

Article

Inverter Efficiency Analysis Model Based on Solar Power Estimation Using Solar Radiation

Chul-Young Park ¹, Seok-Hoon Hong ¹, Su-Chang Lim ¹, Beob-Seong Song ²,
Sung-Wook Park ³, Jun-Ho Huh ^{4,*} and Jong-Chan Kim ^{3,*}

¹ TEF Co., Ltd., 115-105 Apgok-Gil, Seo-Myeon, Suncheon-City, Jeollanam-do 57903, Korea; naksu21@sunchon.ac.kr (C.-Y.P.); seokhoon79@naver.com (S.-H.H.); suchangLim@sunchon.ac.kr (S.-C.L.)

² Iumict Co., Ltd., 255 Jungang-Ro, Suncheon-City, Jeollanam-do 57922, Korea; iumict.co.ltd@gmail.com

³ Department of Computer Engineering, Suncheon National University, 255 Jungang-ro, Suncheon-City, Jeollanam-do 57922, Korea; park7231654@gmail.com

⁴ Department of Data Informatics, Korea Maritime and Ocean University, Busan 49112, Korea

* Correspondence: 72networks@pukyong.ac.kr (J.-H.H.); seaghost@sunchon.ac.kr (J.-C.K.);
Tel.: +82+41-410-4371 (J.-H.H.); +82-61-750-3620 (J.-C.K.)

Received: 12 August 2020; Accepted: 29 September 2020; Published: 1 October 2020



Abstract: The photovoltaic (PV) industry is an important part of the renewable energy industry. With the growing use of PV systems, interest in their operation and maintenance (O&M) is increasing. In this regard, analyses of power generation efficiency and inverter efficiency are very important. The first step in efficiency analysis is solar power estimation based on environment sensor data. In this study, solar power was estimated using a univariate linear regression model. The estimated solar power data were cross-validated with the actual solar power data obtained from the inverter. The results provide information on the power generation efficiency of the inverter. The linear estimation model developed in this study was validated using a single PV system. It is possible to apply the coefficients presented in this study to other PV systems, even though the nature and error rates of the collected data may vary depending on the inverter manufacturer. To apply the proposed model to PV systems with different power generation capacities, reconstructing the model according to the power generation capacity is necessary.

Keywords: PV system; linear model; linear regression; inverter efficiency; PV power generation forecasting; Smart Grid; Micro Grid

1. Introduction

Renewable energy is attracting worldwide attention. Wind power generation and solar power generation are typically used. Recently, there has been a growing global interest in the use of wind energy, which accounts for about 10% of Europe's energy consumption structure and more than 15% of energy consumption in the United States and Spain. In addition, the cumulative installed wind power capacity worldwide reached nearly 591.55 GW at the end of 2018 [1].

The operation of wind power systems is vulnerable to stochastically unstable wind speed and can adversely affect energy transport and power grid operations. It is essential to improve the accuracy and consistency of wind speed forecasts. Accurate forecast results and dispatch departments can help to easily and effectively coordinate programs. Wind farms can minimize the negative impact on the grid and maximize the use of wind power in the global electricity market [2].

The method of maximizing the efficiency of wind power generation is to detect defects in the wind turbine system in advance. Wind turbine systems are complex industrial systems with harsh operating conditions. Research is underway to develop techniques for diagnosing and classifying

failures based on the Fast Fourier Transform (FFT) and uncorrelated multi-linear principal component components for wind turbine systems [3].

The solar power industry is in development as an essential core of the field, and the use of photovoltaic (PV) systems is on the rise [4,5]. The Korean government has set up policies prioritizing the installation of PV systems on unused sites (BIPV: building-integrated photovoltaics) such as rooftops and parking lots. Moreover, with the recent exponential spread of PV systems, the importance of their long-term performance and quality is drawing attention.

The Korean solar industry generally helps households with their power bills or creates revenue by commercially offering power from the PV systems to the operators of existing power grids. Investors invest in the installation of PV systems with the goal of long-term profits. To ensure the feasibility of such investments, the PV facilities need to guarantee operational efficiency for at least 25 years, for which the certification system is currently only applied to the level of products such as PV modules and inverters. The power output actually delivered to the grids connected to households varies according to the efficiency of the overall system, which is outside the scope of product certification. Moreover, without investments in O&M (operation and maintenance), the profits of investors may be hampered due to reduced power output.

The current determination of inverter efficiency depends on the data measured by the inverter, whose type may affect the reference level of efficiency. Thus, there is a need for performance diagnosis technology that is capable of accurately identifying performance from the solar input energy to the system output as well as any loss and fault in the process [1]. This has triggered interest in the O&M of PV systems, which ensures their maximum performance while minimizing energy loss up to the end of their lifecycle [6,7].

The O&M of PV systems is significant in terms of the economic feasibility of guaranteeing revenue. PV systems inevitably suffer from the constant reduction of power output, not just due to the natural aging of the entire system including the inverter but also owing to the deterioration of the solar modules exposed to UV rays. From the perspective of O&M, inverter efficiency is an essential consideration directly linked to the cost [8]. If reliable power output estimates are available on-site, a reduction of power output can be preemptively identified by comparing it with the actual output of the PV facilities. Likewise, the data can be applied to individual related facilities to maintain optimal conditions by predicting and identifying faults while detecting performance deterioration. In addition, the real-time estimation of reference solar power output in current technical states can help with diagnosing any error of the output system, against which immediate measures can be taken to improve performance.

One of the methods of identifying an issue with the inverter is to estimate power output based on meteorological data [9–11]. Solar power output is greatly affected by meteorological conditions that include but are not limited to the ambient temperature, solar radiation, the UV index, humidity, the total cloud amount, and wind velocity, and power output is the one most closely related to solar radiation. Thus, it can be assumed that solar radiation and power output have a strong correlation. We established the hypothesis that an issue with the inverter can be diagnosed by extracting the pattern of differences between its actual power output and the output estimated based on solar radiation sensor data.

Measuring inverter efficiency requires estimating power output. There have been many studies on power output estimation by applying data analysis or those applying new analysis methods such as O&M-related inverter efficiency analysis [12–14]. In addition, the current wide acceptance of PV systems initiated studies to develop efficient O&M methods [15–17]. Especially, industry and communities have taken an interest in studies on boosting the accuracy of the algorithm for the analysis of inverter efficiency or power output efficiency [18,19].

The current determination of inverter efficiency depends on the data measured by the inverter, whose type may affect the reference level of efficiency. This study proposes a model capable of real-time efficiency measurement through reference models for diverse types of inverters while improving the accuracy of efficiency measurement by fitting a linear model based on solar radiation sensors.

The PV system selected for this study is a grid-connected system consisting of a 10 kW single-phase inverter. It collects and transmits to the cloud system the inverter status data from the real-time unit (RTU) and data from the environment sensors. We applied the data refinement methods described in Chapter 4 to the transmitted data and used them in the experiment.

Univariate data were employed for the goal of measuring inverter efficiency by estimating real-time power output. It is based on the belief that the univariate data will allow us to obtain the desired results without having to resort to complicated models such as neural networks.

The RTU is limited in terms of processor resources and is difficult to apply to highly complex models such as neural networks. Likewise, the linear regression model with a low calculation capacity was used, as it was deemed sufficient for estimating univariate data.

We used the univariate linear regression model for estimating power output based on data from environmental sensors, and the fitted model was subsequently used to measure efficiency by making comparisons with the data from the inverter. After setting up the linear regression model, we verified the algorithm for measuring inverter efficiency in order to provide the optimal system for O&M while making real-time estimates of output reduction from the PV system. We also confirmed the trend of power output through statistical verification methods, based on which we found that the real-time diagnosis of inverter efficiency can be made.

The contents of this paper are organized as follows. In Section 2, the PV system overview, solar power estimation, and linear regression algorithm used in this study for solar power estimation are described. The PV monitoring system, which is the basic element of this study, is defined in Section 3. Section 4 explains the proposed linear regression model analysis method, describing an overview of the analysis method, data cleansing, refinement, linear modeling, and validation. In Section 5, the performance of the analysis method is evaluated using objective indicators, and the final section concludes the paper.

2. PV System and Solar Power Estimation Model

2.1. PV System Overview

PV systems directly convert solar energy to electricity. When sunlight is directed onto semiconductors (silicon-N type and P type), electrons move between the connected electrodes, generating an electric current [20,21]. Presently, most power generation systems use fossil fuels for power generation and require complex machinery. They also emit carbon dioxide due to the use of fossil fuels. PV systems use semiconductors for converting solar energy into electrical energy. PV systems are very eco-friendly because their construction is relatively simple, and unlike fossil fuels, they do not emit pollutants. They are also cost effective, as their operation and maintenance are straightforward.

Figure 1 shows the configuration of a PV system. The system is divided into “Collect”, “Storage”, and “Distribute and Use” sections. The “Collect” section is a solar cell that directly produces electricity and consists of a module and an array. The “Storage” section stores the extra power that remains after use or sale and the power produced during the daytime for power supply at night. The “Distribute and Use” section converts direct current (DC) into alternating current (AC).

PV systems are classified into stand-alone, grid-connected, building-integrated, and hybrid PV systems. Stand-alone PV systems operate independently of the commercial electric utility grid [22]. The shortcoming of these systems is that the efficiency of the storage batteries is lowered in the process of using the batteries to supply power at night, or on days when power generation is impossible. The service life of storage batteries is only 3–5 years, while that of solar cells is approximately 25 years. Grid-connected PV systems are designed to complement the stand-alone PV systems [23]. They are divided into two types: those with a battery backup device and those with an integrated energy storage device in the form of a battery as shown in Figure 2.

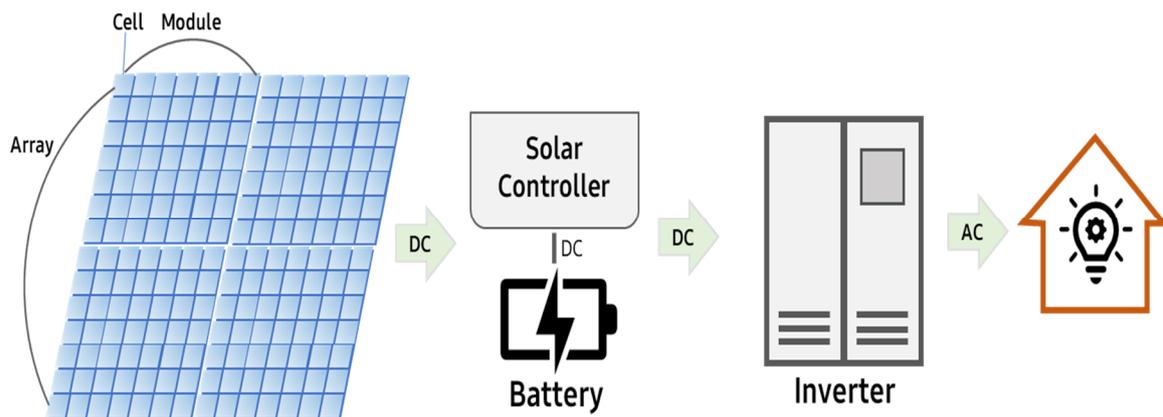


Figure 1. Photovoltaic (PV) system overview.

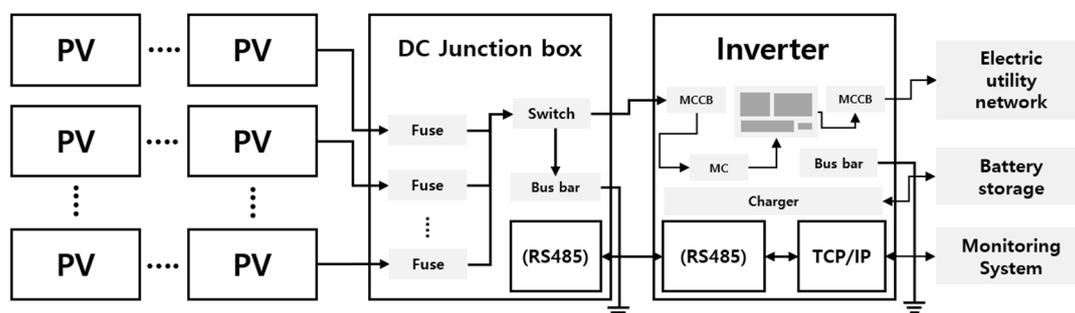


Figure 2. Grid-connected PV system [24].

Grid-connected PV systems supply power to the commercial power grid after converting the produced power into AC and use power from the public grid if necessary. Building-integrated PV systems (BIPVs) produce power by installing PV modules on building exteriors such as rooftops, curtain walls, balconies, sunshades, and panels, and directly supply power to the building [25]. Hybrid PV systems are equipped with auxiliary means such as diesel generators to ensure a stable power supply when solar power generation is not possible due to a lack of sunlight.

2.2. Solar Power Estimation and Inverter Efficiency Analysis

The electricity produced by solar cells is DC. The produced electricity is not constant over time, as the intensity of sunlight is dependent on weather conditions. As the magnitude of energy generated through PV systems frequently changes, it is difficult to directly use the power obtained from PV modules. In general, the characteristics of the output from solar cells depend on the solar radiation, surface cleanliness of the solar cells, and attributes of the environment of the solar cell array such as the surface operating temperature of the cell [26]. The energy loss in PV systems is a major factor affecting energy output. It is most important to convert the uneven DC power into stable DC or AC.

A key element in stable power conversion is the inverter. The inverter is responsible for the uniform output of power from PV modules. The efficiency of the inverter can be said to be good when there is no significant difference between the input power and the output power. However, owing to the nature of the device, power losses are inevitable. System losses are losses that occur in the course of converting DC, which is produced by PV panels and solar cells, into AC, a consumable form.

All the energy losses that occur in the inverter are considered system losses. In the case of DC-to-AC conversion, no converter can achieve 100% efficiency. This means that the output (AC) energy is not as high as the input (DC) energy. The efficiency of the inverter generally ranges from 95 to 98%. The efficiency may vary depending on the DC input power and voltage. Research is being conducted to maintain the efficiency of the inverter by extracting the maximum power from the PV panel using the Maximum Power Point Tracking (MPPT) algorithm [27,28]. To determine the actual efficiency of the

inverter rather than rely on the efficiency provided by the manufacturer, the produced solar power must be estimated first [29].

Power output varies mainly with environmental factors. Solar radiation is a primary factor affecting power output. Some studies are ongoing with the goal of estimating solar radiation to predict future power output [30–32]. There are also studies on power output estimation based on ambient temperature, wind velocity, and incident light [33–35]. The method is based on historical weather data; there is a high correlation between the weather conditions in the present or past, and the solar power generation in the future. For solar power estimation, artificial neural networks, support vector machine (SVM), and machine learning have been utilized [36–38]. Techniques that use Long Short-Term Memory (LSTM), a time-series analysis method for weather data [39,40], as well as techniques that use both the past and current weather data, have also been proposed. Other studies introduce methods that use the adaptive linear time series model, and a technique for applying both past data and forecasts to the fuzzy decision tree model [41–43].

There have also been other studies based on the linear regression model used in this study, for solar power estimation. For example, a multiple linear regression model was constructed for system-life evaluation and solar power estimation, using data such as the number of solar panels, the number of inverters, and geographic space (GIS), as well as solar radiation data [44,45].

2.3. Linear Regression Model

A linear model is capable of relatively fast calculations, due to the small system load [46,47]. Linear regression is an algorithm for finding the parameters w (weight) and b (bias) that minimize the mean squared error for the training set. Linear regression based on one explanatory variable is referred to as simple linear regression, and that based on two or more explanatory variables is referred to as multiple linear regression [48]. Linear regression uses a linear estimation function to model a regression equation and estimates unknown parameters from available data [49].

A linear regression model assumes that there is a linear correlation between one or more independent variables X and dependent variables Y , and finds a relationship from the given data. A linear regression model assumes the independent and dependent variables to be related as shown in Equation (1).

$$y = \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

where ε is an independent error term that follows a normal distribution whose mean is zero and standard deviation is σ . For regression analysis, the given data can be defined as the matrix of independent variables X and the vector of dependent variables Y ; β values representing regression coefficients can be expressed as a vector as shown in Equation (2). p is the number of independent variables, and n is the number of data.

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{n1} & \cdots & x_{np} \end{bmatrix}, Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad (2)$$

Applying Equation (2) to Equation (1) yields the following.

$$Y = X\beta + \varepsilon \quad (3)$$

The regression coefficients that minimize the error are estimated as follows.

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4)$$

The equation for estimating the value of \hat{Y} with the estimated regression coefficient $\hat{\beta}$ is given by,

$$\hat{Y} = X\hat{\beta} \quad (5)$$

The analysis of variance (ANOVA) must be conducted to test the statistical significance of the regression model. The sum of squares of total (SST), the sum of squared errors of prediction (SSE), and the sum of squares error (SSR) required to perform ANOVA using the observed value Y , the mean \bar{Y} of the observed value, the estimated value \hat{Y} , and the mean $\bar{\hat{Y}}$ of the estimated value are as follows:

$$SST = \|Y - \bar{Y}\|^2 \quad (6)$$

$$SSE = \|Y - \hat{Y}\|^2 \quad (7)$$

$$SSR = \|\hat{Y} - \bar{\hat{Y}}\|^2 \quad (8)$$

where $\| \cdot \|$ is L2-Norm.

The suitability of the linear regression model for predicting the actual model must be determined considering various statistics derived through ANOVA. Representative criteria include the coefficient of determination (R^2), adjusted coefficient of determination (adjusted R^2), standard error of the estimate from the regression equation, and statistical significance of the regression equation (t statistics of each regression coefficient).

The coefficient of determination is an index that represents the ratio of the variance explained by the regression model to the total variance of the dependent variables. The coefficient of determination has a value between 0 and 1. A value closer to 1 indicates that the regression model has higher predictive power. Below is the coefficient of determination.

$$R^2 = \frac{SSR}{SST} \quad (9)$$

The adjusted R -squared is an estimate for the coefficient of determination of the population, which reflects the degree of freedom for correcting the tendency of the coefficient of determination of the sample data to become larger than that of the population. The adjusted R -squared R_{adj}^2 is given by

$$R_{adj}^2 = 1 - \frac{(n-1)(1-R^2)}{n-p-1} \quad (10)$$

The standard error represents the size of the deviation that is not explained by the regression model, and is expressed as follows.

$$SE = \sqrt{\frac{SSE}{n-p-1}} \quad (11)$$

The hypothesis for testing whether the estimated regression equation is statistically significant is stated in the following equation.

$$\begin{aligned} H_0 &: \beta_1 = \beta_2 = \dots = \beta_p = 0 \\ H_1 &: H_0 \end{aligned} \quad (12)$$

H_0 is the null hypothesis, whereas H_1 stands for the alternative hypothesis. Here, the null hypothesis is that all the regression coefficients are 0, and the alternative hypothesis is that the regression equation is meaningful. With a significance level of 0.05, the null hypothesis will be rejected when the p -value is lower than 0.05. It means that the selection of the alternative hypothesis provides significance to the regression equation.

The F-Test statistic for testing the hypothesis of Equation (12) is given below.

$$F_0 = \frac{SSR/p}{SSE/(n-p-1)} \quad (13)$$

where F is an abbreviation for the F -test.

It is possible to test whether the regression coefficient of an independent variable is statistically significant; the hypothesis for the i -th independent variable is as follows.

$$\begin{aligned} H_0 &: \beta_i = 0 \\ H_1 &: \beta_i \neq 0 \end{aligned} \quad (14)$$

The T -Test statistic for testing the hypothesis of Equation (14) is shown in Equation (15). c_{ii} is the i -th diagonal element of $(X^T X)^{-1}$.

$$t_i = \frac{\hat{\beta}_i}{SSE \sqrt{c_{ii}}} \quad (15)$$

where t is an abbreviation for the T -test.

The indicators for the statistical significance of the regression model are the statistics of Equations (10), (11), (13), and (15) listed above. A statistically significant model must satisfy the above conditions.

3. Yield from TEF Energy Network Monitoring System Implementation

3.1. PV System Monitoring

The PV system (Figure 3a,b) for solar power estimation and analysis is a grid-connected system consisting of a fixed array-type 10 kW single-phase inverter (HexPower, Hex Power System, Seoul, Republic of Korea) and an environment sensor (MRT, MRT System, Kwangju, Republic of Korea). It is installed on the rooftop of the “M” building in Yeonhang-dong, Suncheon-si, Jeollanam-do (Figure 3c). The annual mean time of solar power generation of the area was 3.5 h, with an average wind speed of 28 m/s and meridian altitude of 78.5° (summer) and 31.5° (winter).

In this study, the solar power of the 10 kW inverter was analyzed using the vertical solar radiation, module temperature, horizontal solar radiation, and outside temperature. Figure 4 shows the configuration of the PV system and monitoring system.

The data collection targets of the PV system are the inverter and environment sensors; the data were transmitted to the cloud server after data preprocessing in the real-time unit (RTU). The data collected from the inverter included the real-time status information, accumulated solar power, and inverter error information. The environment sensors for weather observation recorded the vertical solar radiation, module temperature, horizontal solar radiation, and outside temperature. The cloud provided PV system, inverter, and weather observation information through an external access interface for data extraction and linkage (REST API, Representational State Transfer Application Programming Interface). Table 1 shows the detailed specifications of the PV monitoring system.

Table 1. Specifications of photovoltaic monitoring system.

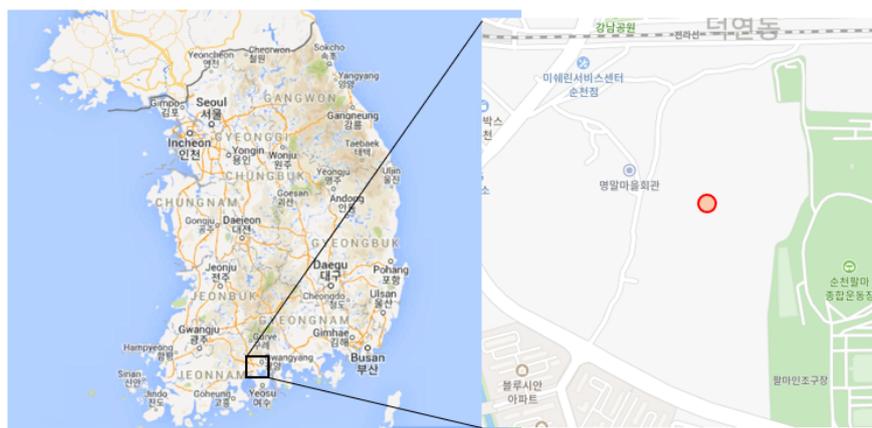
Type	Description
Operating System	Ubuntu 18.04 LTS
Hardware	Intel Xeon E5530 2.4 GHz 8 MB L3/8 GB RAM/8 TB
RTU	Windows 7/Intel J5005/4 GB RAM/1 TB
Protocol	TCP/IP/RS485, MODBUS



(a) Environment sensor and PV panel.



(b) Inverter and junction box.



(c) Region of installed PV System outside Suncheon City, Jeollanam-do, used as a case study.

Figure 3. Photovoltaic monitoring system components.

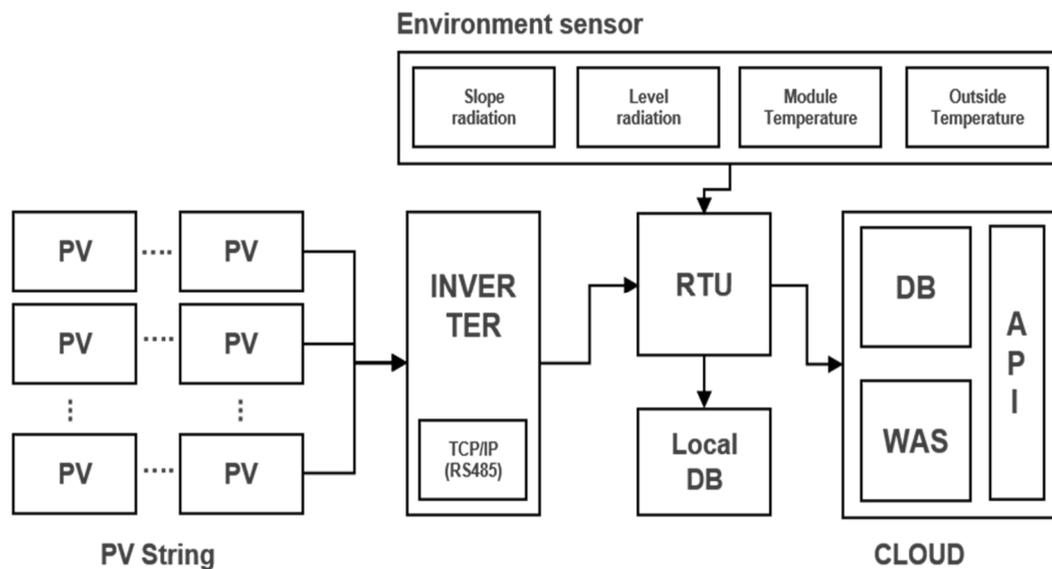


Figure 4. Photovoltaic monitoring system architecture.

3.2. Collection of PV System Data

The inverter status information, accumulated solar power, error information, and environment sensor data for weather observation were collected. The inverter data were collected every minute, and the environment sensor data were collected every five minutes and stored in a database. The data used in this study were collected from August 2017 to February 2019. Table 2 shows a table schema of the collected environment sensor data.

Table 2. Environment data schema for photovoltaic monitoring system.

Column	Data Type	Default	Comment
env_index	INT(11)	AUTO_INCREMENT	Index
env_date	DATE	NULL	DATE
env_time	TIME	NULL	TIME
env_horizonsolar	FLOAT(12)	NULL	horizontal solar radiation
env_vertsolar	FLOAT(12)	NULL	vertical solar radiation
env_modtemp	FLOAT(12)	NULL	module temperature
env_airtemp	FLOAT(12)	NULL	outside temperature

Table 3 shows a table schema of the collected inverter data.

Table 3. Inverter data schema for photovoltaic monitoring system.

Column	Data Type	Default	Comment
pow_index	INT(11)	AUTO_INCREMENT	Index
pow_id	INT(11)	"1"	Inverter id
pow_date	DATE	NULL	DATE
pow_time	TIME	NULL	TIME
pow_dcv	INT(11)	NULL	dc voltage
pow_dca	DOUBLE(22,0)	NULL	dc ampere
pow_dcp	DOUBLE(22,0)	NULL	dc power (W)
pow_acvr	INT(11)	NULL	ac voltage (R)
pow_acvs	INT(11)	NULL	ac voltage (S)
pow_acvt	INT(11)	NULL	ac voltage (T)
pow_acar	FLOAT(12)	NULL	ac ampere (R)
pow_acas	FLOAT(12)	NULL	ac ampere (S)

Table 3. Cont.

Column	Data Type	Default	Comment
pow_acat	FLOAT(12)	NULL	ac ampere (T)
pow_acp	DOUBLE(22,0)	NULL	ac power (W)
pow_pf	DOUBLE(22,0)	NULL	power factor
pow_totpower	DOUBLE(22,0)	"0"	cumulative power (W)
pow_freq	FLOAT(12)	NULL	frequency (Hz)

4. Proposed Method of Analysis

4.1. Data Preprocessing

The data used in this study were collected from the grid-connected system with a single 10 kW inverter installed on the rooftop of building M located in Suncheon City, Jeollanam-do, from August 2017 to February 2019. Figure 5 shows the flowchart of the proposed analysis method.

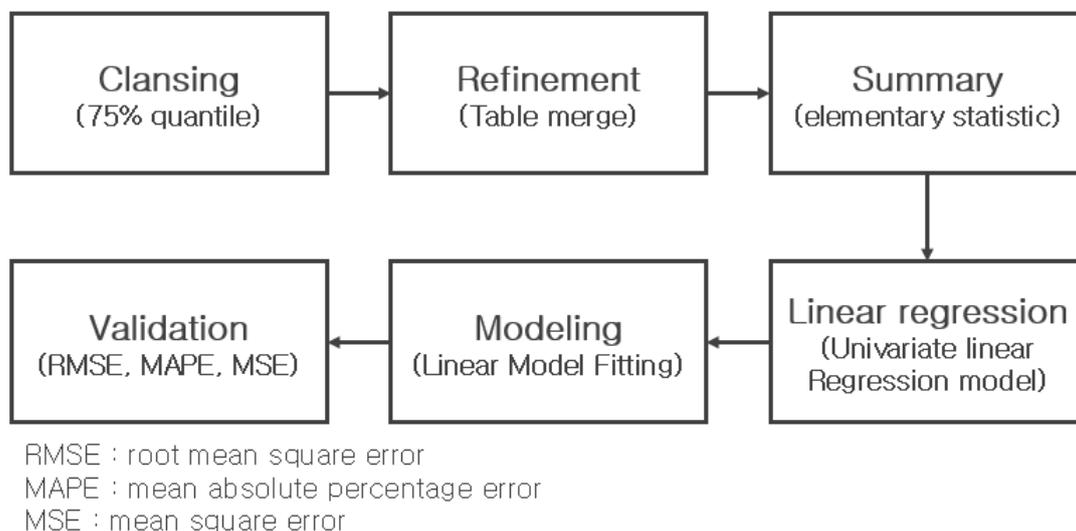


Figure 5. Flowchart for data analysis process.

Data samples containing outliers caused by system problems or collection equipment errors were subjected to cleansing by calculating the 75% quantile.

$$Quartile3 + 1.5 \times IQR \quad (16)$$

The interquartile range (IQR) of the DC power (pow_dcp) data in Table 3 was calculated. If the solar power was higher than that provided by Equation (16), it was determined as an outlier and removed. In the PV monitoring system, the sampling period for the inverter data was seven minutes and that for the weather observation data was one minute. In the data refinement process, the sampling periods for all the collected data were converted to 15 min so that the inverter and environment sensor data could have the same sampling period, enabling the integration of the tables.

4.2. Linear Correlation Analysis

From a PV system O&M perspective, the efficiency of the inverter is a very important consideration, as it is directly related to cost. The conventional method of determining inverter efficiency is based on the data measured from the inverter, and the reference values of efficiency measurement may vary depending on the inverter type. This study proposes a model capable of real-time efficiency

measurement, due to the increased accuracy of efficiency measurement through linear model fitting based on solar radiation sensors, thereby suggesting a reference model for all types of inverters.

In the preliminary step of validating the application of the linear model, the basic statistics for the data were analyzed, and the results are shown in Table 4.

Table 4. Elementary statistics for photovoltaic monitoring system data.

Data (Unit)	Minimum	Median	Mean	Maximum	Standard Deviation
DC Power (W)	200	4900	4680	11,200	2125.33
AC Power (W)	100	4800	4568	11,100	2135.06
Vertical solar Radiation (W/m ²)	9.733	518.533	572.657	1026.677	201.08
Horizontal solar Radiation (W/m ²)	9.733	572.200	572.857	1211.933	271.37

Table 4 shows the basic statistics; the median values are higher than the mean values, and the standard deviation values are 2125 and 2135, respectively, for DC and AC power. DC and AC power, however, are skewed left towards a normal distribution, with no significant difference between the median and mean values. Figures 6 and 7 show the linear relationships between the vertical and horizontal solar radiations, and DC power, which is unrelated to the inverter conversion efficiency.

Figure 6 shows the scatter plot for the linear relationship between the vertical solar radiation and DC power. The correlation coefficient for vertical solar radiation and DC power is 0.907067 as shown in Table 5, indicating a positive linear correlation.

Figure 7 shows the scatter plot for the linear relationship between horizontal solar radiation and DC power. The correlation coefficient for horizontal solar radiation and DC power is 0.937929 as shown in Table 5, indicating a positive linear correlation.

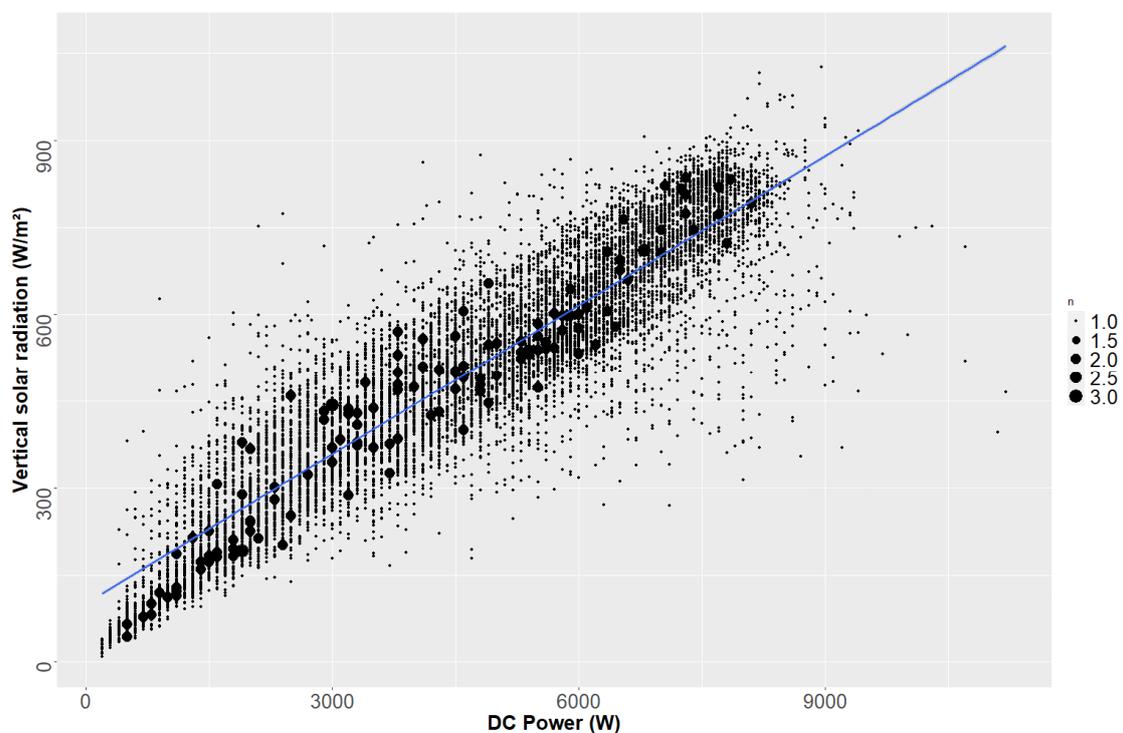


Figure 6. Scatter plot for the linear relationship between Direct Current (DC) power (W) and vertical solar radiation (W/m²).

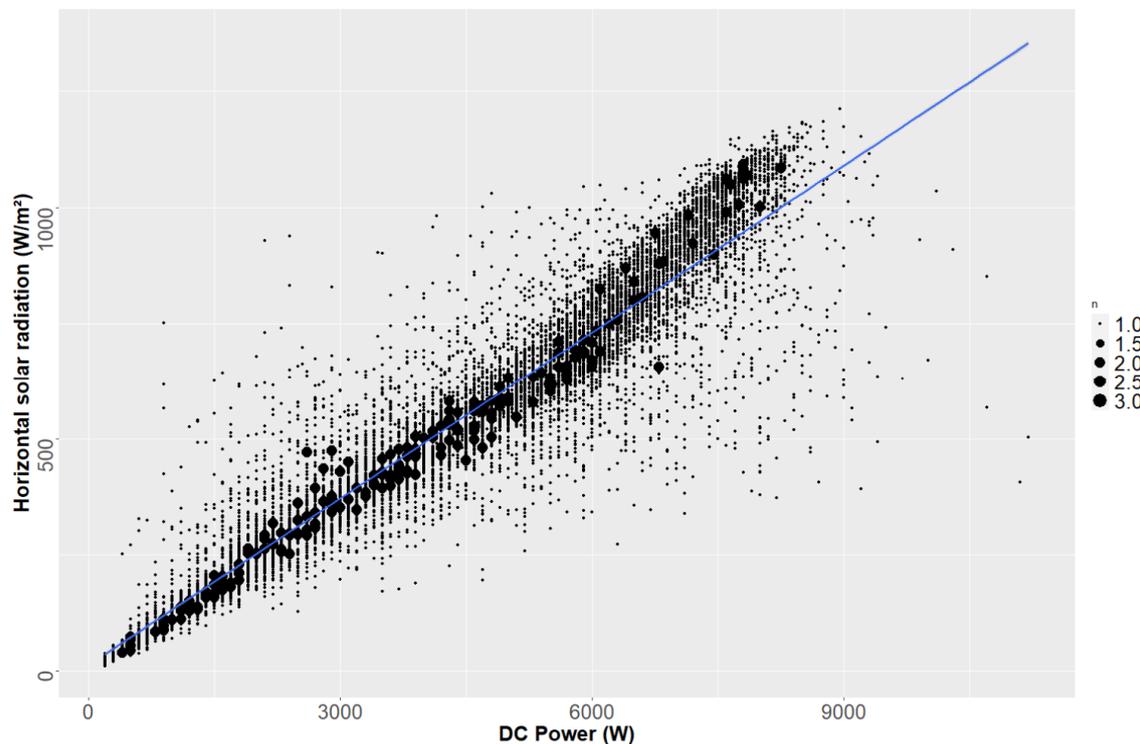


Figure 7. Scatter plot for the linear relationship between DC power (W) and horizontal solar radiation (W/m^2).

Table 5. Correlation coefficients for power and solar radiation.

Corr.	Vertical Solar Radiation	Horizontal Solar Radiation
DC Power	0.907067	0.937929
AC Power	0.906583	0.937743

Based on Figures 6 and 7 as well as Tables 4 and 5, linear modeling was performed using the horizontal solar radiation, which exhibited a high correlation with DC power, which is not related to the inverter conversion efficiency, as a parameter.

5. Results and Discussion

5.1. Target PV Monitoring System

Based on the method proposed in Sections 3 and 4, data cleansing, refinement, basic statistical analysis, and linear correlation analysis were carried out. Linear model fitting showed a higher correlation between the DC power data from the inverter, and the horizontal solar radiation data from the environment sensors, as compared to other parameters.

Therefore, the basic expression of the linear estimation model is as given by Equation (17) below.

$$\text{Estimate} \leftarrow \text{lm}(\text{DC power} \sim \text{Horizontal solar radiation}) \quad (17)$$

“lm” means the linear estimated model, whereas the values in parentheses are the data entered into the estimation model. We used “~” to express the correlation between DC power and horizontal solar radiation. Consequently, the value obtained from lm is displayed as the result of “Estimate” to the left.

The results of linear model fitting are shown in Table 6.

Table 6. Fitting results for linear model between DC power (W) and horizontal solar radiation.

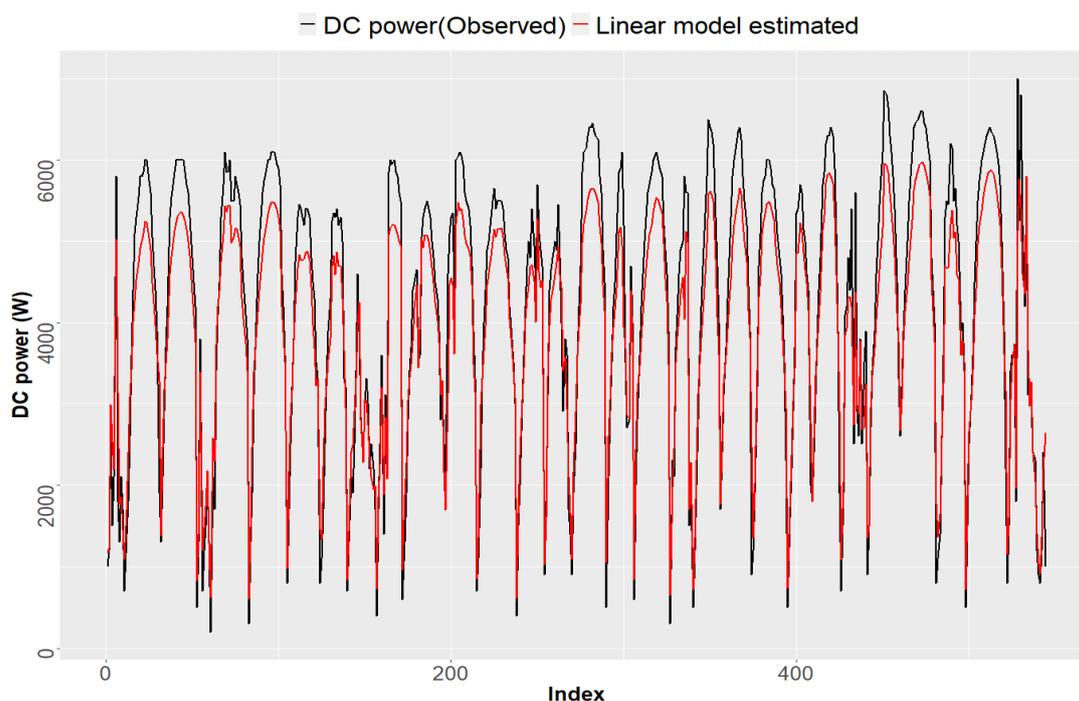
Variable	Vertical Solar Radiation	Horizontal Solar Radiation
(intercept)	471.84924	$< 2.2 \times 10^{-16}$
Horizontal solar radiation	7.34580	$< 2.2 \times 10^{-16}$
R^2		0.8228

Solar radiation (W/m²).

Table 6 shows the results of linear model fitting that used DC power and horizontal solar radiation as parameters. The significance probability (p -value) of the parameter used (horizontal solar radiation) is much smaller than the significance level of 0.05 (5%). Therefore, the alternative hypothesis was adopted, indicating that the independent and dependent variables were highly correlated. The correlation coefficient (R^2) of the model was 0.8228, and thus, its goodness of fit (explained) was 82%. The constructed linear model is expressed by Equation (18).

$$y_i = 7.34580x_i + 471.84924 \quad (18)$$

In Equation (18), x_i is the horizontal solar radiation, an independent variable. y_i stands for the solar power generation calculated by applying horizontal solar radiation to the linear model. The solar power in January 2019 was estimated using the model constructed with the equation, and the results are shown in Figure 8.

**Figure 8.** Comparison of observed and linear-model estimated values of DC power.

In Figure 8, the black solid line represents DC power, and the red solid line shows the results of the linear estimation model that used the horizontal solar radiation data collected through the environment sensor, as a parameter. Table 7 shows that the results of running the linear estimation model closely follow the trend of the observed data.

Table 7 shows the comparison of the solar power generation estimated with the linear model proposed by this work and the data collected from August 2017 to February 2019. The “DC power” column is the actual solar power generation according to the collected data, whereas the “linear model estimated” column means the solar power generation estimated by utilizing the proposed model

and the horizontal solar radiation data. We also extracted 544 pieces of data at random from all the horizontal solar radiation data to estimate the power output.

Table 7. Comparison between observed and linear-model estimated values of DC power.

Index	DC Power (Observed)	Linear Model Estimated	Difference
1	1000	1253.4424	−253.4424
125	1100	1417.9883	−317.9883
249	4400	4073.9147	326.0853
342	2000	2384.0528	−384.0528
435	2900	3055.9967	−155.9967
544	1000	1448.8406	−448.8406

5.2. Model Validation

For the validation of the estimation model, which is the final step in the flowchart of the analysis method shown in Figure 5, the root mean square error (RMSE) and mean absolute percentage error (MAPE) were determined, and residual validation was performed [50]. The equations for the RMSE and MAPE are expressed as follows:

$$\sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - f_t)^2} \quad (19)$$

$$\frac{100}{T} \sum_{t=1}^T \frac{|y_t - f_t|}{y_t} \quad (20)$$

Here, t is the time index, whereas y_t is the observed value and f_t is the estimate. The MAPE is used as the index to present the ratios of errors in the estimates. The smaller it is, the better the model.

Table 8 shows the RMSE and MAPE calculated for different tolerance values or margins of error, for the inverter data. The linear estimation model exhibited a 12% error rate when there was no tolerance, and a 9.5% error rate when a 10% tolerance or margin of error was considered.

Table 8. Validation results for estimation model.

Margin of Error	RMSE	Difference
0%	500.0695	0.1213013
5%	384.4624	0.0991206
10%	369.0405	0.0951060

Figure 9 shows the residuals between the observed data and the values from the estimation model, along with the standard deviation indicated by dotted lines. The correlation coefficient (R^2) of the constructed linear estimation model was 0.8228. The MAPE was found to be 12%. Figure 10 shows the frequency distribution of the residuals. The distribution followed the form of a normal distribution, with a standard deviation of 453.71 and a mean of 211.89.

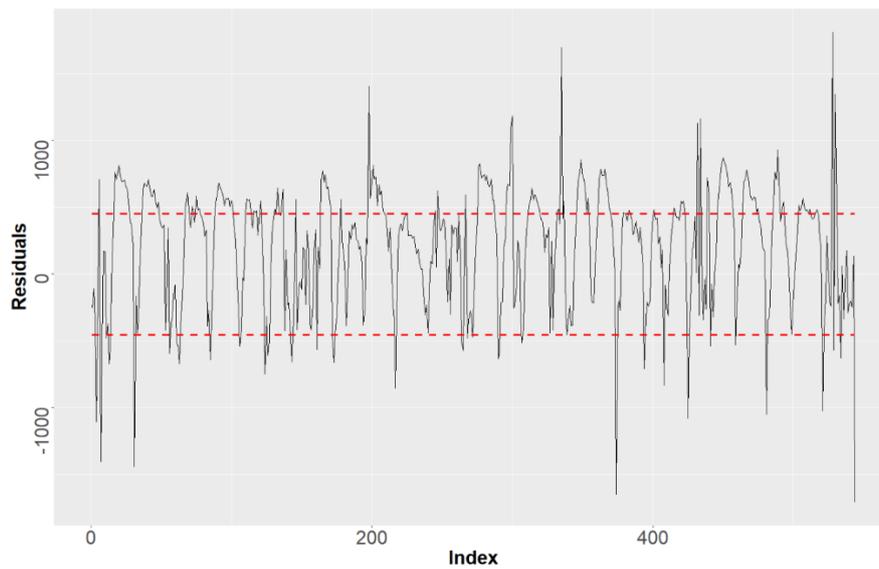


Figure 9. Residuals between observed and linear model estimation.

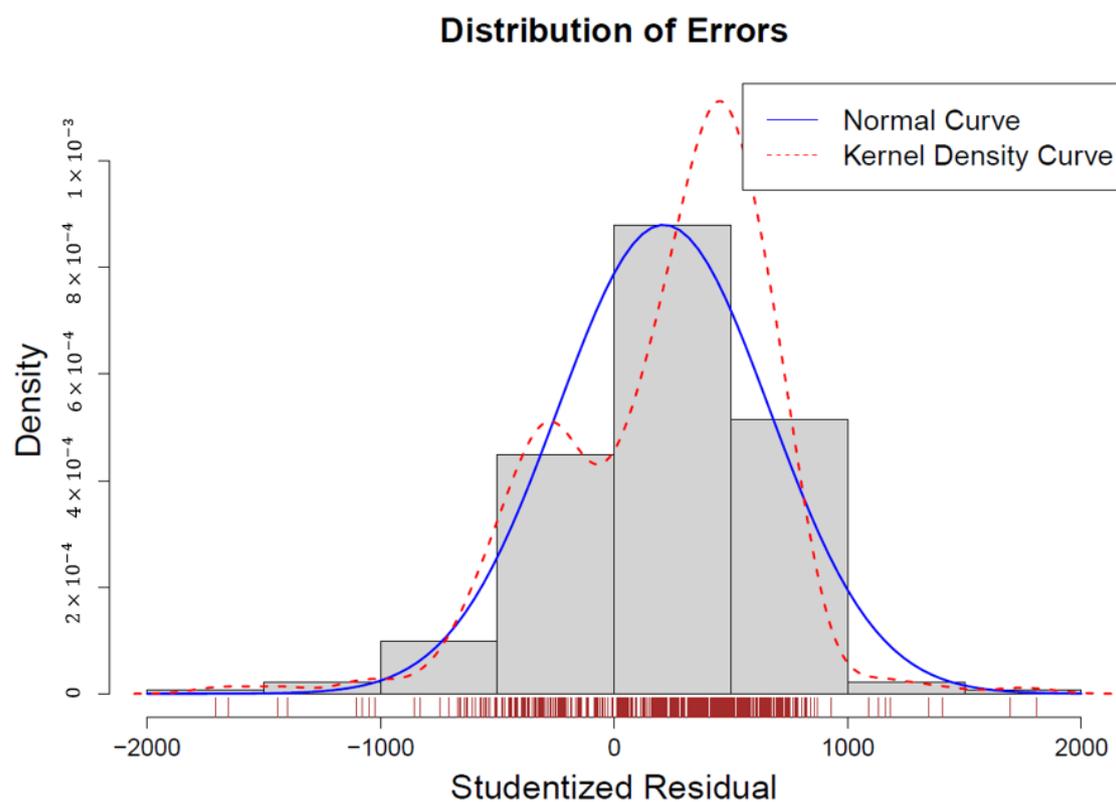


Figure 10. Histogram for residuals between observed values and linear model estimation values.

5.3. Discussion

We proposed a linear model to predict the power output of an inverter. Data from the PV system under commercial operation (Tables 2 and 3) were employed for estimation. We identified correlations by applying statistical methods to vertical solar radiation, horizontal solar radiation, and DC power data, and they are presented in Figures 6 and 7. There were fluctuations in positive directions. We proved strong correlations from a linear correlation of approximately 0.94 from the horizontal solar radiation data.

Note, however, that such strong correlations do not guarantee that solar radiation completely affects power output. With high solar radiation, power output varies per the temperature of the panels. This study focused on solar radiation.

Table 7 shows the quantity of the observed power output as well as the result of applying a linear model to the 544 solar radiation data pieces randomly extracted from the dataset. The average error from Lines 2 to 5 is approximately 14.3%. Nonetheless, Line 6, the last one, shows an error rate of about 45%. The observed output in Line 6 is identical to that in Line 2. The estimated output in Line 6 is higher than that in Line 2, showing that the solar radiation at the corresponding time is also higher. Still, it can be assumed from the identical outputs that the observed output was affected by external factors not considered in the experiment.

Figure 8 shows trends different from those of the figures in Table 7. There are differences between the observed values and estimates, but the trends of the latter show patterns similar to those of the observed values. Moreover, as shown by the qualitative evaluation of the results and numerical data from statistical verification, the proposed approach is meaningful since the estimated power output follows the trend of the observed output, even though there are differences in numerical data when compared with the latest algorithm.

This study utilized a simple primary linear model, whereas studies for comparison [34,35] used neural networks such as Recurrent Neural Network (RNN) and LSTM. It can be safely said that our study yielded qualitatively good results when comparing the computing power required for the calculation of each model as well as the complexity of the calculations.

6. Conclusions

This study proposes an inverter efficiency analysis method based on solar power estimation, using horizontal solar radiation data collected from an environment sensor. To determine inverter efficiency with high accuracy, in a PV system, solar power estimation based on the environment sensor data must be performed first. The proposed inverter efficiency analysis model is used to evaluate the inverter efficiency in real time during the operation of a PV system, and to determine the maintenance time. The constructed linear estimation model operates in real time based on the environment sensor data from the data collection device (real time unit; RTU) shown in Figure 1 and performs solar power estimation. The solar power estimation information accumulated in the database is continuously compared with the solar power information collected from the inverter, and the power generation efficiency values for the inverter are provided to the operator.

In this study, the solar power of a 10 kW inverter was analyzed; data cleansing was performed on the data obtained from the RTU, by calculating the interquartile range (IQR). In addition, correlation analysis was carried out for the preliminary validation of linear model application, and the validity of the correlation coefficient was confirmed. The results from the linear estimation model showed that the model performed solar power estimation with an 82% goodness of fit, and that the mean absolute percentage error (MAPE) of the model was 12%. The distribution of the residuals followed the form of a normal distribution, with a standard deviation of 453.71 and a mean of 211.8903, thereby establishing the validity of the model.

The linear estimation model was constructed in this study for a single PV system, and the nature and error rates of the collected data may vary depending on the inverter manufacturer. It will be possible, however, to apply the coefficients presented in this study to other PV systems. As the power generation capacity of a PV system varies depending on the installation scale and weather conditions, it is deemed necessary to reconstruct the model according to the power generation capacity. Future possibilities for research include the application of machine learning techniques to automatically perform model fitting according to the accumulated data in the system; it is also necessary to consider factors affecting power generation efficiency such as module characteristics and outside temperatures.

Author Contributions: Conceptualization, C.-Y.P., S.-H.H., S.-C.L., B.-S.S., S.-W.P., J.-H.H., and J.-C.K.; data curation, C.-Y.P., S.-H.H., S.-W.P., and J.-C.K.; formal analysis, C.-Y.P., S.-C.L., and J.-H.H.; funding acquisition, C.-Y.P., S.-H.H., and B.-S.S.; investigation, B.-S.S., J.-H.H., and J.-C.K.; methodology, C.-Y.P., S.-C.L., B.-S.S., S.-W.P., J.-H.H., and J.-C.K.; project administration, S.-H.H. and B.-S.S.; resources, C.-Y.P. and S.-W.P.; software, C.-Y.P., S.-H.H., S.-W.P., and J.-C.K.; supervision, S.-C.L., J.-H.H., and J.-C.K.; validation, S.-H.H., S.-C.L., and B.-S.S.; visualization, S.-H.H., J.-H.H., and J.-C.K.; writing—original draft, C.-Y.P., S.-W.P., J.-H.H., and J.-C.K.; writing—review and editing, J.-H.H. and J.-C.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Korea Institute of Energy Technology Evaluation and Planning (KETEP) grant funded by the Korean government (MOTIE) (20203040010130, Development and demonstration of remote intelligent operation and maintenance (O&M) technology of MW-class solar power plant using 5G technology).

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

References

- Salcedo-Sanz, S.; Pérez-Bellido, Á.M.; Ortiz-García, E.G.; Portilla-Figueras, A.; Prieto, L.; Correoso, F. Accurate short-term wind speed forecasting by exploiting diversity in input data using banks of artificial neural networks. *Neurocomputing* **2009**, *72*, 1336–1341. [\[CrossRef\]](#)
- Dong, Y.; Zhang, L.; Liu, Z.; Wang, J. Integrated forecasting method for wind energy management: A case study in China. *Processes* **2020**, *8*, 35. [\[CrossRef\]](#)
- Fu, Y.; Gao, Z.; Liu, Y.; Zhang, A.; Yin, X. Actuator and sensor fault classification for wind turbine systems based on fast Fourier transform and uncorrelated multi-linear principal component analysis techniques. *Processes* **2020**, *8*, 1066. [\[CrossRef\]](#)
- Liu, B.; Li, K.; Niu, D.D.; Liu, Y. The characteristic analysis of the solar energy photovoltaic power generation system. *IOP Conf. Ser. Mater. Sci. Eng.* **2017**, *164*, 1–6. [\[CrossRef\]](#)
- Manzoor, E.; Ghulam, A.; Irfan, K.; Paul, M.K.; Mashood, N.; Ali, R.; Umar, F. Recent approaches of forecasting and optimal economic dispatch to overcome intermittency of wind and photovoltaic (PV) systems: A review. *Energies* **2019**, *12*, 4392.
- Ahn, H.W.; Cho, Y.S. Study on certification policy for stabilization of photovoltaic industry—A case study of PV power plant O&M-. *J. Clim. Chang. Res.* **2015**, *6*, 105–111.
- Li, P.; Zhang, C.; Long, H. Solar power interval prediction via lower and upper bound estimation with a new model initialization approach. *Energies* **2019**, *12*, 4146. [\[CrossRef\]](#)
- Burger, B.; Kranzer, D.; Stalter, O. Cost reduction of PV-Inverters with SiC-DMOSFETs. In Proceedings of the 5th International Conference on Integrated Power Electronics Systems, Nuremberg, Germany, 11–13 March 2008; pp. 1–5.
- 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.* **2018**, *8*, 28. [\[CrossRef\]](#)
- Brecl, K.; Topič, M. Photovoltaics (PV) system energy forecast on the basis of the local weather forecast: Problems, uncertainties and solutions. *Energies* **2018**, *11*, 1143. [\[CrossRef\]](#)
- Cha, W.C.; Park, J.H.; Cho, U.R.; Kim, J.C. Design of generation efficiency fuzzy prediction model using solar power element data. *J. Trans. Korean Inst. Electr. Eng.* **2014**, *63*, 1423–1427.
- Yang, Z.; Mourshed, M.; Liu, K.; Xu, X.; Feng, S. A novel competitive swarm optimized RBF neural network model for short-term solar power generation forecasting. *Neurocomputing* **2020**, *397*, 415–421. [\[CrossRef\]](#)
- Wang, J.; Qian, Z.; Wang, J.; Pei, Y. Hour-ahead photovoltaic power forecasting using an analog plus neural network ensemble method. *Energies* **2020**, *13*, 3259. [\[CrossRef\]](#)
- Nespoli, A.; Ogliari, E.; Leva, S.; Massi Pavan, A.; Mellit, A.; Lughì, V.; Dolara, A. Day-ahead photovoltaic forecasting: A comparison of the most effective techniques. *Energies* **2019**, *12*, 1621. [\[CrossRef\]](#)
- Kwon, O.H.; Lee, K.S. Photovoltaic system energy performance analysis using meteorological monitoring data. *J. Korean Sol. Energy Soc.* **2018**, *38*, 11–31.
- Lee, S.Y.; Cha, B.H.; Kim, W.S.; Lee, Y.M. The solar power forecasting based on weather forecasting and statistics analysis. In Proceedings of the Fall Conference of the Korean Society for New and Renewable Energy, Jeju, Korea, 13–14 November 2019; p. 317.
- Zdyb, A.; Gulkowski, S. Performance assessment of four different photovoltaic technologies in Poland. *Energies* **2020**, *13*, 196. [\[CrossRef\]](#)

18. Das, U.K.; Tey, K.S.; Seyedmahmoudian, M.; Mekhilef, S.; Idris, M.Y.I.; Deventer, W.V.; Horan, B.; Stojcevski, A. Forecasting of photovoltaic power generation and model optimization: A review. *Renew. Sustain. Energy Rev.* **2018**, *81*, 912–928. [[CrossRef](#)]
19. Chung, M.H. Comparison of estimation methods by different photovoltaic software and performance evaluation. *J. Korea Inst. Ecol. Archit. Environ.* **2019**, *19*, 93–99.
20. Abraham, B.; Ledolter, J. *Statistical Methods for Forecasting*; Wiley: New York, NY, USA, 2009; Volume 234.
21. Potts, W.J. *Neural Network Modeling: Course Notes*; SAS Institute Inc.: Cary, NC, USA, 2000.
22. Hossain, C.A.; Chowdhury, N.; Longo, M.; Yaïci, W. System and cost analysis of stand-alone solar home system applied to a developing country. *Sustainability* **2019**, *11*, 1403. [[CrossRef](#)]
23. Roman, E.; Alonso, R.; Ibañez, P.; Elorduizapatarietxe, S.; Goitia, D. Intelligent PV module for grid-connected PV systems. *IEEE Trans. Ind. Electron.* **2006**, *53*, 1066–1073. [[CrossRef](#)]
24. Rashid, M.H. *Power Electronics Handbook*; Butterworth-Heinemann: Oxford, UK, 2017.
25. Cerón, I.; Caamaño-Martín, E.; Neila, F.J. State-of-the-art of building integrated photovoltaic products. *Renew. Energy* **2013**, *58*, 127–133. [[CrossRef](#)]
26. Jang, S.T.; Park, Y.M.; Sung, T.K.; Jung, C.B.; Kim, B.C.; Kim, M.S. Analysis of power conversion efficiency of inverter for photovoltaic power generation system. In Proceedings of the Winter Conference of the Korean Institute of Electrical Engineers, Daejeon, Korea, 28 November 2014; pp. 421–424.
27. Tian, A.-Q.; Chu, S.-C.; Pan, J.-S.; Liang, Y. A novel pigeon-inspired optimization based MPPT technique for PV systems. *Processes* **2020**, *8*, 356. [[CrossRef](#)]
28. Gohar Ali, H.; Vilanova Arbos, R.; Herrera, J.; Tobón, A.; Peláez-Restrepo, J. Non-linear sliding mode controller for photovoltaic panels with maximum power point tracking. *Processes* **2020**, *8*, 108. [[CrossRef](#)]
29. Yoon, Y. Integrated management system to improve photovoltaic operation efficiency. *J. Internet Broadcast. Commun.* **2019**, *19*, 113–118.
30. McCandless, T.; Dettling, S.; Haupt, S.E. Comparison of implicit vs. explicit regime identification in machine learning methods for solar irradiance prediction. *Energies* **2020**, *13*, 689. [[CrossRef](#)]
31. Moncada, A.; Richardson, W., Jr.; Vega-Avila, R. Deep learning to forecast solar irradiance using a Six-Month UTSA SkyImager dataset. *Energies* **2018**, *11*, 1988. [[CrossRef](#)]
32. Mpfumali, P.; Sigauke, C.; Bere, A.; Mulaudzi, S. Day ahead hourly global horizontal irradiance forecasting—Application to South African Data. *Energies* **2019**, *12*, 3569. [[CrossRef](#)]
33. Carrera, B.; Kim, K. Comparison analysis of machine learning techniques for photovoltaic prediction using weather sensor data. *Sensors* **2020**, *20*, 3129. [[CrossRef](#)]
34. Rodríguez, F.; Fleetwood, A.; Galarza, A.; Fontán, L. Predicting solar energy generation through artificial neural networks using weather forecasts for microgrid control. *Renew. Energy* **2018**, *126*, 855–864. [[CrossRef](#)]
35. Lotfi, M.; Javadi, M.; Osório, G.J.; Monteiro, C.; Catalão, J.P.S. A novel ensemble algorithm for solar power forecasting based on kernel density estimation. *Energies* **2020**, *13*, 216. [[CrossRef](#)]
36. Dolara, A.; Grimaccia, F.; Leva, S.; Mussetta, M.; Ogliari, E. A physical hybrid artificial neural network for short term forecasting of PV plant power output. *Energies* **2015**, *8*, 1138. [[CrossRef](#)]
37. Ahmad, M.W.; Mourshed, M.; Rezugui, Y. Tree-based ensemble methods for predicting PV power generation and their comparison with support vector regression. *Energy* **2018**, *164*, 465–474. [[CrossRef](#)]
38. Kim, S.G.; Jung, J.Y.; Sim, M.K. A two-step approach to solar power generation prediction based on weather data using machine learning. *Sustainability* **2019**, *11*, 1501. [[CrossRef](#)]
39. AlKandari, M.; Ahmad, I. Solar power generation forecasting using ensemble approach based on deep learning and statistical methods. *Appl. Comput. Inform.* **2020**. [[CrossRef](#)]
40. Suresh, V.; Janik, P.; Rezmer, J.; Leonowicz, Z. Forecasting solar PV output using convolutional neural networks with a sliding window algorithm. *Energies* **2020**, *13*, 723. [[CrossRef](#)]
41. Bacher, P.; Madsen, H.; Aalborg Nielsen, H. Online short-term solar power forecasting. *Sol. Energy* **2009**, *83*, 1772–1783. [[CrossRef](#)]
42. Detyniecki, M.; Marsala, C.; Krishnan, A.; Siegel, M. Weather-based solar energy prediction. In Proceedings of the 2012 IEEE International Conference, Fuzzy Systems, Brisbane, Australia, 10–15 June 2012; pp. 1–7.
43. Abdullah, N.A.; Abd Rahim, N.; Gan, C.K.; Nor Adzman, N. Forecasting solar power using Hybrid Firefly and Particle Swarm Optimization (HFPSO) for optimizing the parameters in a Wavelet Transform-Adaptive Neuro Fuzzy Inference System (WT-ANFIS). *Appl. Sci.* **2019**, *9*, 3214. [[CrossRef](#)]

44. Le, N.T.; Benjapolakul, W. Evaluation of contribution of PV array and inverter configurations to rooftop PV system energy yield using machine learning techniques. *Energies* **2019**, *12*, 3158. [[CrossRef](#)]
45. Abuella, M.; Chowdhury, B. Solar power probabilistic forecasting by using multiple linear regression analysis. In Proceedings of the IEEE Southeastcon, Ft. Lauderdale, FL, USA, 9–11 April 2015; pp. 1–5.
46. Ko, J.P.; Kim, E.J.; Byun, H.R. Normalization of face images subject to directional illumination using linear model. *J. Comput. Sci. Eng.* **2004**, *31*, 54–60.
47. Lim, D.H.; Kim, J.S.; Lee, B.G. Wave information estimation and revision using linear regression model. *J. Korea Multimed. Soc.* **2016**, *19*, 1377–1385. [[CrossRef](#)]
48. Kim, H.S.; Lim, W.T.; Lee, J.C.; Lee, K.W.; Park, K.T. Motion estimation method using multiple linear regression model. *Vis. Commun. Image Process.* **1997**, *3024*, 600–607.
49. Heo, Y.J.; Choi, M.K.; Lee, H.G.; Lee, S.C. Regional projection histogram matching and linear regression based video stabilization for a moving vehicle. *Korean Inst. Broadcast Media Eng.* **2014**, *19*, 798–809. [[CrossRef](#)]
50. Zhang, D.; Gao, Z. Improvement of refrigeration efficiency by combining reinforcement learning with a coarse model. *Processes* **2019**, *7*, 967. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).