Methods for Integrating Extraterrestrial Radiation into Neural Network Models for Day-Ahead PV Generation Forecasting

Seung Chan Jo 1, Young Gyu Jin 2, Yong Tae Yoon 1 and Ho Chan Kim 2,*

1 Department of Electrical and Computer Engineering, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea; tmdcks943@snu.ac.kr (S.C.J.); ytyoon@snu.ac.kr (Y.T.Y.)
2 Department of Electrical Engineering, Jeju National University, 102 Jejudaehak-ro, Jeju-si 63243, Jeju, Korea; ygjin93@jejunu.ac.kr
* Correspondence: hckim@jejunu.ac.kr; Tel.: +82-64-754-3676

Abstract: Variability, intermittency, and limited controllability are inherent characteristics of photovoltaic (PV) generation that result in inaccurate solutions to scheduling problems and the instability of the power grid. As the penetration level of PV generation increases, it becomes more important to mitigate these problems by improving forecasting accuracy. One of the alternatives to improving forecasting performance is to include a seasonal component. Thus, this study proposes using information on extraterrestrial radiation (ETR), which is the solar radiation outside of the atmosphere, in neural network models for day-ahead PV generation forecasting. Specifically, five methods for integrating the ETR into the neural network models are presented: (1) division preprocessing, (2) multiplication preprocessing, (3) replacement of existing input, (4) inclusion as additional input, and (5) inclusion as an intermediate target. The methods were tested using two datasets in Australia using four neural network models: Multilayer perceptron and three recurrent neural network (RNN)-based models including vanilla RNN, long short-term memory, and gated recurrent unit. It was found that, among the integration methods, including the ETR as the intermediate target improved the mean squared error by 4.1% on average, and by 12.28% at most in RNN-based models. These results verify that the integration of ETR into the PV forecasting models based on neural networks can improve the forecasting performance.

Keywords: PV generation forecasting; extraterrestrial radiation; neural network; recurrent neural network; seasonal component; time series forecasting

1. Introduction

Recently, the use of renewable energy sources (RES) for reducing greenhouse gases and consequent sustainable development has been considered inevitable. The integration of RES has been successful to some extent, as various countries have actively implemented policies for wide integration, such as feed-in-tariffs and renewable portfolio standards, and the levelized cost of energy (LCOE) of RES has decreased through technological developments. The International Renewable Energy Agency (IRENA) reported that photovoltaic (PV) power generation tripled from 250 TWh in 2015 to 720 TWh in 2019 [1]. Additionally, it was projected via an International Energy Agency (IEA) sustainable development scenario that approximately 3268 TWh will be produced by PV generation by 2030 [2].

However, such an increase in PV generation causes instability in power systems due to its variability, intermittency, and limited controllability. Specifically, from the perspective of transmission system operators (TSOs), the unfavorable characteristics of RES, including PV generation, can exacerbate the imbalance between power supply and demand [3] and make power system planning and operation more difficult [4]. To address such problems, TSOs need to secure a significant number of flexible resources, which is then followed by an increase in the electricity bills of customers. Its threat to power systems is ongoing due to its unlimited access to the grid. For example, electricity consumption is...
frequently settled at a negative price in Germany, where 32.5% of gross power production is generated by wind and solar. The demand of the power system subtracted by generation of the RESs is called net demand, and their generation cannot be changed, much like usual demand, due to unlimited access to the grid. The reason for negative prices is that significantly reduced net demand, due to a surge in increased supply from RES at an instant, causes in the power system a condition of oversupply, requiring that inflexible generators with high ramp-up and ramp-down costs generate even for negative electricity prices. Therefore, a negative electricity price indicates a severe condition of a power system because its occurrence means the power system cannot remain in balance using only flexible resources [5]. Furthermore, from the perspective of distribution system operators (DSOs), a high penetration of PV generation results in the need for investment in various alternative resources to achieve stable power system operation by addressing problems such as voltage fluctuation, increased network losses, and feeder overloading in the distribution network [6–8].

Increasing the forecasting accuracy of PV generation is one of the most simple and economic solutions to such problems because it can incentivize balance-responsible parties (BRPs) in an attempt to reduce the penalty caused by the imbalance between the scheduled and actual PV generation. Ultra short-term forecasting is used in techniques such as power smoothing and real-time electricity dispatch. Short-term forecasting is useful for power scheduling tasks, such as unit commitment and economic power dispatch, and can also be used in PV-integrated energy management systems [9]. In contrast, medium-term and long-term forecasting are effectively used in power system planning, where network optimization is performed and investment decisions are made [10].

PV forecasting can be classified, according to its methodology, into physical methods, statistical approaches, and machine learning approaches [9]. In physical methods, the radiant energy reaching the Earth’s surface is determined using a physical atmospheric model affecting the solar radiation. In Dolara et al. [11], the PV output was forecasted using an irradiance model considering the transmittance of the atmosphere and an air-glass optical model. In statistical approaches, PV generation is forecasted through the statistical analysis of input variables. For instance, multi-period PV generation was forecasted using autoregressive moving average (ARMA) and autoregressive-integrated moving average (ARIMA) models in Colak et al. [12], and an autoregressive moving average with exogenous inputs (ARMAX) model was applied to statistical PV forecasting by including weather forecasting data in Li et al. [13]. In the last category of machine learning approaches, a forecasting model is trained by updating the model parameters based on existing data. Machine learning has wide applications, such as in dynamical systems with memory [14] and health monitoring [15]. Various machine learning models have been applied to forecasting PV generation, such as a support vector machine (SVM) [16], Bayesian neural network [17] and RNN-based models such as long short-term memory (LSTM) [18] and gated recurrent network (GRU) [19].

There have been attempts to separate the components of a forecasting target in the time series forecasting. In Keles et al. [20], the spot prices in the electricity market were forecasted by separating them into deterministic and stochastic signals. In Zhu et al. [21], Zang et al. [22], and Li et al. [23], time series data were decomposed using wavelet decomposition, and then the decomposed signals were used to teach a neural network. Similarly, time series data PV generation can be regarded as having seasonal and temporal components. The seasonal component results from changes in the altitude of the Sun due to the Earth’s orbit and rotation, while various stochastic factors, such as weather conditions, are the source of the temporal component.

On the other hand, there have been studies where PV generation data of similar dates during a year are combined to forecast the output. As the generations at similar dates have the seasonal component in common, their combination leads to more accurate forecasting of the seasonal component. Thus, a forecasting method using the seasonal component is likely to show a good forecasting performance. PV generation data of adjacent days were
used in the convolution neural network (CNN) [22] and the time correlation method [24]. In Li et al. [25], the seasonal component was forecast using the weighted sum of the two outputs from the CNN with inputs of PV generation at adjacent days and from the LSTM with inputs of total PV generation over one day. The forecasting error can be reduced by indirectly considering seasonal changes through the regularizing effect of an ensemble model. In Wen et al. [26], the ensemble of a back-propagation neural network, radial basis function neural network, extreme learning machine, and Elman neural network was used for forecasting the seasonal component of PV generation, and in Gigoni et al. [27], the combination of the Grey-box model, neural network, k-nearest neighbor, quantile random forest, and support vector regression was the best in terms of forecasting accuracy. Table 1 summarizes prior studies on PV generation forecasting by method.

Table 1. Summary of existing PV generation forecasting methods.

| Method               | Reference | Highlights                                                                 |
|----------------------|-----------|----------------------------------------------------------------------------|
| Physical             | [11]      | Models the physical state and dynamics of the atmosphere.                  |
| Statistical          | [12,13]   | Generates forecasting data with appropriate statistical assumptions.       |
|                      | [16–19]   | Uses machine learning approach which improves the performance automatically by using data. |
| Machine learning     | [21–23]   | Preprocesses time series data by decomposing the data into several sub-signals for machine learning. |
|                      | [22,24,25] | Uses a machine learning model that combines data from similar days.         |
|                      | [26,27]   | Uses an ensemble of different models.                                     |

In contrast to the approaches used in the previous studies, the seasonal component can be determined more accurately by direct geometric modeling, based on the fact that solar radiation changes periodically according to the Earth’s rotation and orbit. Thus, in this study, we propose methods to calculate the seasonal component using the angle of incidence and solar constant for improving forecasting accuracy, and those methods are verified using four different neural network models: MLP and three RNN-based models including Vanilla RNN, LSTM, and GRU. The contribution of this paper is as follows.

- This is the first study that explicitly introduces extraterrestrial radiation, that is, the solar radiation outside the atmosphere, into various neural network models;
- Methods for integrating the extraterrestrial radiation into neural network models for PV generating forecasting are presented;
- To verify the effectiveness of the proposed methods, the methods are applied to four neural network models: MLP and three RNN-based models including Vanilla RNN, LSTM, and GRU. Then, their forecasting performances are examined and compared.

2. Seasonal Changes in Extraterrestrial Radiation

Extraterrestrial radiation (ETR) refers to the power per unit area of sunlight irradiated at the distance between the Sun and the Earth in space. The ETR is radiation that does not consider the effects of the atmosphere and acts as an upper limit of the energy from the Sun reaching the Earth. The ETR has a seasonal characteristic because the Earth’s orbit and rotation are seasonal. The parameters for modeling the seasonality of the ETR include the change in the distance between the Sun and Earth, the change in the solar declination due to the Earth’s orbit, and the circumferential motion of the Sun due to the Earth’s rotation.

Since the Earth’s orbit is not circular but elliptical, the distance between the Sun and the Earth changes according to the orbit, which causes corresponding changes in radiation. In addition, since the Earth orbits the Sun with its rotation axis tilted by 23.5 degrees, the declination of the Sun varies with time, which is associated with the amount of sunlight throughout the year. The Earth’s rotation also affects the radiation throughout the day, that is, there is a large amount of radiation during the day in contrast to the absence of radiation during the night.
In this study, the ETR was geometrically determined as a function of time based on the diurnal motion model. Specifically, the solar constant was used to determine the maximum of daily ETR, and then it was adjusted according to the angle of incidence of the Sun with respect to the PV panel.

2.1. Solar Constant

The solar constant, \( G_{sc} \), is defined as the power per unit area irradiated from the Sun at the average distance between the Sun and Earth. It can be measured from a satellite to eliminate the atmospheric effect and determined as \[ G_{sc} = 1367 \text{ W/m}^2. \] (1)

The radiation on a surface perpendicular to the sunlight on the \( n \)-th day of a year, denoted as \( G_{on} \), is calculated as \[ G_{on} = G_{sc} \times \left( 1 + 0.033 \times \cos \left( \frac{360n}{365} \right) \right). \] (2)

2.2. Angle of Incidence

The ETR parallel to the final radiation incident on the PV panel is obtained considering the angle of incidence \( (\theta) \), which means the angle between the radiation and the orthogonal line to the PV panel. The angle of incidence can be derived as a function of time by geometric modeling. The necessary variables include not only the declination, hour angle, latitude, and longitude but also the tilt angle and azimuth angle, which are the angles defining the geometric configuration of the PV panels.

2.2.1. Declination

As illustrated in Figure 1, the declination \( (\delta) \) is defined as the angle between the lines connecting the Sun and the Equator in the equatorial coordinate system. It has a positive value in the Northern Hemisphere. The declination on the \( n \)-th day of a year can be approximately determined as \[ \delta \approx 23.45 \times \sin \left( 360 \times \frac{284 + n}{365} \right). \] (3)

![Figure 1. Geometrical description of the declination (\( \delta \)) in the equatorial coordinate system compared with the latitude (\( \phi \)).](image)
2.2.2. Hour Angle

The Sun’s position in the celestial coordinate system can be expressed as a function of time within a day by considering the circumferential motion of the Sun due to the Earth’s rotation. However, to accurately integrate the time zone into the geometrical model, the local time needs to be converted into solar time, which is determined based on the Sun, following the rule as follows [28]:

\[
\text{Solar time} = \text{Local time} + \frac{4(L_{st} - L_{loc}) + E}{60}.
\] (4)

where \(L_{st}\) and \(L_{loc}\) are the longitudes of Local Standard Time Meridian and the longitude at a specific location, respectively, and \(E\) is calculated as:

\[
E = 229.2 \times (0.000075 + 0.001868 \cos B - 0.032077 \sin B - 0.014615 \cos 2B - 0.04089 \sin 2B),
\] (5)

\[
B = \frac{(n - 1)360}{365}.
\] (6)

Then, the solar time is further converted into the hour angle \((\omega)\) according to the following relationship as:

\[
\omega = \frac{\text{Solar time} - 12 \text{ hours}}{15^\circ}.
\] (7)

The hour angle is conceptually illustrated in Figure 2.

![Figure 2. Geometrical description of the hour angle (\(\omega\)) in the celestial coordinate system.](image)

2.2.3. Tilt and Azimuth of PV Panels

The PV panels are installed with a slope to maximize the amount of power generated by minimizing the angle of incidence. The parameters defining the geometrical configuration of a PV panel consist of tilt \((\beta)\) and azimuth \((\gamma)\), which are shown in Figure 3. The tilt is the angle between the horizontal plane and the PV panel. The azimuth is the angle measured from the Meridian to the point where the normal vector of the PV panel is projected orthogonal to the horizontal plane, and thus it is positive in the western area and negative in the eastern area [28].

2.2.4. Calculation of the Angle of Incidence

The angle of incidence \((\theta)\) for determining the ETR can be calculated by using the explained parameters in Sections 2.2.1–2.2.3, and is as follows [28]:

\[
\theta = \cos^{-1}(\sin \delta \sin \phi \cos \beta - \sin \delta \cos \phi \sin \beta \cos \gamma + \cos \delta \cos \phi \cos \beta \cos \omega
+ \cos \delta \sin \phi \sin \beta \cos \gamma \cos \omega + \cos \delta \sin \phi \sin \beta \sin \gamma \sin \omega).
\] (8)
2.3. Calculation of the Extraterrestrial Radiation

Once the solar constant and the angle of incidence at a place of interest are calculated according to (2) and (8), the value of ETR, denoted as $G_{ext}$, can be simply calculated as:

$$G_{ext} = G_{an} \times \cos \theta$$

\hspace{1cm} (9)

Figure 3. Geometrical description of the tile ($\omega$) and azimuth ($\gamma$) of a PV panel compared with the angle of incidence ($\theta$).

For instance, the process of calculating the ETR was applied to a place in Yulara, Australia. The resulting values of the ETR for a tilted surface on the first day of each month throughout a year are shown in Figure 4. The values of ETR had the same bell shape as a typical clear-sky PV generation over a day. The place is located in the Southern Hemisphere and, thus, the Sun’s altitude and the corresponding value of $\cos \theta$ are the highest in December and January. Accordingly, the ETR is the greatest in those months. The final radiance on the PV panel had the same bell-shaped pattern, but with more fluctuations depending on weather conditions. Consequently, the ETR can be interpreted as a piece of clean information on radiation by the Sun with noise from the atmosphere removed.

Figure 4. Illustrative results from the process of calculating the ETR throughout a year for a PV panel tilted by 15° at Yulara, Australia.
3. Forecasting Models

In this section, the traditional persistent model and representative neural network models are briefly described, which are used for comparison purposes in the case study later.

3.1. Persistence Model

In the persistence model, PV outputs of yesterday are used as forecasted PV generation. Although the persistence model is simple, it shows a satisfactory forecasting performance, particularly when the weather conditions do not change significantly. The persistence model was used as a reference in a study comparing the forecasting accuracy.

3.2. Multilayer Perceptron

The multilayer perceptron (MLP) is a model where nodes that imitate a human neural network are stacked in series and parallel. The node receives input data, multiplies them by its weights, applies its activation function to the intermediate value, and finally generates the outputs. The wider and deeper the nodes are, the more complex the functions can be modeled.

A learning process is performed by computing the gradient of the loss function and updating the weights of a node using the back-propagation algorithm. If the MLP is deep with many layers, a vanishing gradient problem can occur, which means the gradient is no longer passed to the previous layers. Some activation functions, such as the rectified linear unit (ReLU), can mitigate the vanishing gradient problem. The MLP can be regarded as more effective and flexible than traditional regression models because there is no assumption on the form of the target function to be modeled.

3.3. Recurrent Neural Network

Unlike the MLP, the recurrent neural network (RNN) has a memory function that is implemented by sequentially feeding back the outputs or states in the previous times to the current input. Obviously, the outputs in the previous times contain the information in the past. This property makes the RNN suitable for dealing with time series data. The learning process of the RNN is performed by the technique of back-propagation through time, which unfolds the RNN with respect to time and applies the same back-propagation algorithm as the MLP. However, the weights of the RNN have the risk of divergence, particularly when handling a long sequence of data, because the same weight parameters are updated repetitively during training.

To address the divergence problem, gate structures are proposed for the RNN, called gated RNN, such as long short-term memory (LSTM) and gated recurrent unit (GRU). The gates in the gated RNN determine which information is retained or discarded in the hidden state. Nodes constituting the gates make these decisions. The nodes applies the sigmoid function as their activation function to their intermediate outputs, which are then multiplied with the hidden state. As the sigmoid function is bounded in (0, 1), only a portion of the hidden states is preserved by multiplication, meaning the gate considers the portion important. The parameters of the nodes are added to those of the simple RNN, and they are also updated in the learning process so that the gates make effective decisions. It has been empirically verified that gated RNNs have better characteristics in terms of convergence and accuracy for the forecasting task of a long sequence. Recently, gated RNNs have been widely used for time series forecasting [29,30].

3.3.1. Vanilla RNN

Vanilla RNN is a preliminary RNN structure that stores historic information by a hidden state. The hidden state at time $t$ ($h_t$) is determined by applying the hyperbolic tangent function to the weighted sum of the current input and hidden states in the previous times as follows:

$$h_t = \tanh(W_{h,x}x_t + W_{h,h}h_{t-1} + b_h),$$ (10)
where \( W_{i,x} \) and \( W_{h,h} \) are the weights; \( b_i \) is the bias; \( x_t \) is the input at time \( t \); and \( h_t \) is the output at time \( t \).

### 3.3.2. Long Short-Term Memory

Figure 5 shows the structure of a cell in the LSTM. The cell has the input, forget, and output gates that determine the output through weight update Equations in (11)–(14).

\[
i_t = \sigma(W_{i,x}x_t + W_{i,h}h_{t-1} + b_i),
\]

\[
f_t = \sigma(W_{f,x}x_t + W_{f,h}h_{t-1} + b_f),
\]

\[
o_t = \sigma(W_{o,x}x_t + W_{o,h}h_{t-1} + b_o),
\]

\[
g_t = \tanh(W_{g,x}x_t + W_{g,h}h_{t-1} + b_g),
\]

where \( W_{i,x}, W_{i,h}, W_{f,x}, W_{f,h}, W_{o,x}, W_{o,h}, W_{g,x}, \) and \( W_{g,h} \) are the weights; \( b_i, b_f, b_o, \) and \( b_g \) are the biases; \( x_t \) and \( h_{t-1} \) are the input and output at time \( t \); \( i_t, f_t, \) and \( o_t \) are the input, forget, and output gates at time \( t \), respectively; \( g_t \) is the candidate; \( \sigma(\cdot) \) is the sigmoid function. The function of the three gates of the LSTM cell is implemented as the element-wise multiplication, denoted as \( x \) around a circle in Figure 6. Specifically, the forget gate eliminates unnecessary information from the cell state \((c_{t-1})\) in the previous time; the input gate extracts important information from the input; the output gate generates the selective output from the sigmoid function values \((o_t)\) for the hidden state \((h_{t-1})\) in the previous time.

### 3.3.3. Gated Recurrent Unit

Unlike the LSTM, the cell of the GRU has two gates, that is, a reset gate and an update gate (Figure 6). The update equations associated with the gates are given as:

\[
z_t = \sigma(W_{z,x}x_t + W_{z,h}h_{t-1} + b_z),
\]

\[
r_t = \sigma(W_{r,x}x_t + W_{r,h}h_{t-1} + b_r),
\]

\[
h_t = \tanh(W_{h,x}x_t + W_{h,h}(r_t \odot h_{t-1}) + b_h),
\]

where \( W_{z,x}, W_{z,h}, W_{r,x}, W_{r,h}, W_{h,x}, \) and \( W_{h,h} \) are the weights; \( b_z, b_r, \) and \( b_h \) are the biases; \( x_t \) and \( h_{t-1} \) are the input and output at time \( t \); \( z_t \) and \( r_t \) are the update and reset gates at

\[
\text{Figure 5. Structure of a cell in the LSTM.}
\]
time $t$, respectively; $h_t$ is the candidate; $\odot$ denotes the Hadamard product; and $\sigma(\cdot)$ is the sigmoid function. The functions of the reset and update gates of the GRU are similar to those of the forget and input gates of the LSTM. Therefore, there is no significant difference between the structures of the GRU and LSTM. However, the number of weights of the GRU are fewer than that of the LSTM, which results in higher efficiency in the learning process of the GRU.

Figure 6. Structure of a cell in the GRU.

4. Forecasting Methods with Extraterrestrial Radiation

The ETR has rarely been considered in forecasting applications using data-driven machine learning models. One of the most useful properties of the ETR in a forecasting problem is that it accurately contains the seasonal effect on the target variable to be forecasted. Thus, the forecasting accuracy is expected to improve if the ETR is effectively integrated into existing forecasting methods. In this section, we propose methods to integrate the ETR into a data-driven forecasting method.

4.1. Forecasting Framework of the Base Method

In this subsection, the base day-ahead hourly forecasting framework, which does not consider the ETR, is presented as shown in Figure 7. The procedures are as follows:

1. Imputer imputes day-ahead generation and weather data;
2. The imputed data are arranged in the timeframe of a past interval (15 min) and present interval (1 h);
3. Sequenced data are split for training and testing;
4. Split data are scaled using the MinMaxScaler;
5. A forecasting model is trained and tested;
6. Forecasted results are saved and errors are calculated

The simple imputer imputes omitted data by nearby data. The past interval, which is the interval of the model’s input, was determined to be 15 min. The present interval, which is the interval of the model’s output, was determined to be 1 h as most of the day-ahead electricity market receives bids and offers on the hourly units and unit commitment is solved in an hourly manner. The lower bound of the scaling interval of the MinMaxScaler was determined to be 0.1 to prevent the distortion of data resulting from dividing by a small number.
Figure 7. Framework of neural network models for PV generation forecasting model: (a) base method; (b) proposed methods.

4.2. Forecasting Framework of Proposed Methods

In this subsection, the methods combining the ETR with a data-driven neural network model are described. The framework of the proposed methods is presented in Figure 7. Specifically, five integration methods were developed and are presented: (1) division preprocessing, (2) multiplication preprocessing, (3) replacement of existing input, (4) inclusion as additional input, and (5) inclusion as an intermediate target, which are denoted in order as M1, M2, M3, M4, and M5. In the proposed forecasting framework, the ETR is calculated for given timestamps using the latitude, longitude, tilt angle, and azimuth angle of the panel in the place. Compared with the base method, M1 and M2 perform the additional function in the adjustment/re-adjustment block, which is drawn with a dashed line in Figure 7. Similarly, M3, M4, and M5 perform the function of the input filtering block, which is drawn with the dashed-double dotted line in Figure 7, to modify their input for their method-specific implementation.

4.2.1. Division Preprocessing

ETR is attenuated by the atmosphere in proportion to the total magnitude of ETR. The attenuation ratio of ETR to the solar radiation reaching the land is called the clearness index in meteorological terms. Even if the atmospheric conditions are the same at 9:00 a.m. and 12:00 p.m., the absolute attenuated amount is larger at 12:00 p.m. than the other because of differences in ETR. To accurately reflect the atmospheric conditions in a model, PV generations need to be adjusted for the model to receive the same signal when meteorological circumstances are the same. Dividing the radiation by ETR is one way of...
transforming the radiation to only meteorologically affected signals. Since PV generation is strongly correlated with the radiation by $R^2 = 0.99$ [9], dividing the PV generation by ETR is expected to have the same effect as dividing the radiation. Thus, method M1, which normalizes PV generation using ETR, is proposed. The framework for M1 is shown in Figure 7, and it is implemented by dividing the PV generation by the ETR in the adjustment block and multiplying in the re-adjustment block.

4.2.2. Multiplication Preprocessing

Data-driven neural network models are trained by back-propagating an error. Mean squared error, which calculates the squared difference between the predicted output and the target, is used in general for training a model, and is one of the performance indexes for evaluation. Since MSE is calculated by squaring absolute differences, the outputs around noon with a large ETR significantly affect total MSE due to the fact of their magnitude. For example, if the model predicts poorly around noon and favorably around morning, the MSE will be larger than the opposite situation. If the training objective is to minimize MSE, which is the most likely situation in real applications, it can be assumed desirable to train more accurately in timeslots with a high ETR. Multiplying PV generation by ETR makes squared error in the timeslots with a high ETR large during training, resulting in a high back-propagation signal to input variables in those timeslots to minimize the modified MSE. For this reason, a method that multiplies PV generation by ETR is proposed. The framework for M2 is also shown in Figure 7, and it is implemented by multiplying the PV generation by the ETR in the adjustment block and dividing in the re-adjustment block.

4.2.3. Replacement of Existing Input

Most of the forecasting frameworks use historical PV generation as one of their inputs. In the frameworks, the model receives limited data, such as for one or two days prior, as input to forecast generation at a specific time unless long periods of historical data are inserted in direct ways, such as sequence-to-sequence LSTM, or indirect ways such as prior studies have done. In other words, the model preserves historical data only in a form of model parameters and has limited access to historical generation only as input for forecasting. As meteorological conditions vary from day to day, today’s PV generation can be significantly different from day-ahead generation due to the stochasticity. Performance can be expected to improve if the model receives a clean signal with the stochasticity removed. The ETR is the solar radiation unaffected by meteorological conditions and can be deemed a total sum of historical data since summing every historic generation erases the stochasticity. Therefore, a method that replaces day-ahead generation with the ETR is proposed. The framework for M3 is also shown in Figure 7, and it is implemented by inserting the ETR and weather data through an input filtering block to a sequencing block.

4.2.4. Inclusion as Additional Input

The ETR is the solar radiation without stochasticity and is an upper bound of PV power generation. Day-ahead PV generation includes information about the day-ahead meteorological conditions, which can be correlated with today’s conditions. Specifically, the day-ahead data are especially valuable in a situation where similar weather conditions frequently continue for several days. Both ETR and day-ahead PV generation are meaningful information for forecasting. Therefore, a method that includes both day-ahead generation and ETR is proposed. The framework for M4 is also shown in Figure 7, and it is implemented by inserting day-ahead PV generation, ETR, and weather data through an input filtering block to a sequencing block.

4.2.5. Inclusion as an Intermediate Target

Instead of inserting the raw ETR directly, it would be more effective for the ETR to be scaled by the daily averaged clearness index considering the weather. Therefore, a method that utilizes an MLP model trained separately to predict the clearness index is proposed.
This MLP model is named as the clearness model, as it predicts the clearness index. The training framework for this model is shown in Figure 8 and is as follows:
1. The clearness model takes weather data as input and outputs a clearness index, which is a one-dimensional scalar;
2. ETR is multiplied by the output of the clearness model;
3. MSELoss between the adjusted ETR and the target value is calculated;
4. MSELoss is backpropagated to train the clearness model.

![Figure 8. Training procedures of the method including the ETR as the intermediate target.](image)

The framework for M5 is also shown in Figure 7, and it is implemented by inserting day-ahead PV generation, ETR multiplied by the output of the trained clearness model, and weather data through an input filtering block to a sequencing block.

5. Case Study

5.1. Dataset

In the case study, datasets provided by the Desert Knowledge Australia Solar Centre (DKASC) in Australia were used [31,32]. DKASC operates solar power plants in Alice Springs and Yulara in Australia, and their information is listed in Table 2.

Table 2. Information on the two regions for integrating the ETR in the case study.

| Parameters             | Alice Springs                  | Region        |
|------------------------|--------------------------------|---------------|
| Name                   | BP Solar                       | Desert Gardens|
| Array Rating           | 4.95 kW                        | 1058.4 kW     |
| Array Structure        | fixed, ground mount            | fixed, ground mount |
| Panel Rating           | 165 W                          | 315 W         |
| Number of Panels       | 30                             | 3360          |
| Panel Type             | BP 3165                        | Jinko Solar JKM315PP |
| Tilt Angle             | 20                             | 15            |
| Azimuth Angle          | 0                              | 0             |
| Latitude               | −23.7                          | 133.8         |
| Longitude              | −25.2                          | 130.9         |

The datasets provided by DKASC contain both meteorological information and PV power generation data in the two regions. The specific features in the dataset are listed in Table 3. The feature of wind direction was excluded because it increased the forecasting error. The features of global horizontal radiation and diffuse horizontal radiation were not chosen either, because it strongly correlates with PV generation [9]. Pyranometer measurement was also excluded, as it measures radiation in a specific frequency band.
Table 3. Features of the dataset used in the case study.

| Category       | Alice Springs                  | Region           |
|----------------|-------------------------------|------------------|
| Included Features | wind speed | temperature | daily rainfall | relative humidity | wind speed | temperature | daily rainfall | air pressure |
| Excluded Features | global horizontal radiation | diffuse horizontal radiation | pyranometer measurement | wind direction | global horizontal radiation | wind direction |

The total period of data used in the case study was from May 2016 to January 2021. Among them, the data from May 2016 to January 2020 were used for training; the remaining data over the year from February 2020 to January 2021 were used for the test.

5.2. Models and Performance Index

The effectiveness of the proposed methods was verified and compared using four representative neural network models: MLP, Vanilla RNN, LSTM, and GRU.

To evaluate the forecasting performance, two types of indexes, that is, mean squared error (MSE) and mean absolute error (MAE), were used. The two indexes are defined as follows:

\[
MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - f(x_i))^2,
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - f(x_i)|.
\]

where \( n \) is the number of samples, \( x_i \) is the predictors, \( y_i \) is the target output, and \( f \) is the output forecasted by a model. To determine the hyperparameters of the models, 10-fold cross-validation was conducted by randomly sampling 10 subsets with equal size from the training set and using each of the subsets for validation in each stage of validation. The resulting hyperparameters of the models are listed in Table 4. The Adam optimizer was used in the training. Epoch was determined as 150 because the average of 10-fold validation errors were saturated and oscillated in a small range after being trained 60–90 times, and increased after being trained over 200 times, as shown in Figure 9. Figure 9 also shows the average of a 10-fold training error of the same model, which steadily declined as the epoch increased. This means the model had enough capacity to fit the training data, validating a reasonable choice of the size of the hidden states. The difference in scale between the validation error and training error was because MinMaxScaler was applied during training. Learning rate schedulers were not used in this study, because they can make the model converge into a poor local minimum, which leads to a higher MSE. The RNN-based models were configured to be bidirectional, as shown in Figure 10, because it is more effective to consider future data and predicted outputs for prediction especially for the day-ahead forecasting. A simple two-layer MLP model was applied to the output of the RNN-based models to reduce the dimension of the hidden states to one.

5.3. Results

The proposed methods were applied to the four neural network models described in Section 5.2 for the two datasets. The results are listed in Tables 5 and 6. Each method was trained and tested 20 times for each combination of the model and the dataset to examine its effect on average.

As a result, M5 was the most effective in the majority of the dataset-model combinations. Table 7 shows the improvements M5 achieved relative to the base method in RNN-based models. M5 reduced the MSE by 4.1% on average and, at most, by 12.28% in the RNN-based models. These results imply that integrating ETR into neural network models can improve model performance without additional investment in data collection.
The fact that M5 had consistently higher performance than other methods under different datasets and RNN-based models means that M5 can be expected to be effective when one forecasts PV generation without verifying which method would be the best.

Table 4. Hyperparameters of the AI model in the case study.

| Model                      | Epoch | Hidden State         | Number of Hidden Layers | Hyperparameter | Bias | Batch Size | Learning Rate |
|----------------------------|-------|----------------------|-------------------------|----------------|------|------------|---------------|
| MLP                        | 150   | 192, 192, 192,       | 7                       | -              | True | 64         | $1 \times 10^{-3}$ |
| Vanilla RNN, LSTM, GRU    | 150   | 120                  | 2                       | True           | True | 64         | $1 \times 10^{-3}$ |

Figure 9. The average of 10-fold cross-validation and training error of a LSTM model with the base method as the epoch changes.

Figure 10. The architecture of the RNN-based models.
Table 5. Test errors of each method and model trained using BP Solar dataset of Alice Springs.

| Model     | Method | MSE     | MAE     |
|-----------|--------|---------|---------|
| Persistence | Base   | 0.004682 | 0.0005631 |
| MLP       | Base   | 0.009812 | 0.0006263 |
| MLP       | M1     | 0.006271 | 0.0005512 |
| MLP       | M2     | 0.01120  | 0.0006745 |
| MLP       | M3     | 0.01317  | 0.0005137 |
| MLP       | M4     | 0.01147  | 0.0007031 |
| MLP       | M5     | 0.006418 | 0.0004872 |
| Vanilla RNN | Base | 0.002846 | 0.0003521 |
| Vanilla RNN | M1   | 0.004247 | 0.0003682 |
| Vanilla RNN | M2   | 0.003389 | 0.0005086 |
| Vanilla RNN | M3   | 0.004591 | 0.0005549 |
| Vanilla RNN | M4   | 0.003380 | 0.0003373 |
| Vanilla RNN | M5   | 0.002649 | 0.0003134 |
| LSTM      | Base   | 0.002362 | 0.0003176 |
| LSTM      | M1     | 0.003526 | 0.0003751 |
| LSTM      | M2     | 0.002824 | 0.0003751 |
| LSTM      | M3     | 0.003741 | 0.0003836 |
| LSTM      | M4     | 0.002599 | 0.0003028 |
| LSTM      | M5     | 0.002340 | 0.0002974 |
| GRU       | Base   | 0.002425 | 0.0003216 |
| GRU       | M1     | 0.003513 | 0.0003237 |
| GRU       | M2     | 0.003000 | 0.0004244 |
| GRU       | M3     | 0.004296 | 0.0003926 |
| GRU       | M4     | 0.002469 | 0.0003035 |
| GRU       | M5     | 0.002349 | 0.0002964 |

Table 6. Test errors of each method and model trained using Desert Gardens dataset of Yulara.

| Model     | Method | MSE     | MAE     |
|-----------|--------|---------|---------|
| Persistence | Base  | 0.7960  | 0.0004327 |
| MLP       | Base   | 1.8905  | 0.0006457 |
| MLP       | M1     | 1.8190  | 0.0006359 |
| MLP       | M2     | 1.9097  | 0.0006475 |
| MLP       | M3     | 2.3726  | 0.0008752 |
| MLP       | M4     | 1.6309  | 0.0008511 |
| MLP       | M5     | 1.7905  | 0.0006541 |
| Vanilla RNN | Base | 0.6664  | 0.0003841 |
| Vanilla RNN | M1   | 0.6763  | 0.0003934 |
| Vanilla RNN | M2   | 0.6716  | 0.0003881 |
| Vanilla RNN | M3   | 1.7004  | 0.0007612 |
| Vanilla RNN | M4   | 0.6589  | 0.0003872 |
| Vanilla RNN | M5   | 0.6570  | 0.0003880 |
| LSTM      | Base   | 0.6982  | 0.0004042 |
| LSTM      | M1     | 0.6869  | 0.0003961 |
| LSTM      | M2     | 0.6918  | 0.0003986 |
| LSTM      | M3     | 1.7218  | 0.0007212 |
| LSTM      | M4     | 0.6708  | 0.0003911 |
| LSTM      | M5     | 0.6563  | 0.0003822 |
| GRU       | Base   | 0.6540  | 0.0003888 |
| GRU       | M1     | 0.6493  | 0.0003852 |
| GRU       | M2     | 0.6484  | 0.0003851 |
| GRU       | M3     | 1.7227  | 0.0007383 |
| GRU       | M4     | 0.6489  | 0.0003842 |
| GRU       | M5     | 0.6448  | 0.0003821 |
We analyzed the results of each method, from which important lessons were derived as follows. First, for M1, two points are notable: (1) Comparing the performance of M1 with the persistence model on each dataset to consider the relative error, the performance on the BP Solar dataset was worse than that on the Desert Gardens dataset. (2) Even though M5 was superior most of the time, the performance of M1 was the best when MLP was used on the BP Solar dataset. From point (1), it can be derived that the performance of each method varies considerably between different datasets even though they are not vastly different, as their geographical locations are similar. From point (2), it can be derived that the performance of each method varies across different models. The reason that M1 featured a large error in one of the datasets is inferred as follows. When normalized, PV generation in timeslots with small ETR and large ETR become closer to each other. Then, gradient signals with similar magnitudes are backpropagated in each timeslot during training, the two timeslots being treated equally even though their ETRs are different. During testing, the outputs of the model are multiplied by ETR, which leads to amplification of the error. If trained inappropriately, forecasting in the timeslots with a large ETR can be inaccurate compared with a small ETR, resulting in a higher MSE than the base method. However, as shown in point (2), M1 can be the best method depending on the forecasting configuration. Therefore, rather than using the best method for a different dataset, it is important to verify the competitiveness of each method by validation to choose the best method for the dataset and model used. M2 had better performance than M1 on average. Nevertheless, it is not certainly superior to the base method, implying that adjusting the gradient signal can affect the result inappropriately.

M3, which replaces the historical generation with ETR, performed poorly in general. In particular, the MSE of M3 was higher by 150% than the base method when the Desert Gardens dataset and RNN-based models were used. This result suggests two implications: (1) Assuming that ETR performed as a good baseline for prediction, inaccurate forecasting of M3 means a lack of enough meteorological data to predict the attenuation ratio of the atmosphere. (2) Comparing the situations with and without day-ahead generation, the superior performance of the method that includes day-ahead generation implies that day-ahead PV forecasting is effective by using day-ahead generation as a baseline for the prediction. In other words, the neural network-based models improve their performance from the persistence model by considering meteorological information.

The results of M4 and M5 can be analyzed from the perspective above. M4 and M5, which take day-ahead PV generation as one of the inputs, had additional improvement by having one more feature than the base method with day-ahead generation as a baseline. M4 including raw ETR as one of the inputs saw decent improvement in the Desert Gardens dataset while performing worse in the BP Solar dataset. It was more desirable to adjust ETR using meteorological information most of the time in this experiment.

Although M5 had a general advantage in our experiments, there are various circumstances in which PV generation is forecasted in a day-ahead manner in real world applications, and the effectiveness of each method varies depending on models and datasets. Therefore, it should be noted that for a given model and dataset, various methods should be evaluated by validation to effectively integrate ETR into a forecasting model.

### Table 7. The improvements of M5 in MSE relative to the base method in RNN-based models.

| Dataset                  | Vanilla RNN | LSTM  | GRU  |
|--------------------------|-------------|-------|------|
| Alice Springs/BP Solar   | 12.28%      | 0.82% | 3.14%|
| Yulara/Desert Gardens    | 1.13%       | 6.00% | 1.42%|
6. Conclusions

This study presents a simple and effective method to improve the forecasting accuracy of PV generation for mitigating the problems caused by the inherent characteristics of PV, such as variability, intermittency, and limited controllability. This study focused on the ETR strongly associated with the seasonal component of PV as a means to improve forecasting performance. We selected neural network models as the basic forecasting method. Then, we composed five methods to integrate the ETR into them and examined the effect in terms of forecasting performance. The specific integration methods were (1) division preprocessing, (2) multiplication preprocessing, (3) replacement of existing input, (4) inclusion as additional input, and (5) inclusion as an intermediate target.

The methods were tested on MLP, Vanilla RNN, LSTM, and GRU using the two PV datasets. The results show that combining the ETR with existing models can achieve meaningful improvement in forecasting performance and present a new approach to considering seasonal changes in PV generation. Among the methods, including the ETR as an intermediate target (i.e., M5) showed relatively better results than the other integration methods. However, the combination of M5 with Vanilla RNN was the best for one dataset, but the combination of M5 with LSTM was the best for the other. Thus, a certain neural network model combined with the ETR did not show absolute superiority as usual with the comparison results between AI methods.

This study was limited to a few selected neural network models, even though they are known as the methods that effectively deal with time series data. Thus, the effectiveness of the proposed integration methods of the ETR can be further examined in other neural network models. It is also necessary that the proposed methods be applied to other PV datasets in extended studies. Then, the relative superiority of M5 can be further evaluated and more elaborate advice for combining the ETR with neural network-based models can be given.

Author Contributions: Conceptualization, S.C.J. and Y.G.J.; methodology, S.C.J. and Y.G.J.; implementation, S.C.J.; formal analysis, S.C.J. and Y.T.Y.; investigation, Y.G.J. and Y.T.Y.; data curation, S.C.J.; writing—original draft preparation, S.C.J.; writing—review and editing, Y.G.J., Y.T.Y. and H.C.K.; visualization, S.C.J.; supervision, H.C.K.; funding acquisition, H.C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the research grant of Jeju National University in 2017.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available datasets were analyzed in this study. This data can be found here: http://dkasolarcentre.com.au/download, accessed on 10 February 2021.

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

References
1. The International Renewable Energy Agency (IRENA). Renewable Energy Statistics 2020; The International Renewable Energy Agency (IRENA): Abu Dhabi, United Arab Emirates, 2020.
2. International Energy Agency (IEA). World Energy Model; International Energy Agency (IEA): Paris, France, 2020.
3. Karimi, Y.; Oraee, H.; Guerrero, J.M. Decentralized method for load sharing and power management in a hybrid single/three-phase-islanded microgrid consisting of hybrid source pv/battery units. IEEE Trans. Power Electron. 2017, 32, 6135–6144. [CrossRef]
4. Luna, A.C.; Diaz, N.L.; Graells, M.; Vasquez, J.C.; Guerrero, J.M. Mixed-integer-linear-programming-based energy management system for hybrid PV-wind-battery microgrids: Modeling, design, and experimental verification. IEEE Trans. Power Electron. 2017, 32, 2769–2783. [CrossRef]
5. Nicolosi, M. Wind power integration and power system flexibility—An empirical analysis of extreme events in Germany under the new negative price regime. Energy Policy 2010, 38, 7257–7268. [CrossRef]
6. Mohammadi, P.; Mehraeen, S. Challenges of pv integration in low-voltage secondary networks. IEEE Trans. Power Deliv. 2017, 32, 525–535. [CrossRef]
7. Agalgaonkar, Y.P.; Pal, B.C.; Jabr, R.A. Stochastic distribution system operation considering voltage regulation risks in the presence of PV generation. IEEE Trans. Sustain. Energy 2015, 6, 1315–1324. [CrossRef]
8. Woyte, A.; Thong, V.V.; Belmans, R.; Nijs, J. Voltage fluctuations on distribution level introduced by photovoltaic systems. *IEEE Trans. Energy Convers.* **2006**, *21*, 202–209. [CrossRef]

9. 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]

10. Han, S.; Qiao, Y.H.; Yan, J.; Liu, Y.Q.; Li, L.; Wang, Z. Mid-to-long term wind and photovoltaic power generation prediction based on copula function and long short term memory network. *Appl. Energy* **2019**, *239*, 181–191. [CrossRef]

11. Dolara, A.; Leva, S.; Manzolini, G. Comparison of different physical models for PV power output prediction. *Sol. Energy* **2015**, *119*, 83–99. [CrossRef]

12. Colak, M.I.; Yesilbudak, N.G.; Bayindir, R. Multi-period prediction of solar radiation using ARMA and ARIMA models. In *Proceedings of the IEEE 14th International Conference on Machine Learning and Applications (ICMLA)*, Miami, FL, USA, 9–11 December 2015. [CrossRef]

13. Li, Y.; Su, Y.; Shu, L. An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renew. Energy* **2014**, *66*, 78–89. [CrossRef]

14. Martínez-García, Y.M.; Zhang, K.S.; Zhang, Y.-D. Deep recurrent entropy adaptive model for system reliability monitoring. *IEEE Trans. Ind. Inform.* **2021**, *17*, 839–848. [CrossRef]

15. Martínez-García, Y.M.; Zhang, J.W.; McGinty, J. Visually interpretable profile extraction with an autoencoder for health monitoring of industrial systems. In *Proceedings of the IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*, Toyonaka, Japan, 3–5 July 2019. [CrossRef]

16. 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]

17. Yacef, R.; Benghanem, M.; Mellit, A. Prediction of daily global solar irradiation data using Bayesian neural network: A comparative study. *Renew. Energy* **2012**, *48*, 146–154. [CrossRef]

18. Abdel-Nasser, M.; Mahmoud, K. Accurate photovoltaic power forecasting models using deep LSTM-RNN. *Neural Comput. Appl.* **2019**, *31*, 2727–2740. [CrossRef]

19. Wang, Y.; Liao, W.; Chang, Y. Gated recurrent unit network-based short-term photovoltaic forecasting. *Energies* **2018**, *11*, 2163. [CrossRef]

20. Keles, D.; Genoese, M.; Most, D.; Fichtner, W. Comparison of extended mean-reversion and time series models for electricity spot price simulation considering negative prices. *Energy Econ.* **2012**, *34*, 1012–1032. [CrossRef]

21. Zhu, H.; Li, X.; Sun, Q.; Nie, L.; Yao, J.; Zhao, G. A power prediction method for photovoltaic power plant based on wavelet decomposition and artificial neural networks. *Energies* **2016**, *9*, 11. [CrossRef]

22. Zang, H.; Cheng, L.; Ding, T.; Cheung, K.W.; Wei, Z.; Sun, G. Day-ahead photovoltaic power forecasting approach based on deep convolutional neural networks and meta learning. *Int. J. Electr. Power Energy Syst.* **2020**, *118*, 105790. [CrossRef]

23. Li, P.; Zhou, K.; Lu, X.; Yang, S. A hybrid deep learning model for short-term PV power forecasting. *Appl. Energy* **2020**, *259*, 114216. [CrossRef]

24. Wang, F.; Xuan, Z.; Zhen, Z.; Li, K.; Wang, T.; Shi, M. A day-ahead PV power forecasting method based on LSTM-RNN model and time correlation modification under partial daily pattern prediction framework. *Energy Convers. Manag.* **2020**, *212*, 112766. [CrossRef]

25. Li, G.; Xie, S.; Wang, B.; Xin, J.; Li, Y.; Du, S. Photovoltaic power forecasting with a hybrid deep learning approach. *IEEE Access* **2020**, *8*, 175871–175880. [CrossRef]

26. Wen, S.; Zhang, C.; Lan, H.; Xu, Y.; Tang, Y.; Huang, Y. A hybrid ensemble model for interval prediction of solar power output in ship board power systems. *IEEE Trans. Sustain. Energy* **2021**, *12*, 14–24. [CrossRef]

27. Gigoni, L.; Betti, A.; Crisostomi, E.; Franco, A.; Tucci, M.; Bizzarri, F.;ucci, D. Day-ahead hourly forecasting of power generation from photovoltaic plants. *IEEE Trans. Sustain. Energy* **2018**, *9*, 831–842. [CrossRef]

28. Duffie, J.; Beckman, W. *Solar Engineering of Thermal Processes*, 4th ed.; John Wiley & Sons, Inc.: Hoboken, NJ, USA, 2013.

29. Peng, L.; Zhu, Q.; Lv, S.X.; Wang, L. Effective long short-term memory with fruit fly optimization algorithm for time series forecasting. *Soft Comput.* **2020**, *24*, 15059–15079. [CrossRef]

30. Ma, Y.; Zhang, Z.; Li, D.; Tang, M. Deflated reputation using multiplicative long short-term memory neural networks. *Future Gener. Comput. Syst.* **2021**, *118*, 198–207. [CrossRef]

31. Desert Knowledge Australia Centre. 11 November 2008. BP Solar, 5.0 kW, poly-Si, Fixed, 2008. Alice Springs. Available online: http://dkasolarcentre.com.au/source/alice-springs/dka-m5-c-phase (accessed on 10 February 2021).

32. Desert Knowledge Australia Centre. 3 March 2016. Desert Gardens, 1058.4 kW, poly-Si, Fixed, 2016. Alice Springs. Available online: http://dkasolarcentre.com.au/download?location=yulara (accessed on 10 February 2021).