, demonstrated robust performance, yielding average RMSE and AbsDEV values of 4.157 and 0.116, respectively.

The study in [26] proposed a different hybridization strategy, combining decomposition and machine learning. By applying the Time-Varying Filter Empirical Mode Decomposition (TVF-EMD), historical PV data were separated into multiple intrinsic mode functions (IMFs) spanning low to high frequencies. These IMFs were subsequently modeled using an Extreme Learning Machine (ELM), producing the TVF-EMD-ELM framework. Results confirmed that this model could deliver highly accurate intra-hour forecasts across diverse Algerian locations.

Similarly, [76] examined the predictive capabilities of LSTM and its variant enhanced with a Grid Search Algorithm (GSA). The baseline LSTM was first assessed through sensitivity analysis under three scenarios to identify suitable hyperparameter configurations for different time horizons. The integration of GSA enabled automated hyperparameter tuning, thereby improving forecasting precision and reducing errors. Moreover, correlation analyses using Spearman (SCC) and Pearson (PCC) coefficients were employed to evaluate input–output dependencies and to establish the most effective lag structures, which strongly influenced predictive accuracy. Comparative results demonstrated that the GSA-LSTM achieved up to 28

A complementary perspective was presented in [77], where a hybrid model that combines machine learning techniques with the Theta statistical method (T-MLSHM) was evaluated. Using datasets from Shagaya (Kuwait) and Cocoa (USA), the authors demonstrated that the integration of statistical and machine learning approaches yields superior performance compared to hybrids restricted to machine learning alone.

Other regional case studies highlight the breadth of PV forecasting research. In Suwon, South Korea, an LSTM-based framework produced reliable five-day-ahead forecasts [78]. In China, a wavelet-based neural network enhanced with a genetic algorithm (GA-WNN) exhibited strong accuracy under clear conditions but diminished robustness during cloudy or rainy intervals [79]. Also in China, researchers combined a Deep Convolutional Generative Adversarial Network (DCGAN) with Quantile Regression LSTM (QRLSTM) for one-day-ahead forecasting, attaining high accuracy albeit with increased computational demand [80]. In Saudi Arabia, the Multilayer Feedforward Neural Network optimized via Multiverse Optimization (MFFNN-MVO) was successfully applied to a 4-kW PV system in Shaqra City, achieving strong predictive outcomes [81]. Finally, in Alice Springs, Australia, the Convolutional Gated Recurrent Unit (Conv-GRU) model demonstrated excellent one-day-ahead forecasting capabilities [82].

Tableau 2.1: Overview of recent methods, their characteristics, and key contributions

Ref.	Specification			Main Contribution
	Location	PV Capacity	Horizon	
[83]	China	211.37 kW, Data (7 years)	1 h	Incorporates a hybrid Time2Vec - WDCNN-BiLSTM.
[84]	China	Data (1 year)	15 min	Addresses gradient vanishing issues using hybrid methods.
[85]	USA	Data (6 years)	24 h	Applies a TCM and PDPP to adjust the output of LSTM-RNN.
[86]	Malaysia, Turkey	Data (2 years)	15 min	Proposes two new hybrid models resulting in a higher convergence rate and lower stochastic error.
[66]	Algeria	20 MW, Data (2 years)	15 to 60 min	Implements a NARX model and utilizes a genetic algorithm for gradient-free training.
[87]	Algeria	PVc: 160 W	3 s	Explores the utilization of a FFNN with RPL and PL.
[26]	Algeria	135.1 MW	30 min	Analyzes the implementation of a TVF-EMD with ELM.
[88]	Algeria, Australia	73 MW, Data (2 years)	15 to 60 min	Conceives an ELM model with Recursive Intrinsic Functions decomposition for feature extraction.
[89]	Algeria	6 MW, Data (1 year)	–	Examines the correlation between meteorological variables and photovoltaic output.
[90]	Algeria	9 MW	15 to 60 min	Underscores the VAE's ability to learn high-level features that enhance forecasting accuracy.
[91]	Algeria	9 MWp, Data (1 year)	15 min	Investigates LSTM's effectiveness in handling data dependencies.
[92]	Turkey	Data (1 year)	24 h	Proposes an improvised dynamic group-based cooperative search mechanism with RBNN.
[93]	Morocco	500 kWp	Short term	Uses PCA to reduce PV feature dimensions.
[71]	Australia	8910 Wp	1 to 6 h	Provides a detailed analysis of deterministic and probabilistic forecasting models.

2.3.6   Accuracy assessement

The accuracy of a PV forecast refers to the degree of closeness between the predicted values and the actual observed values. A higher accuracy indicates that the forecasting model can reliably estimate future PV generation with minimal deviation from real measurements. Forecast accuracy is influenced by various factors, including the quality of input data, the complexity of the forecasting method, and the consideration of meteorological and environmental variables.

To assess the accuracy of a forecasting model, several performance parameters provide quantitative measures that help evaluate its efficacy. These metrics not only quantify the errors between predicted and actual values but also enable comparisons between different forecasting approaches, aiding in the selection of the most effective model. In the field of PV forecasting, the correlation coefficient, root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) are widely employed to assess model performance.

2.3.6.1 The Correlation Coefficient (r)

Correlation describes the statistical association between two variables, reflecting both the degree and direction of their interdependence. This relationship is generally quantified through a correlation coefficient, which assumes values in the interval [1,1] [94]. A coefficient of 1 signifies a perfect inverse relationship, meaning that as one variable increases, the other decreases correspondingly. Conversely, a coefficient of +1 denotes a perfect direct relationship, where both variables rise or fall simultaneously. A value of 0 implies the absence of any linear association. Multiple measures exist for quantifying correlation, the most widely used being the Pearson product-moment correlation, the Spearman rank correlation, and Kendall's Tau coefficient [42]:

$$r = \frac{\sum_{i=1}^{n}(\text{Pr}_i - \bar{\text{Pr}})(\text{Ac}_i - \bar{\text{Ac}})}{\sqrt{\sum_{i=1}^{n}(\text{Pr}_i - \bar{\text{Pr}})^2}\sqrt{\sum_{i=1}^{n}(\text{Ac}_i - \bar{\text{Ac}})^2}}$$

where:

$$\text{Pr}_i : \text{Predicted value at time } i$$
$$\text{Ac}_i : \text{Actual value at time } i$$
$$\bar{\text{Pr}} : \text{Mean of predicted values}$$
$$\bar{\text{Ac}} : \text{Mean of actual values}$$
$$n : \text{Number of data points}$$

2.3.6.2 The Determination Coefficient (R^2)

The coefficient of determination (R^2) is another key metric employed to evaluate the performance of photovoltaic forecasting models. It measures the degree to which the predicted outputs correspond to the observed values by evaluating their closeness to the reference 1:1 line. Larger values of R^2 reflect superior predictive performance, as they imply a stronger agreement between estimated and actual data points [95].

$$R^2 = 1 - \frac{\sum\limits_{i=1}^{n} (\text{Ac}_i - \text{Pr}_i)^2}{\sum\limits_{i=1}^{n} (\text{Ac}_i - \bar{\text{Ac}})^2} \tag{2.1}$$

2.3.6.3 The Root Mean Square Error (RMSE)

The RMSE is a statistical metric that quantifies prediction accuracy by computing the square root of the mean of the squared deviations between estimated and observed values. Its formulation inherently emphasizes larger errors, since squaring magnifies the contribution of extreme deviations. This property makes RMSE effective in highlighting substantial discrepancies; however, it may also introduce bias in datasets containing outliers, as such values can dominate the overall error measure [70]. Another version of RMSE, called normalized RMSE (nRMSE), is widely used to provide a standardized evaluation of model performance. By normalizing the RMSE, it enables a fair comparison of prediction errors across datasets with different scales or units, making it particularly useful for comparing models applied to diverse datasets. The normalization process typically involves dividing the RMSE by the range or mean of the actual values, ensuring that the metric is dimensionless and easier to interpret [96].

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} \left(\text{predicted}_i - \text{actual}_i\right)^2} \tag{2.2}$$

$$\text{nRMSE} = \frac{\text{RMSE}}{\max(\text{actual}) - \min(\text{actual})} \quad \text{or} \quad \text{nRMSE} = \frac{\text{RMSE}}{\text{mean}(\text{actual})} \tag{2.3}$$

2.3.6.4 The Mean Absolute Error (MAE)

MAE indicates the average of the absolute differences between predicted and actual values, offering a clear measure of the magnitude of forecasting errors without considering their direction [97]. Unlike R^2 and RMSE, MAE does not allow positive and negative errors to cancel each other out, providing a straightforward indication of the typical error magnitude in predictions. Normalized Mean Absolute Error (nMAE) is widely used to standardize model results for comparison [96].

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} \left|\text{predicted}_i - \text{actual}_i\right| \tag{2.4}$$

$$\text{nMAE} = \frac{\text{MAE}}{\max(\text{actual}) - \min(\text{actual})} \quad \text{or} \quad \text{nMAE} = \frac{\text{MAE}}{\text{mean}(\text{actual})} \tag{2.5}$$

2.3.6.5 Mean Absolute Percentage Error (MAPE)

The MAPE evaluates the mean percentage deviation of predicted values from their corresponding observations. This metric highlights the relative scale of prediction errors, making it particularly useful for comparing model performance across varying time horizons or geographical regions [98].

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} \left| \frac{\text{predicted}_i - \text{actual}_i}{\text{actual}_i} \right| \times 100 \tag{2.6}$$

2.4 Economic aspect of forecasting and potential application

For renewable energy systems that rely on weather conditions, forecasting plays a pivotal role in scheduling and operational planning [99]. Reliable forecasts enable utilities and system operators to align generation with demand, thereby minimizing uncertainties associated with variable resources. Improved predictive accuracy also benefits plant managers, as it supports optimized scheduling of maintenance activities and more accurate market bidding strategies. With the increasing penetration of PV systems, the consequences of inaccurate forecasts on grid stability become more significant. In certain electricity markets, solar producers are subject to penalties when discrepancies between forecasted and actual production exceed a predefined tolerance threshold. Such regulatory mechanisms encourage precise forecasting to reduce risks and maximize financial returns [100].

A number of leading companies are actively involved in advancing PV forecasting solutions. For instance, Schneider Electric, headquartered in Rueil-Malmaison, France [101], and Enphase Energy, located in Fremont, California, USA [102], both provide innovative approaches for solar energy prediction. SteadySun, based in Paris, France [103], and UL Renewables, with headquarters in Atlanta, Georgia, USA [104], focus on irradiance forecasting and performance optimization. Similarly, GreenPowerMonitor, headquartered in Barcelona, Spain [105], delivers integrated monitoring and forecasting services tailored to renewable energy infrastructures, including PV plants. These organizations combine technological expertise with advanced analytical methods to enhance the reliability and efficiency of solar power forecasting.

2.5 Conclusion

With the growing incorporation of solar photovoltaic power into modern electricity networks, operators are confronted with multiple technical challenges arising from the variable nature of the resource. These challenges encompass voltage regulation problems, frequency

deviations, harmonic distortions, complexities in coordinating protection systems, concerns regarding rotor angle stability, and an increasing need for greater grid flexibility. In response to this concern, various strategies have been proposed in the literature to ensure the stability and efficiency of the electrical grid. These include methods such as predicting energy supply and demand, upgrading equipment for better regulation of electrical parameters, integrating storage systems, enhancing reactive power support, adapting protection mechanisms, and combining these approaches for optimal results [25]. This chapter presented a comprehensive examination that enlightens the forecast and prediction processes. The first part of this study outlines the key technical challenges associated with large-scale solar integration. The second section examines the different temporal and spatial resolutions applied in forecasting, followed by a discussion in the third section on the factors that influence predictive accuracy. Building on insights from prior research, the fourth section provides a detailed taxonomy of forecasting methodologies that have been widely explored within the scientific community. A review of recent developments highlights the increasing reliance on hybrid deep learning approaches, while also noting limitations related to computational complexity and processing time. Finally, the last section reported the economic implications of adopting advanced forecasting techniques, emphasizing potential returns on investment and profiling leading industrial players in the field.

Chapter 3

DESCRIPTION OF THE STUDIED PV PLANT

3.1 Introduction

A comprehensive understanding of all PV power plant components is essential for accurately assessing operational conditions and ensuring the reliability of investigative analyses. The performance of a PV system is influenced by several interconnected elements, including photovoltaic modules, inverters, transformers, monitoring systems, and, where applicable, energy storage units. Each of these components plays a critical role in determining the plant's overall efficiency, stability, and energy output.

Beyond the technical architecture, external factors such as meteorological conditions, grid interaction, and system degradation must also be considered, as they directly impact power generation and forecasting accuracy. A rigorous examination of these elements provides a well-defined operational framework, allowing for precise identification of factors affecting energy production. This understanding is fundamental for developing robust forecasting models and optimizing predictive methodologies.

3.2 Geographical localisation

The photovoltaic power plant analyzed in this study, shown in Figure. 1.1, is located in Ain El Bell, a town in Djelfa province, part of Algeria's High Plains. This region falls within a semi-arid zone, characterized by low rainfall and significant temperature variations. The plant's geographical coordinates are 34°40'30" N and 3°15'30" E, with an altitude of 1,036 meters.

Djelfa province has strong photovoltaic potential, as illustrated in Figure. 3.1, which presents the average daily and annual electricity production rates for a grid-connected solar photovoltaic plant with a nominal capacity of 1 kW-peak. These values are based on a

25-year historical dataset [106]. The selection of this power plant was primarily based on the availability of real data, ensuring a solid foundation for building the predictive models and the forecasting analysis.
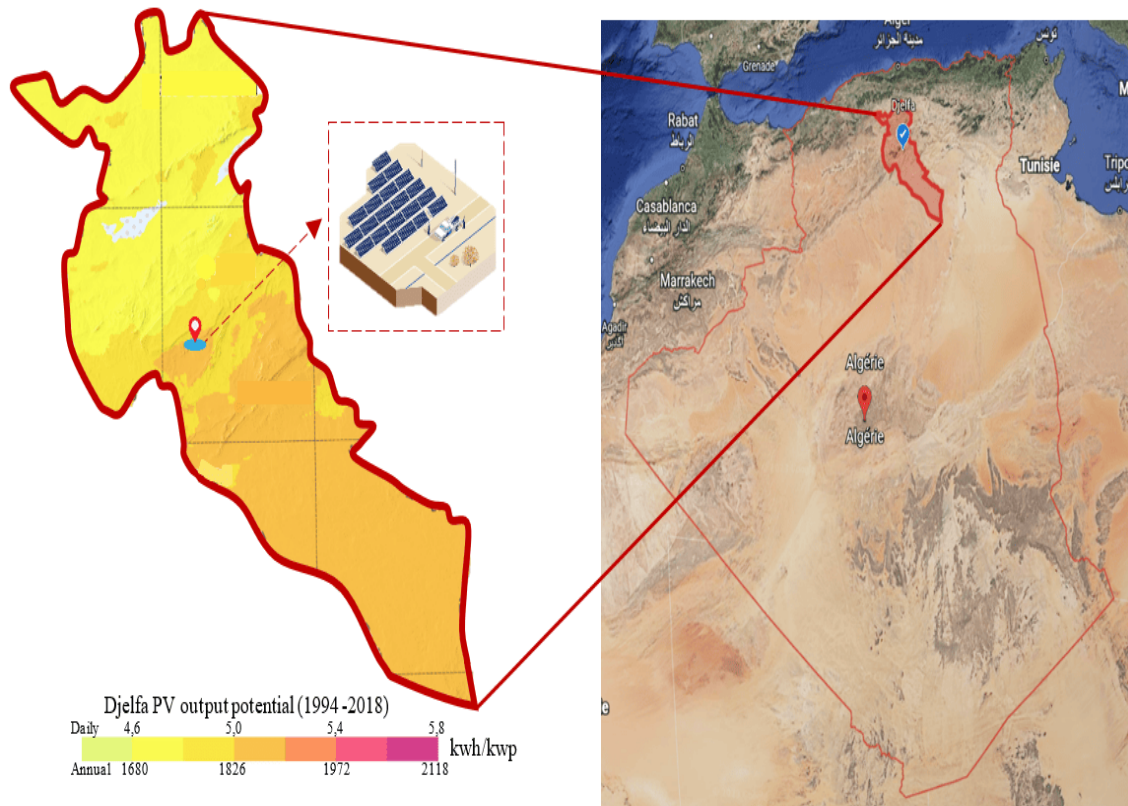
Figure 3.1: Average Daily and Annual PV Generation Potential Based on Historical Data from 1994 to 2018 [106]

The commissioning of the plant commenced on April 6th, 2016. The project is principally owned by SKTM (Shariket Kahraba wa Takat Moutajadida), with a consortium of partner companies, including Yingli Solar, Sinohydro, and Hydrochina, responsible for its construction and development. This photovoltaic power plant is integrated into the high-voltage grid at 60 kV. Initially, upon its commissioning, the station operated with a capacity of 20 MW, known as Djelfa I. However, in 2017, an expansion phase was undertaken, adding an additional 33 MW (Djelfa II), bringing the total installed capacity to 53 MW.

Spanning an area of 120 hectares, the project site accommodates not only the photovoltaic fields but also auxiliary systems, including inverters, transformers, and control units essential for efficient energy conversion and grid integration. The strategic layout of the facility was designed to maximize solar energy harvesting, optimize land use, and ensure operational reliability.Figure. 3.2 provides an overhead satellite image of the plant, obtained through Google Earth, while Figure. 3.3 presents a close-up view of one of the photovoltaic fields, illustrating the configuration and scale of the installation.structures.
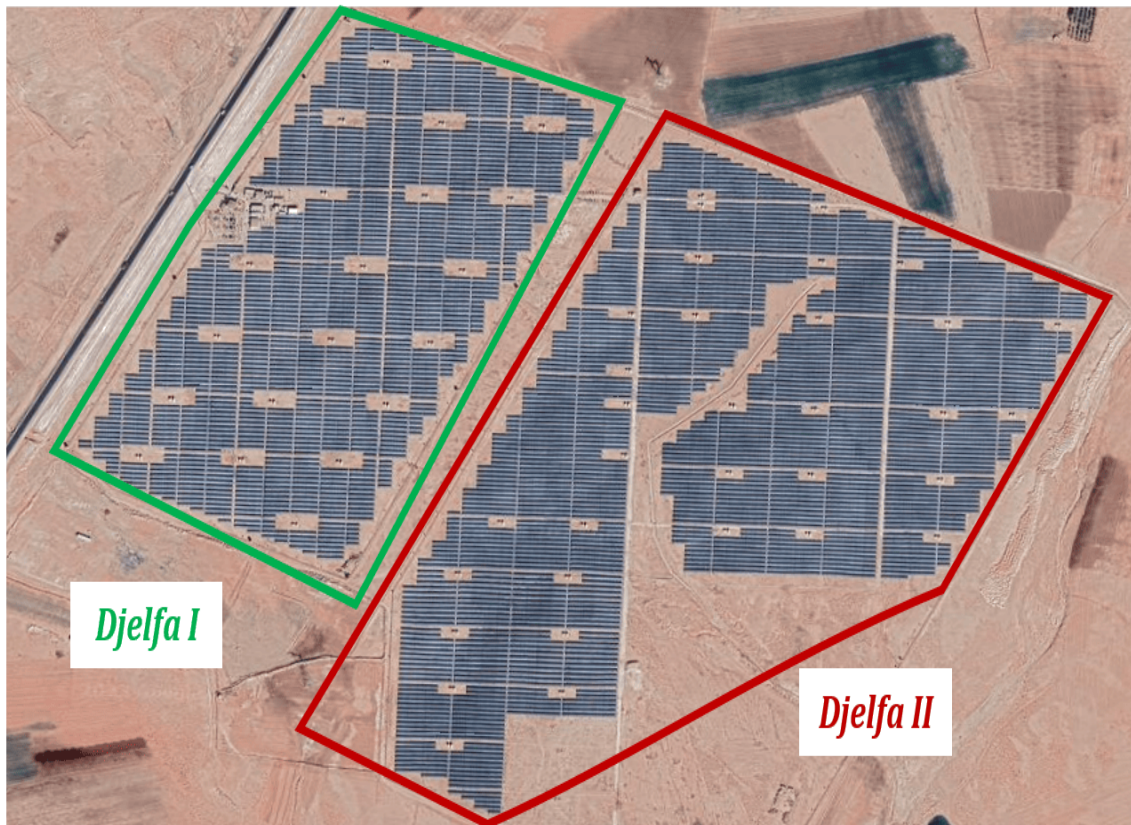
Figure 3.2: Djelfa photovoltaic power plant

Figure 3.3: One of the photovoltaic sub-fields in Djelfa solar plant

3.3 The photovoltaic fields

3.3.1 Photovoltaic modules, characteristics and configuration

Within the power plant, only one type of photovoltaic panel is used, which is of poly-crystalline technology and manufactured by the Chinese "Yingly Solar". For standard test conditions (1000w/m2, 15 °C cell temperature, and 1.5 Air Mass), the module's main characteristics are given in Table 3.1 and Table 3.2 [107].

The total number of PV modules used for this station is 212212 panels, and the optimal tilt angle is 33 degrees.

Table 3.1: Photovoltaic modules characteristic at STC conditions

Module type		YL250P-29b	Unit
Max.power	P_{max}	250	W
Tolerance	ΔP_{max}	0/+5	W
Efficency	n	15.9	%
Max.voltage	V_{max}	29.8	V
Open circuit voltage	V_{oc}	37.6	V
Max. System voltage	$V_{max,syst}$	1000	V
Max. current	I_{max}	8.39	A
Short. current	I_{sc}	8.92	A
Max. series fuse	$I_{max,syst}$	15	A
No. Cells	–	60	
Class	–	A	—
Fire rating	–	C	—

Table 3.2: Module dimensions

	Dimensions		
Width	w	990	mm
Height	h	1650	mm
thickness	t	40	mm

3.3.2 Junction boxes

Junction boxes are located close to the sub-fields. They are used to collect the electricity produced from the photovoltaic field. Mainly, there are two junction boxes; the combiner box and the DC power distributor box.

The power plant consists of 53 sub-fields each producing 1MW. One sub-field contains the following:

- 4004 Photovoltaic modules, spread over 182 strings,
- 91 supporting structures,
- Junction boxes (Combiner Boxes, DC power distribution Boxes).

Each string associates 22 modules and generates a total voltage of 655.5 V. Each eight (08) string are connected in parallel passing by a combiner box, and this results in a current of 67,12 A. After that, each of three (03) combiner boxes are assembled in parallel in the DC power distribution box, giving a current of 201,36 Amperes. This account leads to using 24 combiners and 8 DC power distribution boxes for each 1 MegaWatt. A representation of a single sub-sub-field is shown in Figure. 3.4. However, it is noteworthy that some combiner boxes connect six strings, while others connect four, in order to achieve the precise generated power. Figure. 3.5 and Figure. 3.6 represent the combiner box and the DC power distribution boxes.
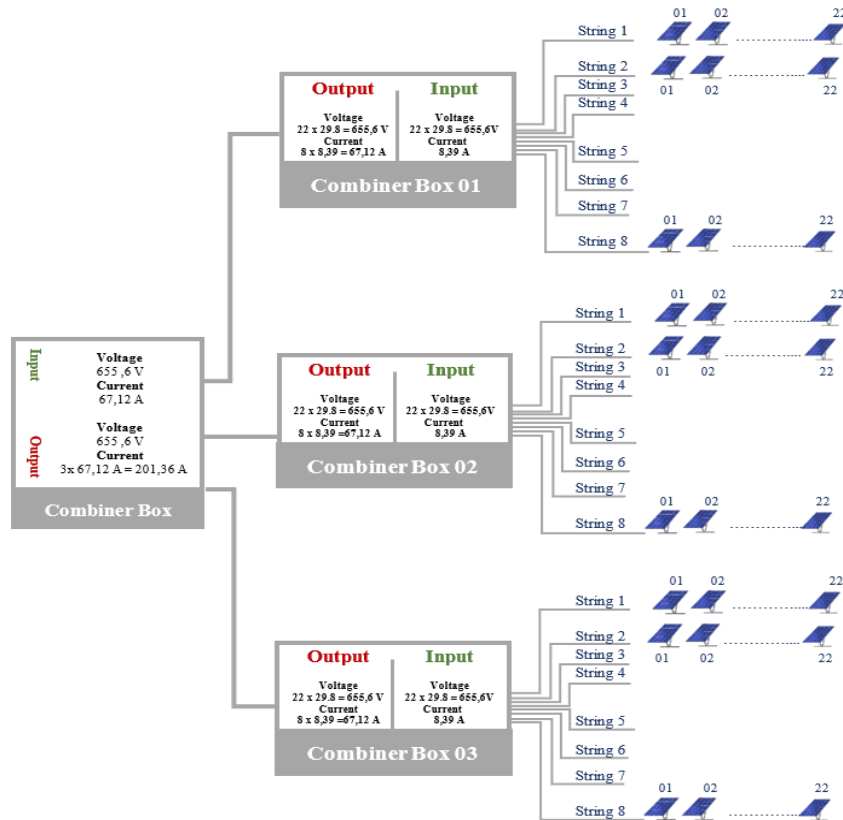
Figure 3.4: A unit component for a sub-sub-field
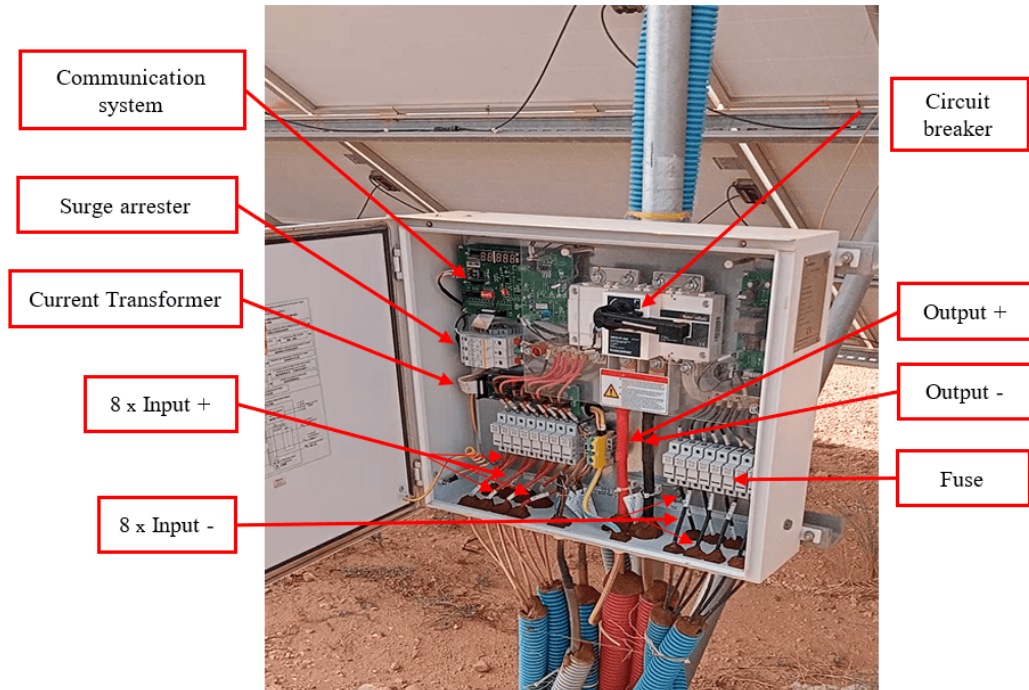
Figure 3.5: SunGrow Combiner box
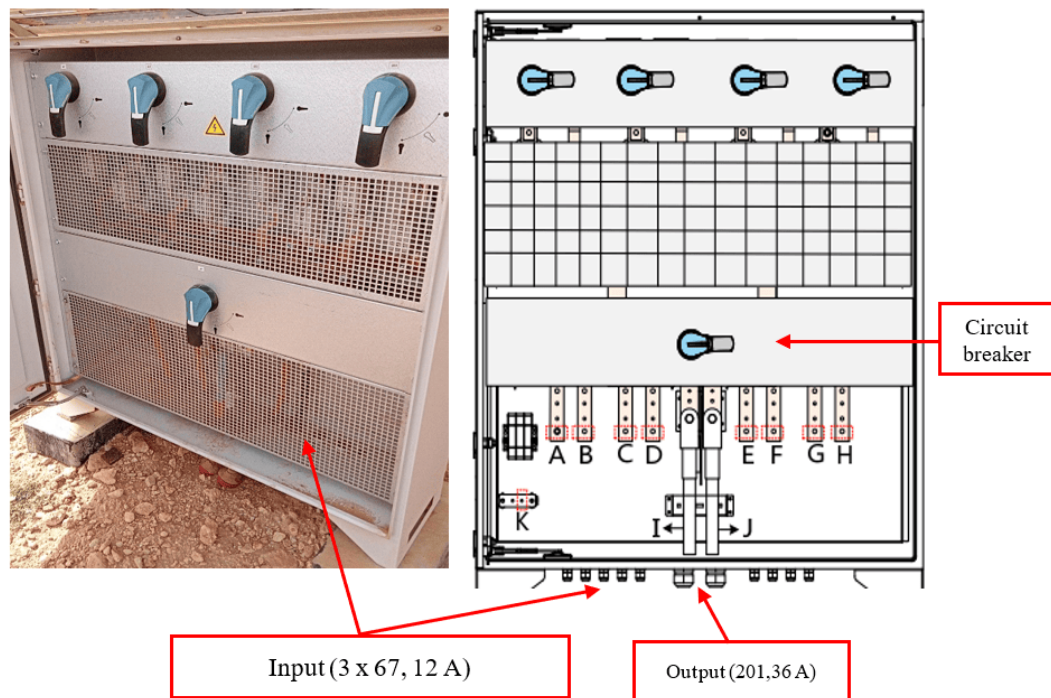
Figure 3.6: SunGrow DC power distribution box

3.4 Inverters

Inverters are electronic devices that transform the direct nature of the photogenerated current into an alternative. A «SUNGROW» inverter model was chosen for this station, having a reference of SG500MX (see Figure. 3.7) and the characteristics reported in Table 3.3 and Table 3.4.

Two inverters (02 x 500 kW) are used for one sub-field to produce 1 Mega Watt, named A and B respectively. Each one of the inverters receives the electricity generated by four (04) DC power distribution boxes. In other words, the input current for each inverter is (04 x 201.36 A), and the input voltage is 655,5 V. After the ripple step, three phases of 315V each are obtained.

Table 3.3: SG500MX input characteristics

	Input	
Max.Power	560	kW
Max.Voltage	1000	V
Max.Current	1200	A
Starting Voltage	500	V
Mpp Voltage	480 - 850	V
No.terminals	8/16	–
Efficiency	98.7	%
Temperature	-30 ~ +55	°C
Humidity	0~ 95	%

Table 3.4: SG500MX output characteristics

	Output	
Max. active Power	500	kW
Max.Apparent Power	550	kVA
Max. current	1008	A
Nominal system voltage	315 (252 ~ 362)	V
Frequency	50	Hz

Figure 3.7: «SG500MX inverter»

3.5 Main transformer and RMU units

The main transformer "SUNTEN 30KV", type SCLB10-1250 is captured in Figure. 3.9. Such a model offers high efficiency, low partial discharge, and high reliability to ensure long-term safe operation. The transformer is small and lightweight because it doesn't require oil usage. Its appropriate weight is half the weight of an oil transformer. Moreover, it is very resistant to sudden short circuits and temperature changes, and it also has a good performance against humidity.

The main transformer increases the voltage to 30kV. The three phases are included in a SafeRing-36kV cell, which is a ring main unit (RMU) with a CCF[1] configuration.(see Figure. 3.10).

This RMU enables the transformer to be connected back to the three phases of the previous sub-field transformer and forward to the three of the next one. In this station, every ten sub-field transformers (i.e each 10MG) are interconnected forming a loop using SafeRing-36kV devices as illustrated in Figure. 3.8. The terminals (the starting and ending point) of each loop are directed into the low-voltage cells that are situated at the discharge point. We can distinguish two terminals shown in Figure. 3.8: AH02, AH03 for a 10MW loop and AH04, AH05, for a second 10MW loop.

One advantage of this loop is to reduce the number of output cables and their equivalent

---

[1]C means cable switches and F refers to switch fuse combination

costs. Another option would be to provide energy in the event that problems occur in one of the sub-fields, which would allow for continuity of service.
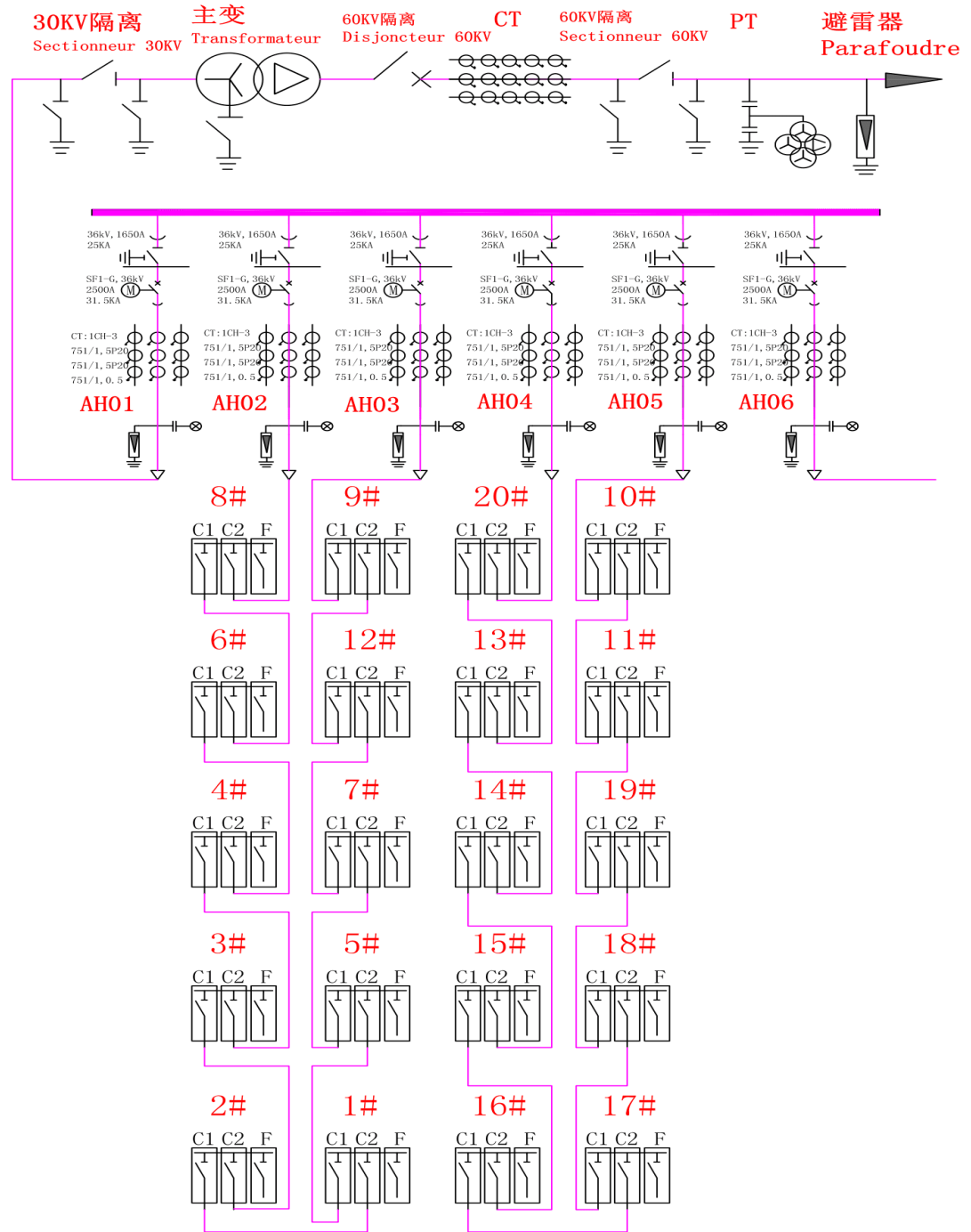
Figure 3.8: «Single line diagram of 20 MG»
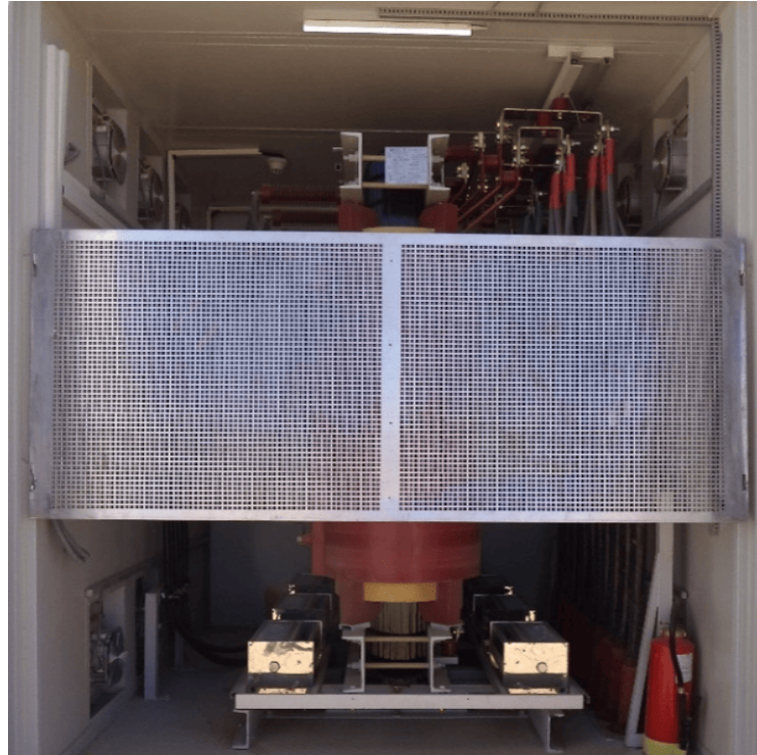
Figure 3.9: Transformer SUNTEN 30 kV

Figure 3.10: SafeRing-36kV

3.6 Discharge point

The discharge point is composed of a series of low-voltage cells that receive the terminals of all the loops, as discussed in the previous section. The global single-line diagram of Djelfa PV plant can be viewed in Figure. 3.11. The low-voltage cells consist of a «CMBGS-0» cabinet (see Figure. 3.12). It is manufactured by «Schneider electric» and is composed of several compartments represented in Figure. 3.13.
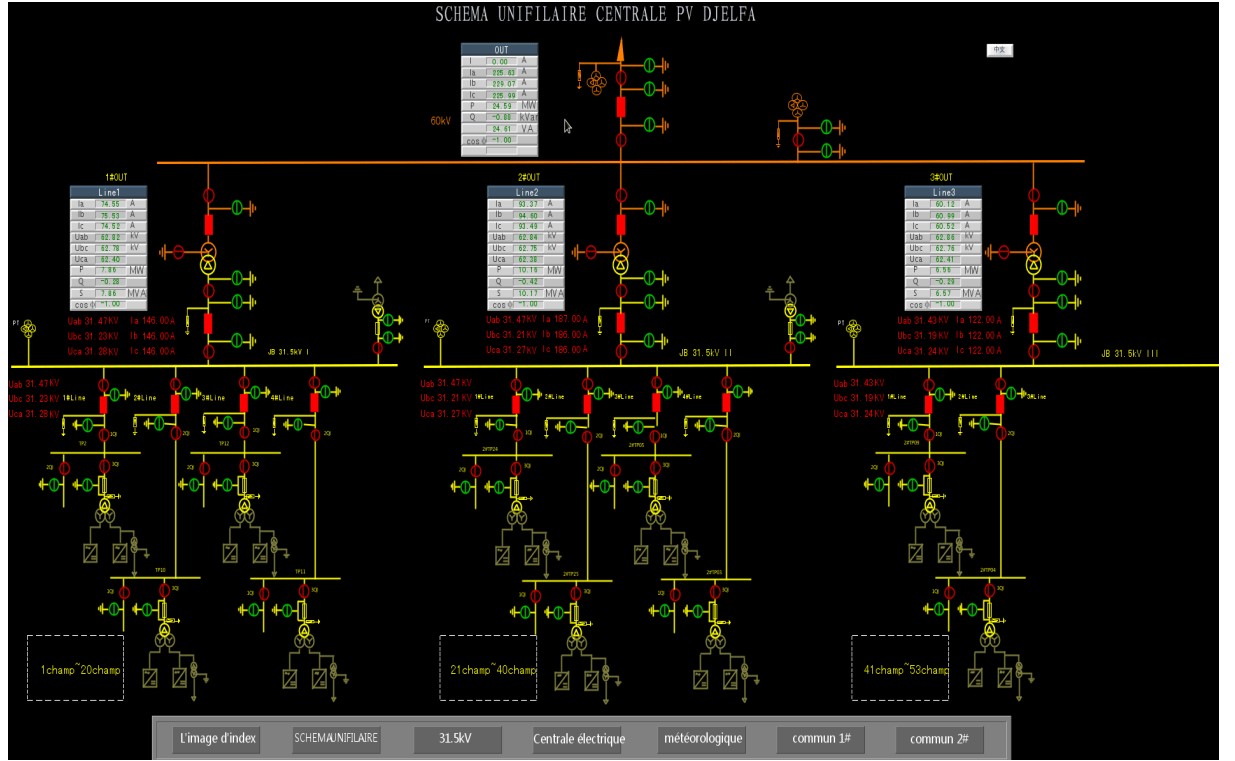
Figure 3.11: Global single line diagram of Djelfa PV plant

At the discharge point, the electricity is finally increased to 60Kv using a second transformer (Figure. 3.14) and then injected into the power grid. This second transformer has the properties reported in Table 3.5.

Besides, the discharge point is considered as the link between the power plant and the external power grid, it allows to:

• Connect/separate the photovoltaic station from the electrical grid,

• The power supply of auxiliary systems,

• Electricity counting using voltage and current transformers,

• In the event of loss of electrical equipment (short circuit, grounded), the interruption of the connection in time with the network to ensure the safety of the plant and the network.

Figure 3.12: «CMBGS-0» MV cell
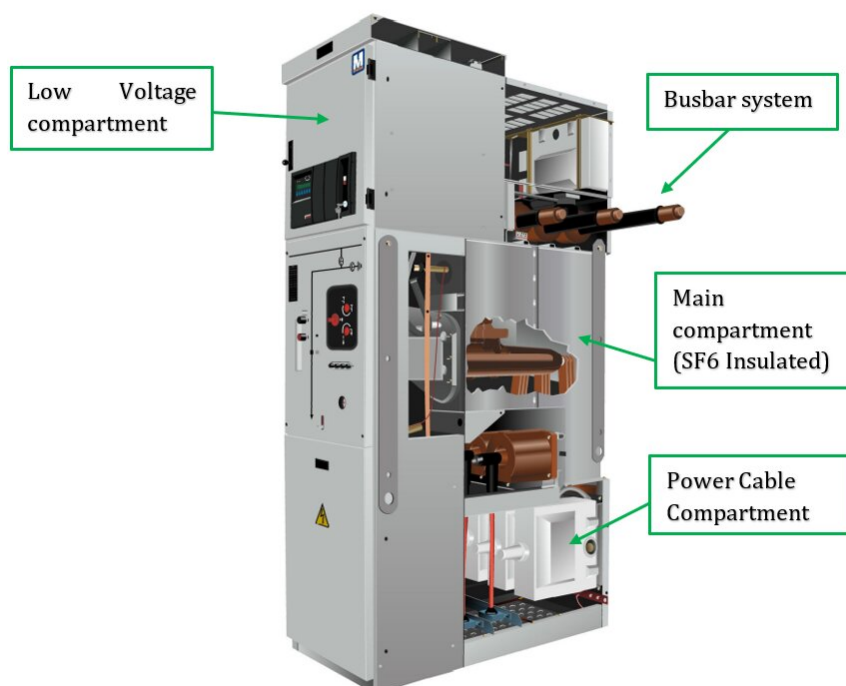
Figure 3.13: «CMBGS-0» main component

Figure 3.14: Transformer type «SFZ11-20000/60»

Table 3.5: Properties of the BETA SFZ11-20000/60 transformer

Type	60kV TEBA	SFZ11-20000/60
Nominal capacity	20000	kVA
Nominal Voltage	(60±8 x1.25%) /31.5	kV
Frequency	50	Hz
Couply type	YNd11	–
Cooling mode	ONAF	–
Atittude	<1100m	m

3.7 Control and monitoring system

The photovoltaic plant is monitored by a NARI «sj30B» server. This device communicates with all the components of the power plant such as the automatic weather station, inverters, Low-voltage cells, auxiliary system transformers, uninterruptible power supply (UPS), and the diesel unit, Hence, control engineers can easily consult any information related to these components, such as power generation, grid injection, auxiliary system consumption, weather data, electrical equipment status (photovoltaic modules, combiner, and DC power distribution boxes, inverters, transformers, etc.) using the SCADA[2] control software. An example of the inverter monitoring is given in Figure. 3.15, and a second example of weather data is given in Figure. 3.16.

Figure 3.15: Monitorin system - Inverter interface-
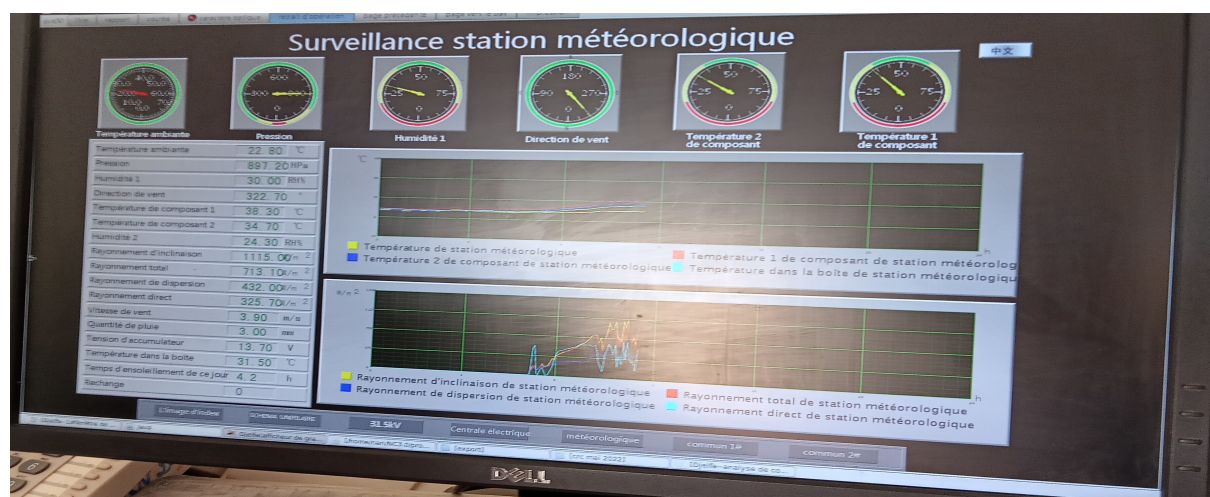
Figure 3.16: Monitoring system - radiometric station interface-

---

[2]Supervisory Control and Data Acquisition

3.8   Auxiliary systems power supply

The photovoltaic plant powers the auxiliary systems during the daytime, e.g., electronic control devices that are integrated into electrical components, air conditioning and lighting of the control room, alarm and fire system, computers and monitoring systems, telephone, and GPS.

During the night or if a power plant issue occurs, an auxiliary system transformer is used to provide the necessary energy. This later consumes from the high-voltage (60kV) network, converts it into 30kV, then 400V, and then powers all the equipment.

Another serious case is when the network encounters a blackout. The uninterruptible power supply "UPS" is then used to provide an emergency load. This system is equipped with batteries (Figure. 3.17) and is only intended for the power supply of the server, and the control room equipment, including computers, the telephone, the internet, the GPS, the alarm, and the fire system. Each inverter has a separate UPS system (Figure. 3.18), which is very important for engineers to monitor the plant's state during the incident.
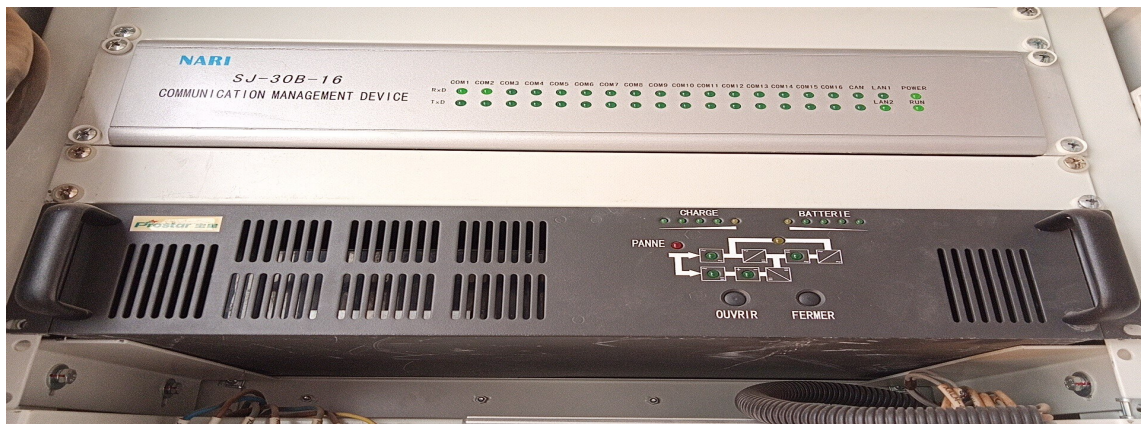
Figure 3.17: Inverter power supply system

Figure 3.18: UPS power supply batteries

3.9 Photovoltaic automatic weather station

The Ain El Bell weather station is fully automatic. It comprises four basic components: the sensor module, the communication module, the data acquisition module, and the power module (See Figure. 3.19). There are several weather sensors in the sensor module, each of which measures different climate data. After the capture is complete, the acquisition device converts analog and digital signals into statistic calculations. Afterward, the data is transmitted to the control room via the communication module. As for the power supply module, it provides the necessary power for the normal operation of all these devices.
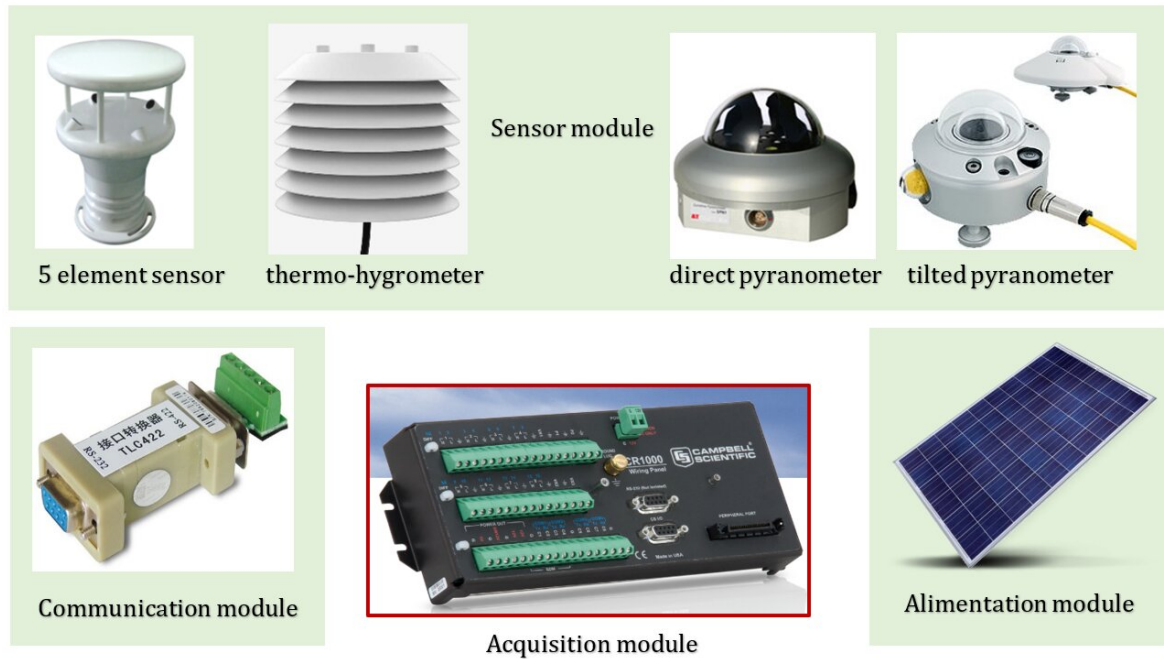
Figure 3.19: Photovoltaic automatic weather station modules

3.9.1 Sensor module

It is composed of the following meteorological sensors: 5 elements sensors, two pyranometers, and one thermo-hygrometer.

The five-element sensor shown in Figure. 3.20 measures five parameters that are: wind speed, wind direction, ambient temperature, relative humidity, and atmospheric pressure. It is installed at the top of the weather station, beside the lightning conductor (See Figure. 3.20). The specific metrics of each measurement are reported in (See Table 3.6):

Table 3.6: Five-element sensor specification

Measurement	Wind Speed (m/s)	Wind direction (degrees°)	Temperature (°C)	Humidity (%)	Atmospheric pressure (hPa)
Type	Ultrasonic	Ultrasonic	Diode voltage	Capacity	–
Range	$0 \sim 60$	$0 \sim 359, 9$	$-40 \sim +80$	$0 \sim 100$	$10 \sim 1100$
Accuracy	±3%	0,1	±0,2	± 3	±0,5 ($at25°C$)
Resolution	0,1	±3	0,1	0,05	0,1

Figure 3.20: Five element sensor

The tilted pyranometer shown in Figure. 3.21 is installed to measure the solar radiation received from the power plant with a tilt angle of 33°. A «CMP6» type is used with a main specification reported in Table 3.7. A second pyranometer of «SPN1» type, manufactured by DELTA-T is used to capture the global solar radiation among the station. This later has the characteristics reported also in Table 3.7 and is illustrated in Figure. 3.22.

Table 3.7: characteristics of the pyranometers

	CMP6	SPN1
ISO CLASS	1	1
Spectral range (50%points)	310 to 2800 nm	400 to 2700 nm
Sensitivity	5 to 16 uV/W/M2	10%
Response time	18s	100ms
Operational Temperature range	-40°C to +80 °C	-40°C to +70 °C
Max.solar irradiance	2000 W/m^2	2000 W/m^2
Field of view	180°	180°

Figure 3.21: CMP6 pyranometer

Figure 3.22: «SPN1» pyranometer

The weather station also incorporates a thermo-hygrometer (Figure. 3.23). It evaluates temperature and relative humidity under shade since the sensor's shelter provides protection against both scattered and direct solar radiation and precipitation.

Figure 3.23: Thermo hygrometer

3.9.2   Acquisition module

The acquisition is done by a CR1000 device. It consists of a measurement and control module and a wiring panel sub. Its standard operating temperature range is -25° to +50 °C, but an optional extended range of -55° to +85 °C is available. For the operating system, the CR1000 has 2 MB of flash memory, and 4 MB of battery-backed RAM for CPU use, program storage, and data storage. The storage capacity of the CR1000 can be increased by using a Compact Flash card. The versatility of the CR1000 allows measurement for several applications that concern; agriculture, soil moisture, air quality, vehicle testing, weather, aquaculture, etc. Figure. 3.24 shows the main compartments in the CR1000.

Input/Output Terminals— Individually configured for ratiometric resistive bridge, thermocouple, switch closure, high frequency pulse, low-level ac, serial sensors, and more.

Removable Power Terminal—simplifies connection to external power supply.

RS-232—provides a 9-pin DCE port for connecting a battery-powered laptop, serial sensors or RS-232 modems.

CS I/O Port—connects with AC-powered PCs and communication peripherals such as phone, RF, short-haul, and multidrop modems.

Peripheral Port—allows data to be stored on a CompactFlash card and/or supports Ethernet communications.

Figure 3.24: CR1000 compartments, made by Campbell Scientific

3.9.3 Communication module

The CR1000 device has only RS232 port, which is characterized by a short transmission distance that limits the range of address choice of the data acquisition chassis. Therefore, a universal 232/485 mini-converter is adopted (Figure. 3.25).

Figure 3.25: Universal 232/485 mini-converters

3.9.4 Power supply module

The CR1000 device is powered with a PS200, PS100, BP12, BP24, or a BPALK. Both the PS200 and PS100 provide a 7 Ah, while BP12 and BP24 provide nominal ratings of 12 and 24 Ah respectively. These batteries are rechargeable and should be connected to a regulated charging source (in this case a solar photovoltaic panel). Finally, The BPALK is composed of eight non-rechargeable D-cell alkaline batteries with a 7.5 Ah rating at 20 °C. An example of a PS200 microcontroller is given in Figure. 3.26. While the whole photovoltaic automatic weather station is shown in Figure. 3.27.

Figure 3.26: PS200 micro-controler

Figure 3.27: Photovoltaic automatic weather station

3.10   Conclusion

Djelfa, a Saharan region of Algeria, possesses significant solar potential, making it an ideal location for PV energy generation. This chapter has provided a comprehensive description of the PV power plant under investigation, detailing its infrastructure, operational characteristics, and key influencing factors. The plant is connected to the 60 kV high-voltage grid, with a nominal capacity of 53 MW, generated by 212,212 photovoltaic panels. The technical specifications and configuration of these panels were outlined in Section 3.3, offering insight into the plant's design and energy conversion process.

The next chapter will introduce a detailed statistical analysis of the PV plant's meteorological parameters alongside PV power production data. Additionally, it will delve deeper into the methodological framework and the building phases of the predictive models. The goal is to establish a robust framework for developing an accurate data-driven forecasting model. By leveraging data mining techniques, this study aims to enhance the predictability of solar energy production, ultimately contributing to more effective energy management and grid integration strategies.

Chapter 4

DEEP LEARNING-BASED PV POWER FORECASTING

4.1 Introduction

For decades, researchers have relied on traditional forecasting methods "statistical models and physics-based approaches " to predict solar energy generation. Yet, solar radiation is anything but predictable. Sudden cloud movements, fluctuating atmospheric conditions, and seasonal variations create nonlinear patterns that rigid mathematical formulations struggle to capture. As a result, these conventional models often fall short in providing accurate and adaptive predictions for photovoltaic (PV) power generation.

This chapter embarks on a journey into the world of deep learning, a transformative approach that has redefined time series forecasting. Unlike traditional methods, DL models do not rely on predefined equations or assumptions; instead, they learn directly from data, uncovering intricate dependencies, long-term correlations, and hidden structures. Their ability to adapt and generalize across diverse fields and conditions has already been demonstrated, but exploring their full potential in PV power forecasting remains a crucial step. By putting these models to the test in this domain, we aim to refine predictive accuracy, uncover new insights, and push the boundaries of solar energy forecasting.

4.2 The proposed methodology framework

The core objective of this study is to develop an advanced deep learning model for PV power forecasting, particularly in the Djelfa province, a high plains region of Algeria. This research integrates Variational Mode Decomposition with advanced machine learning and deep learning techniques to enhance forecasting accuracy. VMD is employed to decompose complex PV power signals into distinct and interpretable modes, providing refined inputs for models such as LSTM, BiLSTM, GRU, BiGRU, CNN, Inception, Residual Networks, ANN,

SVM, and Random Forest.

The proposed methodology is validated using data from a 53 MW grid-connected PV plant in Djelfa, Algeria, covering the period from January 2018 to December 2019. Forecasting performance is evaluated for 30-minute and 3-hour horizons, with monthly assessments offering detailed insights into the effectiveness of the integrated approach. The key steps of the proposed study are as follows:

- **Data Collection and Preprocessing:** The initial phase of the study is focused on detailed data collection and comprehensive processing. Essential activities in this stage encompass rectifying missing values, identifying and removing anomalies, and excluding non-essential nighttime data. Then, data is strategically partitioned into separate subsets for training and testing purposes, 80% and 20%, respectively.

- **Variational Mode Decomposition:** Following the data cleaning process, VMD is employed to extract features from the non-stationary characteristics of the PV generation signal. This technique decomposes the signal into a series of IMFs, each representing distinct frequency components within a limited bandwidth.

- **Training and optimizing the predictive models:** In this stage, the ML and DL modes undergo a rigorous tuning process through a grid search methodology that is appropriate for each model. This process aimed at identifying the most effective hyper-parameters, ensuring the model's ability to deliver precise and reliable PV power forecasts.

- **Seasonal and Monthly Analysis:** The next step is to delve deeper into understanding the model's patterns within the data, particularly during specific times of the year, and to analyze their seasonal variations. This involves not only recognizing recurring trends but also identifying the underlying factors driving these fluctuations across different months. By doing so, we can achieve a more thorough understanding of how the model interacts with the data over time and refine our insights accordingly.

- **Comparative Analysis and Performance Metrics Assessment:** Additional steps include conducting an analysis of the models' forecasting ability using actual data, and evaluating their performances using a range of metrics. This analysis is critical in understanding each model's advantages and disadvantages in real-world scenarios.

4.3   Data mining and processing

Data mining is a fundamental step in any research study, forming the basis for subsequent evaluations and model development. It serves as the initial phase of predictive modeling, focusing on data analysis, correlation identification, and pattern recognition. These processes

are essential for extracting meaningful insights, selecting relevant features, and identifying key relationships among variables.

This section focuses on describing and examining meteorological and electrical data from the Ain El Bell power plant, which is critical for understanding the variability of PV power generation. The analysis of factors such as solar irradiance, temperature, and electrical performance provides a basis for selecting an appropriate forecasting approach and predictive model. By investigating statistical trends and correlations, this study ensures a structured and comprehensive dataset, enhancing the accuracy and reliability of the forecasting framework.

4.3.1   Collection and description

The meteorological data are automatically collected and recorded by the photovoltaic automatic weather station, as referenced in Section 3.9. This station continuously monitors multiple atmospheric and environmental parameters, ensuring precise and consistent data acquisition. The complete list of recorded variables is presented in Table 4.1. However, due to confidentiality restrictions, access to meteorological data for this study is limited to ambient temperature and global solar radiation.

Similarly, in terms of electrical parameters, only the photovoltaic power output has been made available for analysis. While additional electrical variables, such as voltage, current, and efficiency metrics, could provide deeper insights into system performance, their restricted availability necessitates a focus on the accessible data. Despite these limitations, the selected parameters are essential for assessing the influence of meteorological conditions on photovoltaic power generation and formulating a reliable forecasting approach.

Table 4.1: The measured meteorological parameters

Measured parameter	nomenclature	unit
Ambient temperature	T_{amb}	°C
Component temperature	T_{comp}	°C
pressure	P_{atm}	HPa
Relative humidity	R_H	%
Wind speed	W_S	m/s
Wind direction	W_D	degrees (°)
Global solar radiation	G_{global}	w/m^2
Diffused solar radiation	$G_{diffused}$	w/m^2
Direct solar radiation	G_{direct}	w/m^2
Tilted solar radiation	G_{tilt}	w/m^2
Sunshine duration	SS_D	hours
Rainfall amount	R_F	mm

The dataset is systematically organized into a single Excel file comprising two separate sheets, each corresponding to a complete year of recorded data. The first sheet contains measurements spanning from January 1st, 2018, to December 31st, 2018, while the second sheet covers the same period for the year 2019. Each sheet consists of five columns, providing key recorded parameters: date, time, produced photovoltaic power (PV_{prod}), global solar radiation (G_{global}), and ambient temperature (T_{amb}). Data acquisition follows a fixed schedule, with measurements collected at regular 30-minute intervals from 06:00 to 20:00. This temporal resolution allows for a detailed examination of intra-day variations and seasonal trends in PV power generation.

To provide a clearer understanding of the dataset structure, Figure. 4.1 presents an example of a single day's recorded measurements, demonstrating the chronological arrangement of meteorological and electrical parameters. This initial dataset is essential for guiding subsequent analyses and facilitating the methodological steps required in this research.

	A	B	C	D	E
1	Date	TIME	P (MG)	(W/m2)	(·C)
2	01/01/2018	06:00	-0,25	0	9,4
3		06:30	-0,26	0	9,4
4		07:00	-0,26	0	9,3
5		07:30	-0,26	3,3	9
6		08:00	-0,16	9,4	9,1
7		08:30	0,6	23,7	9,2
8		09:00	2,88	78,7	9,4
9		09:30	6,35	147,5	9,7
10		10:00	18,85	280,2	10,2
11		10:30	21,15	319,5	10,4
12		11:00	29,35	410,4	10,6
13		11:30	38,98	596,9	11
14		12:00	40,23	678,3	11,5
15		12:30	47,23	678,8	12,1
16		13:00	41,75	674	12,6
17		13:30	37,46	542,1	12
18		14:00	42,58	665,9	12,8
19		14:30	36,58	542,4	12,7
20		15:00	28,26	382,4	13
21		15:30	23,84	326,9	13
22		16:00	23,85	280,4	12,9
23		16:30	14,68	231,9	13,4
24		17:00	7,97	136,3	12,6
25		17:30	1,25	18,3	11,8
26		18:00	-0,3	4	10,7
27		18:30	-0,25	3,3	10,1
28		19:00	-0,26	7,4	9,6
29		19:30	-0,25	0	9,1
30		20:00	-0,26	0	8,7

2018 | 2019 | ⊕

Figure 4.1: Measured data for January 1st, 2018

4.3.1.1   Global solar radiation

Being far from the Mediterranean coast, the geographical position of Djelfa allows for benefiting from rich solar radiation favorable to photovoltaic applications. This potential varies depending on the daytime and the seasons. The variation in solar radiation levels throughout the day is mainly caused by the angle at which the sun's rays strike the Earth's surface. Hourly solar radiation levels usually increase from sunrise to midday, as the sun rises higher in the sky and its rays pass through less of the earth's atmosphere. The highest levels of solar radiation are typically observed around noon, which is the time when the sun is at its highest point in the sky and its rays have to pass through less of the atmosphere. When the sun starts to set, solar radiation levels begin to decline as its rays pass through more of the earth's atmosphere. The lowest solar radiation levels are typically observed around sunset when the sun is near the horizon. An example of the hourly solar radiation pattern is represented in Figure. 4.2.
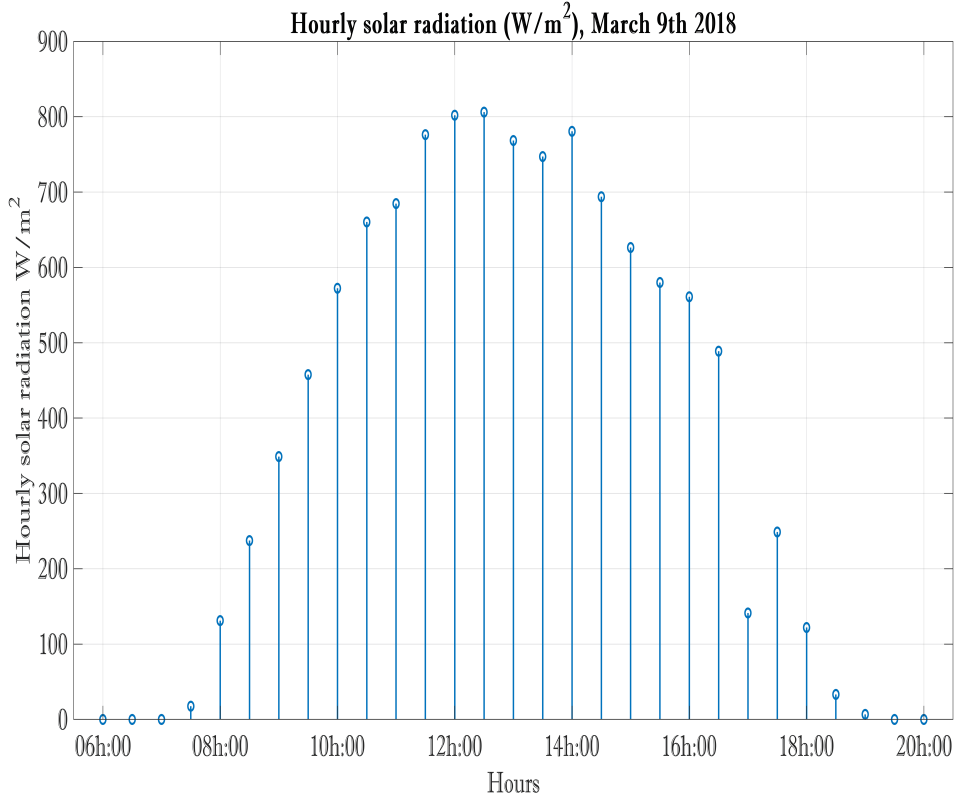
Figure 4.2: Hourly solar radiation, March 9th 2018

Throughout one year, the solar radiation variation is primarily due to the tilt of the Earth's axis, which causes the angle of the sun's rays to change as the Earth orbits around the Sun. During summer months, the sun's rays strike the Earth's surface at a more direct angle, which results in more solar radiation reaching the surface. In contrast, during the winter months, the angle of the sun's rays is more oblique, which results in less solar radiation reaching the surface. For Djelfa region which is located in the northern hemisphere, most solar radiation is received during the summer months, when the northern hemisphere is tilted towards the sun. For example, Figure. 4.4 shows that the highest values of the maximum global solar radiation are recorded during the month of August, and the highest maximal value reached is 1380.3 W.m^2. In contrast, the lowest values of Gmax are recorded during the winter months, especially January and December (666.6 W.m^2 and 700.8 W.m^2 respectively), which are generally characterized by rainy or cloudy days, and short sunshine duration.

But this is not a theorem with no exceptions, solar radiation that a location receives can also be affected by local weather patterns, such as cloud cover or air pollution. Clouds can block or scatter some of the sun's rays, reducing the amount of solar radiation that reaches the Earth's surface, and it arrives where in hot seasons, there are cloudy days, which results in very weak solar radiation recordings. The day of August, 12st shown in Figure. 4.3 is an example of this exception.
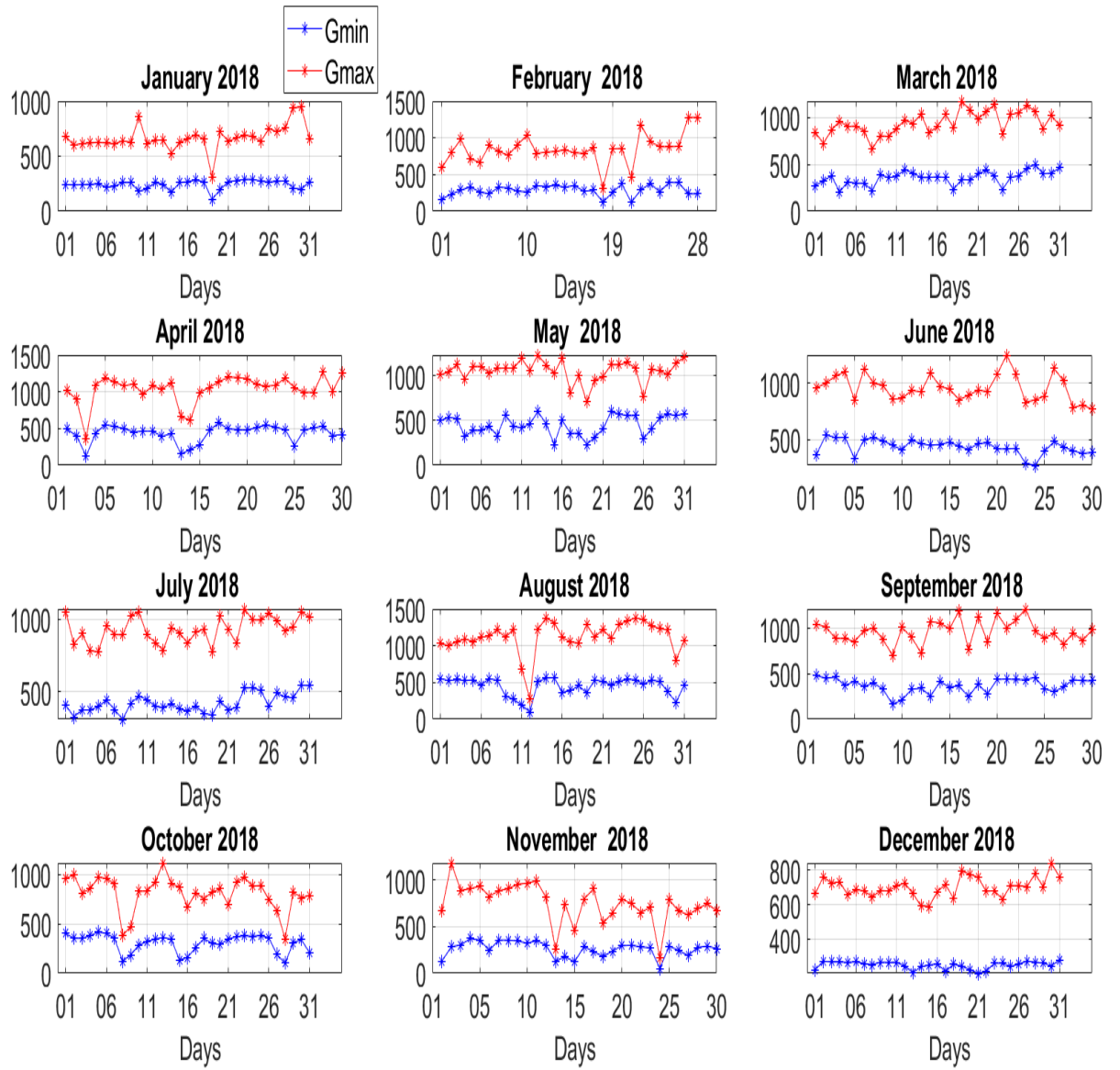
Figure 4.3: Monthly solar radiation variation, 2018
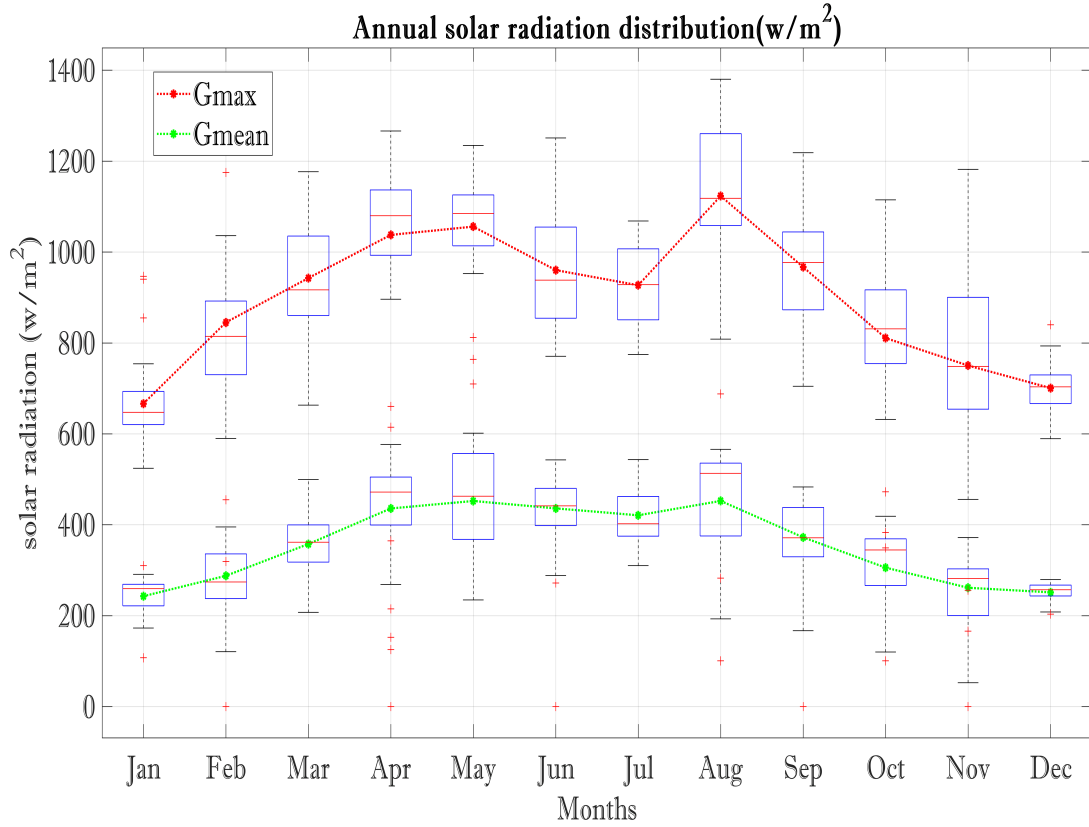
Figure 4.4: Solar radiation variation for the year 2018

4.3.1.2 Temperature

Djelfa has significant temperature fluctuations between day, night, and seasons. This variation is primarily driven by the rotation of the Earth and the way in which the Sun heats the surface of the planet. During the day (See Figure. 4.5), temperatures typically rise as the sun heats up the atmosphere, then increase gradually as the day progresses. The highest temperatures are usually reached in the mid to late afternoon, as the Sun's rays are the most direct and intense at this time. As the sun sets and the atmosphere cools down, temperatures gradually decrease, reaching a low in the early morning hours before the sun rises again.

The temperature levels and patterns are affected by a variety of factors, including the sun's position, cloud cover, wind speed, and humidity levels. For example, the average high temperature in summer (June to August) is around 35-40 degrees Celsius, while the average low temperature is around 18-20 degrees. In winter (December to February), the average high temperature is around 13-16 degrees Celsius, while the average low temperature is around 2-4 degrees Celsius. The monthly and annual temperature evolutions are shown in Figure. 4.6 and Figure. 4.7.
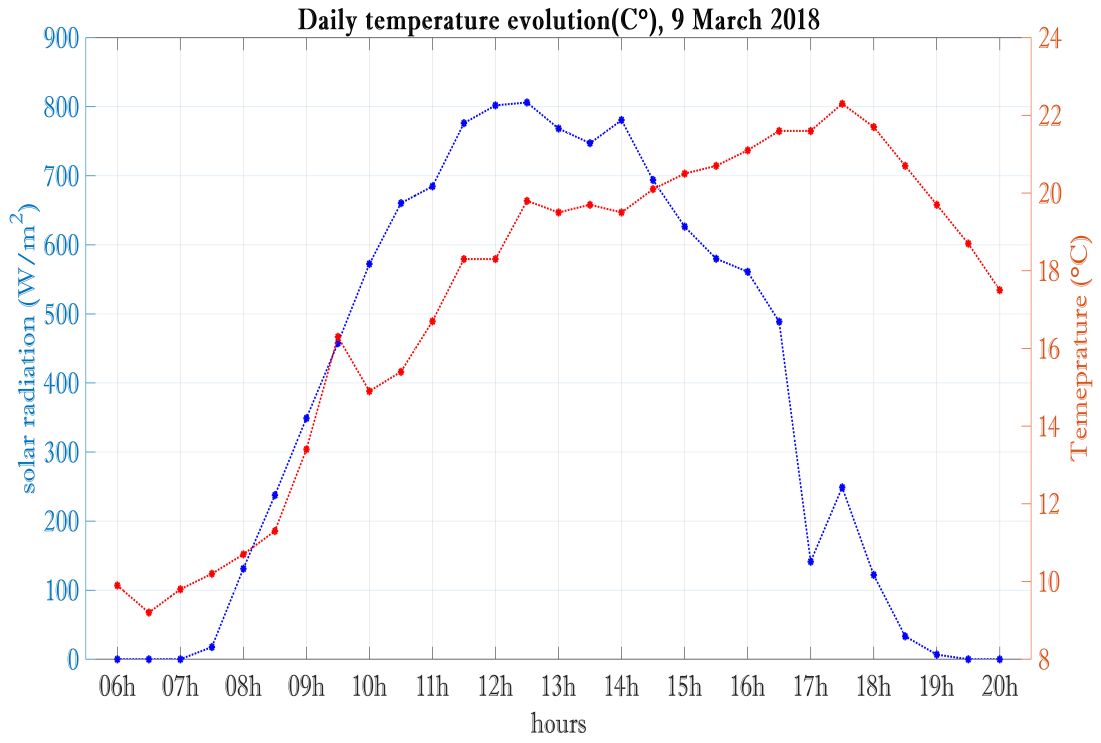
Figure 4.5: Daily temperature evolution (C°), 9 March 2018
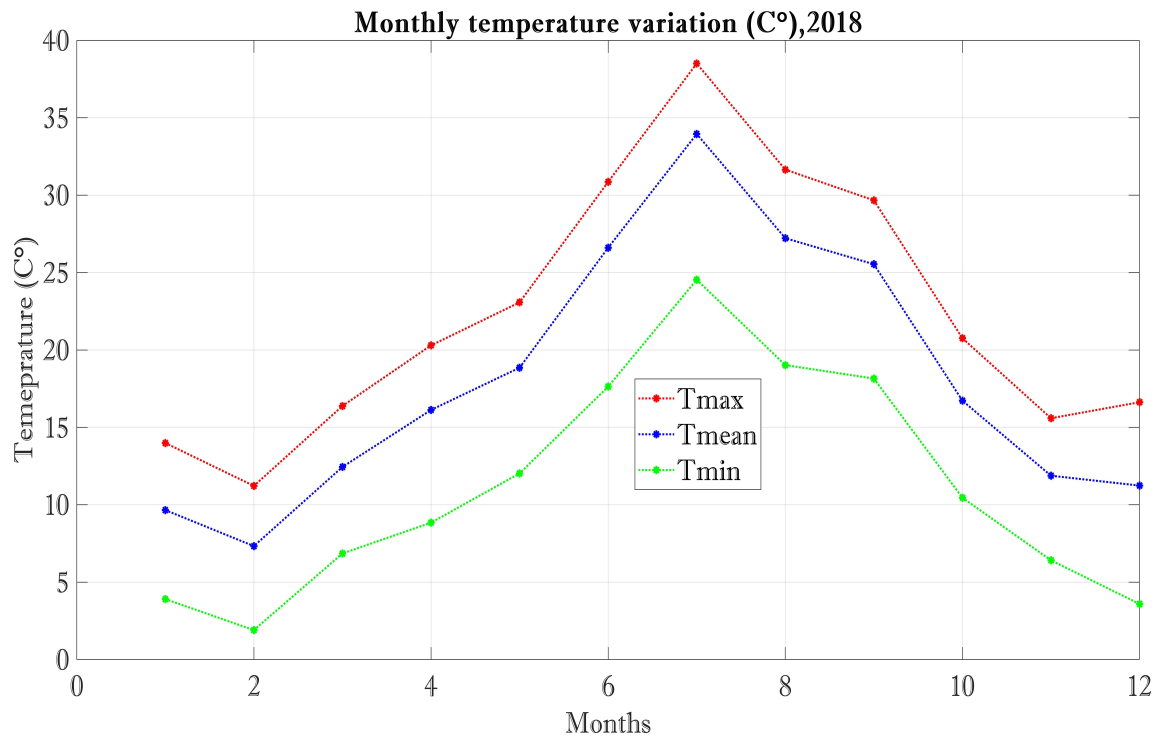
Figure 4.6: Monthly temperature variation (°C), 2018
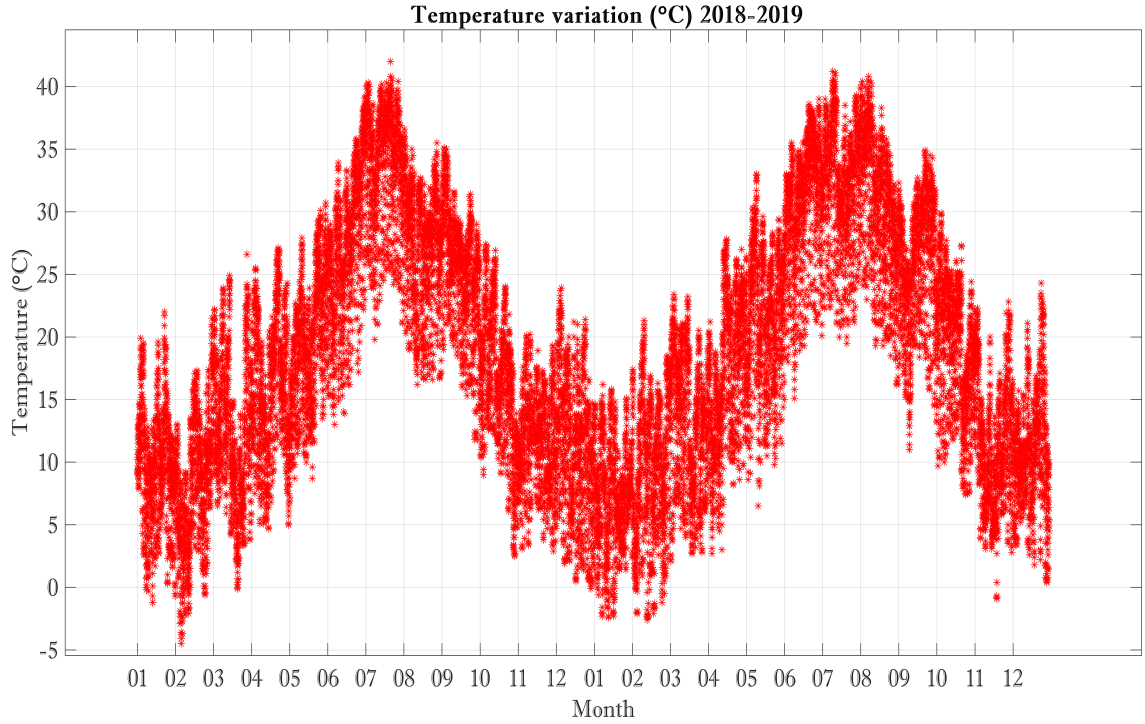
Figure 4.7: Annual temperature evolution(°C) 2018-2019

4.3.1.3   Photovoltaic production

The photovoltaic fields of Djelfa power plant are tilted at 33 Degrees, which is the optimum angle for this localization to receive more direct sunlight and produce more energy. The data shows that the minimum required power for the system to start powering the power grid is approximately 20 W/m^2. Typically, the daily photovoltaic generation evolution follows a curve that gradually increases in the morning, peaks in the midday hours, and then tapers off as the sun sets (see Figure. 4.8). During the morning, when the sun begins to rise, the PV power plant will start generating electricity. As the sun rises higher in the sky, the amount of sunlight hitting the panels increases, and the output of the PV system will gradually increase as well. In the midday hours, when the sun is at its highest point in the sky, the electricity output of the PV system will be at its peak. The output of the PV system will begin to decrease as the sun begins to set in the late afternoon. The daily photovoltaic energy is represented in Figure. 4.9.

The PV generation is variable along the seasons, Figure. 4.10 and Table 4.2, shows that Spring remains the most ideal season for photovoltaic production. The higher maximal PV produced is reached on February,10st at 14h:00 (maxPmax = 58.34 kW/m^2). Additionally, the highest average of Pmax is achieved during March (meanPmax = 46.42 kW/m^2). Whereas a decrease in the maximal production is observed during hot months (June, July, and August). Finally, the lowest value of Pmax is recorded during November, 24st at 15h:00 (minPmax = 7.41 kW/m^2).

hourly and daily PV produced power evolution
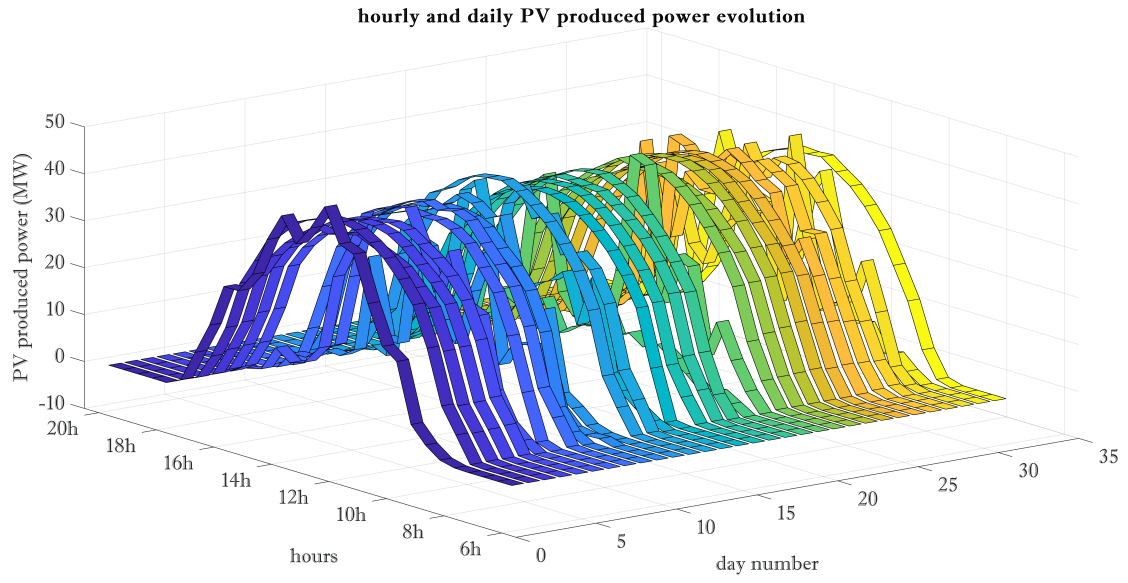
Figure 4.8: hourly and daily PV power output
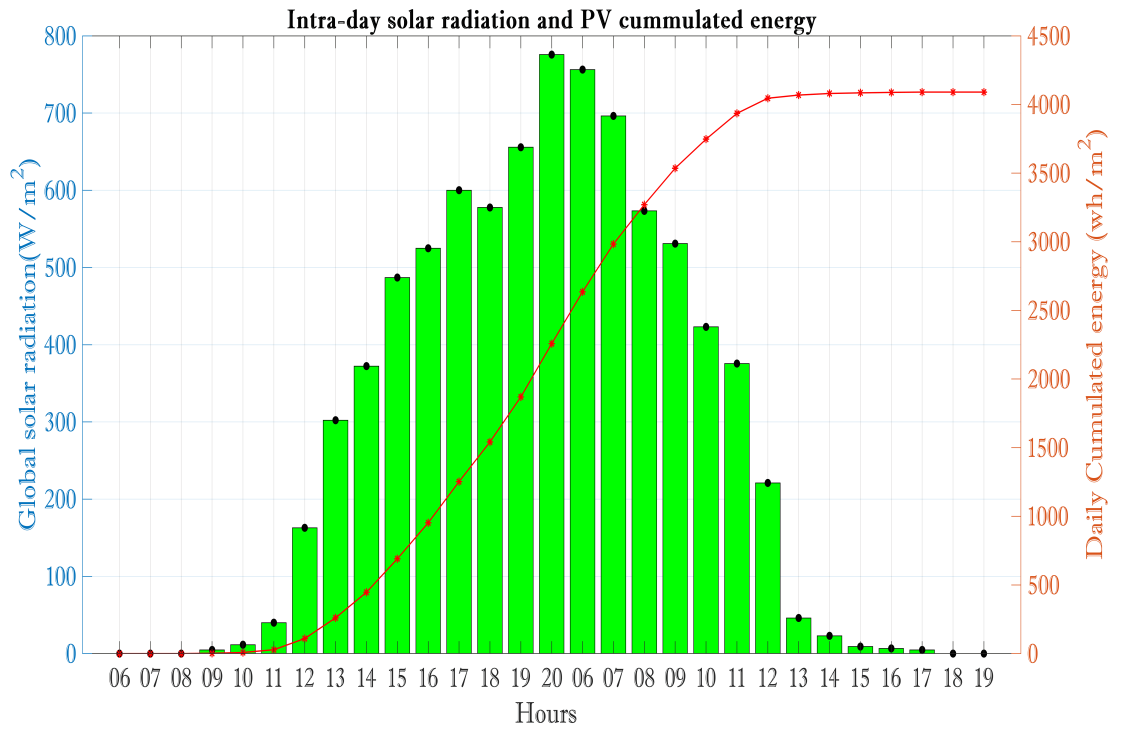
Figure 4.9: Intra-day solar radiation and PV accumulated energy
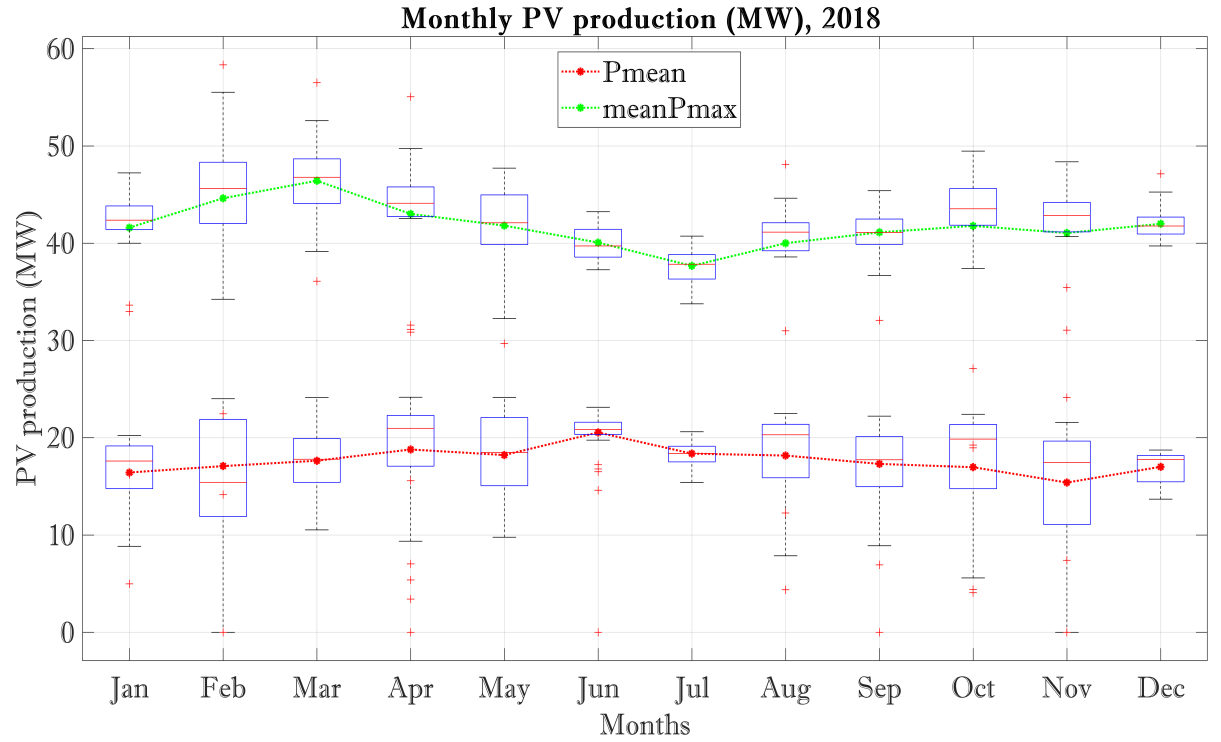
Figure 4.10: Monthly PV production (MW),2018

Table 4.2: Variation of Maximum and Mean PV power

Month Number		01	02	03	04	05	06	07	08	09	10	11	12
Pmax	Mean	41.58	40.31	**46.42**	41.62	41.81	38.78	37.678	40.008	39.8	41.788	39.73	41.99
	Max	47.23	**58.34**	56.50	55.06	47.72	43.24	40.73	48.11	45.4	49.47	48.37	47.14
	Date	01/01	**10/02**	20/03	11/04	01/05	07/06	09/07	10/08	27/09	29/10	14/11	28/12
	Time	12:30	**14:00**	13:00	12:30	12:00	12:30	12:30	13:30	13:30	12:30	13:00	13:30
	Min	16.21	14.15	36.1	15.59	29.68	37.27	33.77	12.26	32.08	18.96	**7.41**	39.72
	Date	19/01	08/02	08/03	03/04	19/05	24/06 30/06	28/07	12/08	09/09	28/10	**24/11**	08/12
	Time	13:00	16:00	14:30	17:30	12:30	12:30 13:00	12:30	14:00	13:00	13:30	**15:00**	14:00
Pmean	Mean	16.43	15.44	17.64	18.19	18.23	**19.88**	18.37	18.17	16.76	16.96	14.90	17.01
	Max	20.24	24.01	24.14	**24.17**	24.16	23.14	20.62	22.50	22.22	22.41	21.58	18.74
	Date	23/01	25/02	27/03	**05/04**	02/05	06/07	30/07	04/08	28/09	18/10	04/11	31/12
	Min	4.99	5.23	10.53	3.41	9.78	14.61	15.41	4.38	6.93	4.01	**2.05**	13.69
	Date	19/01	21/02	24/03	17/04	19/05	01/06	08/07	12/08	09/09	16/10	**24/11**	21/12

4.3.2 Analysis and exploration

In the previous section, it was stated that the data are compiled in a single Excel file. However, storing the database in a block format is not practical for the intended analysis. A more structured format was required to facilitate the examination of existing patterns and correlations among the three parameters with greater precision and detail.

To achieve this, the data were imported into MATLAB, and a program was developed

to automatically reorganize the dataset into a monthly format, where each month contains its corresponding daily measurements. This process necessitated the inclusion of additional columns to specify the year and month. Figure. 4.11 illustrates the reshaped data in Excel format.

	A	B	C	D	E	F	G	H
1	Year	Month	Day	TIME (hours)	TIME (min)	P (MG)	(W/m2)	(·C)
2	2018	1	1	6	0	-0,25	0	9,4
3	2018	1	1	6	30	-0,26	0	9,4
4	2018	1	1	7	0	-0,26	0	9,3
5	2018	1	1	8	30	-0,26	3,3	9
6	2018	1	1	8	0	-0,16	9,4	9,1
7	2018	1	1	8	30	0,6	23,7	9,2
8	2018	1	1	9	0	2,88	78,7	9,4
9	2018	1	1	9	30	6,35	147,5	9,7
10	2018	1	1	10	0	18,85	280,2	10,2
11	2018	1	1	10	30	21,15	319,5	10,4
12	2018	1	1	11	0	29,35	410,4	10,6
13	2018	1	1	11	30	38,98	596,9	11
14	2018	1	1	12	0	40,23	678,3	11,5
15	2018	1	1	12	30	47,23	678,8	12,1
16	2018	1	1	13	0	41,75	674	12,6
17	2018	1	1	13	30	37,46	542,1	12
18	2018	1	1	14	0	42,58	665,9	12,8
19	2018	1	1	14	30	36,58	542,4	12,7
20	2018	1	1	15	0	28,26	382,4	13
21	2018	1	1	15	30	23,84	326,9	13
22	2018	1	1	16	0	23,85	280,4	12,9
23	2018	1	1	16	30	14,68	231,9	13,4
24	2018	1	1	17	0	7,97	136,3	12,6
25	2018	1	1	17	30	1,25	18,3	11,8
26	2018	1	1	18	0	-0,3	4	10,7
27	2018	1	1	18	30	-0,25	3,3	10,1
28	2018	1	1	19	0	-0,26	7,4	9,6
29	2018	1	1	19	30	-0,25	0	9,1

2018 | 2019 | ⊕

Figure 4.11: Reshaped data for January 1st, 2018

4.3.3   Data-Driven Diagnosis

Understanding and optimizing the performance of PV systems has become a priority for ensuring energy reliability and sustainability. Central to this objective is the effective use of operational data, which serves not only as a foundation for diagnosing faults but also as a critical resource for interpreting system behavior under varying environmental and technical conditions. The correlation between PV power output and solar irradiance is a fundamental indicator of system performance, as irradiance directly governs the energy conversion process in PV modules under normal operating conditions. In a healthy PV system, this relationship exhibits a strong, often near-linear correlation, particularly under stable environmental conditions. However, the presence of faults, such as partial shading, soiling, module degradation, or inverter malfunction, can significantly disrupt this correlation.

This section examines the linear relationship between PV power generation and global solar radiation and temperature. This analysis focuses on data recorded simultaneously, varying in parallel with the same measurement step, ensuring consistency in comparison. To quantify this relationship, the Pearson correlation coefficient is employed, as it measures both the strength and direction of the linear association between two continuous variables.

A detailed formulation and explanation of this coefficient are provided in Section 2.3.6 for further reference.

4.3.3.1   PV/ GSR interdependence study

A comparative visual analysis is carried out to investigate the relationship between solar irradiance and photovoltaic power generation across varying temporal scales. The analysis begins with a clear-sky day, serving as a reference for expected system behavior under optimal conditions, and is followed by a day marked by high solar radiation intermittency to highlight potential deviations. This evaluation is then expanded to include a sequence of consecutive days, a complete month, and ultimately an entire year. Such a multiscale approach facilitates a comprehensive assessment of correlation dynamics and supports the identification of performance anomalies or system faults that may manifest intermittently or evolve.

Figure. 4.12 represents the PV generation during a clear sky day. Typically, the daily production of photovoltaic fields follows the same patterns as the solar radiation evolution, as the amount of energy generated by a PV system depends on the amount of sunlight that is available to it. On days with long hours, clear skies, and plenty of sunlight, the electrical production of a PV power plant can be at its highest. This conclusion is confirmed by Figure. 4.13, which shows a strong linear relationship between the two parameters, with a correlation factor equal to 0.978. However, if there is cloud cover or other weather conditions that reduce the amount of sunlight reaching the panels, the daily energy production may be lower, as shown in Figure. 4.14. The linear relationship is still strong, and the correlation factor is also high (R=0.983), as depicted in Figure. 4.15. The same insights are drawn when exploring the correlation for successive days (Figure. 4.16 and Figure. 4.17),
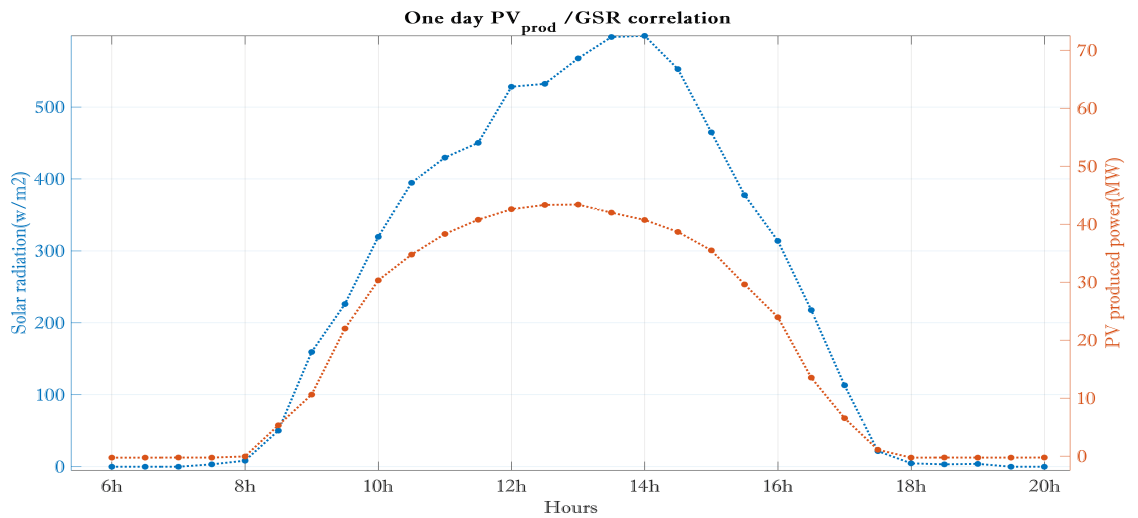
Figure 4.12: Correlation of solar radiation with PV generation for one day
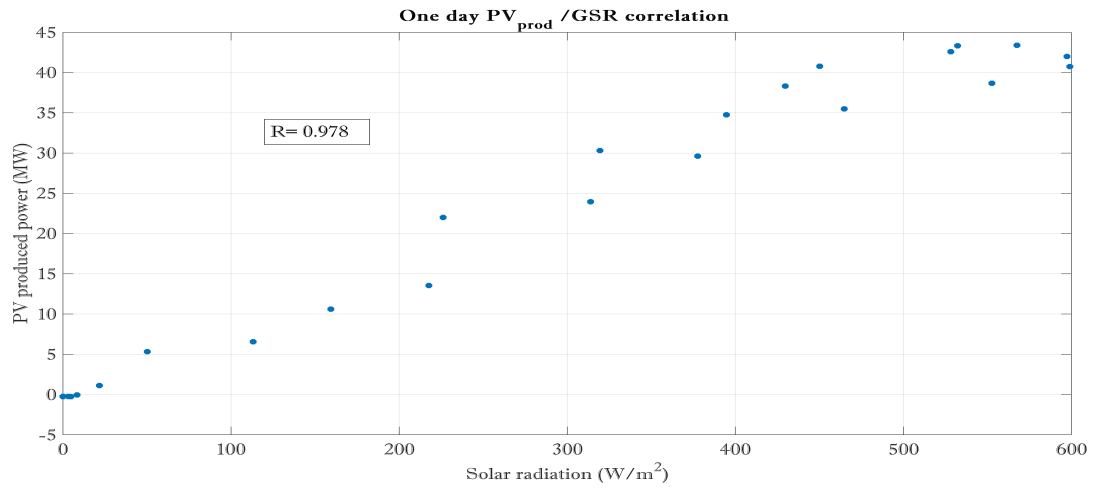
Figure 4.13: Scatter plot of the solar radiation and produced PV for one day
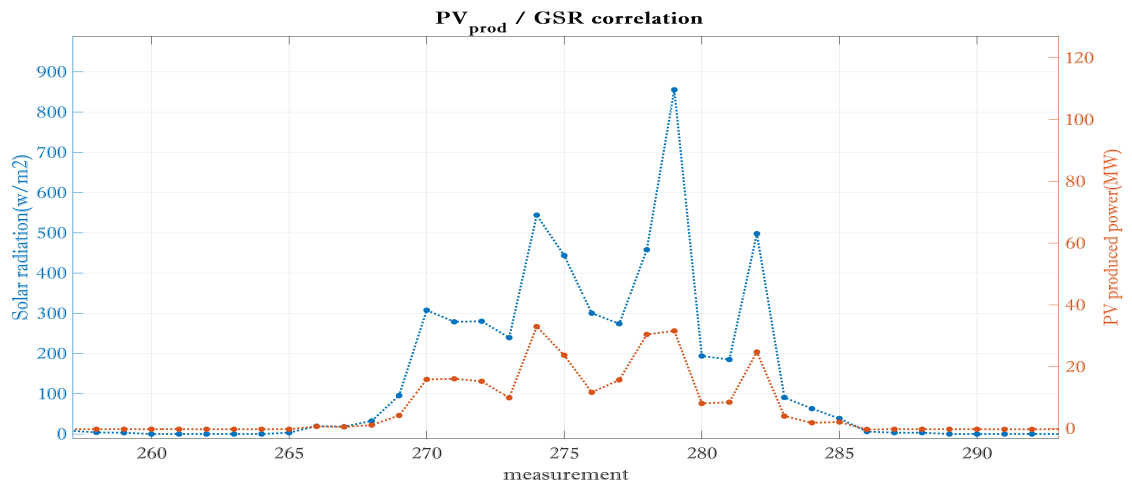
Figure 4.14: Correlation of solar radiation with PV generation during high intermittency
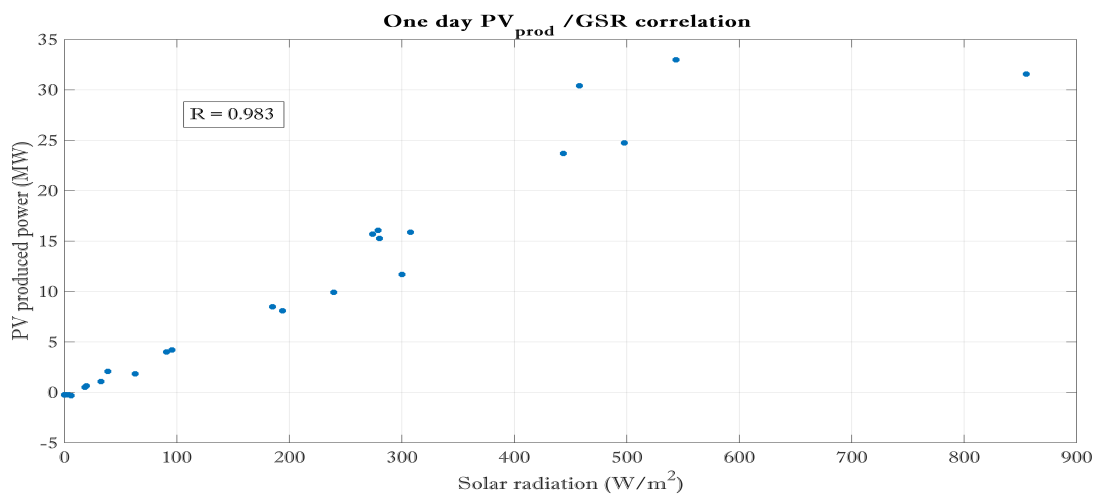
Figure 4.15: Scatter plot of the solar radiation and produced PV during high intermittency
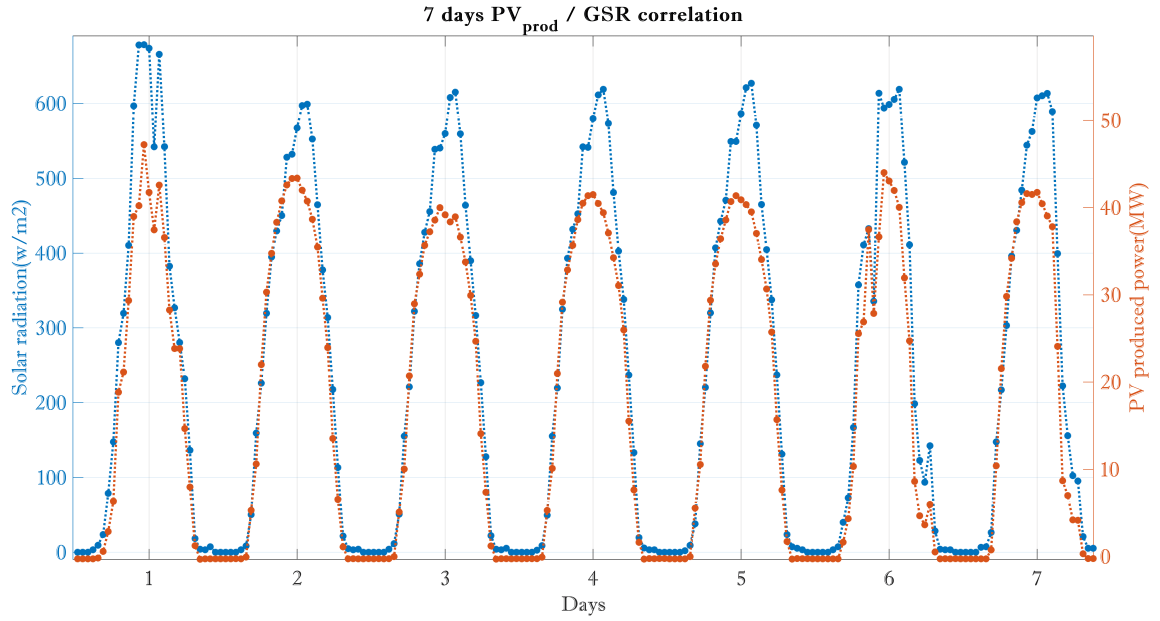
Figure 4.16: TCorrelation of solar radiation with PV generation for seven days
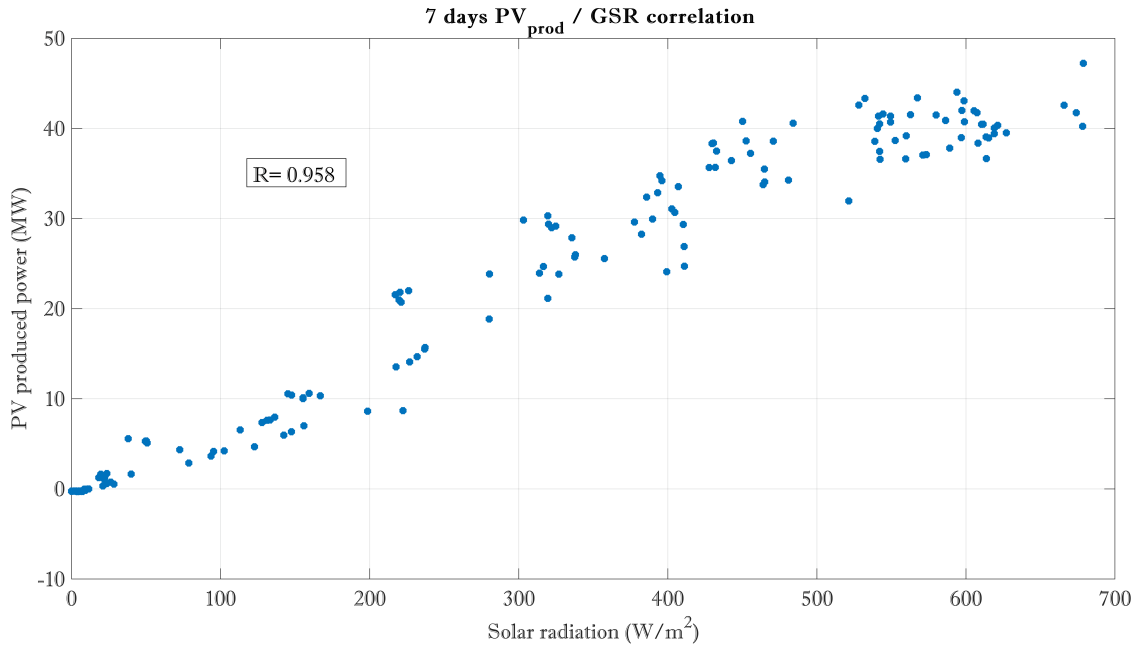
Figure 4.17: Scatter plot of the solar radiation and produced PV for seven days

The correlation becomes particularly evident when the entire month's dataset is considered, as depicted in Figure. 4.18, and is further reinforced by the year-long measurements presented in Figure. 4.20. The pronounced interdependence between the two parameters, along with consistently high correlation coefficients, is clearly illustrated in Figure. 4.19, and Figure. 4.21.
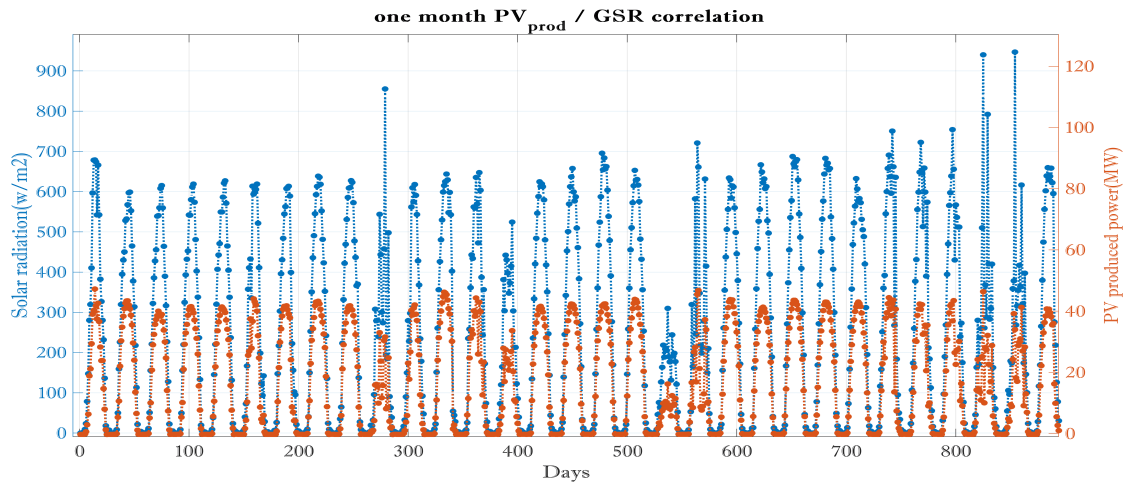
Figure 4.18: Correlation of solar radiation with PV generation for one month
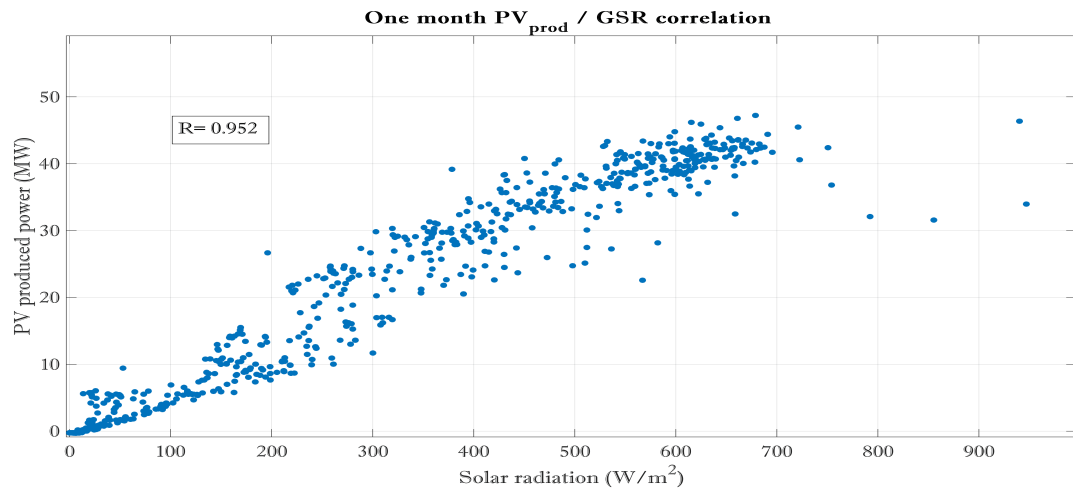
Figure 4.19: Scatter plot of the solar radiation and produced PV for one month
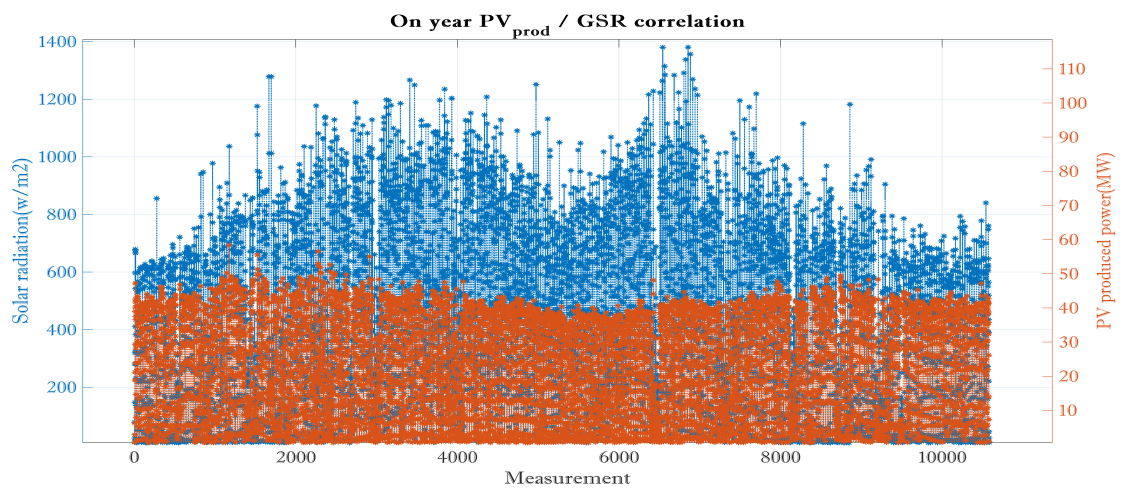
Figure 4.20: Correlation of solar radiation with PV generation for one year
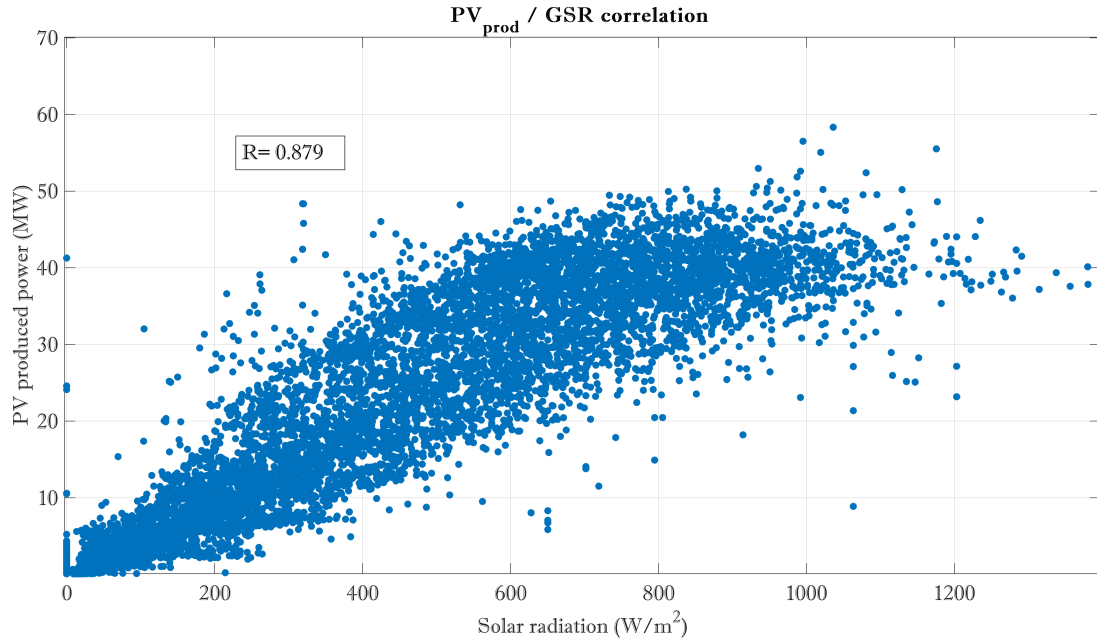
Figure 4.21: Scatter plot of the solar radiation and produced PV for two years

Interestingly, certain periods revealed significant fluctuations in PV power output despite an anomalous and sustained increase in solar irradiance levels. This apparent decoupling between irradiance and power generation strongly suggests the presence of underlying system faults or disturbances. Such behavior may result from various issues, including partial shading, thermal stress, sensor inaccuracies, or inverter inefficiencies, all of which can impair the system's ability to convert solar energy into electrical power effectively. These anomalies disrupt the expected physical correlation between irradiance and power output, serving as strong indicators of degradation or malfunction. Moreover, such deviations become even more pronounced in long-term or annual representations, where excessive and persistent discrepancies can be more easily detected and characterized, reinforcing the diagnostic value of irradiance-power correlation analysis.

It is important to emphasize that the proportionality PV/GSR is not uniform across seasons. As illustrated in Figure. 4.22, the slope characterizing the relationship between generated power and incident solar radiation exhibits slight monthly variations. This seasonal variability is further reflected in the correlation coefficient R, as shown in Figure. 4.23.
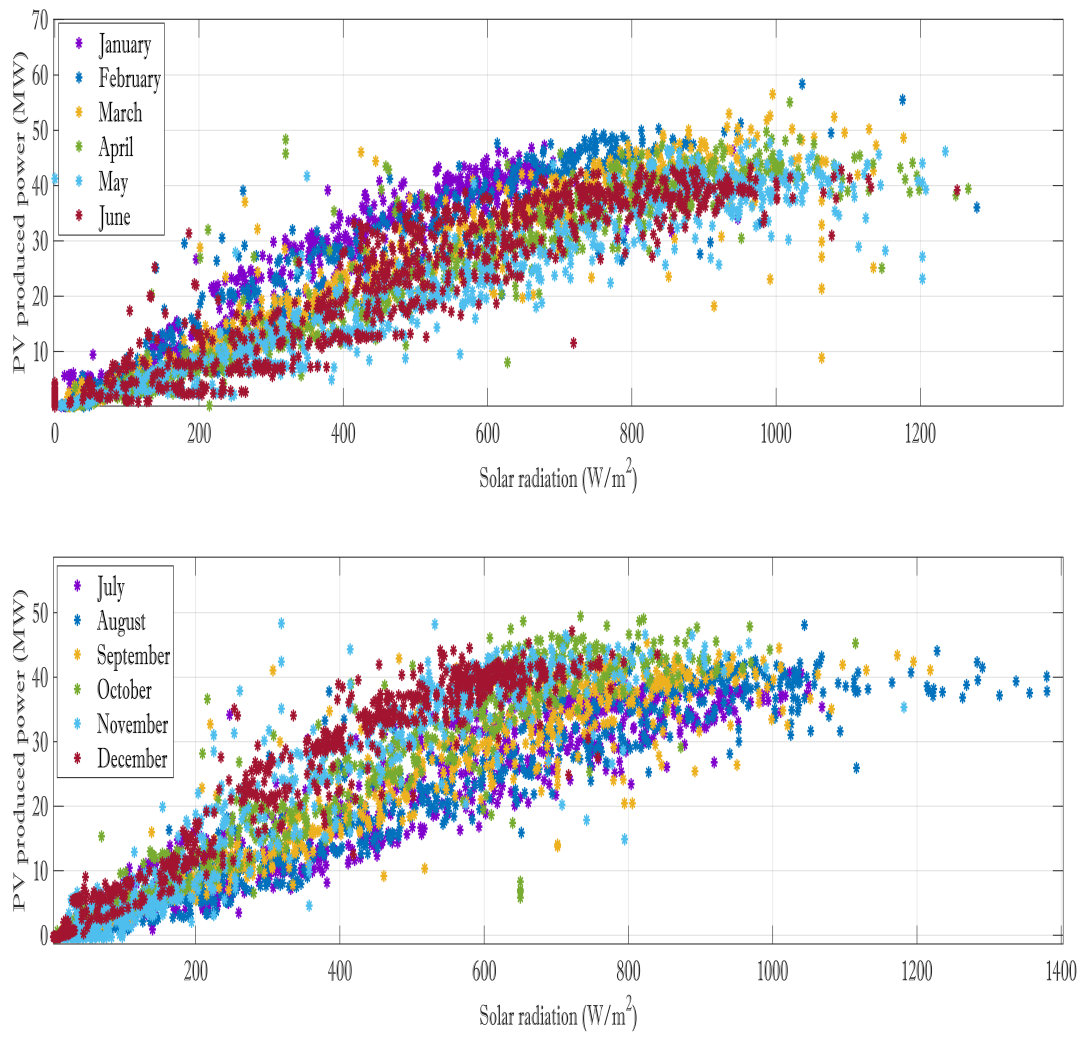
Figure 4.22: Correlation of solar radiation with PV generation for two years
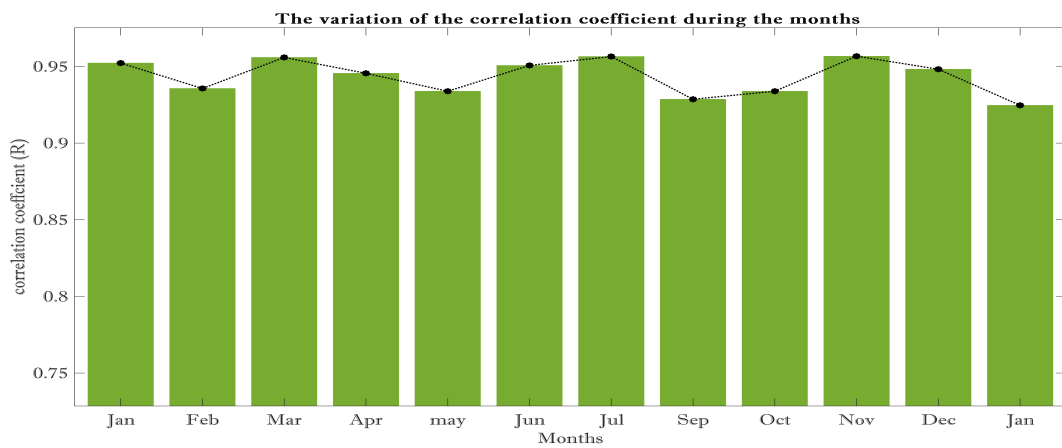
Figure 4.23: The variation of the correlation coefficient during the months

4.3.3.2 PV/ T interdependence study

On the other hand, the correlation between PV output and ambient temperature exhibits variability across different temporal scales. On a daily basis, temperature and PV power often display an inverse relationship during peak irradiance hours, where elevated temperatures reduce module efficiency, resulting in a slight decline in power output despite strong solar input (Figure 4.24). Over a weekly and monthly scale (Figure 4.25 and Figure 4.26), the inverse pattern persists and the correlation remains generally weak. On a yearly basis, Figure (4.27), the trend becomes more nuanced, as seasonal transitions introduce variability in both temperature and solar irradiance. During winter months, lower temperatures may enhance PV efficiency, but reduced solar availability diminishes the overall output. In contrast, heatwaves and summer months offer high irradiance but also impose thermal stress on the modules. As such, Figure 4.28 reflects a complex interaction governed by the interplay between thermal effects and irradiance availability.
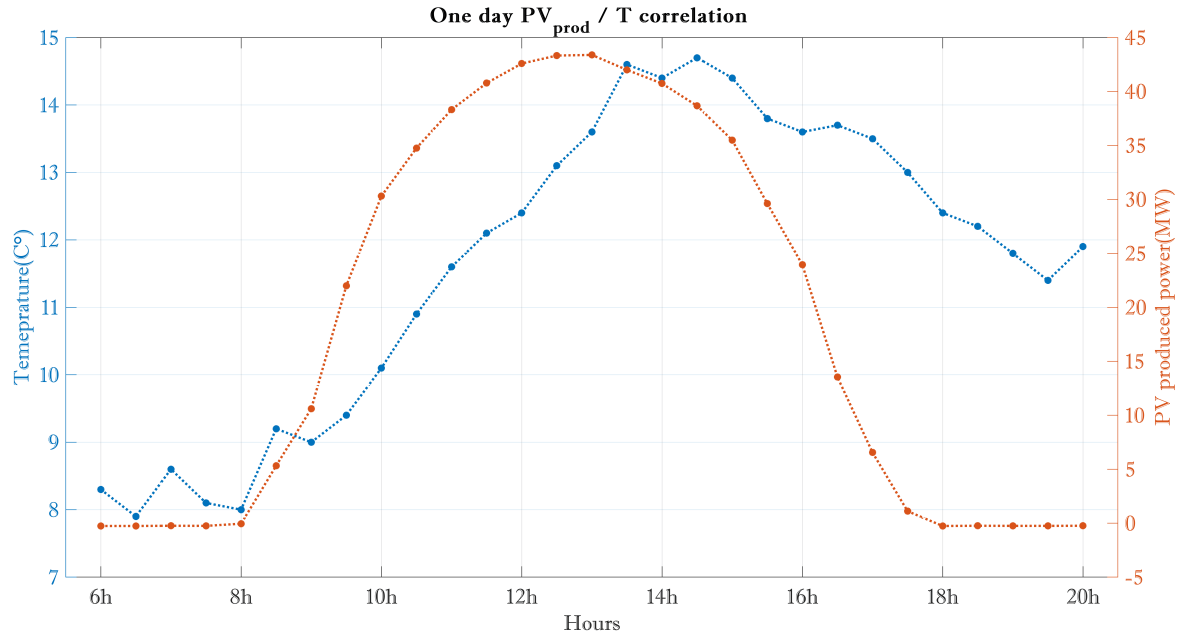
Figure 4.24: PV power correlation with the temperature for one day
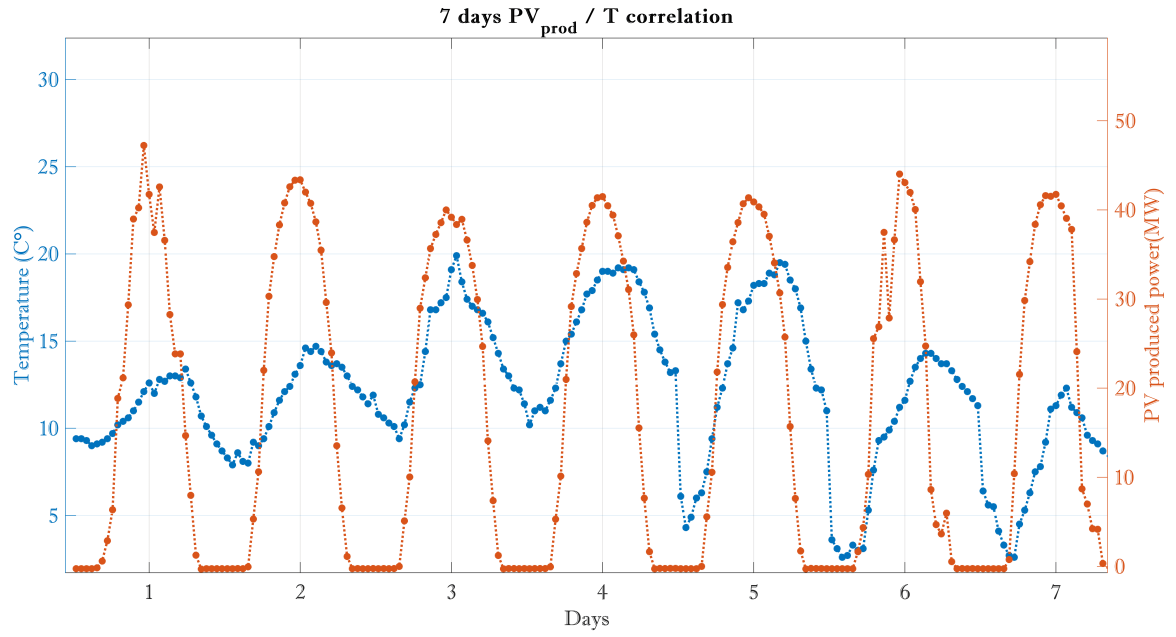
Figure 4.25: PV power correlation with the temperature for seven days
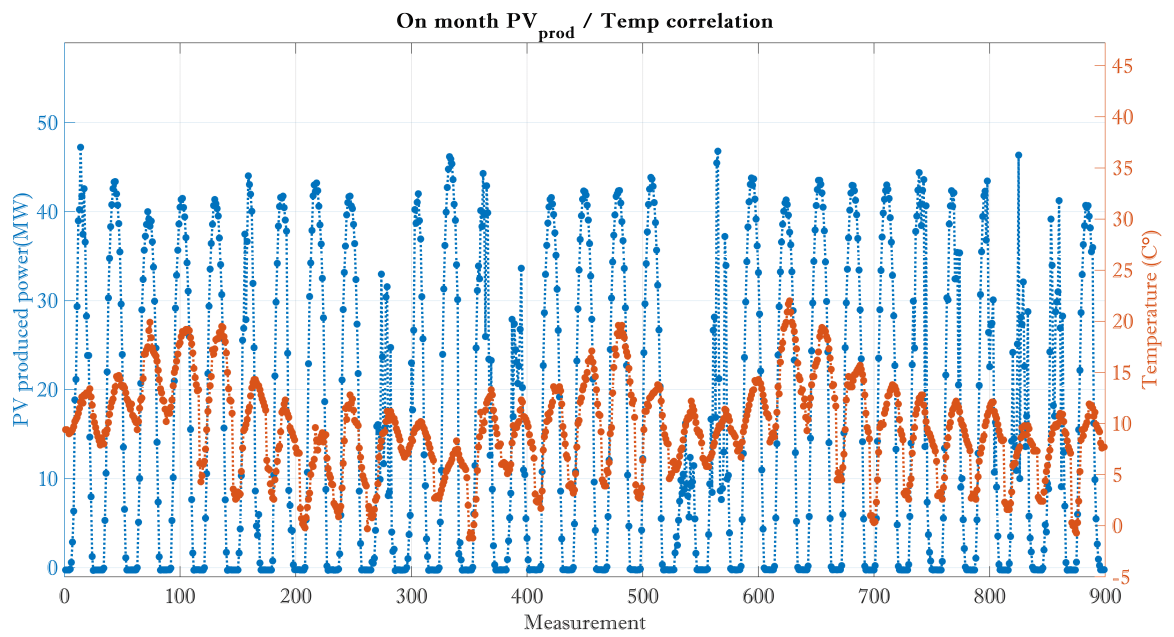
Figure 4.26: PV power correlation with the temperature for one month
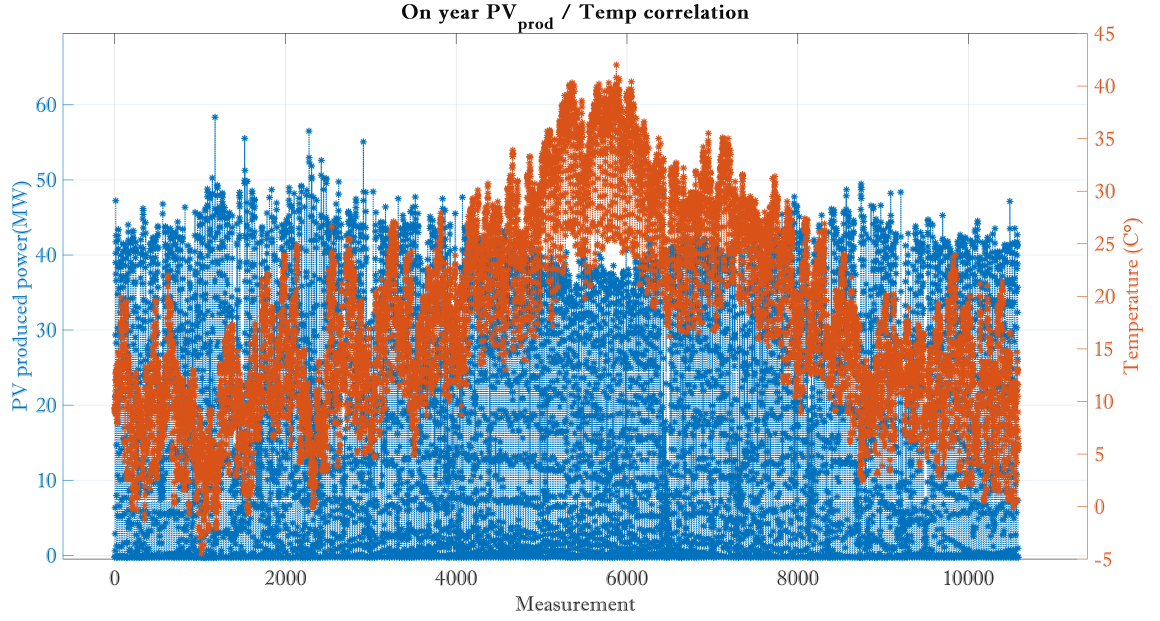
Figure 4.27: PV power correlation with the temperature for one year
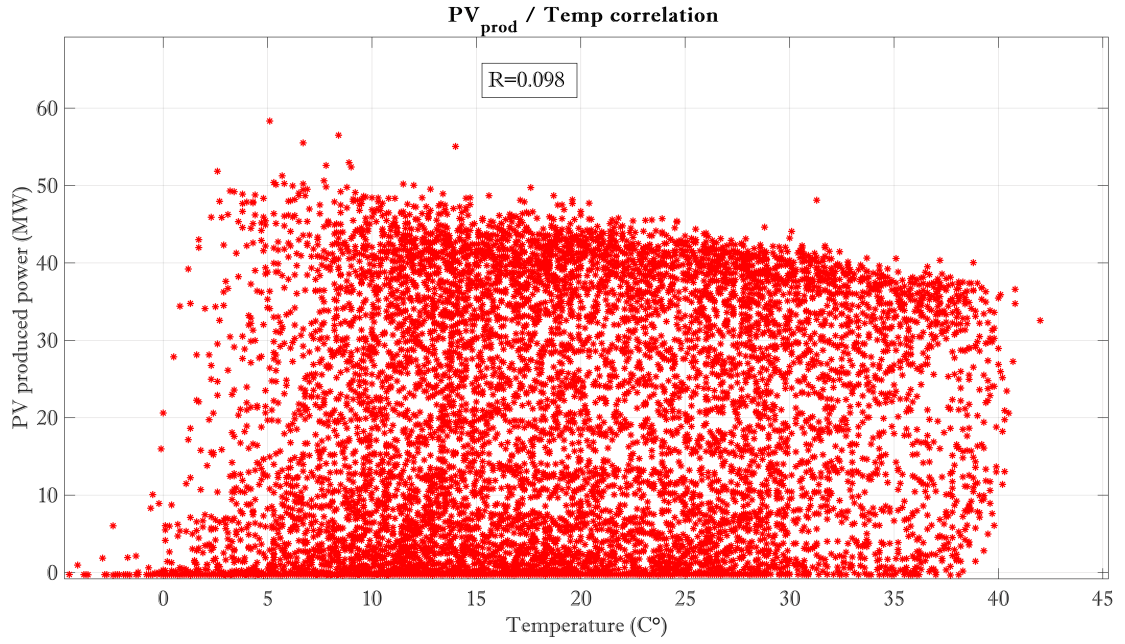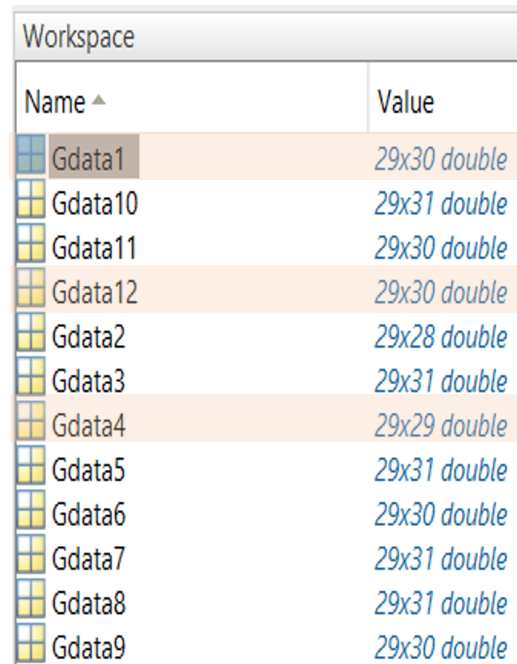
Figure 4.28: PV power correlation with the temperature for two years,

4.3.4 Pre-processing

As the developed model may inherently exhibit biases stemming from the presence of anomalies and outliers (arising from system operational faults such as nonstationarity, spike phenomena, and missing data) these irregularities can significantly compromise the learning process and degrade forecasting performance. Consequently, such faults must be

systematically identified and rigorously addressed, either through correction or elimination, to ensure the reliability and robustness of the predictive modeling framework [54].

During this investigation, the MATLAB-based transformation method described in Section 4.3.2 helped identify missing data in the dataset. For example, the data for January 2018 had a size of 29 × 30, where 29 is the number of daily measurements and 30 is the number of days recorded. Since January normally has 31 days, the expected size should be 29 × 31, meaning that one full day of data was missing. A visual check confirmed that the measurements for January 27th were absent, as shown in Figure 4.29. The missing values constitute approximately 0.55% of the entire dataset, amounting to 21,170 entries over the two years. Given the minimal proportion, no large-scale data reconstruction was undertaken, as the study emphasizes the use of real-time measurements rather than synthesized or interpolated data. Nevertheless, a single missing entry dated December 13th was reconstructed to ensure uninterrupted program execution. This isolated value, representing merely 0.005% of the dataset, was estimated by computing the mean of its two neighboring values.

Workspace	
Name ▲	Value
Gdata1	29x30 double
Gdata10	29x31 double
Gdata11	29x30 double
Gdata12	29x30 double
Gdata2	29x28 double
Gdata3	29x31 double
Gdata4	29x29 double
Gdata5	29x31 double
Gdata6	29x30 double
Gdata7	29x31 double
Gdata8	29x31 double
Gdata9	29x30 double

Figure 4.29: Missing days identification using Matlab

4.3.4.1 Temporal horizon conversion

Recognizing the critical role of intraday forecasting in energy supply systems, the forecasting horizon was rescaled from 30-minute steps to 3-hour steps, as illustrated in Figure 4.30. This modification is intended to align more closely with the operational realities of the PV plant, where power generation is typically evaluated over longer periods to facilitate better planning and grid integration. The shift to 3-hour intervals enhances the ability to assess the intraday capacity of the Djelfa PV plant in a more practical and efficient manner, enabling
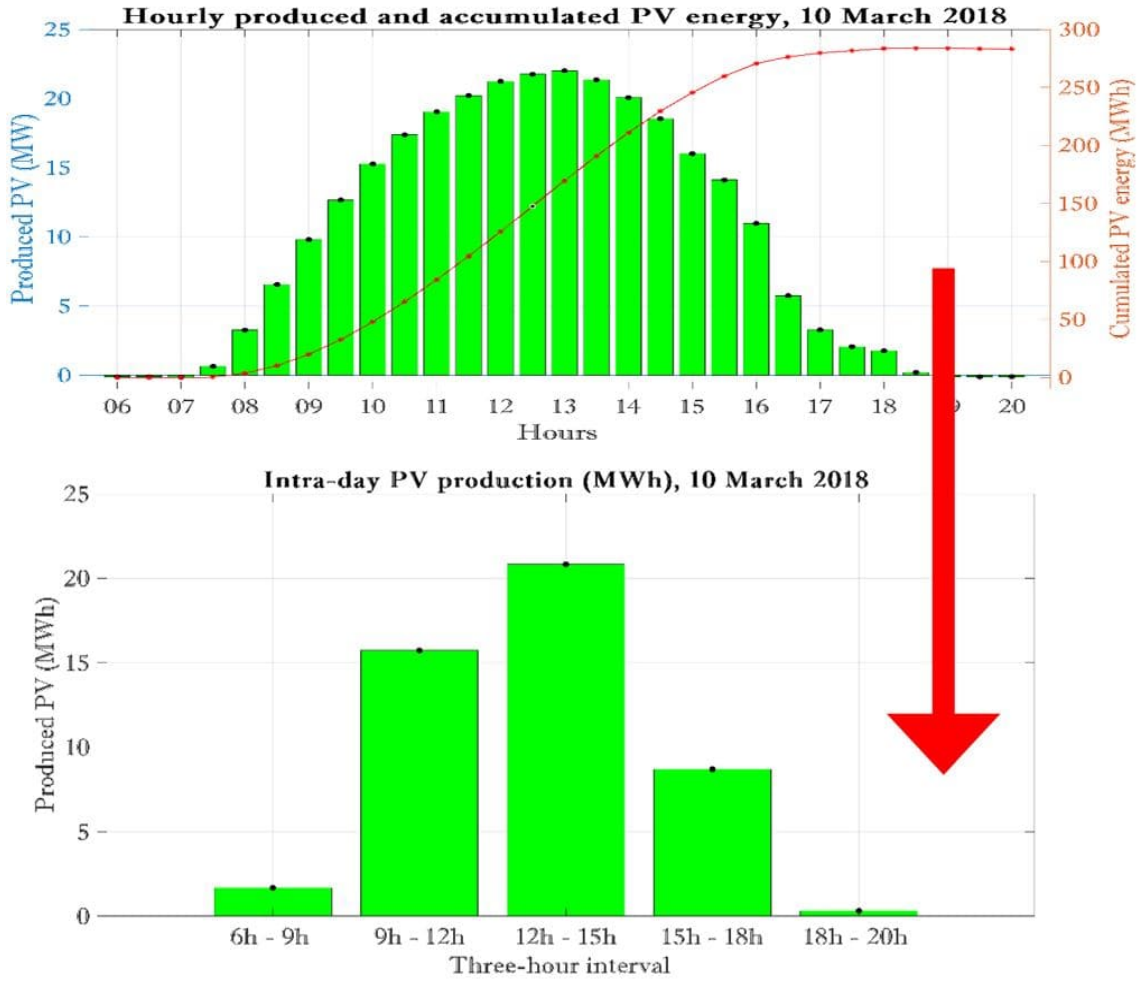
Figure 4.30: Temporal horizon conversion

a more accurate reflection of the plant's energy output in the realm of grid coordination and management. By adapting the temporal resolution in this way, the forecasting process becomes more suited to the operational needs of grid operators, enabling more informed decision-making and advancing the overall effectiveness of power supply systems.

4.3.4.2 Normalization

Discrepancies in input data scaling, attributable to the heterogeneous maximum ranges of features with differing characteristics, may introduce inconsistencies in model training, affecting convergence speed and accuracy. These discrepancies can cause certain features to dominate others, particularly in distance-based algorithms and neural networks, leading to misleading interpretations and suboptimal parameter updates.

These issues can be resolved by applying normalization methods, which standardize input ranges across heterogeneous features and allow the model to learn from all predictors in a balanced and unbiased manner. By transforming the data into a uniform scale, normalization mitigates the influence of extreme values and improves numerical stability, making optimization algorithms more efficient. Moreover, proper normalization enhances the gener-

alizability of the model, particularly when working with datasets containing heterogeneous variables [108].

In this work, the Min–Max scaling approach is employed, as formally expressed in Equation 4.1. This transforms the values of each feature into a predefined interval, commonly within the range [0, 1], preserving the original distribution while eliminating disparities in magnitude across different variables. This method is particularly advantageous for models that rely on activation functions sensitive to input scales, such as neural networks employing sigmoid or tanh functions, where improperly scaled data may result in gradient saturation or vanishing gradients. By applying Min-Max normalization, the dataset is transformed into a more suitable form for predictive modeling, fostering improved convergence behavior and robustness against anomalies.

$$\bar{X}_i = \frac{X_i - X_{\min}}{X_{\max} - X_{\min}} \tag{4.1}$$

Where $\bar{X}_i$ represents the normalized data, X_i indicates original data, and $X_{\min}$ and $X_{\max}$ refer to the minimum and maximum values contained in the dataset.

For all simulations, the data set is divided into two complementary subsets: 80% is allocated for training, while the remaining 20% is reserved for testing, as depicted in 4.31. This division ensures a sufficient amount of data for model learning and evaluation.
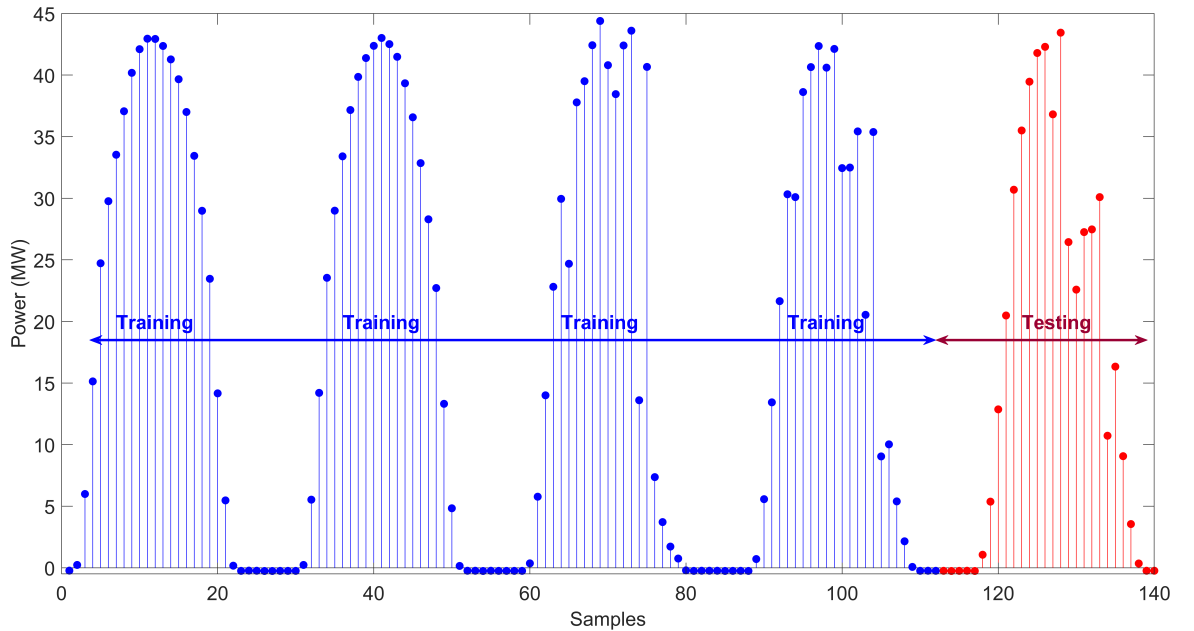
Figure 4.31: Train/Test Ratio

4.4 Models building and parameter tuning

During the development phase of any predictive model, two key considerations must be addressed: what is the minimum amount of data required for the model to function effectively?

and how complex is the model in terms of computational and structural requirements ?

To explore these questions, the minimum amount of input data required for each model to function effectively is first determined using default parameters. A range of temporal sequences, from 1 to 20, was tested, and the results indicated that a sequence length of 4 consistently provided the best balance between simulation time and predictive accuracy across most models. Consequently, A constant time lag of four samples was imposed across all simulations. This means that each prediction (Y_t) is performed using the latest four observations from the input dataset (Y_{t-1}, Y_{t-2}, Y_{t-3}, and Y_{t-4}), ensuring that the models operate with a consistent and relevant historical sequence—an essential factor for accurate forecasting.

The development of a robust predictive model requires meticulous hyperparameter tuning, as performance is influenced by various factors, including training data volume, network architecture, hyperparameter configurations, and optimization algorithms for weight and bias adjustments [64]. To achieve optimal calibration, we conducted an extensive tuning process using the Grid Search algorithm, which systematically explores different hyperparameter combinations to determine the optimal configuration of each model, thereby enhancing predictive accuracy. Key parameters, such as learning rates, the number of layers, the size of batched, and the number of units within the network, were systematically adjusted to realize the highest achievable performance of the model.

This, naturally leads to an important question: should the time delay be determined before hyperparameter tuning, or should the tuning process precede its selection? In reality, this is a trade-off, as both aspects significantly influence the model's overall performance. In our case, we prioritized determining the ideal input sequence first, as PV power forecasting is particularly challenging due to abrupt and sudden fluctuations caused by varying weather conditions, cloud cover, and other external factors. Using the most recent data ensures that the model captures the most up-to-date patterns, making predictions more relevant and responsive to real-world changes. Furthermore, selecting an optimal input sequence length is crucial for balancing model accuracy and computational efficiency. A smaller input sequence serves not only to lower the model's complexity but also improves processing speed by minimizing the number of parameters and computations required during training and inference. This leads to faster convergence, lower memory usage, and enhanced real-time forecasting capabilities, an essential factor when deploying predictive models in energy management systems. By first establishing the optimal sequence length before tuning hyperparameters, we ensured that the grid search process was performed on a well-structured dataset, allowing for more efficient exploration of hyperparameter configurations and further improving the model's predictive performance.

4.4.1 Artificial Neural Networks

Artificial neural networks (ANN) constitute computational frameworks designed to emulate the structural organization and functional mechanisms of the human brain, designed to process and learn from data through interconnected nodes called neurons [109]. These networks are composed of multiple layers: An input layer responsible for acquiring the initial data, followed by one or more latent layers that perform feature abstraction and progressive refinement, and an output layer that generates predictions or classifications [110, 111]. The learning process in ANNs involves two fundamental phases: feedforward and backpropagation [112]. In the feedforward phase, data propagates from the input layer through the hidden layers to the output layer, where computations are performed using weighted connections and activation functions to produce an output [113]. However, the initial weights may not yield optimal results. To improve performance, the backpropagation algorithm is employed, which calculates the error between the predicted and actual values and adjusts the network's weights until a defined criterion is attained [114]. This optimization is typically achieved using gradient descent or its variants, allowing the network to gradually refine its weights and improve prediction accuracy over successive iterations [115]. Through this learning process, ANNs can adapt from examples and effectively handle complex and non-linear problems. A basic ANN architecture for for classification task is depicted in Figure 4.32
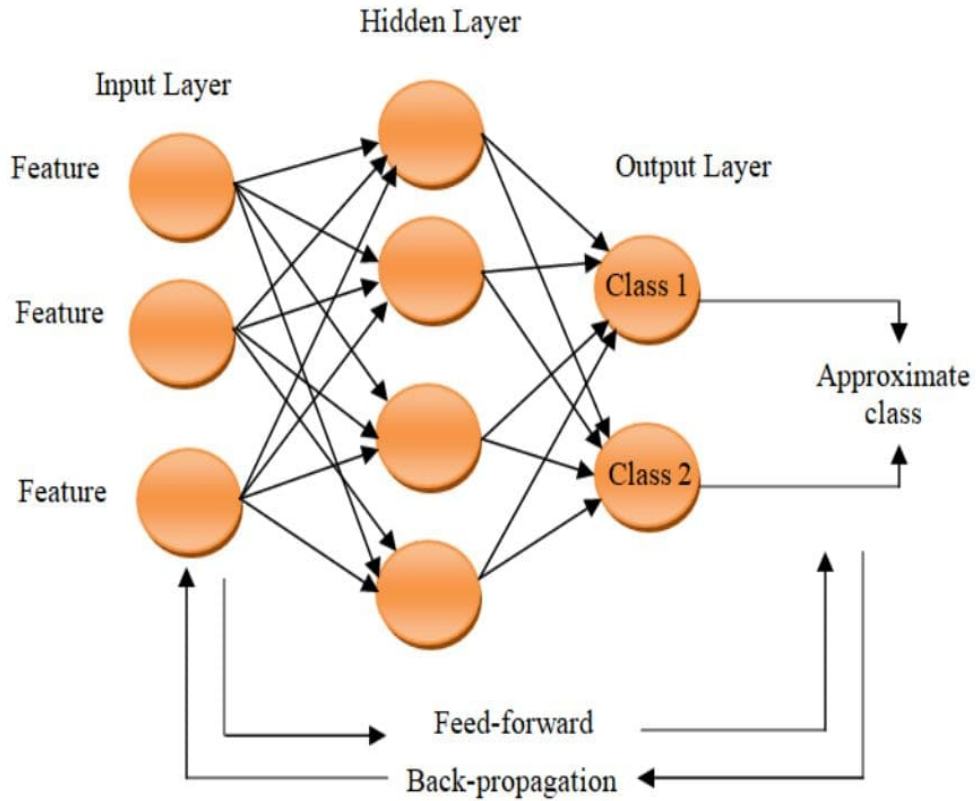
Figure 4.32: ANN architecture for classification task

The tuning parameter process related to RNN in this research focused on varying the number of neurons assigned to each layer (10, 20, 30) and the number of hidden layers (1, 2, 3), while using the Levenberg-Marquardt training algorithm (See Table 4.3). The network followed a feedforward architecture with a ReLU activation function for the hidden layer and a linear activation for the output. The best configuration was determined to be a single hidden layer with 10 neurons.

Tableau 4.3: Hyperparameter tuning step for Neural Networks model

Model	Hyperparameter range	Best preferences
ANN	Neurons number per layer: [10,20,30] Hidden layers number : [1,2,3] Training algorithm: Levenberg-Marquardt Architecture: Feed Forward Hidden Layer activation function: ReLu Output layer: Linear	Neurons number=10; Hidden layers number=1

4.4.2 Long short-term memory networks

Long Short-Term Memory (LSTM), a variant of recurrent neural networks (RNNs), is characterized by an advanced structure that incorporates adaptive weights, internal memory units, and recurrent feedback loops. At the core of the model lies the memory cell, which functions as a long-term storage unit, allowing information to be maintained and transferred across sequential steps. The information flow within these cells is governed by gating mechanisms that control what to store, update, or discard. Unlike conventional RNNs, which struggle with the vanishing gradient issue, LSTMs are capable of sustaining crucial dependencies over extended sequences [116, 117]. The architecture introduces three gates, input, forget, and output (as represented in 4.33) that collectively regulate the learning dynamics. Specifically, the forget gate filters out irrelevant data while preserving useful content for subsequent predictions [118]. The input gate i_t determines the extent to which new information contributes to the current cell state c_t, whereas the output gate o_t modulates how much of the updated state is propagated to the next step in the sequence [119]. These gates operate on values between 0 and 1, thereby functioning as selective regulators of memory. When the forget gate value f_t approaches 1 while the input gate remains near 0, the model preserves long-term memory. Conversely, when these conditions shift, the architecture prioritizes short-term representations [116, 117].
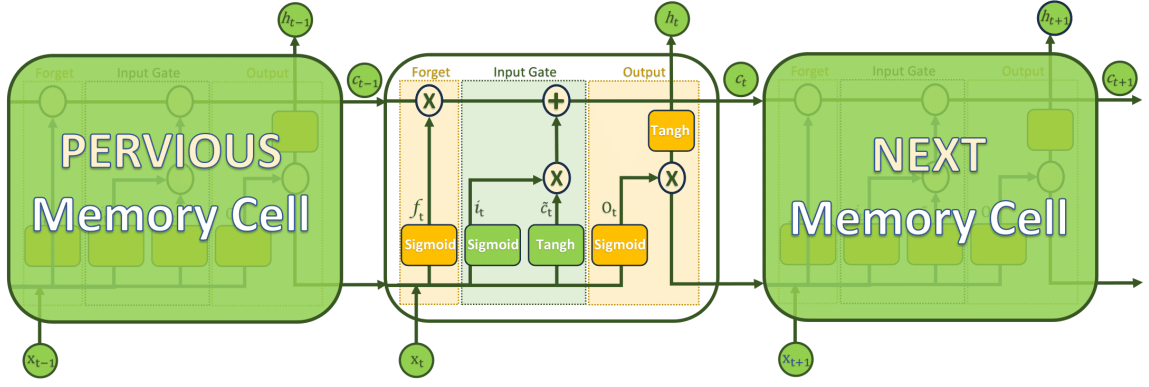
Figure 4.33: Fundamental Structure of an LSTM Network

The LSTM is defined by the following equations [120, 121]:

The gate values

$$f_t = \sigma(W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \tag{4.2}$$

$$i_t = \sigma(W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \tag{4.3}$$

$$o_t = \sigma(W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \tag{4.4}$$

The memory unit:

$$\tilde{C}_t = \tanh(W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \tag{4.5}$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t \tag{4.6}$$

The output unit:

$$h_t = o_t \odot \tanh(C_t) \tag{4.7}$$

The final LSTM Output:

$$y_t = h_t \tag{4.8}$$

The memory cell state at time step t is represented as c_t, while the candidate cell state is denoted by $\tilde{C}_t$. The previous state, c_{t-1}, carries information from the preceding time step. The input vector is indicated as x_t, and the hidden node is given by h_t. The parameter matrices W_i, W_f, W_o, and W_c correspond to the weights applied to the input, forget, output, and candidate cell gates, respectively. Similarly, U_i, U_f, U_o, and U_c denote the recurrent weight matrices associated with the hidden states. The bias terms linked to the gates are expressed as b_i, b_f, and b_o. The non-linear transformations within the network are governed by activation functions, most notably the sigmoid (σ) and the hyperbolic tangent (tanh). In neural networks, an activation function defines the transformation applied at each neuron, enabling non-linear mapping between input and output signals. Among the many available types, the sigmoid, logistic, and hyperbolic tangent functions are the most frequently adopted [122].

4.4.3 Gated recurrent unit

In contrast to the LSTM, which incorporates three separate gating mechanisms, the Gated Recurrent Unit (GRU) employs a more compact design by merging the input and forget gates into a single update gate. Consequently, the GRU architecture operates with only two gates: the update gate and the reset gate. The reset gate determines the extent to which information from the preceding hidden state is ignored, whereas the update gate regulates how much of the new input contributes to the current hidden state. The hidden state is updated as a function of the reset gate, the update gate, and the present input. This reduction in structural complexity allows GRUs to converge more rapidly and alleviates vanishing gradient challenges commonly faced by RNNs [84].

The internal configuration of the GRU is illustrated in Figure 4.34, where h_{t-1} denotes the state transmitted from the previous step, x_t represents the input at time t, and h_t corresponds to the updated hidden output. The variables r_t and z_t indicate the reset and update gates, respectively, while $\tilde{h}_t$ stands for the candidate hidden state. The term $(1 - z_t)$ expresses the complementary contribution of the update gate. The sigmoid activation compresses values into the interval $[0, 1]$, thereby enabling the gating mechanism. When r_t approaches zero, the model prioritizes current input information while disregarding past memory. Conversely, when r_t approaches one, the network retains previous hidden information. The value of z_t, ranging between 0 and 1, controls the balance between retaining historical information and incorporating new data.

The mathematical operations defining GRU units are expressed as follows [84]:

Reset gate:

$$r_t = \sigma(W_{rx}x_t + W_{rh}h_{t-1} + b_r) \tag{4.9}$$

Update gate:

$$z_t = \sigma(W_{zx}x_t + W_{zh}h_{t-1} + b_z) \tag{4.10}$$

Candidate hidden state:

$$\tilde{h}_t = \tanh(W_{hx}x_t + r_t \odot (W_{hh}h_{t-1}) + b_h) \tag{4.11}$$

New hidden state:

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \tag{4.12}$$
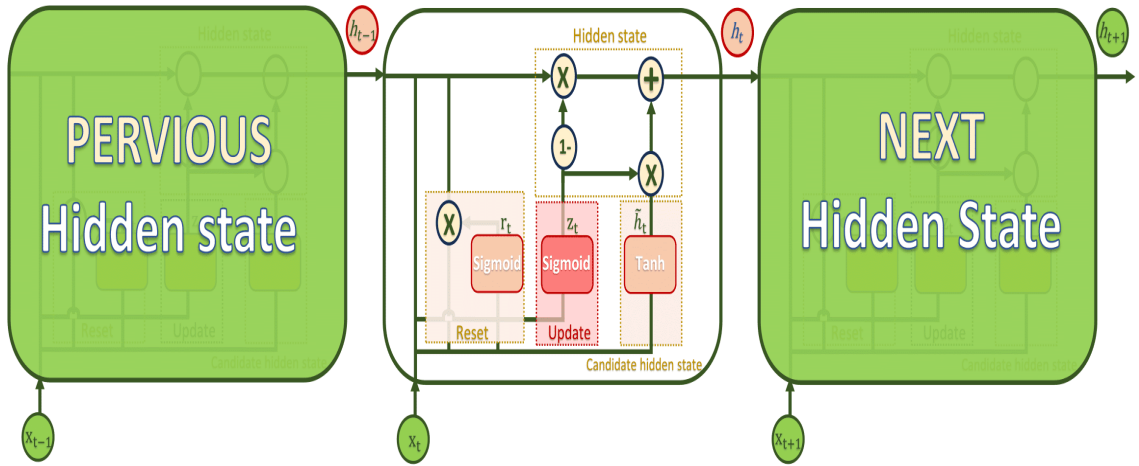
Figure 4.34: Fundamental Structure of a GRU Network

4.4.4 Bidirectional recurrent networks

It is essential to integrate the effects of past and upcoming observations on the current state in order to achieve reliable forecasting outcomes. One effective strategy to address this challenge is the development of RNN models that operate in both forward and backward directions, thereby enabling the capture of intricate temporal relationships spanning previous and subsequent inputs.

Bidirectional Recurrent Neural Networks (BiRNNs), including architectures such as BiL-STM and BiGRU, employ two distinct hidden layers that process information in opposite temporal flows before merging their outputs. Specifically, one layer follows the conventional forward progression of the sequence, while the other traverses it in reverse. This dual processing framework allows the model to simultaneously exploit contextual information from earlier and later time steps. It is important to note that although the forward and backward units operate independently during sequence propagation, the model parameters are optimized jointly during the training process. The schematic representation of a BiRNN structure is provided in Figure 4.35.
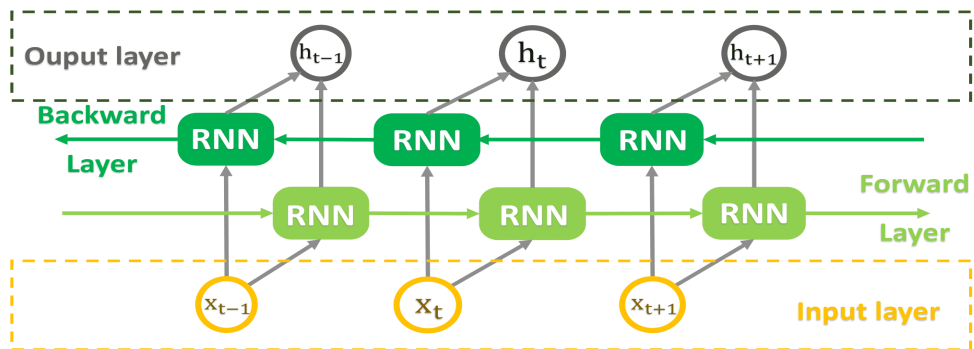
Figure 4.35: Circular Architecture of Bidirectional RNN

$\{x_{t-1}, x_t, x_{t+1}, \ldots, x_n\}$ are the inputs, whereas $\{h_{t-1}, h_t, h_{t+1}, \ldots, h_n\}$ are outputs. To calculate the prediction y_t at time t, the activation function applied is the following:

$$y_t = f(W_y[\overrightarrow{h_t}; \overleftarrow{h_t}] + b_y) \tag{4.13}$$

where:

- W_y is the network's final transformation weight based on the input and volume set.

- $\overrightarrow{h_t}$ and $\overleftarrow{h_t}$ are the forward hidden state and the backward hidden state, respectively.

- $[\overrightarrow{h_t}; \overleftarrow{h_t}]$ are the concatenated hidden states.

- b_y is the bias vector associated with the final transformation layer.

PV power forecasting often requires prediction horizons extending over several hours or even multiple days, which underscores the importance of effectively modeling long-range temporal dependencies. By leveraging both past and forthcoming observations, the forecasting framework is able to deliver reliable outcomes across extended time spans. The bidirectional architecture enhances the model's capacity to discern temporal dynamics, capture seasonal or cyclical patterns, and characterize correlations across time, thereby supporting accurate long-term estimation. Incorporating these temporal attributes significantly strengthens both the accuracy and robustness of PV power forecasting, contributing to advancements in this research field [123].

4.4.5 Recurrent neural network calibration

Table 4.4 outlines the process of hyperparameter tuning for models including LSTM, BiLSTM, GRU, and BiGRU. The tuning process explored different hidden unit sizes (50, 100, 150, 200) and epoch values (50, 100, 200). The training utilized the Adam optimizer with an initial learning rate of 0.01, a dropout rate of 0.2, and a mini-batch size of 64. The optimal configuration was found to be 200 hidden units and 100 epochs. Similarly, GRU and BiGRU models were optimized using the same range of hidden units and epochs. The final choice included a single-layer GRU configuration. This tuning process ensured that each model was adjusted for optimal performance based on the given dataset.

Tableau 4.4: Hyperparameter tuning step for Neural Networks model

Model	Hyperparameter range	Best preferences
LSTM/BiLSTM	Hidden Units= [200, 150, 100, 50] Epochs grid = [200, 100, 100] Training algorithm: Adam Dropout: 0.2 Mini Batch Size: 64 Initial Learning Rate: 0.01	Hidden Unit=200 Max Epochs : 100
GRU/BiGRU	Hidden Unit= [200, 150, 100, 50] Epochs grid = [200, 100, 50] Training algorithm: Adam Mini Batch Size: 64 Initial Learn Rate: 0.01 GruLayer : 1	

4.4.6 Support Vector Machine

Support Vector Machines (SVMs) are predominantly utilized for supervised learning tasks, particularly in binary classification. The method constructs a decision boundary—referred to as a hyperplane—in a transformed feature space to separate data instances belonging to different classes. The optimal hyperplane is determined by maximizing the margin, which represents the minimal distance between the boundary and the nearest training samples from either class [124]. Building upon this foundation, Support Vector Regression (SVR) adapts the same theoretical framework to handle regression problems, where the objective is to approximate continuous target values. SVR preserves the concepts of margin maximization and kernel-based transformations to effectively capture nonlinear relationships [125]. Its mathematical formulation is expressed as [126]:

$$f(x) = w^T \cdot \varphi(x) + b \tag{4.14}$$

where:

- $\varphi(x)$ Serves as kernel function

- x Acts as feature vector input.

- w^T Expresses weight vector w in transposed form.

- b Corresponds to bias element.

- $f(x)$ is the desired output.

Within the framework of SVR, kernel functions define the mapping mechanism that projects the original input data into a higher-dimensional feature domain. Such a projection facilitates the modeling of nonlinear dependencies between explanatory variables and the

response variable. Commonly employed kernel types include linear, polynomial, and radial basis function (RBF) kernels [125, 127, 128]. Based on preliminary evaluation, this study adopts the RBF kernel, whose mathematical formulation is presented as follows:

$$k(x, x_i) = \exp(-\gamma \|x - x_i\|^2) \tag{4.15}$$

where:

- $\gamma = \frac{1}{2\sigma^2}$ represents the kernel coefficient, also called the regularization parameter.

- The reduced γ parameters extend similarity across broader ranges, whereas higher values emphasize proximity among closely located data points.

Hyperparameter tuning in this case involved selecting the appropriate kernel type, regularization parameter gamma (γ), and kernel scale parameter sigma (σ) (See Table 4.5. The RBF kernel was chosen due to its potential to capture non-linear patterns. The search considered different values for γ {0.1, 1, 10} and σ {0.1, 1, 10} to balance model complexity and generalization. After evaluation, the optimal configuration was found to be $\gamma = 0.1$ and $\sigma = 1$, ensuring an effective trade-off between model flexibility and predictive accuracy.

Tableau 4.5: Hyperparameter tuning

Model	Hyperparameter range	Best preferences
SVM	Kernel type: RBF γ values: [0.1, 1, 10] σ values: [0.1, 1, 10]	γ values=0.1 σ values=1

4.4.7 Random Forest

Random Forest (RF) regression constitutes an ensemble-based methodology designed for addressing regression problems, such as PV power forecasting. The approach relies on generating a collection of decision trees during the training phase and subsequently producing predictions by averaging the outcomes of these trees. In each terminal node, the assigned response (R_i) is derived either from the dominant class in classification tasks or from the mean value in regression scenarios. The output of an individual tree for an input X is expressed as $T_i(X)$ [129, 130]:

$$T_i(X) = R_i \tag{4.16}$$

The minimizing algorithm aiming to decrease variance while preserving low bias is given by:

$$MSE(D) = \frac{1}{|D|} \sum_{i \in D} (y_i - \bar{y}_d)^2 \tag{4.17}$$

Where $|D|$ is the instance number in node D, y_i is the target for instance i, and $\bar{y}_d$ is the average target value.

Table 4.6 presents the hyperparameter tuning process for RF model. The optimization focused on selecting the number of trees, minimum leaf size, and validation method to enhance predictive accuracy. The model was implemented using the TreeBagger function for regression, with the number of trees tested at $\{100, 200, 300\}$ and the minimum leaf size varied across $\{1, 5, 10\}$. A 5-fold cross-validation is applied to maintain robustness. Based on the evaluation, the best configuration was identified as 300 trees and a minimum leaf size of 10, striking a balance between performance and computational efficiency.

Tableau 4.6: Hyperparameter tuning step

Model	Hyperparameter range	Best preferences
RF	Function: Tree Bagger N° Trees = [100, 200, 300]; Min. Leaf Size = [1, 5, 10]; Method: Regression k-folds= 5	N° of Trees = 300; Min. Leaf Size = 10;

4.4.8 Convolutional Neural Networks

Convolutional Neural Networks (CNNs) are modeled after the cognitive processes underlying human visual recognition. While two-dimensional CNNs demonstrate outstanding performance in fields such as image classification, face recognition, handwriting analysis, and natural language processing [131], one-dimensional CNNs are increasingly applied in regression tasks owing to their capacity for automated feature extraction [131, 132]. The general structures of 1D, 2D, and 3D CNNs are illustrated in Figure 4.36 [133]. In PV power prediction, CNNs are frequently utilized to capture spatial dependencies embedded in the dataset. A conventional CNN is composed of an input layer, one or more convolutional layers, pooling layers, and an output layer [134]. The convolutional layer functions as the main feature extractor, generating new representations of the input by calculating dot products between the input matrix and a convolutional kernel that systematically slides over its width and height [135], as depicted in Figure 4.37 [136].

In the context of multivariate time series, vertical sliding across height has limited value since dependencies between variables at the same temporal index are relatively sparse. Consequently, a one-dimensional convolutional kernel that moves exclusively along the temporal axis is adopted [137]. The output of this operation is formally defined as:

$$y = x * w, \quad y[i] = \sum_{k=-\infty}^{+\infty} x[i+k]w[k] \tag{4.18}$$

where x is the input vector, and w refers to a filter or kernel.
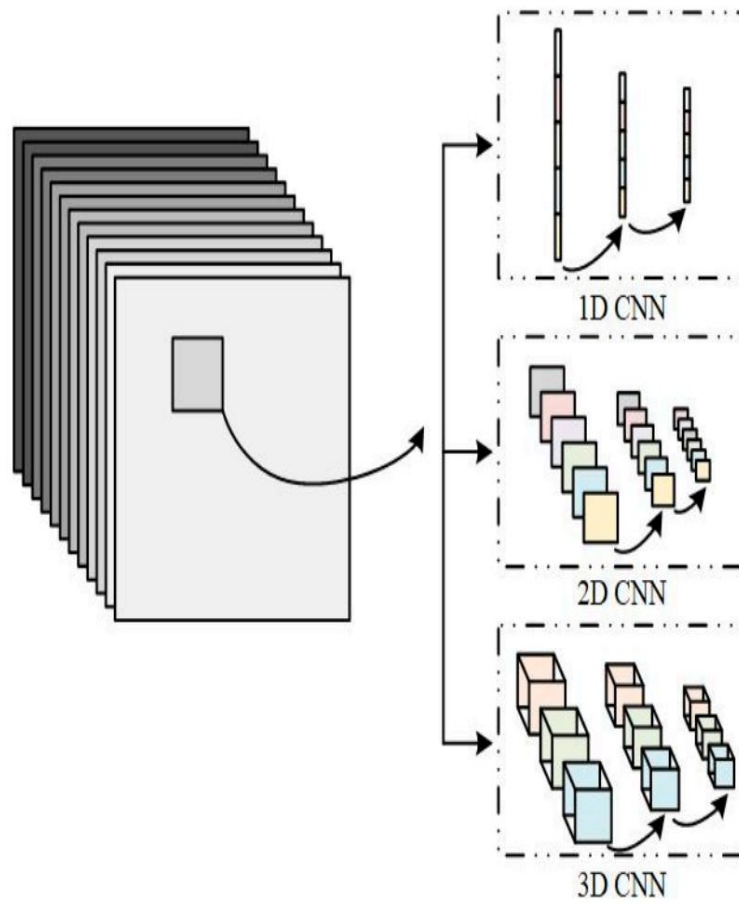
Figure 4.36: Structures of 1D, 2D, and 3D CNNs
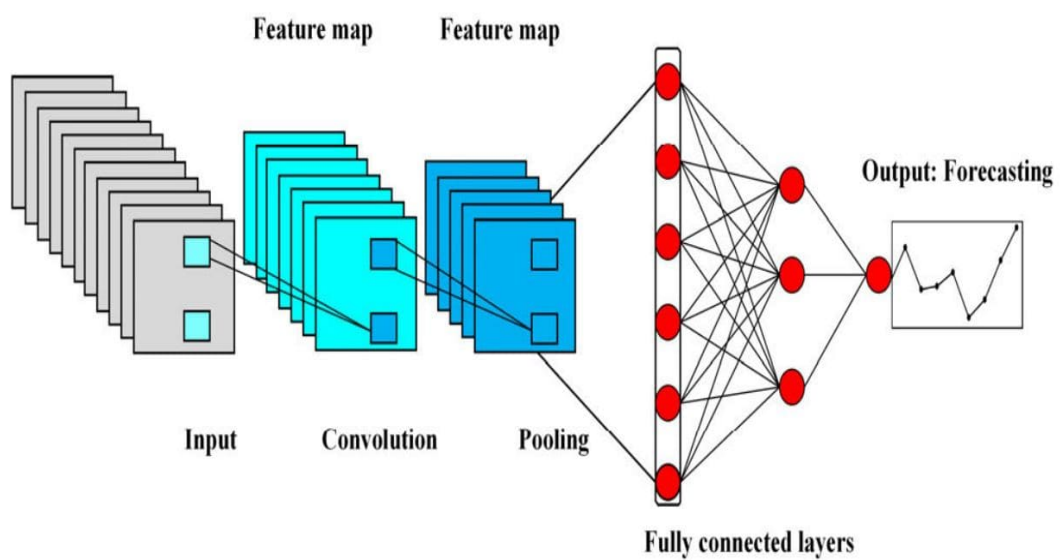
Figure 4.37: Schematic diagram of a basic CNN

Modern deep learning approaches rely on the principle that increasing the depth of convolutional architectures enhances performance by enabling more abstract and discriminative feature representations [138, 139]. However, the training of such deep neural models is often hindered by notable difficulties. Among these, the vanishing gradient problem is particularly critical, arising when gradient signals progressively weaken as they propagate backward through earlier layers. This attenuation can cause gradients to shrink toward negligible values, ultimately resulting in network degradation and susceptibility to overfitting [138, 140]. To mitigate these limitations, advanced architectural strategies such as residual connections and inception modules have been proposed.

4.4.8.1 Residual Modules

The Residual Network (ResNet) was initially proposed in 2016 by Kaiming et al. [141], introducing a paradigm-shifting concept in deep learning architectures. The central idea consists of integrating shortcut connections that allow outputs from earlier layers to be directly forwarded to deeper layers without alteration. These residual links promote more efficient information flow across the network, ensuring that essential features are preserved and propagated. As a result, feature maps are combined more effectively, enhancing predictive accuracy [142]. Moreover, residual pathways contribute to faster convergence and reduced training time [139].

As shown in Figure 4.38(a), ResNet employs two main shortcut module designs. The Identity block bypasses additional convolutions, preserving identical dimensions between input and output. In contrast, the Convolutional block introduces a convolutional layer and batch normalization within the shortcut to adjust dimensional mismatches. When the residual mapping approaches zero, the block approximates an identity transformation, thereby maintaining the integrity of the original information through a direct connection between input and output. This mechanism mitigates performance degradation [143] and simplifies optimization challenges [138].

In practice, residual mapping rarely disappears entirely, allowing the network to iteratively refine residual signals [138, 144]. The corresponding residual formulation can be expressed as:

$$f(x) = h(x) - x \qquad (4.19)$$

where $h(x)$ is the original mapping, x is the input feature, and f is the residual function. Thus, the original mapping can be rewritten as $F(x) + x$.

4.4.8.2 Inception Modules

The Inception architecture introduces the possibility of applying convolutional filters of varying sizes simultaneously, thereby improving representational power without substantially

increasing network depth. This design contrasts with the conventional strategy of deepening CNN models by stacking additional convolutional layers [139]. A key strength of the inception module lies in its ability to capture features across multiple scales and to process information through receptive fields of different dimensions, making it a valuable complement to traditional CNN structures. Within each inception unit, filters are organized in parallel pathways rather than being sequentially connected, allowing diverse feature extractions to occur concurrently [139]. The original inception block incorporates three distinct convolutional kernels, each responsible for extracting features at different scales. This configuration strengthens model robustness and considerably enhances feature representation in time series analysis [145]. For the purposes of this study, the Inception V1 variant is employed, as illustrated in Figure 4.38(b).
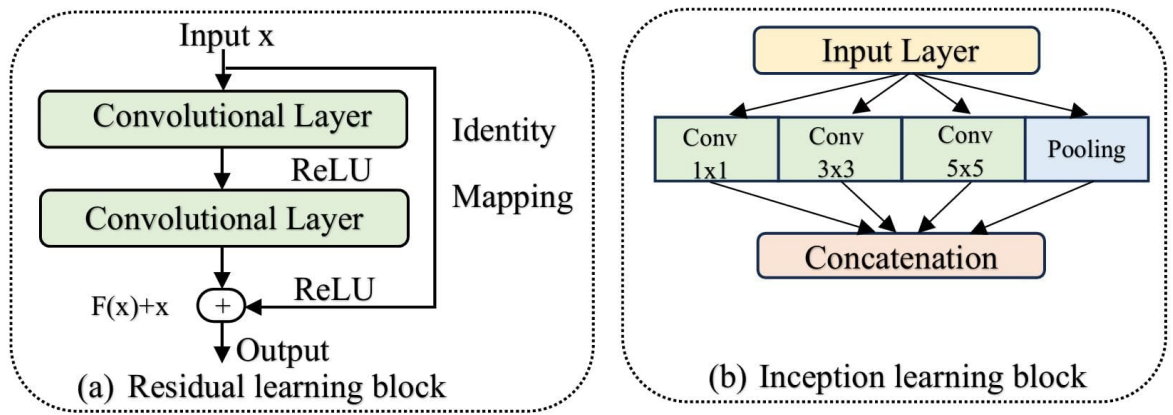
Figure 4.38: Residual and learning modules

An important aspect of the preprocessing stage concerns the structural requirements of CNN-based models. Even when applying one-dimensional convolutions, the input must be reformatted into a three-dimensional tensor to satisfy the architecture's specifications. In this representation, the three axes correspond to the number of samples, the feature dimension (reflecting historical observations or lag values), and the channel dimension, which for 1D CNNs is typically set to one, indicating a single input stream. Such reshaping is critical because CNNs are inherently designed to operate on 3D inputs, thereby enabling efficient extraction of spatial or temporal dependencies within the data. This transformation guarantees that input sequences are appropriately structured, ensuring consistency with the model architecture and facilitating effective learning during both training and evaluation stages of the forecasting process.

4.4.9 Variational Mode Decomposition

A detailed investigation of PV power variability requires the decomposition of time series signals into more tractable and interpretable components [146]. In 2014, Dragomiretskiy and Zosso introduced Variational Mode Decomposition (VMD) [147], a powerful technique

specifically designed for the analysis of nonlinear and nonstationary data. VMD employs the Alternating Direction Method of Multipliers (ADMM) as its optimization backbone, which enhances the decomposition procedure and ensures greater precision and robustness in signal representation. Through this method, the original time series $f(t)$ is separated into K intrinsic mode functions (IMFs), denoted as $u_k(t)$. Each component $u_k(t)$ corresponds to a distinct spectral band, thereby isolating characteristic oscillatory modes of the signal [148].

The estimation of the bandwidth for each intrinsic mode function (IMF) is carried out through a systematic procedure [149, 150]. First, the Hilbert transform is applied to obtain the spectral representation of each u_k. This spectrum is then shifted to the baseband by aligning it with a corresponding estimated center frequency. Subsequently, the bandwidth is quantified by assessing the Gaussian smoothness of the demodulated signal, computed from the squared gradient of its parametric form. The central objective of VMD is to minimize the individual mode bandwidths while ensuring that their collective sum provides an accurate reconstruction of the original input signal. This optimization task is formulated as a constrained variational problem [147], mathematically expressed in the following equation:

$$\min_{\{u_k\},\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \quad \text{subject to} \quad \sum_k u_k(t) = f(t) \quad (4.20)$$

where t represents time, ω_k denotes the central frequency of the kth mode, $f(t)$ is the input signal, and $\delta(t)$ refers to the unit pulse function.

In order to incorporate the constraint into the optimization process, an augmented Lagrangian approach is adopted. This involves introducing a Lagrange multiplier λ, which enables the constraint to be embedded directly within the objective function L [151].

$$
\begin{aligned}
L(\{u_k\}, \{\omega_k\}, \lambda) = {} & \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \\
& + \left\| f(t) - \sum_k u_k(t) \right\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle
\end{aligned}
\quad (4.21)
$$

where α is a quadratic penalty parameter ensuring the fidelity of the signal recovery [151, 152].

An initial test was performed to evaluate different numbers of IMFs, ranging from 1 to 20. Based on the results and a comparison with the literature, five IMFs were found to be sufficient to achieve maximum efficiency without significantly increasing computational time, particularly in the 30-minute forecasting scenario, where the dataset size increases fivefold from 5,288 to 26,440. The extracted IMFs subsequently serve as input features for diverse forecasting models.
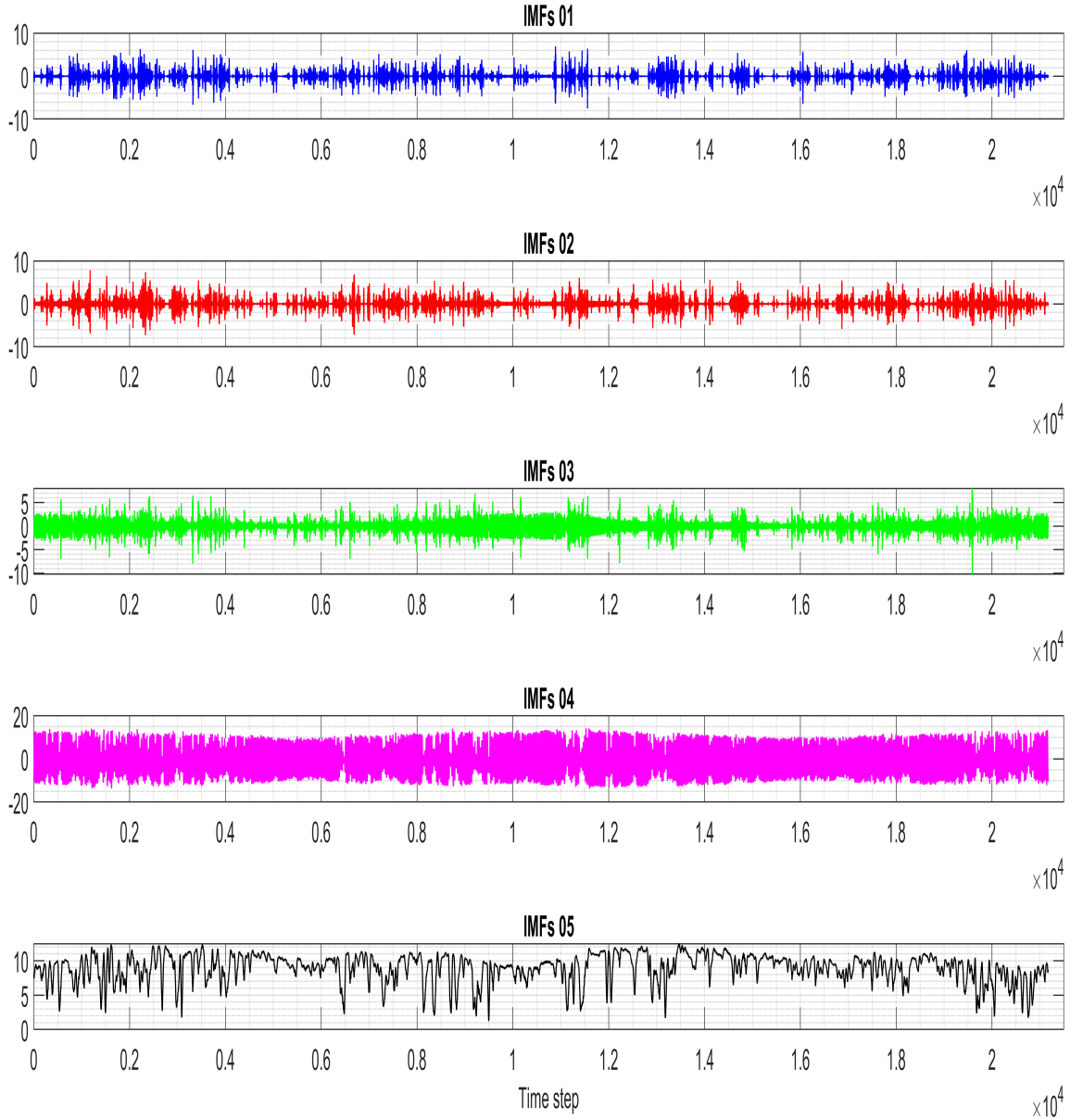
Figure 4.39: The segments produced through VMD

4.5   Accuracy analysis

The final stage of this study focuses on validating the developed model for PV power forecasting to ensure its accuracy and effectiveness. This validation is performed using key performance indicators outlined in Section 2.3.6, namely MAE, nMAE (%), RMSE, nRMSE (%), and R^2. By analyzing these metrics, we assess the model's predictive capabilities, quantify errors, and evaluate its effectiveness in practical forecasting scenarios. This step is essential for verifying the effectiveness of the model before its deployment in real-world energy management applications.

4.6 Results and Discussions

Since the value of any research lies in the depth and relevance of its findings, this section reports the outcomes of our investigation, forming the core of this study and highlighting its major contributions and findings. The results are structured around three fundamental aspects, each yielding important indications about the operational behavior of the models.

The first aspect entails a comparative analysis of stand-alone models, systematically evaluating their predictive performance. This step is crucial for identifying the strengths and weaknesses of different models before incorporating advanced preprocessing techniques.

The second aspect involves the integration of VMD and a detailed examination of its impact on forecasting accuracy. By decomposing the initial time series signal into characteristic IMF elements, VMD enhances feature extraction, reduces noise, and improves the performance of predictive models.

The last aspect, which holds particular significance, consists of a comprehensive seasonal assessment of model performance. This analysis explores temporal variations in forecast accuracy, providing a deeper understanding of how meteorological variables and seasonal trends affect the process.

By examining these factors, the research is directed toward enhancing the interpretability and effectiveness of the predictive models, ultimately contributing to more effective energy management and improved grid stability.

4.6.1 Part 01: Stand-alone model performance analysis

Following the tuning of hyperparameters, two forecasting models were constructed for each algorithm: one dedicated to the short-term horizon (30 minutes) and the other to the medium-term horizon (3 hours). This dual design makes it possible to investigate how different training data volumes influence prediction accuracy and to assess the stability of models across varying temporal horizons. The evaluation of predictive performance relies on statistical indicators computed from normalized datasets. Normalization ensures that results are directly comparable, allowing researchers to benchmark models across diverse contexts such as different plant locations, regional conditions, or capacity scales (kW, MW, GW).

In the case of the **30-minute horizon**, the results in Table 4.7 show that the CNN achieved the best predictive outcomes, characterized by an R^2 of 0.9678, together with the lowest RMSE (0.0471) and nRMSE (0.0545). This superior accuracy highlights the CNN's ability to automatically capture hierarchical patterns within the data through its convolutional filters, which excel at identifying localized dependencies and correlations essential for modeling the complex variability of PV power signals [153, 154]. Meanwhile, the BiLSTM model also demonstrated competitive accuracy, attaining an R^2 of 0.9631. Its bidirectional structure allows simultaneous processing of past and future sequence information, enabling it to uncover temporal relationships more effectively and deliver accurate forecasts [155].

Tableau 4.7: Stand-alone results: 30 min ahead

Horizon	30 min				
Metrics	MAE	nMAE (%)	RMSE	nRMSE (%)	R^2
LSTM	0.0621	7.180	0.0934	10.82	0.8734
BiLSTM	0.0278	3.220	0.0504	5.840	0.9631
GRU	0.0567	6.560	0.0917	10.61	0.8780
BiGRU	0.0498	5.760	0.0878	10.16	0.8882
CNN	0.0251	2.900	0.0471	5.450	0.9678
Resnet	0.0423	4.900	0.0751	8.690	0.9182
Incepnet	0.0364	4.220	0.0573	6.640	0.9523
ANN	0.0529	6.120	0.0963	11.14	0.8656
RF	0.0507	5.870	0.0937	10.85	0.8726
SVR	0.0818	9.470	0.1082	12.52	0.8302

The **Inception module** also delivers strong results, attaining an R^2 of 0.9523. Its effectiveness stems from the ability to process information at multiple scales by applying convolutional filters of varying sizes in parallel [123, 139]. This architectural design enables the model to detect both fine-grained fluctuations and broader temporal patterns within the PV signal, making it particularly suitable for heterogeneous and complex forecasting tasks [156]. The **Residual module**, with an R^2 of 0.9182, benefits from the use of shortcut connections, which alleviate the vanishing gradient problem and preserve the continuity of feature transmission across layers. By maintaining the flow of information, this structure strengthens the capacity to represent long-range dependencies and subtle temporal variations, thereby contributing to its competitive performance [157, 158].

On the other hand, conventional machine learning models such as ANN, SVM, and RF yield noticeably lower accuracy, with R^2 scores of 0.8656, 0.8726, and 0.8302, respectively. Among them, SVR records the weakest predictive outcome (R^2 = 0.8302), underlining its limited ability to generalize and capture the complex nonlinear dynamics that characterize PV power generation. These findings suggest that while traditional methods can extract certain regularities from the data, they fall short in accurately representing the high variability and nonstationary behavior of solar power.

For the **3-hour forecasting horizon**, the processed dataset consists of 725 averaged samples, significantly fewer than the 5,288 observations available for the 30-minute horizon. As presented in Table 4.8, the CNN again demonstrates the strongest predictive capability, reaching an R^2 of 0.9839 with the lowest RMSE (0.0371). Other deep learning models, including BiLSTM, Inception, and Residual networks, also perform consistently well, clearly surpassing the results of traditional machine learning approaches. Their superior outcomes emphasize the advantage of advanced neural architectures in effectively capturing nonlinear structures and temporal dependencies inherent to PV power dynamics.

Tableau 4.8: Stand-alone results : 3 hours ahead

Horizon	3 hours				
Metrics	**MAE**	**nMAE (%)**	**RMSE**	**nRMSE (%)**	R^2
LSTM	0.0760	8.520	0.1058	11.85	0.8687
BiLSTM	0.0362	4.050	0.0520	05.83	0.9682
GRU	0.0710	7.950	0.1000	11.21	0.8826
BiGRU	0.0568	6.360	0.0869	9.730	0.9114
CNN	0.0281	3.150	0.0371	4.150	0.9839
Resnet	0.0469	5.250	0.0645	7.220	0.9512
Incepnet	0.0431	4.820	0.0726	8.140	0.9381
ANN	0.0599	6.710	0.0882	9.880	0.9088
RF	0.0960	5.640	0.0857	9.600	0.9138
SVR	0.0564	6.320	0.1063	11.91	0.8675

Notably, traditional machine learning approaches (ANN, RF, and SVR) exhibit improved accuracy in the extended 3-hour forecasting horizon, with RF achieving the highest performance among them, reflected by an R^2 of 0.9138. The superior outcome of RF can be attributed to its ensemble structure, where the aggregation of multiple decision trees enhances generalization and facilitates the extraction of salient features across a broader predictive window. The ANN model also benefits from its architectural flexibility, which enables it to adapt to less complex temporal dependencies and capture generalized patterns that are more prominent over longer horizons. Likewise, SVR demonstrates stronger predictive accuracy, a result that can be linked to its kernel-based framework, which efficiently models large-scale trends when fine-grained temporal resolution is less critical. Collectively, these findings indicate that machine learning models may achieve comparatively stronger results under scenarios where limited short-term historical information is required.

4.6.2   Part 02: VMD performance exploration

As outlined earlier, the VMD technique is applied to partition the PV power time series into a set of Intrinsic Mode Functions (IMFs). For this investigation, the signal is separated into five individual IMFs, as depicted in Figure 4.39. This section examines the effectiveness of decomposition in providing in-depth signal analysis, facilitating smoother pattern recognition, enhancing data mining, and assessing its impact on forecasting performance.

A preliminary test was performed to evaluate different numbers of IMFs, ranging from 1 to 20. Based on the results and a comparison with the literature, five IMFs were found to be sufficient to achieve maximum efficiency without significantly increasing computational time, particularly in the 30-minute forecasting scenario, where the dataset size increases fivefold—from 5,288 to 26,440. The extracted IMFs are subsequently employed as input features for a range of forecasting models. The corresponding performance outcomes, derived from the decomposed signals, are summarized in Tables 4.9 and 4.10, which report the results

for the 30-minute and 3-hour prediction horizons, respectively.

For the 30-minute forecasting horizon, the integration of VMD markedly enhances the predictive accuracy of the hybrid models when compared with their standalone counterparts (see Table 4.7). By decomposing the signal into intrinsic components, these models gain the ability to disentangle relevant features and uncover hidden structures within the data, thereby capturing intricate temporal behaviors that are often overlooked in the raw series. This refined representation strengthens the models' learning capacity and leads to more reliable forecasts, highlighting the value of VMD as an effective preprocessing strategy for performance enhancement.

Tableau 4.9: VMD based results for 30 min ahead

Horizon	30 min				
Metrics	MAE	nMAE (%)	RMSE	nRMSE (%)	R2
LSTM	0.0281	3.250	0.0434	5.020	0.9727
BiLSTM	0.0243	2.820	0.0373	4.320	0.9798
GRU	0.0153	1.770	0.0251	2.900	0.9909
BiGRU	0.0142	1.650	0.0225	2.600	0.9927
CNN	0.0205	2.370	0.0318	3.680	0.9854
Resnet	0.0291	3.370	0.0414	4.790	0.9752
Incepnet	0.0350	4.050	0.0522	6.050	0.9604
ANN	0.0093	1.08	0.0163	1.890	0.9961
RF	0.0377	4.37	0.0610	7.060	0.9459
SVR	0.0333	3.86	0.0415	4.800	0.9750

While deep learning models initially outperformed traditional machine learning approaches (Table 4.7), the inclusion of VMD preprocessing altered this outcome. In the 30-minute horizon, the ANN model, when combined with VMD, achieved the highest accuracy, with an R^2 of 0.9961, surpassing all other configurations. This was followed by VMD-BiGRU and VMD-GRU, which obtained R^2 values of 0.9927 and 0.9909, respectively. Furthermore, error indicators (RMSE, nRMSE, MAE, and nMAE) consistently decreased under VMD-based frameworks, with the improvements being more pronounced for machine learning models than for deep learning ones.

These results emphasize that machine learning models gain substantial advantages from well-structured feature engineering. In particular, VMD enhances ANN performance by delivering noise-reduced, information-rich inputs that allow the model to generalize more effectively. Although deep learning techniques such as BiGRU and GRU remain highly capable of capturing temporal dependencies, their gains under VMD are relatively moderate, reflecting their inherent reliance on automatic feature extraction. By contrast, the ANN, with its simpler structure, leverages the VMD decomposition to achieve highly efficient learning and accurate forecasting with reduced computational demands.

Overall, these findings highlight the pivotal role of data preprocessing in forecasting tasks.
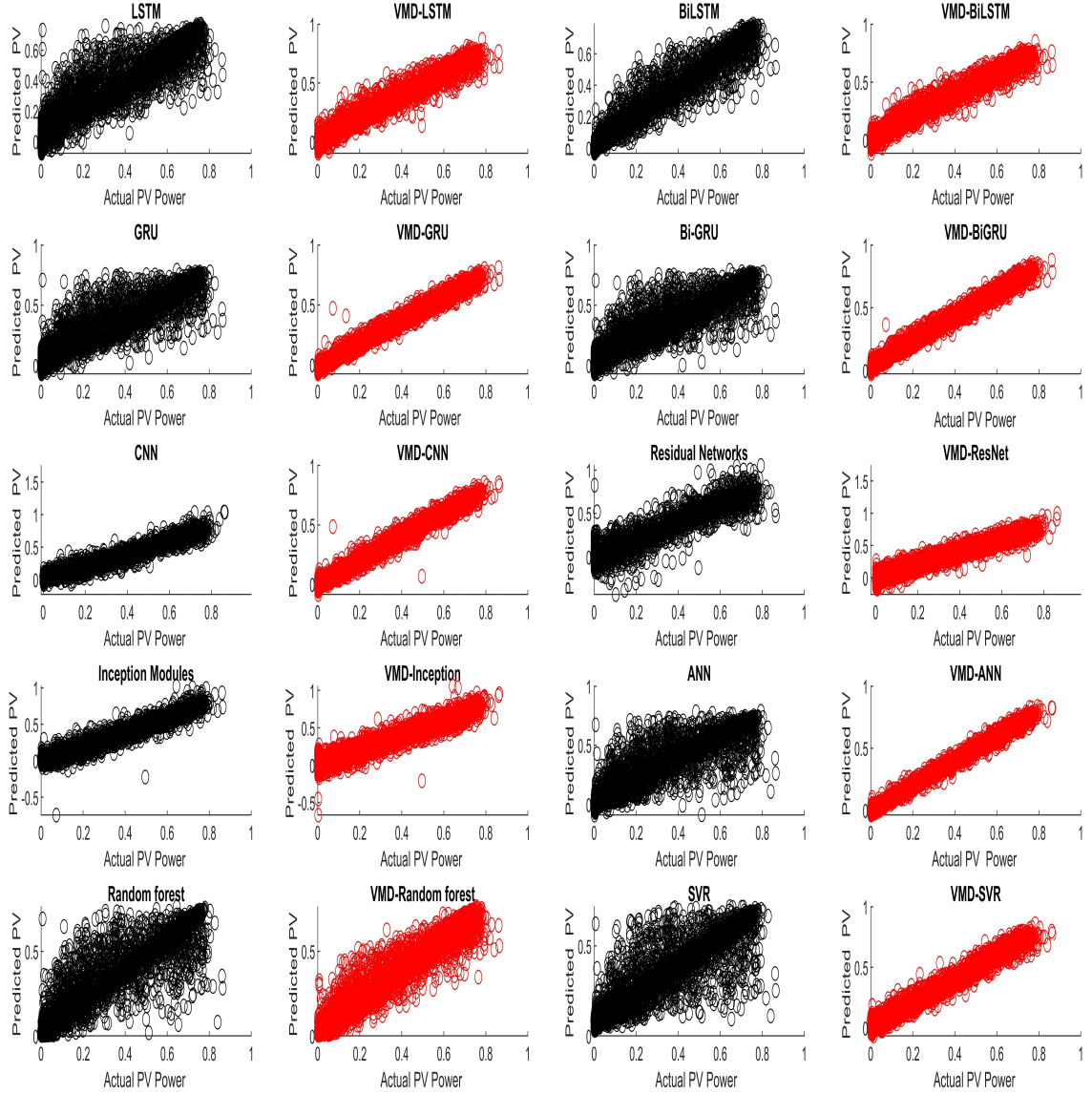
Figure 4.40: Scatter plot for 30min ahead

Careful integration of VMD not only boosts the predictive power of machine learning models but also reduces complexity, reinforcing the importance of aligning preprocessing strategies with model design. The comparative performance is illustrated in the scatter plot shown in Figure 4.40.

Model performance shows substantial gains in the 3-hour ahead forecasting horizon, which relies on a reduced dataset of 725 samples compared to the 5,288 samples used for the 30-minute horizon. The influence of dataset size on forecasting accuracy is particularly striking. Although larger datasets are often assumed to enhance prediction reliability, in this case, the smaller dataset yields superior outcomes. Larger sample sizes can sometimes introduce redundant information, irrelevant variability, or noise that increases the risk of overfitting and complicates the learning process. By contrast, the reduced dataset, once

Tableau 4.10: VMD based results for 3h ahead

Horizon	3 hours				
Metrics	**MAE**	**nMAE (%)**	**RMSE**	**nRMSE (%)**	**R2**
LSTM	0.0272	3.040	0.0367	4.110	0.9842
BiLSTM	0.0286	3.210	0.0389	4.350	0.9822
GRU	0.0161	1.800	0.0218	2.440	0.9944
BiGRU	0.0171	1.910	0.0228	2.550	0.9939
CNN	0.0289	3.240	0.0389	4.360	0.9822
Resnet	0.0356	3.990	0.0518	5.800	0.9686
Incepnet	0.0524	5.700	0.0759	8.500	0.9324
ANN	**0.0090**	**1.100**	**0.0125**	**1.400**	**0.9982**
RF	0.0350	3.920	0.0525	5.880	0.9676
SVR	0.0339	3.800	0.0461	5.160	0.9751

processed through VMD, provides more refined and informative representations, enabling the models to focus on critical features and achieve better generalization. As a result, VMD not only mitigates complexity and computational burden but also contributes to improved forecasting precision.

The scatter plot for the 3-hour horizon (Figure 4.41) underscores the strong effect of VMD-based preprocessing, with performance surpassing that observed in the 30-minute case. The R^2 values are exceptionally high, with the VMD-ANN model achieving near-perfect accuracy ($R^2 = 0.9982$). Deep learning models such as GRU, BiGRU, LSTM, BiLSTM, and CNN also maintain consistently strong performance ($R^2 = 0.9944$, 0.9939, 0.9942, 0.9822, and 0.9822, respectively), demonstrating their resilience in longer prediction horizons when supported by VMD. Error metrics further confirm this trend, as RMSE and nRMSE values decline across all models, with VMD-ANN reaching as low as 0.0125 and 0.0140, respectively. The uniform reduction in errors illustrates how VMD decomposition provides clearer signal representations, facilitating more effective learning. Collectively, these findings affirm that integrating VMD as a preprocessing step significantly strengthens both machine learning and deep learning forecasting models.

Figures 4.42 and 4.43 display the boxplots of absolute errors for the standalone and VMD-enhanced models in both the 30-minute and 3-hour forecasting horizons. The results clearly demonstrate that applying VMD substantially reduces prediction errors, as reflected by the overall shift toward lower values. This effect is particularly evident in the 3-hour horizon, where the combination of a smaller dataset and VMD preprocessing yields markedly improved predictive accuracy. The boxplots further highlight this improvement through tighter interquartile ranges and reduced median values for the VMD-based models, underscoring their enhanced stability and reliability compared to standalone approaches. By decomposing the signal into more tractable components, VMD enables models to concentrate on the most meaningful features, thereby strengthening both accuracy and generalization. The consistent reductions in error across both forecasting horizons reaffirm the critical role of VMD
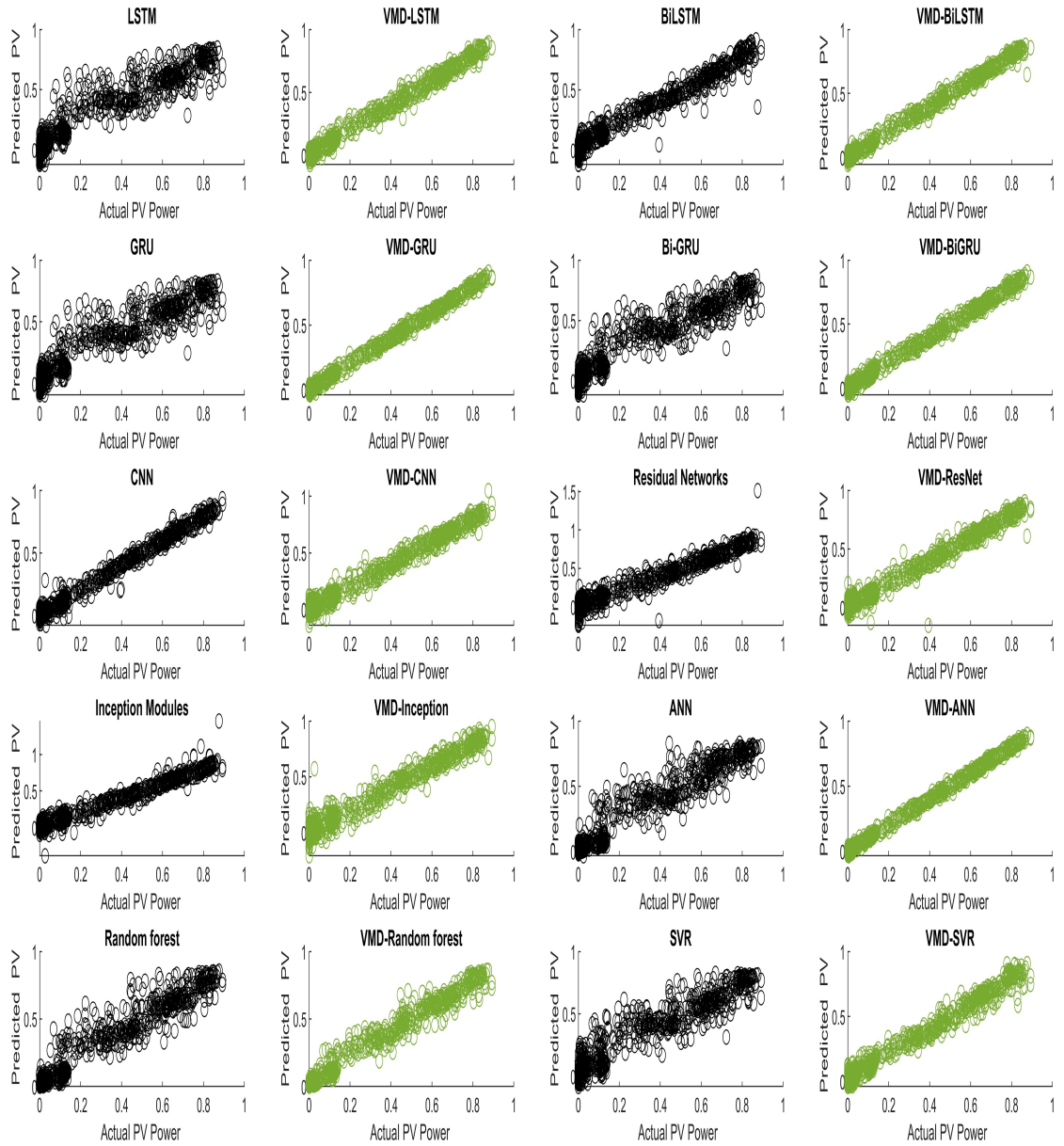
**Figure 4.41:** Scatter plot for 3h ahead forecasting

preprocessing in elevating model performance.

4.6.3 Part 03: Monthly performance assessment

The monthly assessment of forecasting accuracy, illustrated in Figures 4.44, 4.45, and 4.46, demonstrates the effectiveness of VMD-enhanced approaches for photovoltaic power prediction at both the 30-minute and 3-hour horizons. Overall, the VMD framework ensures consistent performance throughout the year, sustaining high determination coefficients and minimal error values despite seasonal fluctuations. This outcome indicates that VMD plays a crucial role in addressing the variability inherent in solar power data across different months.

A closer examination of the monthly distribution identifies June as the period with the most accurate forecasts. For the 3-hour horizon, the R^2 values obtained by VMD-based models range between 0.9834 and 0.9990, reflecting near-perfect predictive capacity. Corresponding RMSE and MAE values remain particularly low, within the intervals 0.0084–0.0342 and 0.0057–0.0253, respectively. Similarly, for the 30-minute horizon, June also exhibits superior results, with R^2 values spanning 0.9850–0.9990, coupled with RMSE and MAE reductions to 0.0077–0.0298 and 0.0239–0.044. These findings suggest that the improved accuracy in June may stem from relatively stable atmospheric conditions and steady solar irradiance, which facilitate more reliable model training and forecasting.

In contrast, November emerges as the least favorable month for prediction. At the 3-hour horizon, the R^2 values decline substantially, ranging from 0.7716 to 0.9970, while RMSE and MAE increase to 0.042–0.1351 and 0.0312–0.1000, respectively. For the 30-minute horizon, the challenge becomes even more apparent, with standalone models achieving only 0.7483–0.9622 in R^2 and producing higher error rates (RMSE: 0.0520–0.1341; MAE: 0.0303–0.1018). Although VMD preprocessing enhances November's results compared to standalone approaches, the month still lags behind others, as reflected in Figures 4.45 and 4.46. The weaker performance can likely be attributed to seasonal factors such as lower solar altitude, shorter daylight duration, and more variable cloud coverage, all of which increase forecasting complexity.
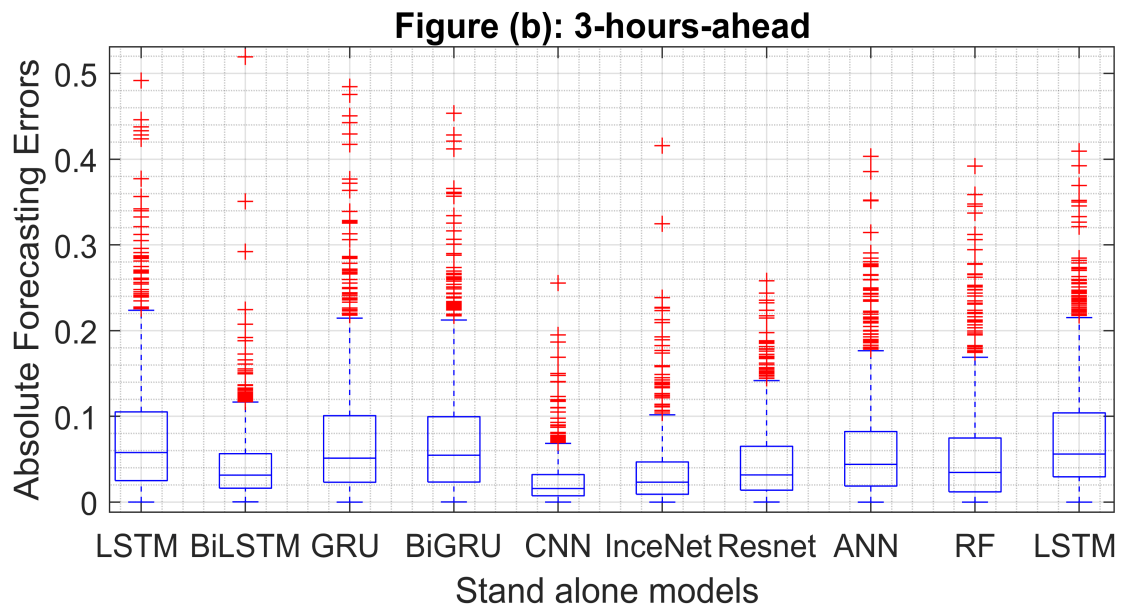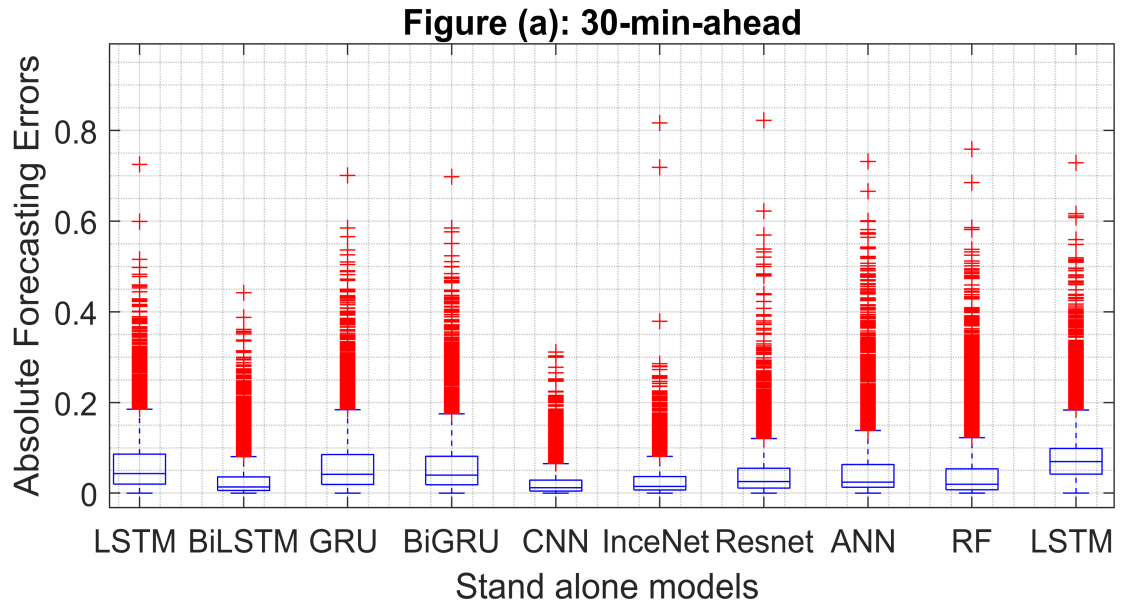
Figure 4.42: Comparative distribution of absolute errors among stand-alone models

Figure (c): 30-min-ahead

Figure (d): 3-hours-ahead
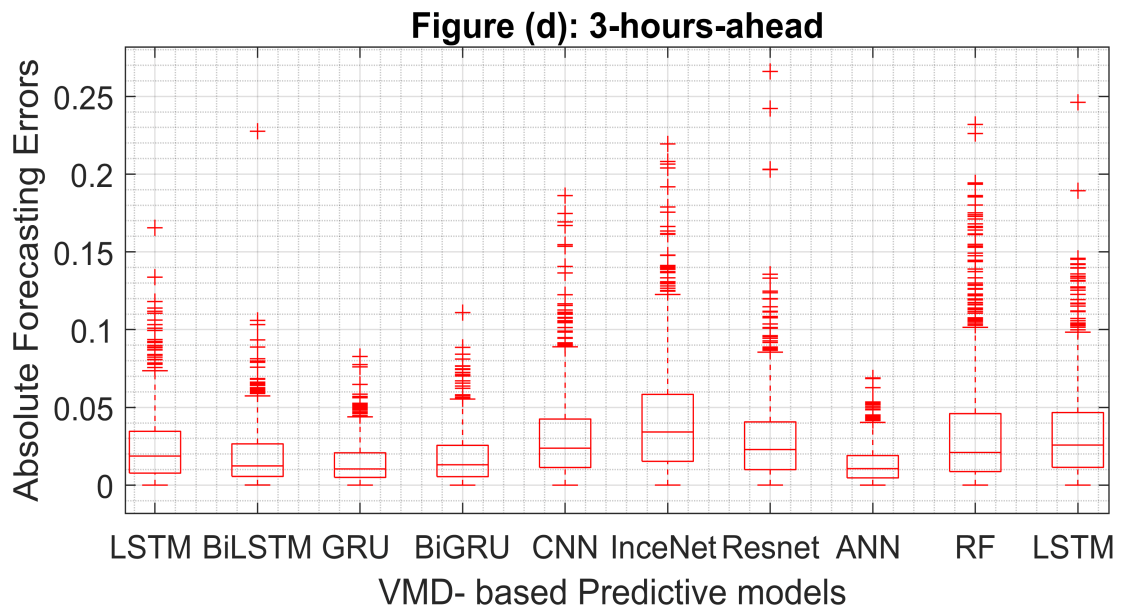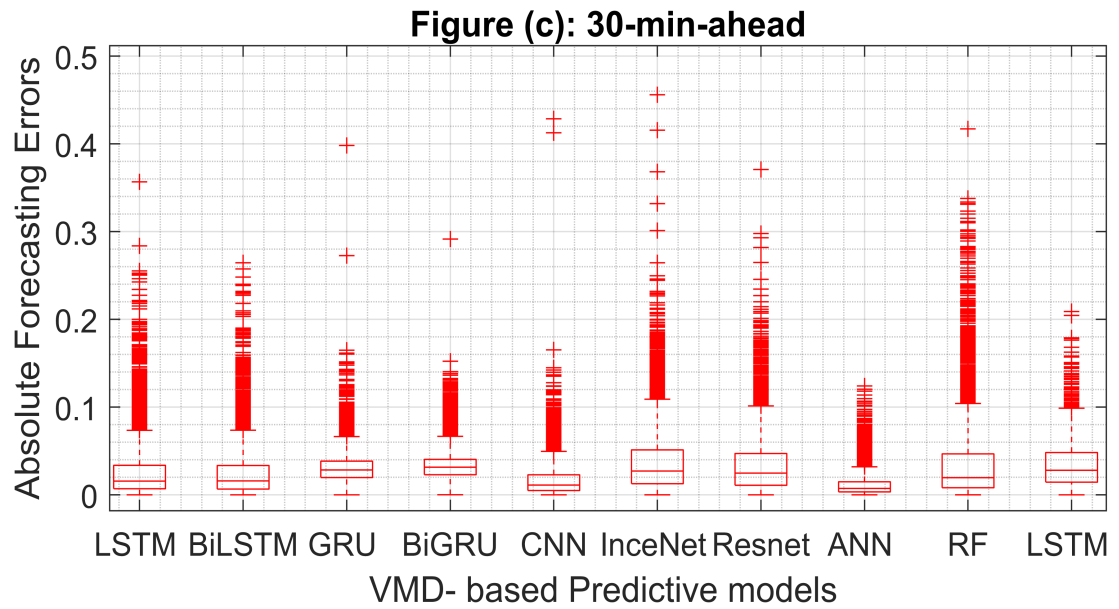
Figure 4.43: Comparative distribution of absolute errors among vmd-based models
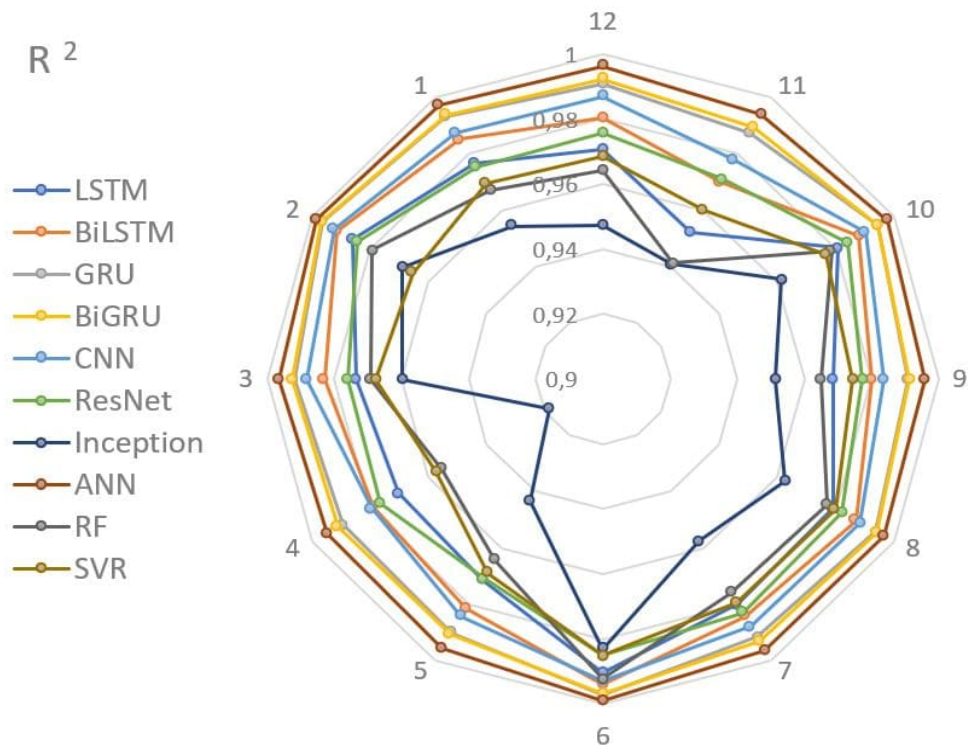
Figure 4.44: Monthly R^2 of 30-minute-ahead models
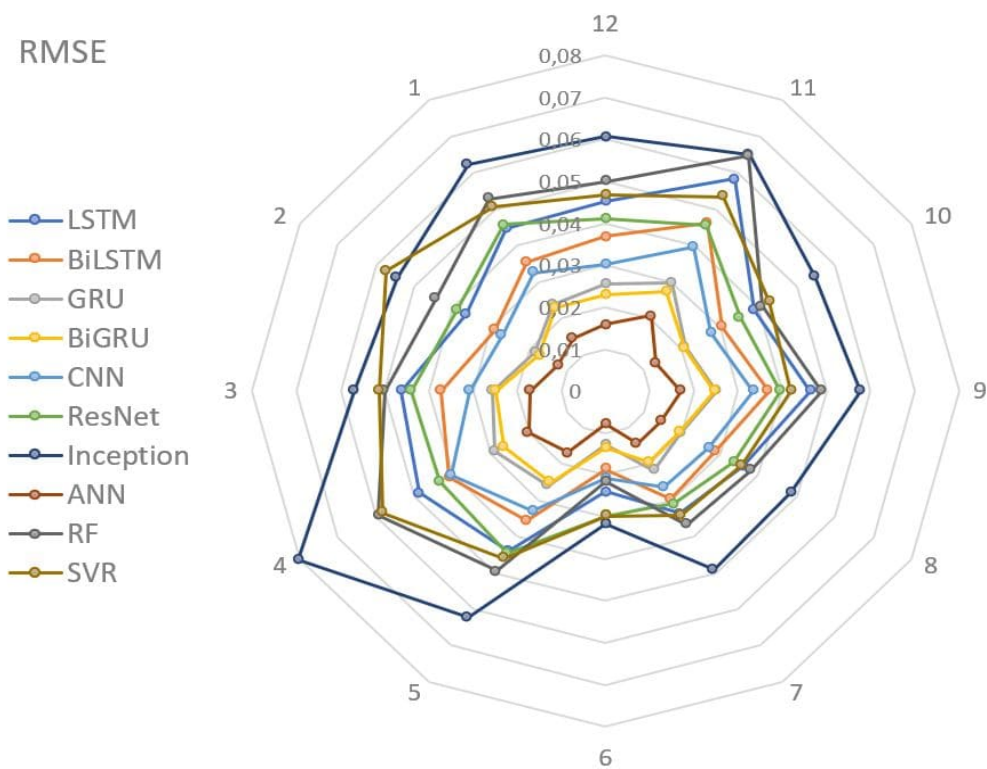
Figure 4.45: Monthly RMSE of 30-minute-ahead models
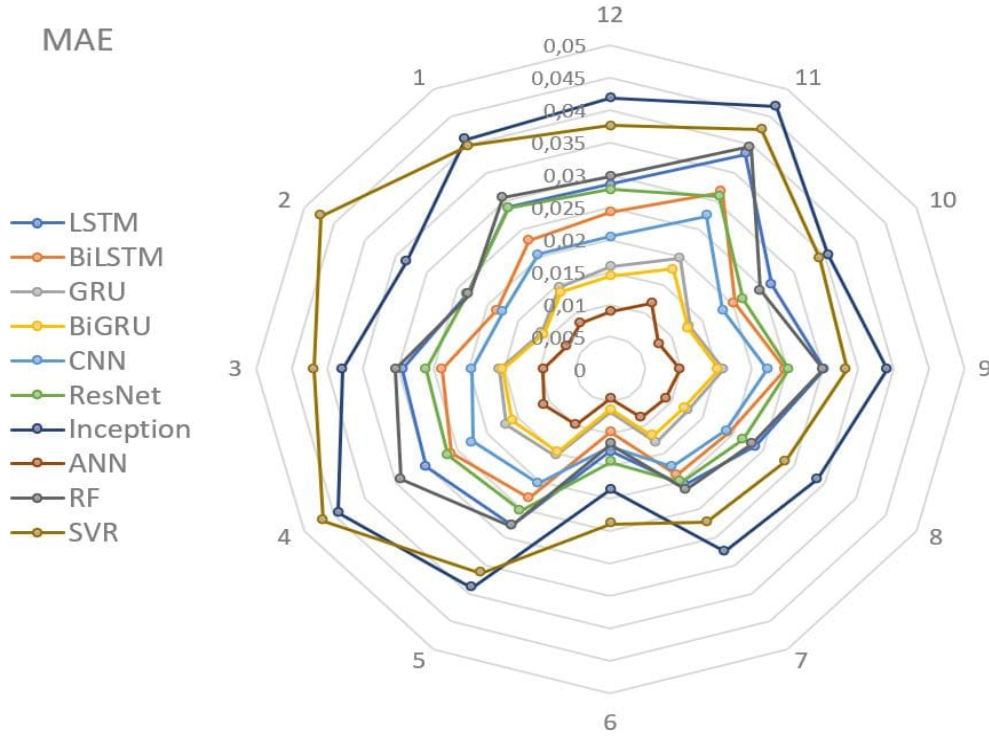
Figure 4.46: Monthly MAE of 30-minute-ahead models

4.7 Conclusions and insights

The simulations assessed the forecasting capabilities of a wide range of machine learning and deep learning techniques, including LSTM, BiLSTM, GRU, CNN, Inception, Residual networks, ANN, SVM, and Random Forest, applied to photovoltaic power prediction. Models were developed and tested under two forecasting horizons—30 minutes and 3 hours—using performance indicators such as RMSE and MAE, with additional monthly analyses to capture seasonal effects. An important component of the study was the integration of Variational Mode Decomposition (VMD) as a preprocessing step for feature extraction, enabling the models to better capture intrinsic signal characteristics. Seasonal investigations were also conducted to explore recurring patterns and the influence of environmental variability. The overall findings demonstrate that aligning suitable model architectures with advanced pre-processing methods can significantly enhance predictive accuracy, thereby supporting more efficient energy management, improved grid stability, and informed decision-making in photovoltaic systems. These contributions are particularly relevant to the global shift toward renewable energy, reinforcing the role of accurate forecasting in ensuring sustainable power integration.

The principal conclusions can be summarized as follows:

- CNN demonstrates the strongest performance among standalone models for both short-term and medium-term horizons, highlighting its ability to capture complex

spatial–temporal structures within the data.

- Inception modules prove effective in multi-scale feature extraction by employing diverse filter sizes within the architecture, enabling the simultaneous modeling of localized variations and broader temporal patterns.

- Residual networks show robustness through skip connections that mitigate vanishing gradient issues and preserve feature transmission across deeper layers.

- Machine learning approaches such as ANN, Random Forest, and SVR achieve relatively higher accuracy in longer-term (3-hour) predictions, particularly when historical data availability is constrained. Conversely, deep learning methods maintain strong accuracy in larger datasets, underlining their suitability for learning intricate nonlinear dependencies.

- VMD preprocessing considerably improves forecasting outcomes, especially for machine learning algorithms, which gain the most from enhanced feature engineering. While deep learning models inherently perform automated feature extraction, VMD further benefits both ML and DL approaches by decomposing signals into cleaner and more informative subcomponents.

- Seasonal analysis confirms the robustness of VMD-based models, with June emerging as the most accurately predicted month for both horizons, while November consistently records the lowest accuracy. These findings emphasize the role of meteorological conditions in shaping model performance and the necessity of careful preprocessing and tuning for reliable photovoltaic forecasting.

GENERAL CONCLUSION

As the world transitions toward cleaner energy, solar power stands at the forefront of this shift. With its vast solar potential, Algeria is embracing this transition, expanding its photovoltaic infrastructure and integrating renewables into the national grid. Nevertheless, the inherent intermittency of solar energy introduces significant challenges, which can be effectively mitigated through precise PV power forecasting.

This research takes a crucial step in this direction, focusing on the short-term forecasting of PV generation in Algeria's High Plains. Through the application of state-of-the-art machine learning and deep learning methodologies, the study enhances forecasting precision, providing valuable guidance for energy management and supporting large-scale solar integration. A comprehensive analysis was conducted across various forecasting models, revealing valuable insights into their performance across different time horizons and seasonal variations.

Key findings highlight the superiority of DL models, particularly CNN and Inception, in capturing intricate patterns in PV power generation. The introduction of VMD as a preprocessing step significantly improved model accuracy, particularly for machine learning approaches. Furthermore, seasonal analysis underscored the importance of adapting forecasting models to temporal fluctuations, with June yielding the highest accuracy and November presenting greater challenges.

By refining predictive models and advancing feature extraction methods, this study not only contributes to the ongoing research in solar forecasting but also offers practical solutions for grid operators and policymakers. These insights are important for enhancing energy planning, improving grid stability, and accelerating the transition toward a more sustainable and resilient power system. Ultimately, this work serves as a foundation for future advancements in photovoltaic energy forecasting, reinforcing the role of AI-driven solutions in shaping the global energy landscape.

Bibliography

[1] Slimane Sefiane, Mahi Nour El Hoda, and Henni Ahmed. Financial or socio-economic feasibility? potential assessment of renewable energy investment in algeria. *Journal of Asian Energy Studies*, 6:48–58, 8 2022.

[2] International Renewable Energy Agency. *World Energy Transitions Outlook 2023: 1.5°C Pathway*. 2023.

[3] Algeria - renewable energy. *International Trade Admission*, 2023.

[4] Enr: 567 mw de capacité totale installée au niveau national à fin 2021. *Algerie Presse Service*, 2022.

[5] Energies nouvelles, renouvelables et maitrise de l'energie. *The Ministery of Energy and mines*.

[6] Evolution du marché des enr en algérie. *Algeria Green Energy Cluster*, 2023.

[7] Ministère de la Transition Energerique et des Energies Renouvelables (MTEER). Avis d'appel d'offres a investisseurs n°01/mteer/2021. 11 2021.

[8] SONELGAZ-EnR. National and international open to tender n°01:ceeg/2023. *BAOSEM EDITION PUBLICITE*, 2 2023.

[9] Oktoviano Gandhi, Dhivya Sampath Kumar, Carlos D. Rodríguez-Gallegos, and Dipti Srinivasan. Review of power system impacts at high pv penetration part i: Factors limiting pv penetration. *Solar Energy*, 210:181–201, 11 2020.

[10] Harish Sharma and Joe Grappe, editors. *IEEE Standard for Measurement and Limits of Voltage Fluctuations and Associated Light Flicker on AC Power Systems*. IEEE STANDARS ASSOCIATION, 1453 - 2022 edition, 6 2022.

[11] Licheng Wang, Ruifeng Yan, and Tapan Kumar Saha. Voltage regulation challenges with unbalanced pv integration in low voltage distribution systems and the corresponding solution. *Applied Energy*, 256, 12 2019.

[12] Waleed Alabri and Dilan Jayaweera. Voltage regulation in unbalanced power distribution systems with residential pv systems. *International Journal of Electrical Power and Energy Systems*, 131, 10 2021.

[13] Shuxia Yang, Xiongfei Wang, Jiayu Xu, Mingrun Tang, and Guang Chen. Distribution network adaptability assessment considering distributed pv "reverse power flow" behavior - a case study in beijing. *Energy*, 275, 7 2023.

[14] Abdullah Alshahrani, Siddig Omer, Yuehong Su, Elamin Mohamed, and Saleh Alotaibi. The technical challenges facing the integration of small-scale and large-scale pv systems into the grid: A critical review, 12 2019.

[15] The grid code issue 6 revision 16 the grid code, 2023.

[16] Andreas Ulbig, Theodor S Borsche, Göran Andersson, and Eth Zurich. Impact of low rotational inertia on power system stability and operation. The International Federation of Automatic Contro, 2014.

[17] Jitendra Thapa and Sanjeev Maharjan. Impact of penetration of photovoltaic on rotor angle stability of power system. *International Journal of Engineering and Applied Sciences (IJEAS)*, 6, 4 2019.

[18] Kanchan Jha and Abdul Gafoor Shaik. A comprehensive review of power quality mitigation in the scenario of solar pv integration into utility grid. *e-Prime - Advances in Electrical Engineering, Electronics and Energy*, 3:100103, 3 2023.

[19] Aida Fazliana Abdul Kadir, Tamer Khatib, and Wilfried Elmenreich. Integrating photovoltaic systems in power system: Power quality impacts and optimal planning challenges, 2014.

[20] Enkhtsetseg Munkhchuluun, Lasantha Gunaruwan Meegahapola, and Arash Vahidnia. Reactive power assisted frequency regulation scheme for large-scale solar-pv plants. *International Journal of Electrical Power and Energy Systems*, 146, 3 2023.

[21] Abdelaziz Salah Saidi, Fahad Alsharari, Emad M. Ahmed, Saad F. Al-Gahtani, Shaik Mohammad Irshad, and Sami Alalwani. Investigating the impact of grid-tied photovoltaic system in the aljouf region, saudi arabia, using dynamic reactive power control. *Energies*, 16, 3 2023.

[22] Saad Alqahtani, Abdullah Shaher, Ali Garada, and Liana Cipcigan. Impact of the high penetration of renewable energy sources on the frequency stability of the saudi grid †. *Electronics (Switzerland)*, 12, 3 2023.

[23] International Renewable Energy Agency. Power system flexibility for the energy transition, part 1: Overview for policy makers, 2018.

[24] Commission for renewable energy and energy efficiency. Energy transtion in algeria, 2020.

[25] Dhivya Sampath Kumar, Oktoviano Gandhi, Carlos D. Rodríguez-Gallegos, and Dipti Srinivasan. Review of power system impacts at high pv penetration part ii: Potential solutions and the way forward. *Solar Energy*, 210:202–221, 11 2020.

[26] Reski Khelifi, Mawloud Guermoui, Abdelaziz Rabehi, Ayoub Taallah, Abdelhalim Zoukel, Sherif S.M. Ghoneim, Mohit Bajaj, Kareem M. AboRas, and Ievgen Zaitsev. Short-term pv power forecasting using a hybrid tvf-emd-elm strategy. *International Transactions on Electrical Energy Systems*, 2023, 2023.

[27] Naiqing Li, Longhao Li, Fan Zhang, Ticao Jiao, Shuang Wang, Xuefeng Liu, and Xinghua Wu. Research on short-term photovoltaic power prediction based on multi-scale similar days and esn-kelm dual core prediction model. *Energy*, 277, 8 2023.

[28] A Botterud, J Wang, and G Conzelmann. A quick guide to wind power forecasting: State-of-the-art 2009, 2009.

[29] Audun Botterud, Z. Zhi, Jianhui Wang, Ricardo Bessa, H. Keko, J. Mendes, Jean Sumaili, and V. Miranda. Use of wind power forecasting in operational decisions. 01 2011.

[30] Hadja Maïmouna Diagne. *Gestion intelligente du réseau électrique réunionnais. Prévision de la ressource solaire en milieu insulaire*. Theses, Université de la Réunion, April 2015.

[31] Amit Dhoke, Rahul Sharma, and Tapan Kumar Saha. Pv module degradation analysis and impact on settings of overcurrent protection devices. *Solar Energy*, 160:360–367, 1 2018.

[32] T Erge, V.U Hoffmann, and K Kiefer. The german experience with grid-connected pv-systems. *Solar Energy*, 70:479–487, 2001.

[33] Aref Eskandari, Jafar Milimonfared, and Mohammadreza Aghaei. Line-line fault detection and classification for photovoltaic systems using ensemble learning model based on i-v characteristics. *Solar Energy*, 211:354–365, 11 2020.

[34] Shibo Lu, B.T. Phung, and Daming Zhang. A comprehensive review on dc arc faults and their diagnosis methods in photovoltaic systems. *Renewable and Sustainable Energy Reviews*, 89:88–98, 6 2018.

[35] Fouzi Harrou, Ying Sun, Bilal Taghezouit, Ahmed Saidi, and Mohamed-Elkarim Hamlati. Reliable fault detection and diagnosis of photovoltaic systems based on statistical monitoring approaches. *Renewable Energy*, 116:22–37, 2 2018.

[36] Olatz Arriaga Arruti, Luca Gnocchi, Quentin Jeangros, Christophe Ballif, and Alessandro Virtuani. Potential-induced degradation in bifacial silicon heterojunction solar modules: Insights and mitigation strategies. *Progress in Photovoltaics: Research and Applications*, 32:304–316, 5 2024.

[37] Dirk C. Jordan, Michael G. Deceglie, and Sarah R. Kurtz. Pv degradation methodology comparison — a basis for a standard. pages 0273–0278. IEEE, 6 2016.

[38] Muhammad Qamar Raza, Mithulananthan Nadarajah, and Chandima Ekanayake. On recent advances in pv output power forecast, 10 2016.

[39] M. Lipperheide, J.L. Bosch, and J. Kleissl. Embedded nowcasting method using cloud speed persistence for a photovoltaic power plant. *Solar Energy*, 112:232–238, 2015.

[40] Vincent P.A. Lonij, Adria E. Brooks, Alexander D. Cronin, Michael Leuthold, and Kevin Koch. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Solar Energy*, 97:58–66, 2013.

[41] Gbemou Shab. Prévision multi-horizon de l'éclairement global horizontal pour la gestion intelligente du réseau électrique de distribution en région occitanie. 4 2022.

[42] Bruce Ratner. The correlation coefficient: Its values range between 1/1, or do they. *Journal of Targeting, Measurement and Analysis for Marketing*, 17:139–142, 6 2009.

[43] Edward Baleke Ssekulima, Muhammad Bashar Anwar, Amer Al Hinai, and Mohamed Shawky El Moursi. Wind speed and solar irradiance forecasting techniques for enhanced renewable energy integration with the grid: A review, 7 2016.

[44] Dazhi Yang. Making reference solar forecasts with climatology, persistence, and their optimal convex combination. *Solar Energy*, 193:981–985, 11 2019.

[45] Richard Perez, Sergey Kivalov, James Schlemmer, Karl Hemker, David Renné, and Thomas E. Hoff. Validation of short and medium term operational solar radiation forecasts in the us. *Solar Energy*, 84(12):2161–2172, 2010.

[46] Utpal Kumar Das, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski. Forecasting of photovoltaic power generation and model optimization: A review, 1 2018.

[47] Martin János Mayer and Dazhi Yang. Pairing ensemble numerical weather prediction with ensemble physical model chain for probabilistic photovoltaic power forecasting. *Renewable and Sustainable Energy Reviews*, 175:113171, 2023.

[48] Wenting Wang, Yufeng Guo, Dazhi Yang, and Jan Kleissl. Solar power forecasting based on numerical weather prediction and physical model chain for day-ahead power system dispatching. In *2022 4th International Conference on Smart Power Internet Energy Systems (SPIES)*, pages 2081–2086, 2022.

[49] Gang Zhang, Dazhi Yang, George Galanis, and Emmanouil Androulakis. Solar forecasting with hourly updated numerical weather prediction. *Renewable and Sustainable Energy Reviews*, 154:111768, 2022.

[50] Francesco Nicoletti and Piero Bevilacqua. Hourly photovoltaic production prediction using numerical weather data and neural networks for solar energy decision support. *Energies*, 17(2), 2024.

[51] Zhiyuan Si, Ming Yang, Yixiao Yu, and Tingting Ding. Photovoltaic power forecast based on satellite images considering effects of solar position. *Applied Energy*, 302:117514, 2021.

[52] Fei Wang, Xiaoxing Lu, Shengwei Mei, Ying Su, Zhao Zhen, Zubing Zou, Xuemin Zhang, Rui Yin, Neven Duić, Miadreza Shafie-khah, and João P.S. Catalão. A satellite image data based ultra-short-term solar pv power forecasting method considering cloud information from neighboring plant. *Energy*, 238:121946, 2022.

[53] Han Seung Jang, Kuk Yeol Bae, Hong-Shik Park, and Dan Keun Sung. Solar power prediction based on satellite images and support vector machine. *IEEE Transactions on Sustainable Energy*, 7:1255–1263, 7 2016.

[54] Muhammad Qamar Raza, Mithulananthan Nadarajah, and Chandima Ekanayake. On recent advances in pv output power forecast, 10 2016.

[55] Jonathan D Cryer and Kung-Sik Chan. Springer texts in statistics time series analysis with applications in r second edition.

[56] Vishal Kushwaha and Naran M. Pindoriya. A sarima-rvfl hybrid model assisted by wavelet decomposition for very short-term solar pv power generation forecast. *Renewable Energy*, 140:124–139, 9 2019.

[57] Mohammed Farsi, Doreswamy Hosahalli, B. R. Manjunatha, Ibrahim Gad, El Sayed Atlam, Althobaiti Ahmed, Ghada Elmarhomy, Mahmoud Elmarhoumy, and Osama A. Ghoneim. Parallel genetic algorithms for optimizing the sarima model for better forecasting of the ncdc weather data. *Alexandria Engineering Journal*, 60:1299–1316, 2 2021.

[58] Jing Huang, Małgorzata Korolkiewicz, Manju Agrawal, and John Boland. Forecasting solar radiation on an hourly time scale using a coupled autoregressive and dynamical system (cards) model. *Solar Energy*, 87:136–149, 1 2013.

[59] Kahina Dahmani, Rabah Dizene, Gilles Notton, Christophe Paoli, Cyril Voyant, and Marie Laure Nivet. Estimation of 5-min time-step data of tilted solar global irradiation using ann (artificial neural network) model. *Energy*, 70:374–381, 6 2014.

[60] Laidi Maamar, Abdellah El Hadj Abdallah, Hanini Salah, and Rezrazi Ahmed. Prediction and extrapolation of global solar irradiation on tilted surfaces from horizontal ones using an artificial neural network. Institute of Electrical and Electronics Engineers Inc., 3 2014.

[61] S Benkaciali, M Haddadi, A Khellaf, K Gairaa, M Guermoui, Ecole Nationale Polytechnique, and El Harrach. Evaluation of the global solar irradiation from the artificial neural network technique, 2016.

[62] Reski Khelifi, Mawloud Guermoui, Abdelaziz Rabehi, and Djemoui Lalmi. Multi-step-ahead forecasting of daily solar radiation components in the saharan climate. *International Journal of Ambient Energy*, 41:707–715, 5 2020.

[63] Mawloud Guermoui and Abdelaziz Rabehi. Soft computing for solar radiation potential assessment in algeria. *International Journal of Ambient Energy*, 41:1524–1533, 11 2020.

[64] Mawloud Guermoui, Abdelaziz Rabehi, Kacem Gairaa, and Said Benkaciali. Support vector regression methodology for estimating global solar radiation in algeria. *European Physical Journal Plus*, 133, 1 2018.

[65] Mawloud Guermoui, Rabehi Abdelaziz, Kacem Gairaa, Lalmi Djemoui, and Said Benkaciali. New temperature-based predicting model for global solar radiation using support vector regression. *International Journal of Ambient Energy*, 43:1397–1407, 2022.

[66] Muhammed A. Hassan, Nadjem Bailek, Kada Bouchouicha, and Samuel Chukwujindu Nwokolo. Ultra-short-term exogenous forecasting of photovoltaic power production using genetically optimized non-linear auto-regressive recurrent neural networks. *Renewable Energy*, 171:191–209, 6 2021.

[67] A. Takilalte, S. Harrouni, and J. Mora. Forecasting global solar irradiance for various resolutions using time series models - case study: Algeria. *Energy Sources, Part A: Recovery, Utilization and Environmental Effects*, 44:1–20, 2022.

[68] Abdelaziz Rabehi, Mawloud Guermoui, and Djemoui Lalmi. Hybrid models for global solar radiation prediction: a case study. *International Journal of Ambient Energy*, 41:31–40, 1 2020.

[69] Khalil Benmouiza and Ali Cheknane. Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models. *Energy Conversion and Management*, 75:561–569, 2013.

[70] Khalil Benmouiza and Ali Cheknane. Clustered anfis network using fuzzy c-means, subtractive clustering, and grid partitioning for hourly solar radiation forecasting. *Theoretical and Applied Climatology*, 137:31–43, 7 2019.

[71] Jayesh Thaker and Robert Höller. Hybrid model for intra-day probabilistic pv power forecast. *Renewable Energy*, 232:121057, 2024.

[72] Yongning Zhang, Xiaoying Ren, Fei Zhang, Yulei Liu, and Jierui Li. A deep learning-based dual-scale hybrid model for ultra-short-term photovoltaic power forecasting. *Sustainability*, 16(17), 2024.

[73] T Sana Amreen, Radharani Panigrahi, and N.R. Patne. Solar power forecasting using hybrid model. In *2023 5th International Conference on Energy, Power and Environment: Towards Flexible Green Energy Technologies (ICEPE)*, pages 1–6, 2023.

[74] Muhammad Naveed Akhter, Saad Mekhilef, Hazlie Mokhlis, and Noraisyah Mohamed Shah. Review on forecasting of photovoltaic power generation based on machine learning and metaheuristic techniques, 2019.

[75] Lining Wang, Mingxuan Mao, Jili Xie, Zheng Liao, Hao Zhang, and Huanxin Li. Accurate solar pv power prediction interval method based on frequency-domain decomposition and lstm model. *Energy*, 262, 1 2023.

[76] Delnia Sadeghi, Aminabbas Golshanfard, Shahab Eslami, Kiyarash Rahbar, and Ryan Kari. Improving pv power plant forecast accuracy: A hybrid deep learning approach compared across short, medium, and long-term horizons. *Renewable Energy Focus*, 45:242–258, 6 2023.

[77] Mariam Alkandari and Imtiaz Ahmad. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Applied Computing and Informatics Emerald Publishing Limited*, 2019.

[78] Eun Gyeong Kim, M. Shaheer Akhtar, and O. Bong Yang. Designing solar power generation output forecasting methods using time series algorithms. *Electric Power Systems Research*, 216, 3 2023.

[79] Cheng Zhang and Maomao Zhang. Wavelet-based neural network with genetic algorithm optimization for generation prediction of pv plants. *Energy Reports*, 8:10976–10990, 11 2022.

[80] Zhenhao Wang, Chong Wang, Long Cheng, and Guoqing Li. An approach for day-ahead interval forecasting of photovoltaic power: A novel dcgan and lstm based quantile regression modeling method. *Energy Reports*, 8:14020–14033, 11 2022.

[81] M. Talaat, Taghreed Said, Mohamed A. Essa, and A. Y. Hatata. Integrated mffnn-mvo approach for pv solar power forecasting considering thermal effects and environmental conditions. *International Journal of Electrical Power and Energy Systems*, 135, 2 2022.

[82] Mohamed Abdel-Basset, Hossam Hawash, Ripon K. Chakrabortty, and Michael Ryan. Pv-net: An innovative deep learning approach for efficient forecasting of short-term photovoltaic energy production. *Journal of Cleaner Production*, 303, 6 2021.

[83] Donghan Geng, Bo Wang, and Qi Gao. A hybrid photovoltaic/wind power prediction model based on time2vec, wdcnn and bilstm. *Energy Conversion and Management*, 291, 9 2023.

[84] Fan Wu, Shunli Wang, Donglei Liu, Wen Cao, Carlos Fernandez, and Qi Huang. An improved convolutional neural network-bidirectional gated recurrent unit algorithm for robust state of charge and state of energy estimation of new energy vehicles of lithium-ion batteries. *Journal of Energy Storage*, 82, 3 2024.

[85] Fei Wang, Zhiming Xuan, Zhao Zhen, Kangping Li, Tieqiang Wang, and Min Shi. A day-ahead pv power forecasting method based on lstm-rnn model and time correlation modification under partial daily pattern prediction framework. *Energy Conversion and Management*, 212, 5 2020.

[86] Majad Mansoor, Adeel Feroz Mirza, Muhammad Usman, and Qiang Ling. Hybrid forecasting models for wind-pv systems in diverse geographical locations: Performance and power potential analysis. *Energy Conversion and Management*, 287, 7 2023.

[87] Abdelhak Keddouda, Razika Ihaddadene, Ali Boukhari, Abdelmalek Atia, Müslüm Arıcı, Nacer Lebbihiat, and Nabila Ihaddadene. Solar photovoltaic power prediction using artificial neural network and multiple regression considering ambient and operating conditions. *Energy Conversion and Management*, 288, 7 2023.

[88] Mawloud Guermoui, Kada Bouchouicha, Nadjem Bailek, and John W. Boland. Forecasting intra-hour variance of photovoltaic power using a new integrated model. *Energy Conversion and Management*, 245, 10 2021.

[89] Abderrezzaq Ziane, Ammar Necaibia, Nordine Sahouane, Rachid Dabou, Mohammed Mostefaoui, Ahmed Bouraiou, Seyfallah Khelifi, Abdelkrim Rouabhia, and Mohamed Blal. Photovoltaic output power performance assessment and forecasting: Impact of meteorological variables. *Solar Energy*, 220:745–757, 2021.

[90] Abdelkader Dairi, Fouzi Harrou, Ying Sun, and Sofiane Khadraoui. Short-term forecasting of photovoltaic solar power production using variational auto-encoder driven deep learning approach. *Applied Sciences (Switzerland)*, 10:1–20, 12 2020.

[91] Fouzi Harrou, Farid Kadri, and Ying Sun. *Forecasting of Photovoltaic Solar Power Production Using LSTM Approach*. IntechOpen, 2020.

[92] Muhammad Hamza Zafar, Noman Mujeeb Khan, Majad Mansoor, Adeel Feroz Mirza, Syed Kumayl Raza Moosavi, and Filippo Sanfilippo. Adaptive ml-based technique for renewable energy system power forecasting in hybrid pv-wind farms power conversion systems. *Energy Conversion and Management*, 258, 4 2022.

[93] Gökhan Sahin, Gültekin Isik, and Wilfried G.J.H.M. van Sark. Predictive modeling of pv solar power plant efficiency considering weather conditions: A comparative analysis of artificial neural networks and multiple linear regression. *Energy Reports*, 10:2837–2849, 11 2023.

[94] Mihriban Sari and Müslüm Arıcı. *Global Solar Radiation Forecasting with Artificial Neural Networks*. 10 2023.

[95] Ferial ElRobrini, Syed Muhammad Salman Bukhari, Muhammad Hamza Zafar, Nedaa Al-Tawalbeh, Naureen Akhtar, and Filippo Sanfilippo. Federated learning and non-federated learning based power forecasting of photovoltaic/wind power energy systems: A systematic review. *Energy and AI*, 18:100438, 12 2024.

[96] Enrico Borgogno-Mondino, Alessandro Farbo, Vittorino Novello, and Laura de Palma. A fast regression-based approach to map water status of pomegranate orchards with sentinel 2 data. *Horticulturae*, 8(9), 2022.

[97] Gilles Notton, Kahina Dahmani, Dizene Rabah, Rabah Dizene, Marie-Laure Nivet, Cyril Voyant, and Christophe Paoli. Application of ann methods for solar radiation estimationestimation. 2016.

[98] Xinbo He, Yong Wang, Yuyang Zhang, Xin Ma, Wenqing Wu, and Lei Zhang. A novel structure adaptive new information priority discrete grey prediction model and its application in renewable energy generation forecasting. *Applied Energy*, 325, 11 2022.

[99] International Renewable Energy Agency. *Advanced forecasting of variable renewable power generation: Innovation Landscape Brief*. 2020.

[100] J. Antonanzas, N. Osorio, R. Escobar, R. Urraca, F. J. Martinez de Pison, and F. Antonanzas-Torres. Review of photovoltaic power forecasting, 10 2016.

[101] Solar irradiance and solar power forecasts | schneider electric nigeria.

[102] Harness the sun to make, use, save, and sell your own power. | enphase.

[103] Solar, wind weather forecasts. solar irradiance data. consulting.

[104] Grid solutions, ul-solutions.

[105] Metodología , greenpowermonitor.

[106] Atlas global solar.

[107] Yingli solar yl p-35b series installation and user manual (page 10 of 11) | manualslib.

[108] Ferial El Robrini and Badia Amrouche. Support vector regression for predicting intra-day photovoltaic generation in algeria. 11 2023.

[109] Chaouki Ghenai, Fahad Faraz Ahmad, and Oussama Rejeb. Artificial neural network-based models for short term forecasting of solar pv power output and battery state of charge of solar electric vehicle charging station. *Case Studies in Thermal Engineering*, 61:105152, 2024.

[110] Younes Ledmaoui, Asmaa El Fahli, Abdellah Chehri, Adila Elmaghraoui, Mohamed El Aroussi, and Rachid Saadane. Monitoring solar energy production based on internet of things with artificial neural networks forecasting. *Procedia Computer Science*, 225:88–97, 2023. 27th International Conference on Knowledge Based and Intelligent Information and Engineering Sytems (KES 2023).

[111] Soteris A. Kalogirou. Artificial neural networks in renewable energy systems applications: a review. *Renewable and Sustainable Energy Reviews*, 5(4):373–401, 2001.

[112] Ranya Ghoniem. A novel bio-inspired deep learning approach for liver cancer diagnosis. *Information*, 11:8articlearticle, author = Ghoniem, Ranya, year = 2020, month = 01, pages = 80, title = A Novel Bio–Inspired Deep Learning Approach for Liver Cancer Diagnosis, volume = 11, journal = Information, doi = 10.3390/info11020080 0, 01 2020.

[113] Soteris A. Kalogirou. Artificial neural networks in renewable energy systems applications: a review. *Renewable and Sustainable Energy Reviews*, 5(4):373–401, 2001.

[114] Chaouki Ghenai, Fahad Faraz Ahmad, Oussama Rejeb, and Maamar Bettayeb. Artificial neural networks for power output forecasting from bifacial solar pv system with enhanced building roof surface albedo. *Journal of Building Engineering*, 56:104799, 2022.

[115] Aadil Ganie and Samad Dadvandipour. From big data to smart data: a sample gradient descent approach for machine learning. *Journal of Big Data*, 10, 10 2023.

[116] Ali Agga, Ahmed Abbou, Moussa Labbadi, and Yassine El Houm. Short-term self consumption pv plant power production forecasts based on hybrid cnn-lstm, convlstm models. *Renewable Energy*, 177:101–112, 11 2021.

[117] Yisheng Cao, Gang Liu, Donghua Luo, Durga Prasad Bavirisetti, and Gang Xiao. Multi-timescale photovoltaic power forecasting using an improved stacking ensemble algorithm based lstm-informer model. *Energy*, 283, 11 2023.

[118] Qi Liu, Oscar Famous Darteh, Muhammad Bilal, Xianming Huang, Muhammad Attique, Xiaodong Liu, and Amevi Acakpovi. A cloud-based bi-directional lstm approach to grid-connected solar pv energy forecasting for multi-energy systems. *Sustainable Computing: Informatics and Systems*, 40, 12 2023.

[119] Dheeraj Kumar Dhaked, Sharad Dadhich, and Dinesh Birla. Power output forecasting of solar photovoltaic plant using lstm. *Green Energy and Intelligent Transportation*, 2, 10 2023.

[120] Mohammad Ahmad A. Al-Ja'afreh, Geev Mokryani, and Bilal Amjad. An enhanced cnn-lstm based multi-stage framework for pv and load short-term forecasting: Dso scenarios. *Energy Reports*, 10:1387–1408, 11 2023.

[121] Delnia Sadeghi, Aminabbas Golshanfard, Shahab Eslami, Kiyarash Rahbar, and Ryan Kari. Improving pv power plant forecast accuracy: A hybrid deep learning approach compared across short, medium, and long-term horizons. *Renewable Energy Focus*, 45:242–258, 6 2023.

[122] Maged Mamdouh, Mostafa Ezzat, and Hesham Hefny. Improving flight delays prediction by developing attention-based bidirectional lstm network. *Expert Systems with Applications*, 238, 3 2024.

[123] Adeel Feroz Mirza, Majad Mansoor, Muhammad Usman, and Qiang Ling. Hybrid inception-embedded deep neural network resnet for short and medium-term pv-wind forecasting. *Energy Conversion and Management*, 294, 10 2023.

[124] Suleyman Gokhan Taskin, Ecir Ugur Kucuksille, and Kamil Topal. Detection of turkish fake news in twitter with machine learning algorithms. *Arabian Journal for Science and Engineering*, 47:2359–2379, 2 2022.

[125] Hwanjo Yu and Sungchul Kim. *SVM tutorial-classification, regression and ranking*, volume 1-4, pages 479–506. Springer Berlin Heidelberg, 1 2012.

[126] Andi A.H. Lateko, Hong Tzer Yang, and Chao Ming Huang. Short-term pv power forecasting using a regression-based ensemble method. *Energies*, 15, 6 2022.

[127] Nahed Zemouri and Hassen Bouzgou. *Ensemble of Support Vector Methods to Estimate Global Solar Radiation in Algeria*, volume 35, pages 155–163. Springer, 2018.

[128] Bo Xiao, Hai Zhu, Sujun Zhang, Zi Ouyang, Tandong Wang, and Saeed Sarvazizi. Gray-related support vector machine optimization strategy and its implementation in forecasting photovoltaic output power. *International Journal of Photoenergy*, 2022, 2022.

[129] Daniel Borup, Bent Jesper Christensen, Nicolaj Søndergaard Mühlbach, and Mikkel Slot Nielsen. Targeting predictors in random forest regression. *International Journal of Forecasting*, 39:841–868, 4 2023.

[130] Taha Ait tchakoucht, Badr Elkari, Yassine Chaibi, and Tarik Kousksou. Random forest with feature selection and k-fold cross validation for predicting the electrical and thermal efficiencies of air based photovoltaic-thermal systems. *Energy Reports*, 12:988–999, 12 2024.

[131] Jose-M. Bueno-Barrachina, Yiyao Ye-Lin, Felix Nieto del Amor, and Vicente Fuster-Roig. Inception 1d-convolutional neural network for accurate prediction of electrical insulator leakage current from environmental data during its normal operation using long-term recording. *Engineering Applications of Artificial Intelligence*, 119:105799, 3 2023.

[132] Serkan Kiranyaz, Onur Avci, Osama Abdeljaber, Turker Ince, Moncef Gabbouj, and Daniel J. Inman. 1d convolutional neural networks and applications: A survey. *Mechanical Systems and Signal Processing*, 151:107398, 4 2021.

[133] Jinxiang Liu, Tiejun Wang, Andrew Skidmore, Yaqin Sun, Peng Jia, and Kefei Zhang. Integrated 1d, 2d, and 3d cnns enable robust and efficient land cover classification from hyperspectral imagery. *Remote Sensing*, 15(19), 2023.

[134] Jianbo Yu and Xingkang Zhou. One-dimensional residual convolutional autoencoder based feature learning for gearbox fault diagnosis. *IEEE Transactions on Industrial Informatics*, 16:6347–6358, 10 2020.

[135] Seyed Mahdi Miraftabzadeh and Michela Longo. High-resolution pv power prediction model based on the deep learning and attention mechanism. *Sustainable Energy, Grids and Networks*, 34:101025, 6 2023.

[136] Rafiq Asghar, Francesco Riganti Fulginei, Michele Quercio, and Assia Mahrouch. Artificial neural networks for photovoltaic power forecasting: A review of five promising models. *IEEE Access*, 12:90461–90485, 2024.

[137] Yanrui Li and Chunjie Yang. Multi time scale inception-time network for soft sensor of blast furnace ironmaking process. *Journal of Process Control*, 118:106–114, 10 2022.

[138] Kefei Zhang, Xiaolin Yang, Hua Cao, Jesse Thé, Zhongchao Tan, and Hesheng Yu. Multi-step forecast of pm2.5 and pm10 concentrations using convolutional neural network integrated with spatial–temporal attention and residual learning. *Environment International*, 171:107691, 1 2023.

[139] Linfei Yin and Mingshan Zhao. Inception-embedded attention memory fully-connected network for short-term wind power prediction. *Applied Soft Computing*, 141:110279, 7 2023.

[140] Mehdi Neshat, Meysam Majidi Nezhad, Seyedali Mirjalili, Davide Astiaso Garcia, Erik Dahlquist, and Amir H. Gandomi. Short-term solar radiation forecasting using hybrid deep residual learning and gated lstm recurrent network with differential covariance matrix adaptation evolution strategy. *Energy*, 278:127701, 9 2023.

[141] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. pages 770–778. IEEE, 6 2016.

[142] Ceyhun Yildiz, Hakan Acikgoz, Deniz Korkmaz, and Umit Budak. An improved residual-based convolutional neural network for very short-term wind power forecasting. *Energy Conversion and Management*, 228:113731, 1 2021.

[143] Ruonan Liu, Fei Wang, Boyuan Yang, and S. Joe Qin. Multiscale kernel based residual convolutional neural network for motor fault diagnosis under nonstationary conditions. *IEEE Transactions on Industrial Informatics*, 16:3797–3806, 6 2020.

[144] Qian Chen, Wenyu Zhang, Kun Zhu, Di Zhou, Hua Dai, and Quanquan Wu. A novel trilinear deep residual network with self-adaptive dropout method for short-term load forecasting. *Expert Systems with Applications*, 182:115272, 11 2021.

[145] Yuchao Han, Xiangqian Tong, Shuyan Shi, Feng Li, and Yaping Deng. Ultra-short-term wind power interval prediction based on hybrid temporal inception convolutional network model. *Electric Power Systems Research*, 217:109159, 4 2023.

[146] Khaled Ferkous, Mawloud Guermoui, Sarra Menakh, Abderahmane Bellaour, and Tayeb Boulmaiz. A novel learning approach for short-term photovoltaic power forecasting - a review and case studies. *Engineering Applications of Artificial Intelligence*, 133:108502, 7 2024.

[147] Konstantin Dragomiretskiy and Dominique Zosso. Variational mode decomposition. *IEEE Transactions on Signal Processing*, 62:531–544, 2 2014.

[148] Mohammad Mustafa Taye. Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. *Computation*, 11:52, 3 2023.

[149] Yishun Liu, Chunhua Yang, Keke Huang, and Weihua Gui. Non-ferrous metals price forecasting based on variational mode decomposition and lstm network. *Knowledge-Based Systems*, 188:105006, 1 2020.

[150] Mumtaz Ali, Ramendra Prasad, Yong Xiang, Mohsin Khan, Aitazaz Ahsan Farooque, Tianrui Zong, and Zaher Mundher Yaseen. Variational mode decomposition based random forest model for solar radiation forecasting: New emerging machine learning technology. *Energy Reports*, 7:6700–6717, 11 2021.

[151] Lakhdar Nadjib Boucetta, Youssouf Amrane, Aissa Chouder, Saliha Arezki, and Sofiane Kichou. Enhanced forecasting accuracy of a grid-connected photovoltaic power plant: A novel approach using hybrid variational mode decomposition and a cnn-lstm model. *Energies*, 17, 4 2024.

[152] Mahima Sivakumar, Jeba Priya S, S. Thomas George, M.S.P. Subathra, Rajasundrapandiyan Leebanon, and Nallapaneni Manoj Kumar. Nine novel ensemble models for solar radiation forecasting in indian cities based on vmd and dwt integration with the machine and deep learning algorithms. *Computers and Electrical Engineering*, 108:108691, 5 2023.

[153] Moez Krichen. Convolutional neural networks: A survey. *Computers*, 12, 8 2023.

[154] Ehtisham Lodhi, Nadia Dahmani, Syed Muhammad Salman Bukhari, Sujan Gyawali, Sanjog Thapa, Lin Qiu, Muhammad Hamza Zafar, and Naureen Akhtar. Enhancing microgrid forecasting accuracy with saq-mtclstm: A self-adjusting quantized multi-task convlstm for optimized solar power and load demand predictions. *Energy Conversion and Management: X*, 24:100767, 10 2024.

[155] Yikang Li, Wei Huang, Keying Lou, Xizheng Zhang, and Qin Wan. Short-term pv power prediction based on meteorological similarity days and ssa-bilstm. *Systems and Soft Computing*, 6:200084, 12 2024.

[156] Muhammad Hamza Zafar, Syed Muhammad Salman Bukhari, Mohamad Abou Houran, Majad Mansoor, Noman Mujeeb Khan, and Filippo Sanfilippo. Deeptimenet: A novel architecture for precise surface temperature estimation of lithium-ion batteries across diverse ambient conditions. *Case Studies in Thermal Engineering*, 61:105002, 9 2024.

[157] Mengqing Bian, Xiaodong Chu, and Hua Ye. Short-term load forecastiong based on residual graph convolutional neural network. pages 1–5. IEEE, 9 2023.

[158] Muhammad Hamza Zafar, Syed Kumayl Raza Moosavi, and Filippo Sanfilippo. Hierarchical recurrent-inception residual transformer (hrirt) for multidimensional hand force estimation using force myography sensor. *IEEE Sensors Letters*, 8:1–4, 9 2024.