Article

Refining Long Short-Term Memory Neural Network Input Parameters for Enhanced Solar Power Forecasting

Linh Bui Duy [1], **Ninh Nguyen Quang** [1,2,*], **Binh Doan Van** [1,2], **Eleonora Riva Sanseverino** [3],
Quynh Tran Thi Tu [2,4], **Hang Le Thi Thuy** [2], **Sang Le Quang** [2], **Thinh Le Cong** [2] **and Huyen Cu Thi Thanh** [2]

[1] Vietnam Academy of Science and Technology, Graduate University of Science and Technology,
Hanoi 11307, Vietnam; buiduylinh.vn@gmail.com (L.B.D.); doanvanbinh@istee.vast.vn (B.D.V.)

[2] Institute of Science and Technology for Energy and Environment, Vietnam Academy of Science and
Technology, Hanoi 11307, Vietnam; tranthituquynh@istee.vast.vn (Q.T.T.T.);
lethithuyhang@istee.vast.vn (H.L.T.T.); lequangsang@istee.vast.vn (S.L.Q.); lecongthinh@istee.vast.vn (T.L.C.);
cuthithanhhuyen@istee.vast.vn (H.C.T.T.)

[3] Engineering Department, University of Palermo, 90128 Palermo, Italy;
eleonora.rivasanseverino@community.unipa.it

[4] Hawaii Natural Energy Institute, University of Hawaii at Manoa, Honolulu, HI 96822, USA

* Correspondence: nguyenquangninh@istee.vast.vn; Tel.: +84-989255901

Abstract: This article presents a research approach to enhancing the quality of short-term power output forecasting models for photovoltaic plants using a Long Short-Term Memory (LSTM) recurrent neural network. Typically, time-related indicators are used as inputs for forecasting models of PV generators. However, this study proposes replacing the time-related inputs with clear sky solar irradiance at the specific location of the power plant. This feature represents the maximum potential solar radiation that can be received at that particular location on Earth. The Ineichen/Perez model is then employed to calculate the solar irradiance. To evaluate the effectiveness of this approach, the forecasting model incorporating this new input was trained and the results were compared with those obtained from previously published models. The results show a reduction in the Mean Absolute Percentage Error (MAPE) from 3.491% to 2.766%, indicating a 24% improvement. Additionally, the Root Mean Square Error (RMSE) decreased by approximately 0.991 MW, resulting in a 45% improvement. These results demonstrate that this approach is an effective solution for enhancing the accuracy of solar power output forecasting while reducing the number of input variables.

Keywords: long short-term memory; clear sky irradiance; large-scale photovoltaic power plant; forecasting PV power; PV power plant; artificial intelligence

1. Introduction

Power output forecasting for solar power plants is the process of determining the amount of electricity that a photovoltaic system could generate during a certain time period. Solar power capacity forecasting relies on various factors such as solar radiation, geographical location, climate, and other elements that can affect the system's performance.

Solar power capacity forecasting is critical in nations with high solar energy development potential because it shapes renewable energy development strategies and ensures electricity supply reliability.

Accurate solar power capacity forecasting enables efficient solar power system management and effective renewable energy integration into the grid. Vietnam, with its abundant sunlight and diverse terrain, has great potential for solar energy development. Supported by favorable policies and legal frameworks, solar energy is becoming a major source of electricity. Solar power capacity forecasts in Vietnam support system operators and renewable energy investors in making informed decisions regarding operations and long-term development plans.

The process of solar power capacity forecasting often involves the use of mathematical models, statistical methods, or artificial intelligence techniques to estimate expected capacity. Furthermore, estimating solar power capacity requires the analysis and processing of big data. Climate data, geographic location data, and historical data on solar system performance are utilized to determine suitable forecasting models and algorithms. However, accurate and reliable solar power capacity forecasting still requires high precision and reliability, especially for large and complicated systems. Researchers and experts are continuously striving to develop and improve forecasting methods to enhance accuracy and dependability in solar power capacity forecasts.

Forecasting methods for solar power plant generation are classified based on the approach they adopt.

Recently, the authors have delved into applying machine learning or deep learning to address the solar power forecasting problem. Consequently, the recent trend considers machine learning methods as an independent branch alongside the traditional statistical approach. There are four main groups of forecasting techniques: physical models, time series statistical methods, machine learning methods, and hybrid methods (ensemble or combination methods) [1].

Solar power plant generation forecasting systems have progressed and improved over time. Initially, physical model-based forecasting was used, and parameters such as solar irradiance, temperature, and solar panel orientation were employed to estimate power generation. However, these methods often only considered basic physical factors, resulting in inaccurate forecasts in many real-world scenarios. To enhance power forecasting, time series statistical methods were applied. These methods forecast power generation by analyzing and modeling cyclic patterns, seasonal models, and other weather conditions. Techniques such as linear regression, ARIMA (Autoregressive Integrated Moving Average), and SARIMA (Seasonal ARIMA) were utilized to forecast solar power plant generation based on patterns and trends in time-series data.

Over time, with technological advancements and growing computational capabilities, machine learning methods have become a powerful choice for forecasting solar power plant generation. Among various machine learning methods, artificial neural networks (ANNs) have proven to be highly effective in forecasting complex time-series data, such as solar power generation. Machine learning methods apply algorithms and learning models from historical data to forecast solar power plant capacity. Methods such as Support Vector Regression (SVR) [2], Random Forest (RF) [3,4], and Neural Networks (NNs) [5,6] have been used to develop forecasting models based on non-linear relationships and the ability to learn from data. Hybrid methods combine the strengths of different forecasting approaches to increase accuracy and reliability in predictions. Ensemble methods such as Bagging (a technique where multiple models are trained on different subsets of the training data, and their predictions are combined to improve overall performance), Boosting (an ensemble technique where weak learners are trained sequentially, each correcting the errors of its predecessor, leading to a strong model), and Stacking (which involves training multiple models and combining their predictions using another model (meta-model) to potentially enhance predictive accuracy), as well as hybrid models like Hybrid ARIMA-NN [7] (a combination of ARIMA (a time series forecasting model) and Neural Networks) and Hybrid SVR-NN [8] (a combination of Support Vector Regression (SVR) and Neural Networks), have been utilized to leverage the advantages of individual methods.

Henceforth, it can be observed that numerous methods have been and are being applied for short-term solar power output forecasting. Nevertheless, novel techniques are ceaselessly studied and enhanced to improve the effectiveness of prediction. Therefore, understanding the essence, pros, and cons of each method, as well as making comparisons to identify the most efficient approach, leads to proposing new directions for application.

The research [9] proposes a predictive interval (PI) forecasting method for photovoltaic (PV) generator output power, using meteorological parameters. The model integrates a deep learning (DL) baseline model with mathematical theorems and a t-Student PDF to compute

the interval. Using real solar irradiation data from Vitoria-Gasteiz, Spain, the proposed forecaster outperforms computed benchmark models in terms of reliability and interval width, achieving higher accuracy and narrower intervals. The methodology, validated with 2017 data, demonstrates its potential applicability in providing additional information for power systems' decision-makers, maximizing PV generators' profits. This is a study that makes an indirect forecast of radiation and then calculates it converted into output capacity, not a direct forecast of solar power capacity. Zhen et al. [10] address the challenge of accurate short-term output prediction for distributed photovoltaic (PV) power plants in micro-grids, especially when meteorological data is lacking due to cost constraints. A novel ultra-short-term PV power prediction model, GA-BiLSTM, is proposed, leveraging an improved bidirectional long short-term memory model with a genetic algorithm. Multiple output series from adjacent PV plants are ingeniously utilized as inputs to enhance prediction accuracy. Through sensitivity analysis and comparative studies, the GA-BiLSTM model demonstrates superior performance in ultra-short-term forecasting, achieving the lowest RMSE values of 0.438, 0.806, and 1.118 for 5 min, 15 min, and 30 min ahead predictions, respectively. The authors also pointed out the practical fact that the meteorological data of PV power plants is not that easy to acquire, due to the high cost of monitoring equipment. In [11] of Qu et al., the study proposes a hybrid forecasting model (ALSM) combining CNN-LSTM with an attention mechanism, along with a multiple relevant and target variables prediction pattern (MRTPP). This approach addresses the limitations of traditional time series and AI modeling methods, achieving higher accuracy in day-ahead hourly photovoltaic power forecasting. Experimental findings demonstrate superior accuracy compared to statistical and neural network models, with suggested optimal memory lengths for various forecasting horizons. The ALSM model provides better stability and accuracy, although it's noted that it's most suitable for day-ahead hourly predictions, with limitations in longer forecast ranges. In [10,11], the author only uses inputs that are the output power or meteorological features but does not include inputs indicating time to increase the accuracy of the method. Jebli et al. [12] introduce a methodology for solar energy forecasting, employing machine and deep learning techniques to enhance solar power plant competitiveness and reduce fossil fuel dependency. Utilizing data from 2016 to 2018 in Errachidia, Morocco, the research evaluates the effectiveness of various models. Random Forest (RF) and Artificial Neural Network (ANN) yielded superior accuracy compared to Linear Regression (LR) and Support Vector Regression (SVR), particularly in real-time predictions. The findings highlight ANN's robustness and potential for real-time and short-term forecasting, with plans to extend the research to diverse climates and explore additional deep-learning models. Scott in [13] evaluates machine learning algorithms (MLA) for forecasting a localized photovoltaic (PV) system's output on an operational university campus in Manchester, UK. Benchmark algorithms including Random Forest (RF), Neural Networks (NN), Support Vector Machines (SVM), and Linear Regression (LR) are compared. Results show RF achieves the lowest average error (32.0 RMSE), outperforming SVM, LR, and NN. Data quality is crucial for algorithm performance, with RF requiring less data and demonstrating higher accuracy. This research aids in optimizing MLA selection and dataset requirements for PV generation forecasting in buildings, contributing to carbon emissions reduction efforts. These studies [12,13] only use the simplest AI model, the ANN network, without using recent current AI predictive models such as recurrent neural networks or LSTMs. In [14], an advanced deep learning ensemble method, DSE-XGB, is proposed, combining ANN, LSTM, and XGBoost. DSE-XGB outperformed individual models by integrating strong base learners, capturing solar PV generation dependencies with ANN, and repetitive trends with LSTM. XGBoost acted as a meta-learner, correcting errors, and enhancing prediction accuracy. The approach manages uncertainty effectively, providing consistent and stable predictions across varied datasets, with improvements in R2 values up to 12%. This stacking ensemble algorithm shows promise for broader applications in fields like medicine and finance. The LSTM model used in this study is a basic LSTM model with only one layer, the authors only

use grid search with batch size, epochs, activation, and optimizer factors, and do not use validation techniques.

The above studies all show that solar power capacity forecasting methods using AI are more accurate than traditional methods in the past. In a remarkable 2023 study on solar power capacity forecasting, Tsai et al. [15] synthesized 70 studies published in prestigious journals covering from 2020 to 2023. The authors concluded that deep learning-based forecasting methods are dominant with 34% of studies using this method. Among the application branches of machine learning methods, the most common deep learning network model used is the long-term memory regression neural network model or Long-Short Term Memory (LSTM), which has been shown to be effective in predicting solar power in the short term [16].

According to [17–19], the key input parameters for forecasting solar power include relative humidity, ambient temperature, wind velocity, solar radiation, and temporal indications. Many research works have utilized temporal indicators as significant input data in the models they construct [20–27].

The paper [28] conducts a case study on forecasting using RF (Random Forest), SVR (Support Vector Regressor), CNN (Convolutional neural network), LSTM (Long Short-Term Memory), and Hybrid models. The training dataset configurations include (1) Raw data (Global Horizontal Irradiance, temperature, humidity, wind speed, etc.), (2) Extended data consisting of raw data and computed data derived from the raw data, (3) Original data supplemented with extended computed data and the addition of Hour of the Day as an indicator for time. The conclusion drawn is that the third data set yields the best results, demonstrating the significance of enhancing forecasts by including the Hour of the Day indicator. The supplementation of temporal indicators was also carried out by the authors in [1]. The study incorporated not only Hour of the Day but also the inputs Day of the Year and Minute of the Hour to develop a highly accurate forecasting model.

The description of these temporal indicators is presented in Table 1. Due to solar energy's cyclical nature, by the Earth's daily rotation and annual orbit around the sun, indicators such as the day of the year and the hour of the day are commonly selected as temporal labels. While labels indicating the month or week of the year are rarely considered.

Table 1. Commonly used temporal labels in solar power forecasting problems.

Temporal Labels	Meaning	Use
Day of the year	Reflects the order of days within a year, ranging from 1 to 365	Day 1/1 is typically labeled as 1, while day 31/12 is usually labeled as 365
Hour of the day	Determines values from 1 to 24 h within a day	Can be labeled from 0 to 23 or 1 to 24
Minute of the hour	Determines values from 1 to 60. Depending on the data resolution, it can be labeled accurately by minutes or grouped into intervals	For data with a resolution of 15 min, it can be labeled 0, 15, 30, 45, or labeled as 0, 1, 2, 3

As shown in Figure 1 below, when analyzing the use of such temporal indicators in forecasting, input signals with a sawtooth waveform appear, while the output power profile of solar systems typically follows a bell-shaped curve. It can be noticed that the temporal indicators change in a linear form while the output power changes in a nonlinear form, which might cause a reduction in efficiency when these inputs are chosen for training to build a model to forecast the output power.

Figure 1 below clearly illustrates the value of the inputs when utilizing Minute, Hour, and Day features in comparison to the solar power output during a two-day period.

The utilization of solar irradiance in forecasting as input has also been explored by several researchers. In [29], the authors conducted experiments employing various clear-sky solar irradiance models as input for a solar radiation forecasting model using a persistence approach (a simple forecasting approach that takes a value at the same time as the previous day to serve as a forecast value for the next day).
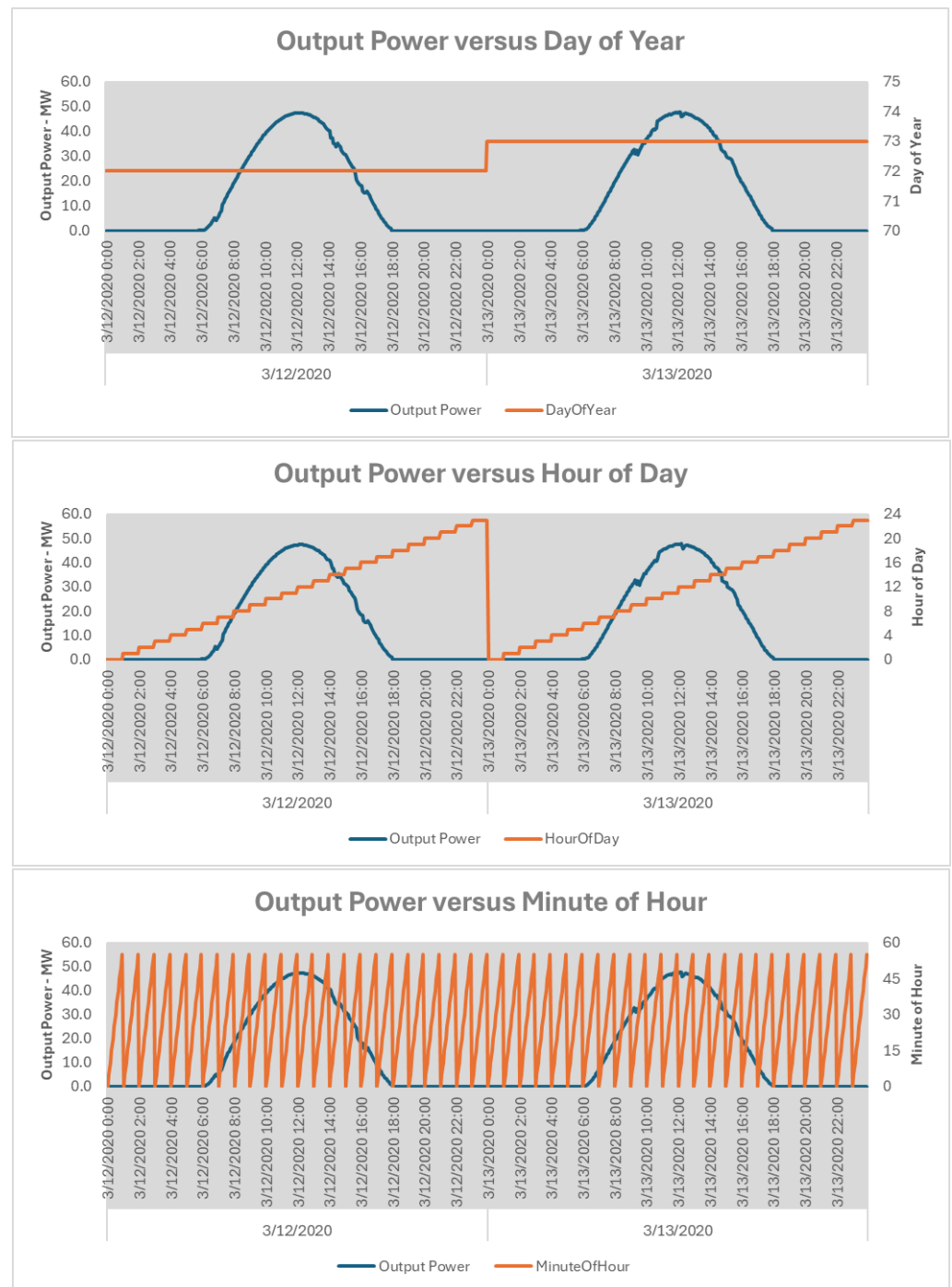
Figure 1. Temporal indicators for a duration of 2 days.

The conclusion drawn was that incorporating solar irradiance as input improved the quality of the forecast. However, this assertion has only been tested with forecasting models utilizing the simplest method, namely the persistence approach, as compared to modern techniques such as artificial neural networks. In [30], the authors employed a neural model to predict input parameters for the BIRD model (a clear sky solar irradiance calculation model) and then utilized conversion formulas to compute the power output. The obtained result indicated relatively good accuracy of the model, although no comparison with other models was provided. In [31], the authors proposed a novel computational method for the clear-sky index of solar systems and subsequently utilized it for forecasting solar systems in neighboring regions. In [32], the authors investigated a model integrating

aerosol and online radiation modules, subsequently assimilating AOD (Aerosol Optical Depth) data from Himawari 8 using 3DVAR (Three-Dimensional Variational) to optimize the forecasting process of solar irradiance in specific regions of China. In [33], the authors suggested utilizing a modified clear-sky model to calculate the solar irradiance of a power plant, using it as input for power forecasting via statistical time series methods. The outcome revealed an approximate 3% improvement in short-term forecast error. Thus, solar irradiance emerges as a potential input to enhance the quality of forecasting models.

However, the state-of-the-art focuses on relatively simple forecasting models, predominantly employing time-series statistical models. As a result, the short-term forecast quality improved by about 3%.

Our research question focuses on whether there is an alternative input that can be used effectively to replace the temporal indicators, such as sets of date and time values, which are commonly employed. In this study, we will propose a methodology that utilizes clear sky solar radiation as a substitute for the usual indices like day, hour, and minute, aiming to enhance the forecast model's effectiveness and reduce forecast errors.

This radiation is not derived from plant operational data but rather stems from the inherent location of the plant and the time of the year. Clear sky radiation can be calculated in advance for any time of the year using a pre-existing computational model that only requires the geographical coordinates of the plant. This value fluctuates based on both the day of the year and the time of day, making it a potential replacement for all three indicators (day, hour, and minute).

Compared to existing approaches using clear sky radiation, this method introduces a novel paradigm shift in solar power forecasting by leveraging clear sky solar radiation as a holistic substitute for traditional temporal indicators such as day, hour, and minute.

Such substitution brings improvements in the accuracy of the forecast and several other positive aspects. Firstly, by utilizing clear sky radiation, the model captures inherent solar energy potential directly linked to geographical coordinates and time, bypassing the need for plant operational data. This not only simplifies data acquisition but also ensures consistency and reliability in forecasting across different locations and timeframes. Secondly, clear sky radiation encapsulates the cyclic nature of solar energy influenced by Earth's orbit and rotation, thereby inherently encompassing temporal dynamics without the need for explicit temporal indicators. This eliminates the linear-versus-nonlinear mismatch observed when using traditional temporal features, potentially enhancing model efficiency and accuracy. Moreover, the use of clear sky radiation allows for the pre-calculation of input values for any time of the year, facilitating proactive forecasting and planning.

This approach opens new perspectives for future research about solar power forecasting because of the proposition of a new input that can replace the temporal indicators inputs, which is clear sky radiation.

Research gaps:

- Identify gaps in existing solar power forecasting models, especially in terms of their capability to take into consideration localized solar irradiance variations and the impact of clear sky conditions.
- Optimize temporal data handling for better forecasting accuracy, especially in areas with high variability in solar radiation.

Main contributions:

- Introduce an innovative approach to significantly enhance forecasting accuracy by using clear sky solar radiation data, replacing traditional temporal inputs such as time of day.
- Develop a methodology that can be adapted to other geographical locations, offering a flexible solution that can improve solar power forecasting globally.
- Validate the proposed model against actual solar power output data, and provide a robust comparison that demonstrates the model's effectiveness.

2. Data and Method

2.1. Dataset

It is crucial to clarify that our approach does not overlook instantaneous cloud effects, which are indeed inherent in practical scenarios. Clear sky radiation, though pre-calculated, serves as a foundational input, and we account for dynamic cloud cover by preserving the actual measurements of GHI. Therefore, our methodology considers both the advantages of utilizing clear sky radiation and the need to incorporate real-time atmospheric conditions, ensuring a comprehensive evaluation of the forecasting model's performance under practical circumstances.

The dataset is derived from a 48MWAC solar power facility situated in Vietnam, spanning from May 2019 to the end of May 2020. The information obtained via the plant's metering system encompasses the power generated by the facility, denoted as P in [MW]; solar radiation, referred to as GHI in [W/m^2]; ambient temperature, indicated as TEMP in [°C]; WIND [m/s] is wind velocity at the power plant location; and HUM [%] is relative humidity at the plant site. The data possesses a granularity of 5 min, wherein each data point within the dataset represents the average measurement over a 5 min interval for each of the mentioned parameters. The comprehensive dataset is effectively illustrated in Figure 2.
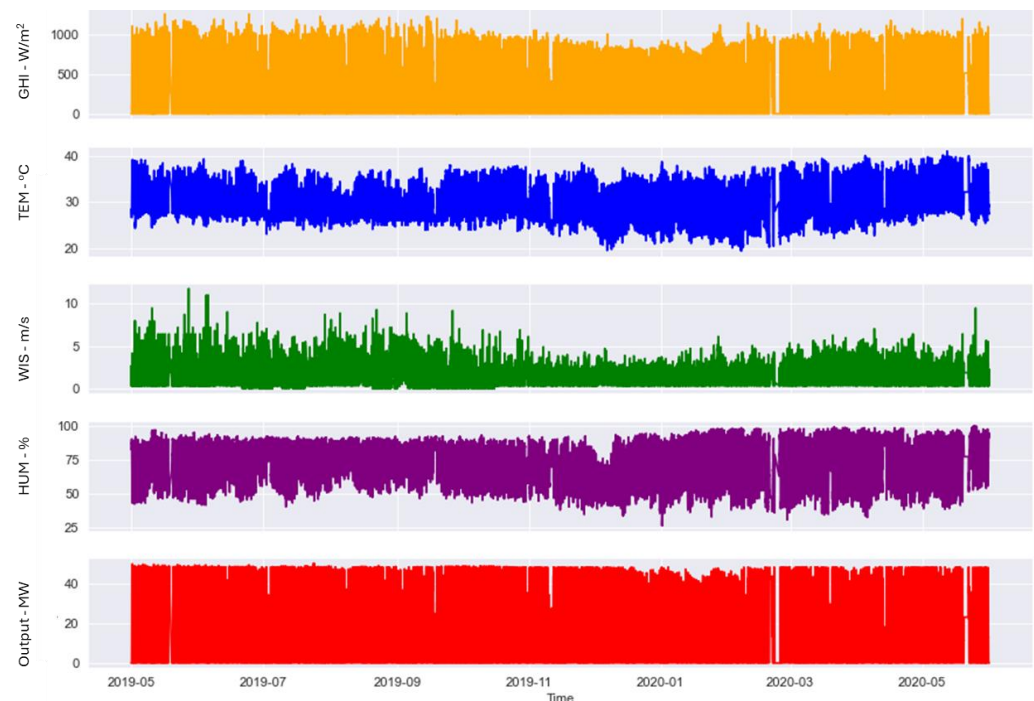

Figure 2. The measured dataset of the solar power plant.

2.2. Solar Radiation

Solar radiation is defined as "the amount of electromagnetic energy incident on a surface per unit time and unit area" [34]. The measurement unit of solar radiation is W/m^2.

The energy emitted by the sun travels through space until it is intercepted by planets, celestial bodies, or gases and particles between the stars. The intensity of solar radiation incident on these objects is governed by the inverse square law. When sunlight reaches the Earth, it is partially blocked by the outer layer of the atmosphere, resulting in a certain portion being reflected back. The amount of solar radiation received on the surface of a solar panel at ground level varies depending on different weather conditions, but on average, it is approximately 40% [34].

As solar radiation passes through the Earth's atmosphere, it interacts with its constituents. Clouds, water droplets, and dust particles contribute to reflection, while com-

ponents such as ozone, oxygen, carbon dioxide, and water vapor significantly absorb radiation within specific frequency ranges. These interactions result in the separation of solar radiation incident on the receiving object, in this case, solar panels, into distinct and discernible components.

Direct radiation, also known as Beam Radiation, is formed by the unreflected or unscattered rays that travel in a straight line from the sun to the surface of the solar panel. This radiation is referred to as DNI (Direct Normal Irradiance). Diffuse radiation, received by the surface of the solar panel, results from indirect light sources from the sky and does not include direct radiation. Albedo radiation is the reflected radiation from the Earth's surface.

The overall synthesized radiation received by the surface of the solar panel consists of DNI, Diffuse Radiation, and Albedo, forming the GHI measured in W/m^2.

Figure 3 provides a detailed depiction of the radiation components that a solar panel receives.
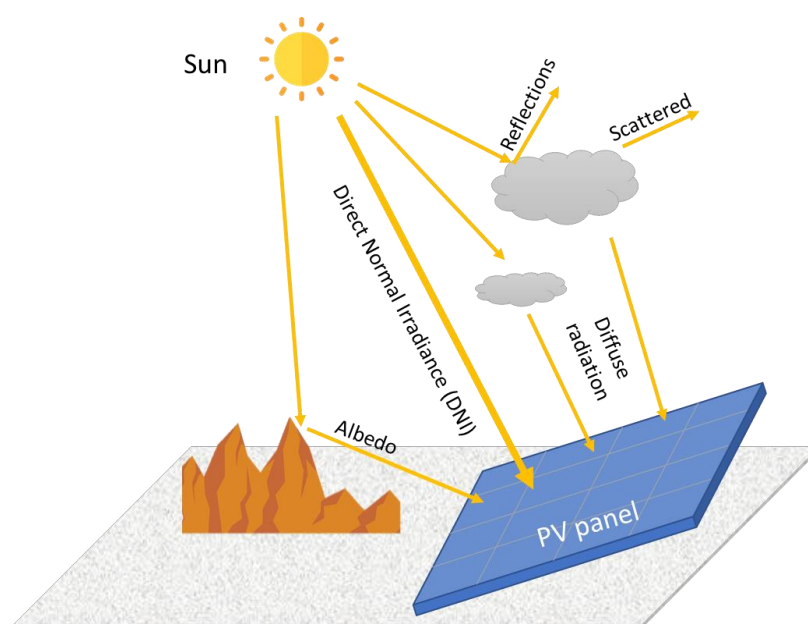

Figure 3. The radiation components received by a solar panel.

2.3. Clear Sky Solar Radiation

Clear sky solar radiation refers to the amount of solar radiation that reaches the Earth's surface under unobstructed, cloud-free conditions. It represents the maximum potential solar radiation that can be received at a specific location on Earth. Calculating clear sky solar radiation at a solar power plant installation site involves employing various models and methods [35]. The research [36] presents a comparative assessment of various models. Some notable models include:

- The Ineichen model, which estimates clear sky radiation based on atmospheric parameters such as the water vapor content, ozone concentration, and aerosol optical depth. This model considers the position of the sun, geographical location, and time of year to calculate the clear sky radiation [37,38].
- The Haurwitz model, which is another approach used to calculate clear sky solar radiation. It utilizes the solar zenith angle, latitude, and time of year to estimate the radiation. This model considers the position of the sun in relation to the site and accounts for the Earth's curvature. According to one report [39] on clear sky solar radiation models, the Haurwitz model demonstrates superior performance among the models that solely rely on the zenith angle.
- The simplified Solis model, which is employed for estimating clear sky radiation. This model takes into account the solar zenith angle and the site's latitude to calculate the

clear sky radiation. The model's accuracy has been reported as 15, 20, and 18 W/m^2 for the components of GHI and direct normal irradiance (DNI) [40].

- The Bird Clear Sky Model, which combines the Bird solar radiation model with clear sky conditions to estimate the Clear Sky GHI. It considers variables such as atmospheric water vapor concentration, ozone concentration, and spectral depth of airborne particles [35].

The calculated values provide vital information for evaluating the maximum power generation potential of a solar power plant at any geographical location throughout the year. According to the comparative analysis report on clear sky radiation models, the Ineichen/Perez model demonstrates exceptional performance with minimal input data [39]. Therefore, this study will employ this model to compute the clear sky solar radiation at the plant's location.

2.4. Long Short-Term Memory Networks (LSTM) Method

The Long Short-Term Memory (LSTM) is a distinctive type of Recurrent Neural Network (RNN) capable of capturing both short-term and long-term dependencies [41]. Unlike feed-forward networks, RNNs are designed to capture the variability of data over time. This variability is managed through feedback loops that affect the architecture of an RNN. For these networks, the number of neurons depends on the number of past sample data points used to make the prediction. Information only flows in one direction in a feed-forward neural network: from the input layer to the output layer via the hidden layers. The data flows directly across the network.

Feed-forward (FF) neural networks are poor at anticipating future events and lack memory of the data they receive. They do not understand the order of inputs over time as they only consider the current input. They are incapable of remembering anything beyond what they have been explicitly taught. In contrast, RNNs loop information through the network. When making a choice, an RNN takes into account both the current input and the information learned from previous inputs. Figure 4 below describes this difference.
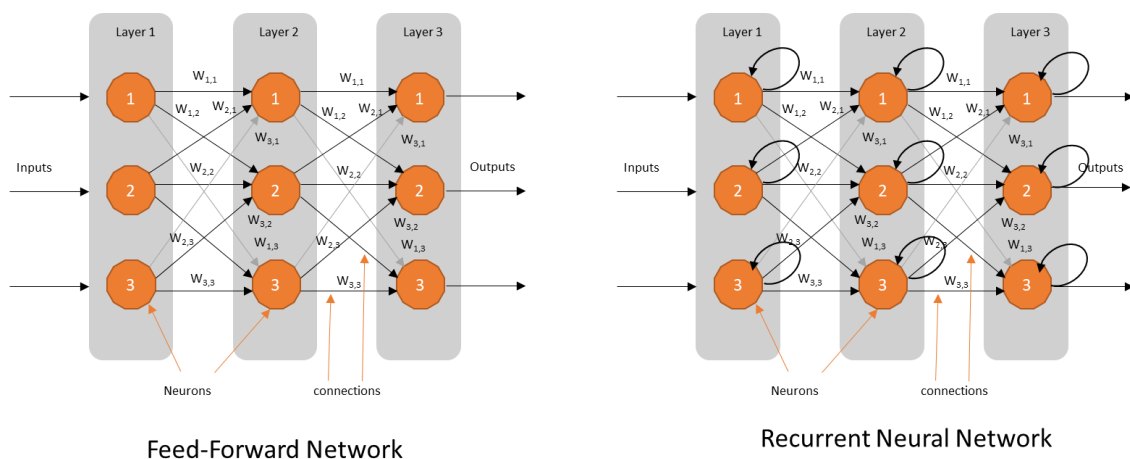

Feed-Forward Network   Recurrent Neural Network

Figure 4. Architectures of RNN vs. FF neural networks.

An RNN typically has a short-term memory. They possess both long-term and short-term memory when implemented as LSTM. An RNN's internal memory allows it to recall short-term data. It generates output, duplicates it, and then feeds it back into the network.

FF ANNs and RNNs can use backpropagation and gradient-based learning. However, when the values of a gradient are too small, the model may stop learning or take way too long to produce meaningful results. This phenomenon is called a 'vanishing gradient'. LSTMs have been engineered to address the latter challenge in traditional RNNs. This problem indeed makes the training process tougher and harder, and can sometimes hinder further advancement in training the network.

In the same way, the LSTM also follows the sequence structure of the above RNN, but the LSTM block's repeating module is more complex. To bring forward even older memories, the LSTM blocks have memory cells connected across layers. Additionally, they are made up of various interacting layers that regulate the blocks' selective information flow. Figure 5 illustrates the structure of a standard LSTM block.
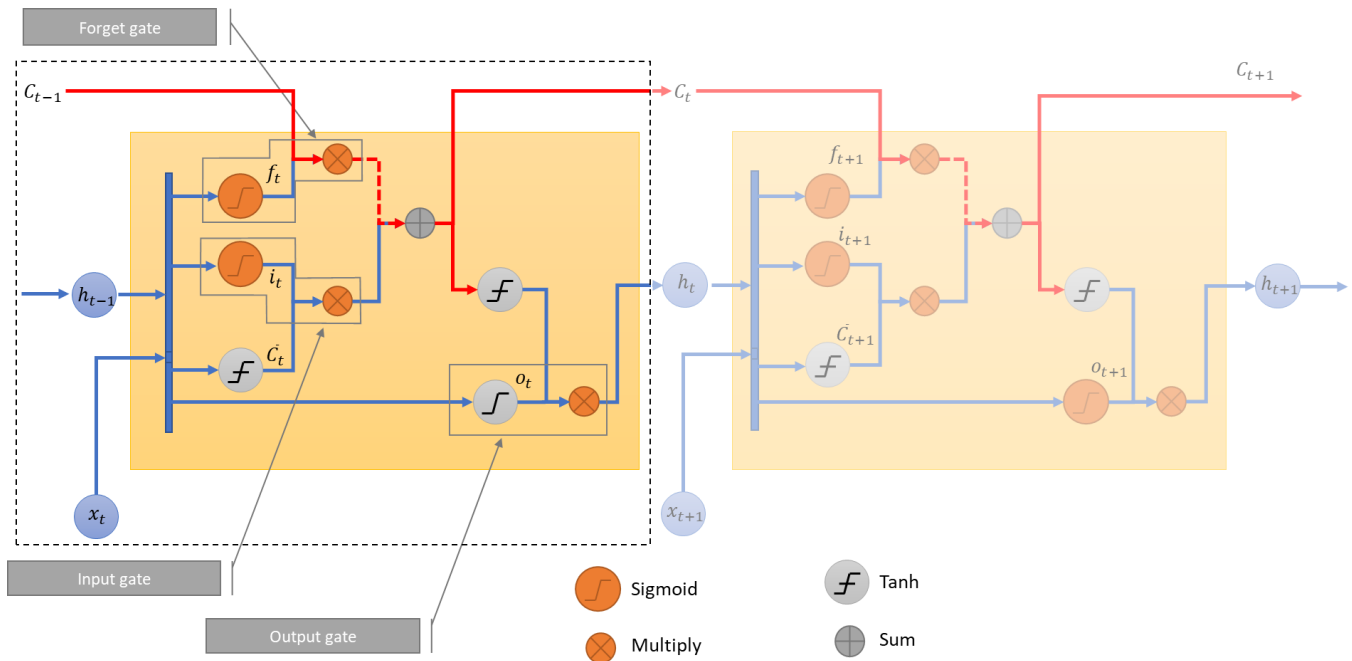
Figure 5. Structure of a standard LSTM block.

As already mentioned, the LSTM network is designed to address the vanishing gradient problem by utilizing gates to regulate the amount of information transmitted throughout the network. Each LSTM block consists of a memory cell and gates. These gates allow the LSTM unit to decide which information should be retained in the memory, which information should be discarded, and which information should be passed on to other units within the network. The forget gate determines which information should be removed from the memory cell to forget irrelevant history, as some information may be deemed unimportant. The input gate determines which portion of the new input information is relevant and should be stored in its memory cell. The output gate determines the block's output based on the current input information and the memory state. Each gate consists of a sigmoid network layer and a multiplication operation aimed at filtering the information passing through that gate.

- The input to the sigmoid network layer of each gate consists of the previous step's internal output state, h_{t-1} and the current step's input value, x_t.
- The output of the sigmoid network layer for each gate undergoes a multiplication operation with a different piece of information to obtain the final output result, specifically:
  - Forget Gate: The output of the sigmoid function is multiplied by the previous step's memory cell value, C_{t-1}.
  - Input Gate: The output of the sigmoid function is multiplied by the value $\overline{C}_t$ (which is obtained from a *tanh* function with inputs h_{t-1} and x_t).
  - The value of the memory cell, C_t is updated by adding the output values of the forget gate and the input gate.
  - Output Gate: The output of the sigmoid function, o_t, is multiplied by the value of the memory cell (after being processed by the *tanh* function). The final result obtained, h_t, represents the output state of the current step.

One of the most important features of LSTM is that its various gating and updating processes work together to create the internal cell state *C*, which allows gradients to pass through smoothly over time. This functions as a kind of expressway for cell states, alleviating and mitigating the vanishing gradient issue. The LSTM unit's information flow can be expressed as follows:

$$f_t = sigmoid\Big(W_f[x_t, h_{t-1}] + b_f\Big) \tag{1}$$

$$i_t = sigmoid(W_i[x_t, h_{t-1}] + b_i) \, , \tag{2}$$

$$\overline{C}_t = tanh(W_C[x_t, h_{t-1}] + b_C) \tag{3}$$

$$C_t = \big(i_t \otimes \overline{C}_t\big) \oplus f_t \otimes C_{t-1} \tag{4}$$

$$o_t = sigmoid(W_o[x_t, h_{t-1}]) + b_o \tag{5}$$

$$h_t = o_t \otimes tanh(C_t) \tag{6}$$

where:

x_t: input value of time step t
h_t, h_{t-1}: output value of time step t
f_t: output of the forget gate
W_f, b_f: weight function of forget gate
i_t: output of the input gate
W_i, b_i: weight function of the input gate
$\overline{C}_t$: vector of new candidate values for time step t
W_C, b_C: weight function of tanh layer to calculate $\overline{C}_t$
C_t, C_{t-1}: cell state of time step t, t − 1
o_t: output of the sigmoid layer
W_o, b_o: weight function of the sigmoid layer

In recent studies, authors typically employ inputs comprising measured Global Horizontal Irradiance (GHI), along with other meteorological factors, in addition to time-related indicators. In our research, we depart from this convention by replacing only the time-related indicators with pre-calculated clear sky radiation, while retaining the other meteorological factors unchanged. This substitution allows us to assess the effectiveness of utilizing clear sky radiation in lieu of time-related indicators.

3. Simulation

3.1. Simulation Setup

In this study, the simulation conditions were carefully selected to reflect realistic solar power generation scenarios. The following key parameters were included:

- Location of solar power plant: The solar power plant is located in a southern province of Vietnam. This specific location was chosen due to high solar irradiance and the availability of data.
- Meteorological parameters:
    - Data source: The study used Global Horizontal Irradiance (GHI), temperature, wind speed, and humidity data obtained from ground-based measurements.
    - Measurement period: The data covers a period from May 2019 to the end of May 2020, ensuring comprehensive coverage of various weather conditions.

These parameters were chosen to ensure that the simulations closely match real-world conditions, thereby enhancing the accuracy and applicability of the solar power forecasting model.

3.2. Data Preprocessing

The dataset is acquired from a solar power plant within the southern region of Vietnam. The length of the dataset is 13 months from May 2019 to June 2020, which includes the features mentioned in Section 2.1.

Following the acquisition of the raw data, the next phase involves data preprocessing, which is a critical step in the data mining process for constructing the forecasting model [42]. This preprocessing stage includes the following primary tasks: feature extraction, data cleaning, and feature transformation.

3.2.1. Feature Extraction

In forecasting solar power generation, the extracted data for model construction includes actual power output and real meteorological data. Additionally, static data pertaining to the solar power plant, such as installed capacity and geographical location, are also utilized in the calculation of input variables.

3.2.2. Data Cleaning

To ensure the quality of the input data, we performed rigorous data cleaning. This involved removing any outliers and handling missing values. By refining the dataset, we minimized the risk of inaccurate predictions caused by erroneous or incomplete data. Historical datasets used for solar power forecasting often contain temporary gaps and grid incidents. These outliers, which lack a trend and are influenced by random events, significantly impact the forecast. Furthermore, the data can sometimes be erroneous or incomplete due to sensor errors or signal transmission issues. Therefore, preprocessing the input data to address errors through techniques such as outlier detection, interpolation, or seasonal adjustment (i.e., data cleaning and restructuring) is of utmost importance [43].

In this study, error detection techniques were employed based on previously published methods utilized in studies [17,44]. These techniques include the following:

- Outlier detection using the Interquartile Range (IQR) technique, combined with the incorporation of new features (the ratio P/GHI, where P is Output Power and GHI is Global Horizontal Irradiance), and the GHI clustering technique to enhance the effectiveness of anomaly detection.
- Utilizing the Q2 (median) to replace abnormal or missing values.

The effectiveness of the method on the dataset is presented in Figure 6, where (a) represents the data before cleaning, and (b) represents the data after cleaning. The results clearly show that the data has been cleaned effectively, with most of the abnormal data points being removed and processed, resulting in a more reliable dataset for further analysis.

It can be observed from the plot that anomalous points, where the irradiance is high but the power output is low (lower half of the chart), or where the irradiance is low but the power output is unusually high (upper half of the chart), have been successfully removed.

In evaluating the effectiveness of data cleaning represented in Figure 6b, the authors carefully examined data points where solar irradiance and power output showed unexpected differences. Specifically, the authors focused on instances where high irradiance levels corresponded to unusually low power output (as seen in the lower half of the chart) and vice versa, where low irradiance levels coincided with abnormally high power output (as observed in the upper half of the chart).

The successful removal of these anomalous data points indicates that the cleaning process efficiently addressed potential inaccuracies or outliers. This thorough approach ensures that the dataset used for subsequent analysis and modeling is more reliable and representative of the actual dynamics of solar power generation.
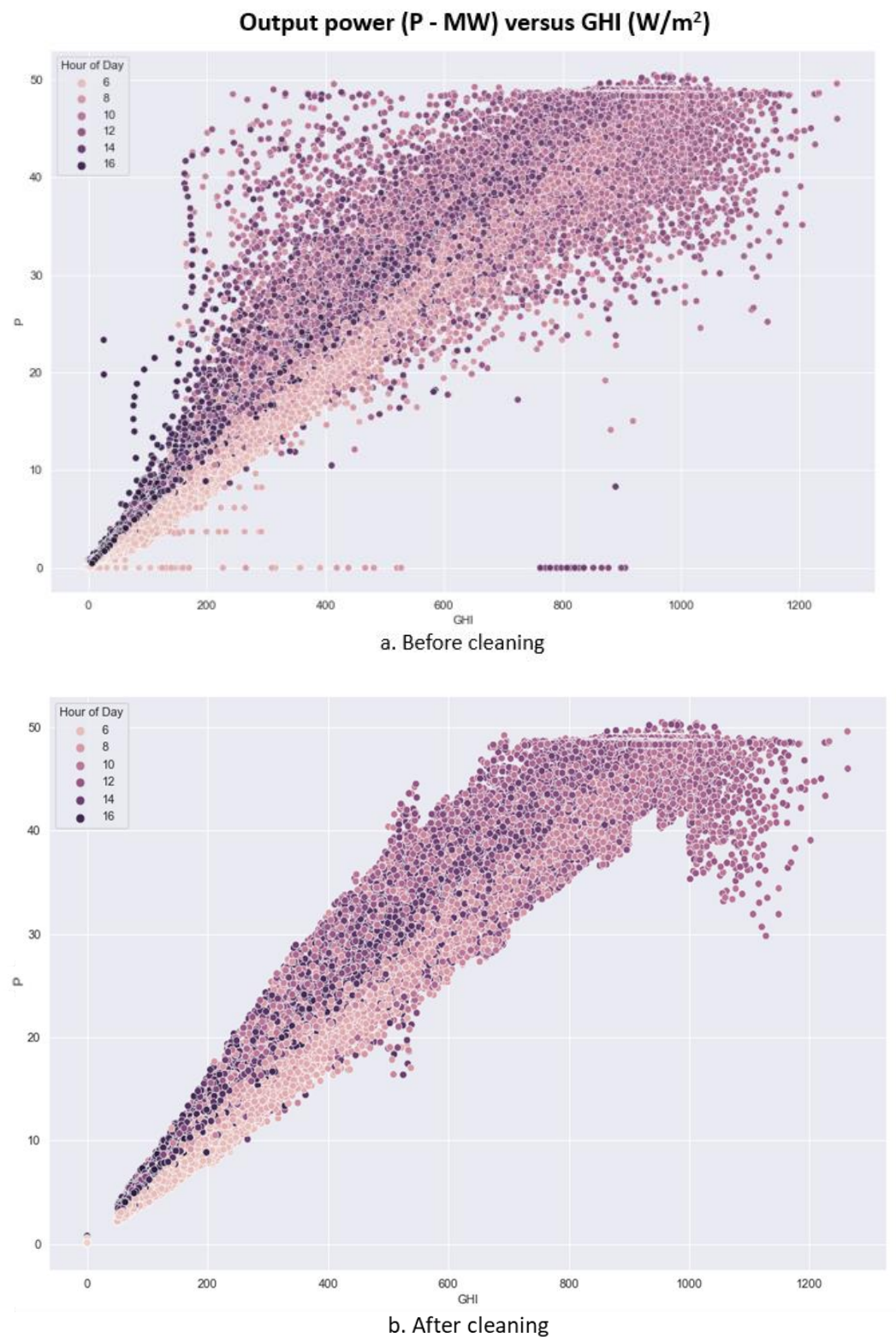
Output power (P - MW) versus GHI (W/m²)

a. Before cleaning

b. After cleaning

Figure 6. The effectiveness of the cleaning method on the dataset.

3.2.3. Feature Transformation

We created additional features from the raw data, which are known to significantly impact solar power generation. These engineered features enrich the model by providing it with more relevant information, leading to improved forecasting performance. From the selected basic features, some studies have performed transformations on those features to

achieve improvements. Examples include the study [45], which used statistical tools to remove seasonal trends from solar radiation data, and the studies [46,47] which utilized the Fourier series transformation to capture the cyclic pattern in solar radiation data. There have also been techniques proposed to determine the trend in solar radiation data, as it is complex to accurately determine the daily solar radiation trend due to daily weather behavior. Although clear sky GHI at a fixed location may be considered as fixed, the rest of the input data used in computing the forecast power output is affected by the error in the forecast meteorological data used in computing prediction.

A notable research direction is the incorporation of weather conditions classification as an input factor to enhance the forecasting quality. However, recent studies have yet to reach a consensus and consistency on how to classify daily weather conditions. In terms of the impact on forecasting results, studies [48–51] suggest that better forecasting results can be achieved, especially for the next-day (24 h ahead) forecast, with a carefully designed weather condition classification system. The reason, as explained, is that individual forecasting models that do not utilize weather condition classification are significantly influenced by historical data from the past three days and lack the ability to forecast weather conditions one day ahead.

In this study, as detailed in the introduction, the focus is on replacing the conventional time-based indicator dataset with an alternative input dataset, specifically solar irradiance. Instead of using traditional time-based indicators, this research incorporates calculated clear sky solar irradiance from the Ineichen/Perez model at the plant's geographical location as the forecasting input.

The tool used for calculating solar irradiance is PVLIB version 0.10.4 (19 March 2024), an open-source software library in the field of solar energy. PVLIB provides tools for computing and simulating the performance of solar power systems, including parameters such as electricity generation, solar radiation, and sun angle. Developed in Python, PVLIB is used for analyzing and forecasting the performance of photovoltaic (PV) systems [52].

A specific example of computing GHI at the solar power plant location compared to the measured GHI using the plant's devices is presented in Figure 7. It illustrates that the calculated clear sky GHI obtained from PVLIB is a reasonable approximation when compared to the measured GHI.
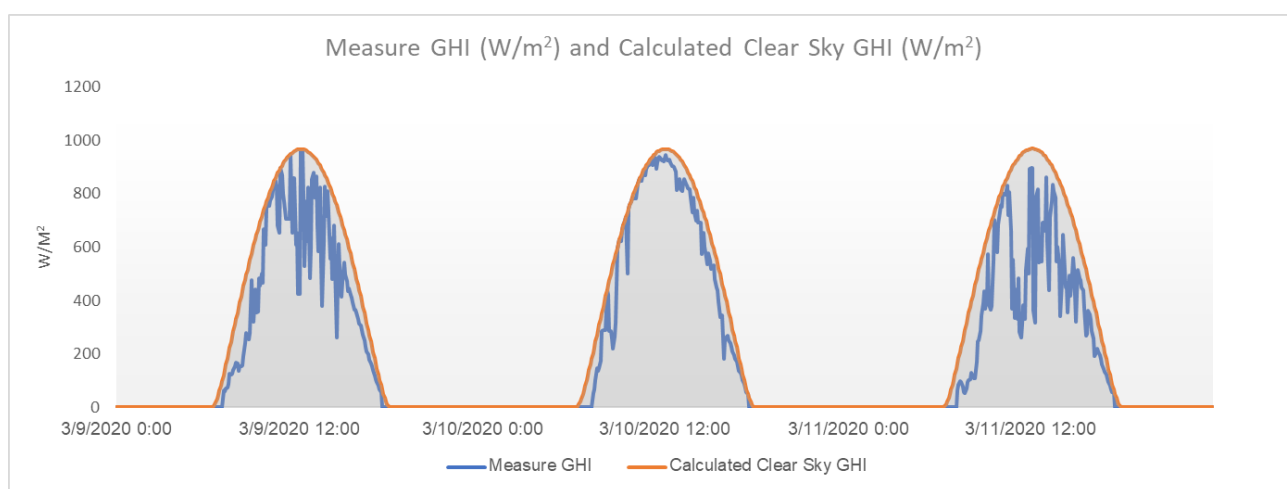

Figure 7. Measure GHI (W/m^2) and Calculated Clear Sky GHI (W/m^2).

As mentioned previously, there are numerous models published by scientists for calculating clear sky GHI, each with different inputs and methodologies. Due to limitations in the available data, the authors chose the Ineichen/Perez model with the simplest inputs, incorporating only the geographical location of the solar plant, while other parameters were assumed within the Ineichen/Perez model. The purpose of presenting this estimated

GHI alongside the measured GHI is to demonstrate that the estimation is reasonable and suitable for use as an alternative to traditional time-related indicators.

This study opted for clear sky solar irradiance over traditional time-related parameters due to several key reasons that enhance the model's accuracy and applicability:

- Relevance and direct impact: Clear sky irradiance directly measures the potential solar energy available at a specific location and time, making it a more relevant predictor for solar power output. Unlike traditional time-related parameters, which are indirect indicators and may not capture local environmental variations, clear sky irradiance provides a direct assessment of solar potential.
- Reducing model complexity: By using clear sky irradiance, we reduce the complexity of the model. Time-related parameters often require additional processing to capture cyclical patterns effectively (e.g., hour of the day and day of the year). In contrast, clear sky irradiance can be used as a continuous variable that inherently includes time-of-day and seasonal effects on solar radiation, streamlining the model architecture.
- Enhancing model performance: Our empirical tests showed that models using clear sky irradiance as an input consistently outperformed those using traditional time-related parameters. Specifically, the Mean Absolute Percentage Error (MAPE) and Root Mean Square Error (RMSE) were significantly lower, indicating more accurate predictions.
- Generalizability: Clear sky irradiance models, such as the Ineichen/Perez model used in our study, are adaptable to different geographic locations without substantial recalibration. This makes the approach more scalable and applicable in diverse settings compared to models heavily reliant on specific time patterns that may vary greatly across regions.

3.3. Train Set and Test Set

The collected dataset is partitioned into two distinct segments as follows:

- Training set: Data spanning from May 2019 to April 2020 is utilized to train the forecasting models.
- Test set: Data from May 2020 is employed to compute the forecasted output capacity and compare it against the measured output, thereby assessing the efficacy of the forecasting model.

3.4. Evaluation of Error

Numerous studies have been conducted to develop forecasting models for solar power capacity. However, evaluating the effectiveness of different forecasting methods remains relatively complex, as these studies employ various metrics as individual criteria. Evaluation and measuring the magnitude of errors play a vital role in assessing the accuracy of forecasting models. From several different studies [53–56], commonly used groups of criteria can be observed, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), and Root Mean Square Percentage Error (RMSPE). According to [18,56–58], the formulas used for evaluating these error criteria are as follows:

$$MAE = \frac{1}{N}\sum_{k=1}^{N}(P_M - P_P) \tag{7}$$

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{|P_M - P_P|}{P_{Rate}} \times 100\% \tag{8}$$

$$MSE = \frac{1}{N}\sum_{k=1}^{N}(P_M - P_P)^2 \tag{9}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{k=1}^{N}(P_M - P_P)^2} \tag{10}$$

$$RMSPE = \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}(P_{db} - P_t)^2}}{P_{Rate}} \qquad (11)$$

where P_M, P_P, and P_{Rate} refer to the actual power output collected through measurement devices, the forecasted output power, and the rated capacity of the plant, respectively. N is the number of sample points.

To align with the methodology used for calculating error metrics as presented in our previous paper in 2021, in this study, errors will be computed exclusively for daytime hours, spanning from 5:00 a.m. to 6:00 p.m.

3.5. Experiments Setup

Besides the architecture of LSTM neurons (number of layers and number of hidden units per layer), the structure of the input data is also a critical aspect of the predictive model's architecture. Altering the structure of the input data by varying the number of features results in different predictive models. While the LSTM network configuration remains unchanged, the training techniques exhibit some improvements across the different models.

According to [1], the authors conducted experiments using data from an industrial-scale solar power plant and developed an approach utilizing the LSTM artificial neural network, along with methods for parameter selection. In this study, we replicated these experimental conditions and introduced new alternative inputs for temporal data. Figure 8 details the configurations for training the three models using the same training data: Basic Model, Model 1, and Model 2. The Basic Model is the one proposed in [1], while Model 1 and Model 2 are tested in this work. Table 2 illustrates the actual measured data used during the experimental process. Details of the data employed for model training and forecasting are presented in Table 3.
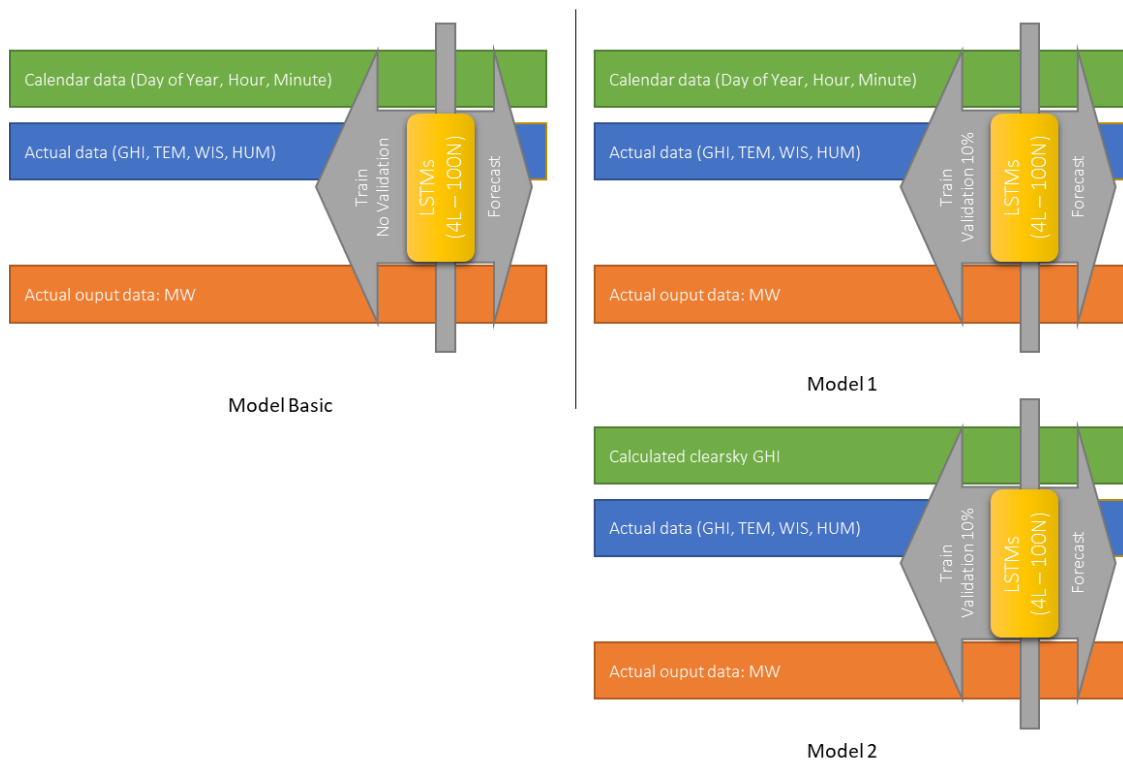

Figure 8. Description of Model Basic, Model 1, and Model 2.

Table 2. Details of actual measured data.

Index	GHI	TEMP	WIS	HUM	Output
count	114,336	114,336	114,336	114,336	114,336
mean	222.73	29.21	1.45	73.94	12.34
std	310.49	4.00	1.18	16.14	16.65
min	0	19.41	0	25.97	0
25%	0	26.1	0.33	59.22	0
50%	0.26	28.11	1.25	78.39	0.13
75%	424.49	32.41	2.16	87.95	24.75
max	1263.53	41.02	11.77	100	48

GHI: Global Horizontal Irradiance (W/m^2), TEMP: Temperature (°C), WIS: Wind speed (m/s), HUM: Humidity (%), std: Standard Deviation.

Table 3. Details of the data employed for model training and forecasting.

Model Name	_Basic Model [1]	Model_1	Model_2
Data	The dataset consists of 13 months of solar power plant data. The first 12 months are utilized for model training, while the final month is kept for testing.	The same dataset as used for the Basic Model.	The same dataset as used for the Basic Model
Input and Output	The input data comprise calendar data (day, hour, minute) describing time t and meteorological data (GHI, TEMP, WIS, HUM) at time t, while Outputs are the data points at t + 5 min.	The same features for training and testing as in study [1] were used for the Basic Model.	The inputs are Clearsky GHI at time t and Meteorological Data (GHI, TEMP, WIS, HUM) at time t, while Outputs are the data points at t + 5 min.
LSTM network configuration	4-layer LSTM network with 100 hidden nodes per layer (4L 100N).	4-layer LSTM network with 100 hidden nodes per layer (4L 100N).	4-layer LSTM network with 100 hidden nodes per layer (4L 100N).
Training	Training using MAE (Mean Absolute Error) loss function, Rectified Linear Unit (ReLU) activation function, Adaptive Moment Estimation (Adam) optimizers, 50 epochs.	Same as the Basic Model but used Early Stopping technique to reduce training time.	Same as the Basic Model but used Early Stopping technique to reduce training time.
Training	No Validation	Validation 10%	Validation 10%
Evaluation Indicator	MSE, RMSE, MAE, MAPE	MSE, RMSE, MAE, MAPE	MSE, RMSE, MAE, MAPE

LSTM: Long Short-Term Memory, MSE: Mean Square Error, RMSE: Root Mean Square Error, MAE: Mean Absolute Error, MAPE: Mean Absolute Percentage Error.

3.6. Comparision Results

Figures 9 and 10 show the loss (Mean Absolute Error, MAE) during the training of Model 1 and Model 2. In these figures, "Train loss" represents the loss (MAE) during the training phase of the models, while "Valid loss" represents the loss during the validation phase [27]. The comparison of prediction errors on the test set is reported in Table 4.

It can be seen that the Model 1 and Model 2 results are relatively good.

- Model 1: MAPE reduces from 3.491% to 3.08% which is 12% better than the Basic Model's performance. The RMSE value obtained is lower by about 0.879 MW, an improvement of 29%, reflecting that the large-value errors of Model 1 were less than those in the Basic Model.

- Model 2: MAPE reduces from 3.491% to 2.766% which is 24% better than the Basic Model's performance. The RMSE value obtained is lower by about 0.991 MW, an improvement of 45%, reflecting that the large-value errors of Model 2 were less than those in the Basic Model and in Model 1.

While it is true that the Basic Model has been optimized, the proposed models, particularly Model 2, achieve considerable improvements of 24% in Mean Absolute Percentage Error (MAPE) and 45% in Root Mean Squared Error (RMSE).
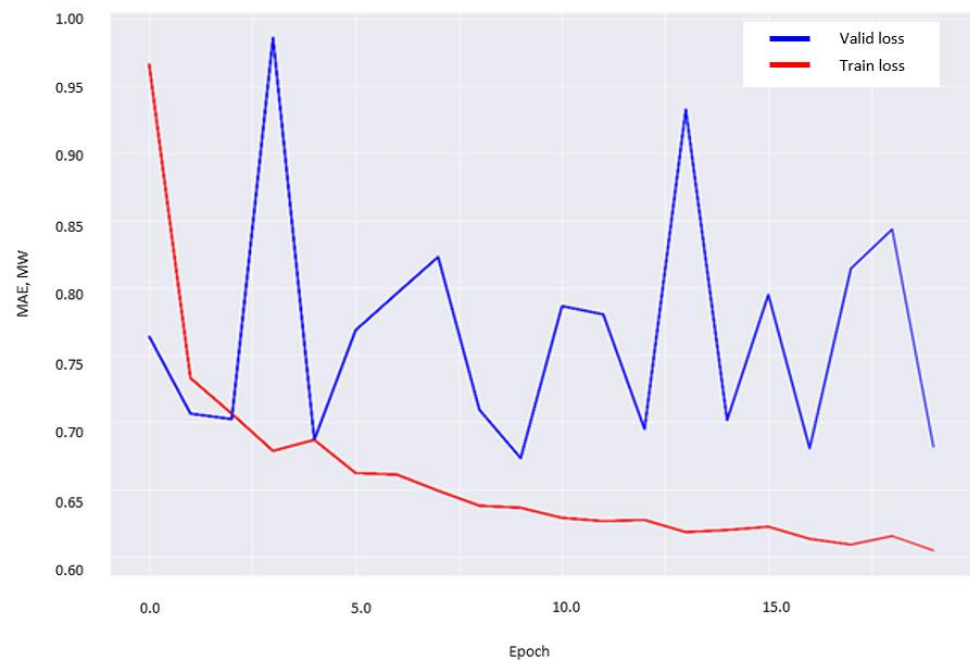
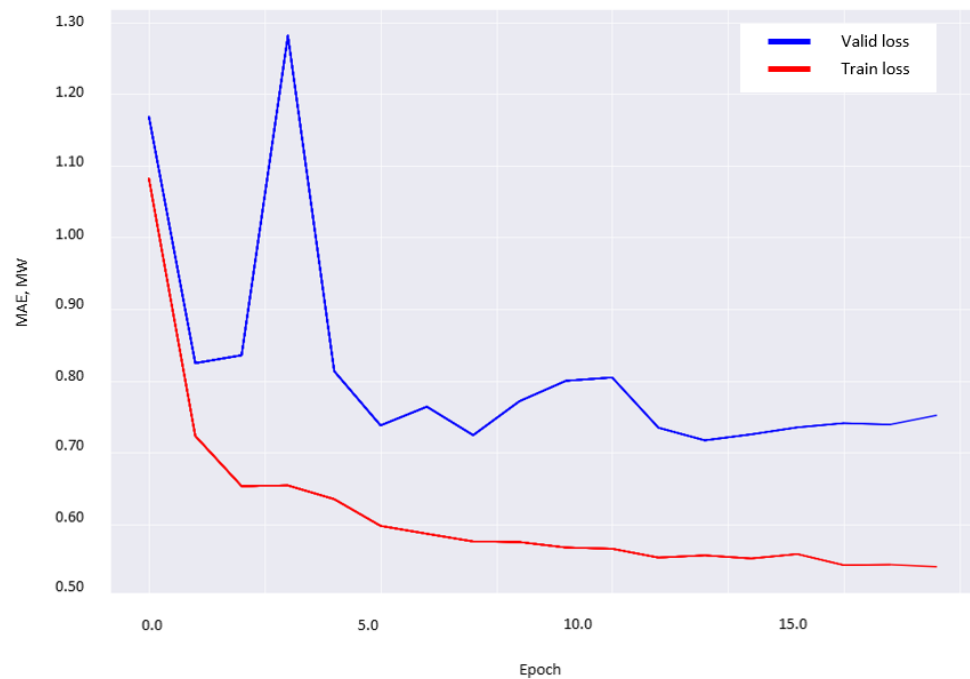**Figure 9.** Training and valid loss (MAE) during training epochs of Model 1.

Figure 10. Training and valid loss (MAE) during training epochs of Model 2.

Table 4. Comparison of prediction errors on the test set.

Error Type	MAE	MAPE	MSE	RMSE
Measurement Unit	MW	%	MW^2	MW
Basic Model	1.676	3.491	9.499	3.082
Model 1	1.478	3.08	4.853	2.203
Model 2	1.328	2.766	4.371	2.091

In real-world scenarios, even seemingly small percentage improvements in forecasting accuracy can translate into meaningful operational and economic benefits. The solar power

industry is particularly sensitive to accurate forecasting due to the intermittent nature of renewable energy sources. While simulation studies provide controlled environments for testing, the real-world implications of improved accuracy extend beyond the laboratory. Our aim is not only to enhance the forecasting model's performance but also to contribute to the broader goals of increasing the reliability and efficiency of solar power generation. We believe that the observed improvements, when applied in practical settings, will have tangible and positive effects on the operational dynamics of solar power plants and the broader energy ecosystem.

Figure 11 illustrates the error distributions on the test set among the forecast results of the models. The detailed power outputs of each 5 min interval of the measured value and forecast value from Model 2 are as in Figure 12.
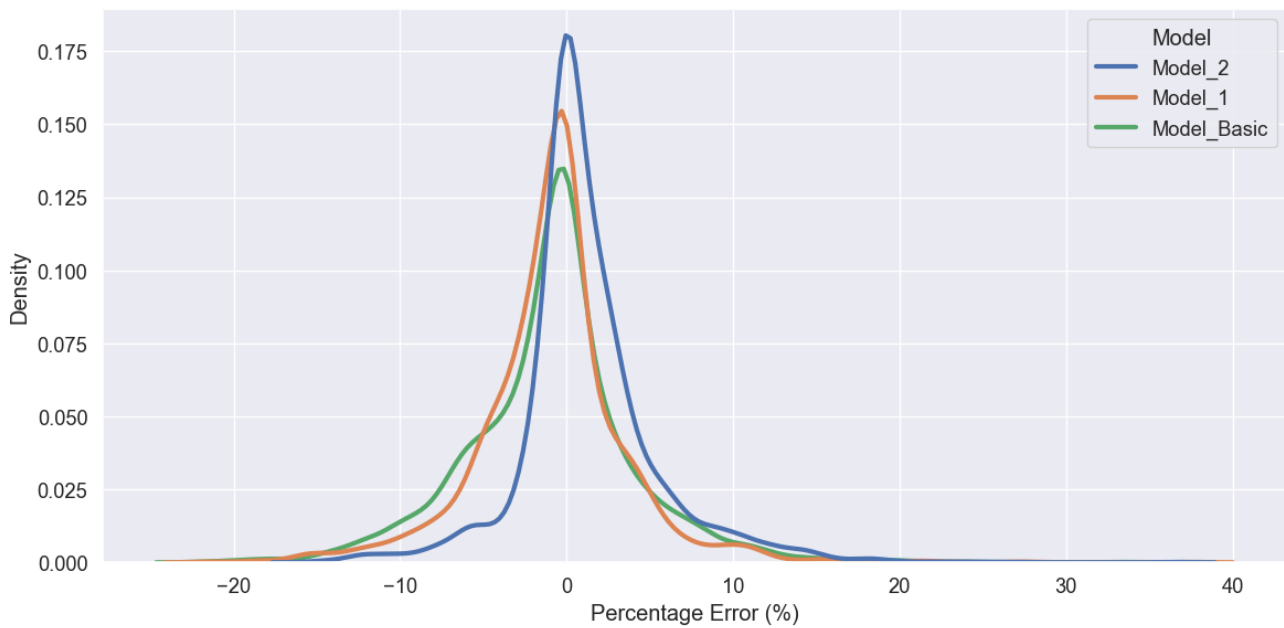

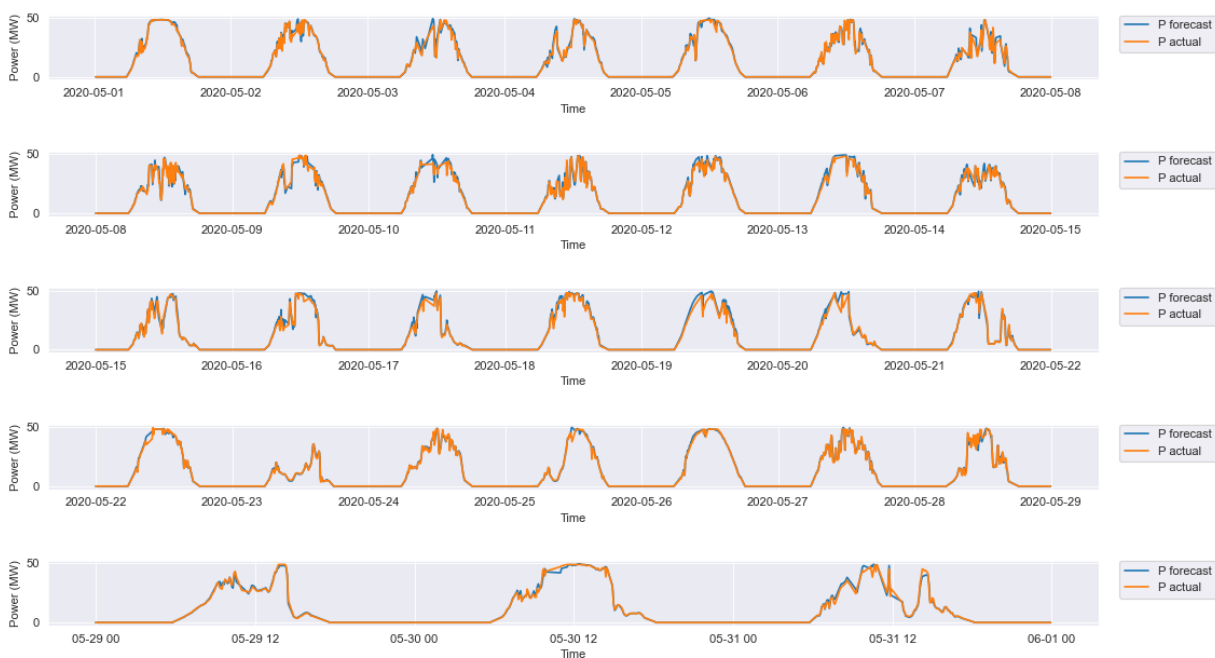**Figure 11.** Distribution of percentage error among forecast results of models.

Figure 12. Measured power versus forecasted power of Model 2.

4. Conclusions

This research successfully addressed the initial research question posed at the outset: whether an alternative input exists capable of effectively replacing conventional temporal indicators, such as date and time values, commonly utilized in short-term power output forecasting models for solar power plants. The methodology proposed in this study leveraged clear sky solar radiation as a substitute for traditional indices like date, hour, and minute. The primary objective was to enhance the effectiveness of the forecast models and minimize forecast errors. The results underscore the conclusion that the integration of clear sky solar irradiance inputs can significantly contribute to enhancing the accuracy of solar power output forecasts.

Moreover, this study does not merely present promising experimental outcomes but also provides valuable insights into the potential applicability of the proposed models in real-world scenarios. The research aligns with the broader objective of advancing solar power forecasting methodologies. Despite these achievements, it is essential to acknowledge certain limitations inherent in the current methodology and offer suggestions for future improvements. By doing so, we contribute to the ongoing dialogue in the field, fostering a deeper understanding of the proposed methodology's robustness and potential for broader application.

Author Contributions: Conceptualization, N.N.Q., L.B.D. and B.D.V.; methodology, N.N.Q., L.B.D. and B.D.V.; software, N.N.Q., L.B.D. and Q.T.T.T.; validation, N.N.Q., E.R.S. and H.L.T.T.; formal analysis, H.L.T.T. and S.L.Q.; investigation, L.B.D. and Q.T.T.T.; resources, L.B.D., T.L.C., S.L.Q. and H.C.T.T.; data curation, L.B.D., T.L.C., S.L.Q. and H.C.T.T.; writing—original draft preparation, N.N.Q., L.B.D. and E.R.S.; writing—review and editing, N.N.Q., L.B.D., H.L.T.T. and E.R.S.; visualization, N.N.Q. and Q.T.T.T.; supervision, B.D.V. and E.R.S.; project administration, N.N.Q. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Nguyen, N.Q.; Duy Bui, L.; Van Doan, B.; Sanseverino, E.R.; Di Cara, D.; Nguyen, Q.D. A new method for forecasting energy output of a large-scale solar power plant based on long short-term memory networks a case study in Vietnam. *Electr. Power Syst. Res.* **2021**, *199*, 107427. [CrossRef]
2. Shang, C.; Wei, P. Enhanced support vector regression based forecast engine to predict solar power output. *Renew. Energy* **2018**, *127*, 269–283. [CrossRef]
3. Ali, M.; Prasad, R.; Xiang, Y.; Khan, M.; Farooque, A.A.; Zong, T.; Yaseen, Z.M. Variational mode decomposition based random forest model for solar radiation forecasting: New emerging machine learning technology. *Energy Rep.* **2021**, *7*, 6700–6717. [CrossRef]
4. Liu, D.; Sun, K. Random forest solar power forecast based on classification optimization. *Energy* **2019**, *187*, 115940. [CrossRef]
5. Etxegarai, G.; López, A.; Aginako, N.; Rodríguez, F. An analysis of different deep learning neural networks for intra-hour solar irradiation forecasting to compute solar photovoltaic generators' energy production. *Energy Sustain. Dev.* **2022**, *68*, 1–17. [CrossRef]
6. Almaghrabi, S.; Rana, M.; Hamilton, M.; Rahaman, M.S. Multivariate solar power time series forecasting using multilevel data fusion and deep neural networks. *Inf. Fusion* **2023**, *104*, 102180. [CrossRef]
7. Fernandez-Jimenez, L.A.; Muñoz-Jimenez, A.; Falces, A.; Mendoza-Villena, M.; Garcia-Garrido, E.; Lara-Santillan, P.M.; Zorzano-Alba, E.; Zorzano-Santamaria, P.J. Short-term power forecasting system for photovoltaic plants. *Renew. Energy* **2012**, *44*, 311–317. [CrossRef]
8. Ghimire, S.; Bhandari, B.; Casillas-Pérez, D.; Deo, R.C.; Salcedo-Sanz, S. Hybrid deep CNN-SVR algorithm for solar radiation prediction problems in Queensland, Australia. *Eng. Appl. Artif. Intell.* **2022**, *112*, 104860. [CrossRef]
9. Rodríguez, F.; Galarza, A.; Vasquez, J.C.; Guerrero, J.M. Using deep learning and meteorological parameters to forecast the photovoltaic generators intra-hour output power interval for smart grid control. *Energy* **2022**, *239*, 122116. [CrossRef]

10. Zhen, H.; Niu, D.; Wang, K.; Shi, Y.; Ji, Z.; Xu, X. Photovoltaic power forecasting based on GA improved Bi-LSTM in microgrid without meteorological information. *Energy* **2021**, *231*, 120908. [CrossRef]

11. Qu, J.; Qian, Z.; Pei, Y. Day-ahead hourly photovoltaic power forecasting using attention-based CNN-LSTM neural network embedded with multiple relevant and target variables prediction pattern. *Energy* **2021**, *232*, 120996. [CrossRef]

12. Jebli, I.; Belouadha, F.Z.; Kabbaj, M.I.; Tilioua, A. Prediction of solar energy guided by pearson correlation using machine learning. *Energy* **2021**, *224*, 120109. [CrossRef]

13. Scott, C.; Ahsan, M.; Albarbar, A. Machine learning for forecasting a photovoltaic (PV) generation system. *Energy* **2023**, *278*, 127807. [CrossRef]

14. Khan, W.; Walker, S.; Zeiler, W. Improved solar photovoltaic energy generation forecast using deep learning-based ensemble stacking approach. *Energy* **2022**, *240*, 122812. [CrossRef]

15. Tsai, W.C.; Tu, C.S.; Hong, C.M.; Lin, W.M. A Review of State-of-the-Art and Short-Term Forecasting Models for Solar PV Power Generation. *Energies* **2023**, *16*, 5436. [CrossRef]

16. Lee, D.; Kim, K. Recurrent Neural Network-Based Hourly Prediction of Photovoltaic Power Output Using Meteorological Information. *Energies* **2019**, *12*, 215. [CrossRef]

17. Quang, N.; Duy, L.; Van, B.; Dinh, Q. Applying Artificial Intelligence in Forecasting the Output of Industrial Solar Power Plant in Vietnam. *EAI Endorsed Trans. Energy Web* **2021**, *8*, 36. [CrossRef]

18. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Van Deventer, W.; Horan, B.; Stojcevski, A. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [CrossRef]

19. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [CrossRef]

20. Yin, S.; Wang, J.; Li, Z.; Fang, X. State-of-the-art short-term electricity market operation with solar generation: A review. *Renew. Sustain. Energy Rev.* **2021**, *138*, 110647. [CrossRef]

21. Michael, N.E.; Mishra, M.; Hasan, S.; Al-Durra, A. Short-Term Solar Power Predicting Model Based on Multi-Step CNN Stacked LSTM Technique. *Energies* **2022**, *15*, 2150. [CrossRef]

22. Liu, J.; Huang, X.; Li, Q.; Chen, Z.; Liu, G.; Tai, Y. Hourly stepwise forecasting for solar irradiance using integrated hybrid models CNN-LSTM-MLP combined with error correction and VMD. *Energy Convers. Manag.* **2023**, *280*, 116804. [CrossRef]

23. Gaboitaolelwe, J.; Zungeru, A.M.; Yahya, A.; Lebekwe, C.K.; Vinod, D.N.; Salau, A.O. Machine Learning Based Solar Photovoltaic Power Forecasting: A Review and Comparison. *IEEE Access* **2023**, *11*, 40820–40845. [CrossRef]

24. Gupta, P.; Singh, R. Forecasting hourly day-ahead solar photovoltaic power generation by assembling a new adaptive multivariate data analysis with a long short-term memory network. *Sustain. Energy Grids Netw.* **2023**, *35*, 101133. [CrossRef]

25. Wang, L.; Mao, M.; Xie, J.; Liao, Z.; Zhang, H.; Li, H. Accurate solar PV power prediction interval method based on frequency-domain decomposition and LSTM model. *Energy* **2023**, *262*, 125592. [CrossRef]

26. Huang, Z.; Huang, J.; Min, J. SSA-LSTM: Short-Term Photovoltaic Power Prediction Based on Feature Matching. *Energies* **2022**, *15*, 7806. [CrossRef]

27. Wentz, V.H.; Maciel, J.N.; Ledesma, J.J.G.; Junior, O.H.A. Solar Irradiance Forecasting to Short-Term PV Power: Accuracy Comparison of ANN and LSTM Models. *Energies* **2022**, *15*, 2457. [CrossRef]

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

29. Yang, D. Choice of clear-sky model in solar forecasting. *J. Renew. Sustain. Energy* **2020**, *12*, 026101. [CrossRef]

30. Li, J. Short-term Photovoltaic Power Prediction Based on Moderate-resolution Imaging Spectroradiometer Clear Sky Data. In Proceedings of the 2020 Chinese Automation Congress, CAC 2020, Shanghai, China, 6–8 November 2020; Institute of Electrical and Electronics Engineers Inc.: Piscataway, NJ, USA, 2020; pp. 930–934. [CrossRef]

31. Engerer, N.A.; Mills, F.P. KPV: A clear-sky index for photovoltaics. *Sol. Energy* **2014**, *105*, 679–693. [CrossRef]

32. Wang, S.; Dai, T.; Li, C.; Cheng, Y.; Huang, G.; Shi, G. Improving Clear-Sky Solar Power Prediction over China by Assimilating Himawari-8 Aerosol Optical Depth with WRF-Chem-Solar. *Remote Sens.* **2022**, *14*, 4990. [CrossRef]

33. Ma, Y.; Zhang, X.; Mei, S.; Zhen, Z.; Gao, R.; Zhou, Z. Ultra-short-term solar power forecasting based on a modified clear sky model. In Proceedings of the 39th Chinese Control Conference, Shenyang, China, 27–29 July 2020.

34. Sumathi, S.; Ashok Kumar, L.; Surekha, P. *Solar PV and Wind Energy Conversion Systems*; Springer: Cham, Switzerland, 2015. [CrossRef]

35. Antonanzas-Torres, F.; Urraca, R.; Polo, J.; Perpiñán-Lamigueiro, O.; Escobar, R. Clear sky solar irradiance models: A review of seventy models. *Renew. Sustain. Energy Rev.* **2019**, *107*, 374–387. [CrossRef]

36. Ineichen, P. Validation of models that estimate the clear sky global and beam solar irradiance. *Sol. Energy* **2016**, *132*, 332–344. [CrossRef]

37. Perez, R.; Ineichen, P.; Moore, K.; Kmiecik, M.; Chain, C.; George, R.; Vignola, F. A new operational model for satellite-derived irradiances: Description and validation. *Sol. Energy* **2002**, *73*, 307–317. [CrossRef]

38. Ineichen, P.; Perez, R. A New Airmass Independent Formulation for the Linke Turbidity Coefficient. 2002. Available online: www.elsevier.com/locate/solener (accessed on 5 May 2024).

39. Reno, M.J.; Hansen, C.W.; Stein, J.S. Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis. 2012. Available online: https://energy.sandia.gov/wp-content/gallery/uploads/SAND2012-2389_ClearSky_final.pdf (accessed on 5 May 2024).

40. Ineichen, P. A broadband simplified version of the Solis clear sky model. *Sol. Energy* **2008**, *82*, 758–762. [CrossRef]

41. Hochreiter, S.; Schmidhuber, J. Long Short-Term Memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]

42. Aggarwal, C.C. *Data Mining*; Springer International Publishing: Cham, Switzerland, 2015. [CrossRef]

43. Yang, Z.; Wang, J. A hybrid forecasting approach applied in wind speed forecasting based on a data processing strategy and an optimized artificial intelligence algorithm. *Energy* **2018**, *160*, 87–100. [CrossRef]

44. Bui, L.D.; Nguyen, N.Q.; Van Doan, B.; Sanseverino, E.R. Forecasting energy output of a solar power plant in curtailment condition based on LSTM using P/GHI coefficient and validation in training process, a case study in Vietnam. *Electr. Power Syst. Res.* **2022**, *213*, 108706. [CrossRef]

45. Reikard, G. Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Sol. Energy* **2009**, *83*, 342–349. [CrossRef]

46. Boland, J. Time Series Modelling of Solar Radiation. In *Modeling Solar Radiation at the Earth's Surface*; Springer: Berlin/Heidelberg, Germany, 2008.

47. Badescu, V. *Modeling Solar Radiation Modeling Solar Radiation at the Earth's Surface*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 1.

48. Wang, F.; Zhen, Z.; Wang, B.; Mi, Z. Comparative study on KNN and SVM based weather classification models for day ahead short term solar PV power forecasting. *Appl. Sci.* **2017**, *8*, 28. [CrossRef]

49. Yang, H.T.; Huang, C.M.; Huang, Y.C.; Pai, Y.S. A weather-based hybrid method for 1-day ahead hourly forecasting of PV power output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [CrossRef]

50. Shi, J.; Lee, W.J.; Liu, Y.; Yang, Y.; Wang, P. Forecasting power output of photovoltaic systems based on weather classification and support vector machines. *IEEE Trans. Ind. Appl.* **2012**, *48*, 1064–1069. [CrossRef]

51. Chen, C.; Duan, S.; Cai, T.; Liu, B. Online 24-h solar power forecasting based on weather type classification using artificial neural network. *Sol. Energy* **2011**, *85*, 2856–2870. [CrossRef]

52. Holmgren, W.F.; Hansen, C.W.; Mikofski, M.A. pvlib python: A python package for modeling solar energy systems. *J. Open Source Softw.* **2018**, *3*, 884. [CrossRef]

53. Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [CrossRef]

54. Hu, Q.; Zhang, R.; Zhou, Y. Transfer learning for short-term wind speed prediction with deep neural networks. *Renew. Energy* **2016**, *85*, 83–95. [CrossRef]

55. Qureshi, A.S.; Khan, A.; Zameer, A.; Usman, A. Wind power prediction using deep neural network based meta regression and transfer learning. *Appl. Soft Comput.* **2017**, *58*, 742–755. [CrossRef]

56. Möhrlen, C.; Dk, W.; Zack, J.; Messner, J.; Analytics, A.; Browell, J. IEA Wind Task 36-Recommended Practice on Renewable Energy Forecast Solution Selection. 2019. Available online: https://www.ieawindforecasting.dk/publications/recommendedpractice (accessed on 2 May 2023).

57. Behera, M.K.; Majumder, I.; Nayak, N. Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Eng. Sci. Technol. Int. J.* **2018**, *21*, 428–438. [CrossRef]

58. Sobri, S.; Koohi-Kamali, S.; Rahim, N.A. Solar photovoltaic generation forecasting methods: A review. *Energy Convers. Manag.* **2018**, *156*, 459–497. [CrossRef]