

Article

Cloud Cover Forecast Based on Correlation Analysis on Satellite Images for Short-Term Photovoltaic Power Forecasting

Yongju Son ¹, Yeunggurl Yoon ¹, Jintae Cho ² and Sungyun Choi ^{1,*}

¹ School of Electrical Engineering, Korea University, Seoul 02841, Korea; 93yjson@korea.ac.kr (Y.S.); tigeryoon98@korea.ac.kr (Y.Y.)

² Korea Electric Power Research Institute, Daejeon 34056, Korea; jintae.cho@kepco.co.kr

* Correspondence: sungyun@korea.ac.kr

Abstract: Photovoltaic power generation must be predicted to counter the system instability caused by an increasing number of photovoltaic power-plant connections. In this study, a method for predicting the cloud volume and power generation using satellite images is proposed. Generally, solar irradiance and cloud cover have a high correlation. However, because the predicted solar irradiance is not provided by the Meteorological Administration or a weather site, cloud cover can be used instead of the predicted solar radiation. A lot of information, such as the direction and speed of movement of the cloud is contained in the satellite image. Therefore, the spatio-temporal correlation of the cloud is obtained from satellite images, and this correlation is presented pictorially. When the learning is complete, the current satellite image can be entered at the current time and the cloud value for the desired time can be obtained. In the case of the predictive model, the artificial neural network (ANN) model with the identical hyperparameters or setting values is used for data performance evaluation. Four cases of forecasting models are tested: cloud cover, visible image, infrared image, and a combination of the three variables. According to the result, the multivariable case showed the best performance for all test periods. Among single variable models, cloud cover presented a fair performance for short-term forecasting, and visible image presented a good performance for ultra-short-term forecasting.

Keywords: correlation analysis; satellite image; photovoltaic forecast; cloud cover



Citation: Son, Y.; Yoon, Y.; Cho, J.; Choi, S. Cloud Cover Forecast Based on Correlation Analysis on Satellite Images for Short-Term Photovoltaic Power Forecasting. *Sustainability* **2022**, *14*, 4427. <https://doi.org/10.3390/su14084427>

Academic Editors: Yi Liang, Dongxiao Niu, Wei-Chiang Hong and Mengjie Zhang

Received: 21 February 2022

Accepted: 6 April 2022

Published: 8 April 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 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 (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

1.1. Motivation and Aims

According to the Paris Climate Agreement, more than 200 countries, or the countries responsible for 87% of the world's carbon emissions, are implementing agreements aimed at reducing greenhouse gas emissions. To this end, we are attempting to reduce the use of fossil fuels, which are the biggest cause of greenhouse gas generation, and naturally reducing the number of power plants using fossil fuels has become an important task. Accordingly, the proportion of power plants using renewable energy is increasing rapidly. In particular, the solar facility capacity installed in 2020 was 707.5 GW, which was 21.5% higher than the previous year [1].

However, as the proportion of renewable energy generation increases, volatility and intermittency problems play a role in deteriorating the stability of the power system [2]. The power generation of a power plant is generally adjusted by the predicted load because the system needs to maintain a balance between power generation and load. However, because the output of variable renewable energy sources cannot be controlled, the larger the proportion of renewable energy, the more likely it is to break the balance between the power generation and load [3]. To flexibly respond to these inequalities and variability, it is necessary to predict the amount of renewable energy generation. This is because, by predicting the generation amount of each volatile resource in advance, the system operator

can plan a dispatching schedule and reserve power of other power sources as per the load demand [4,5]. This study focused on analyzing and predicting the characteristics of solar power generation, whose proportion of power generation has been increasing among renewable energy resources.

1.2. Literature Survey

Various data and methodologies have been studied to predict photovoltaic (PV) power. The methods for improving the performance of the prediction can be divided into two categories; the first category applies a prediction model that can learn the characteristics of data. In early days, linear regression (LR) or statistical models, such as auto-regressive moving average (ARMA), were frequently used, and in recent years, machine learning or deep learning-based forecast models have been significantly presented [6]. For example, some studies have proven high-performance results using artificial intelligence methodologies, such as artificial neural network (ANN) [7–10], convolutional neural network (CNN) [11], and support vector machine (SVM) [12–15]. The second category makes up high-quality data. No matter how excellent the performance of the forecast model, it is difficult to perform accurate predictions if there are a lot of errors and outliers in the given data. For example, according to [16], using high-correlated data can improve the accuracy. Many studies related to PV forecast use various meteorological data as input variables [17–21]. Among these variables, solar irradiance is used as the principal variable because the correlation between solar power generation and solar irradiance is high [22,23].

According to the many studies, the correlation between solar irradiance and PV power generation can be proved to be high. In these models, the solar irradiance must be input under actual prediction. However, the solar irradiance forecast is not provided by the Meteorological Administration or the weather forecast sites. There are two methods to solve this problem. The first is a method of predicting the amount of insolation using a time-series model. However, this method has the disadvantage that only ultra-short-term solar irradiance trends can be predicted, and fluctuations due to external factors, such as clouds, cannot be predicted. The second is a method of analyzing cloud cover. Escrig et al. [24] proposed three tasks viz. cloud detection, cloud classification, and cloud motion determination to forecast the cloud cover. Infrared and visible satellite images were used to check the cloud height and opacity. Dazhi et al. [25] proposed three models to forecast global horizontal irradiance, with the model using ground-based cloud cover data to forecast the solar irradiance being the most significantly useful.

1.3. Contributions and Organization of the Paper

This study proposes the use of correlation analysis of satellite images to predict solar power generation. Visible images and infrared images are selected to predict the solar power generation through image processing operations and an ANN model. The following are the contributions of this paper:

- Correlation analysis between satellite images with respect to time and space.
- Extraction of the cloud value of the target area in the correlation-based satellite image.
- Presentation of methodology for data performance comparison.

Section 2 describes the data characteristics and the data processing, and Section 3 describes the process for predicting solar power and the technique used in the prediction model. Section 4 presents the simulation results, including the evaluation criteria, simulation method, results, and discussion, and Section 5 presents the conclusion.

2. Characteristics of Dataset

2.1. Meteorological Data

Weather data can be obtained from the Meteorological Administration. These data were collected using the Automated Surface Observing System at meteorological observation stations scattered throughout Korea. The system provides data such as air pressure, dryness, and wind direction, as well as major weather variables such as temperature,

humidity, and wind speed. However, only a few observation stations provide all variables. Therefore, the three stations located in Heuksando, Mokpo, and Yeosu were selected to collect cloud cover and irradiance data. The location of each observation station is shown in Figure 1. Based on the amount of solar irradiance and cloud cover measured at each observation station, the correlation between the PV power generation in Jeollanam-do and the meteorological data will be analyzed in Section 4.

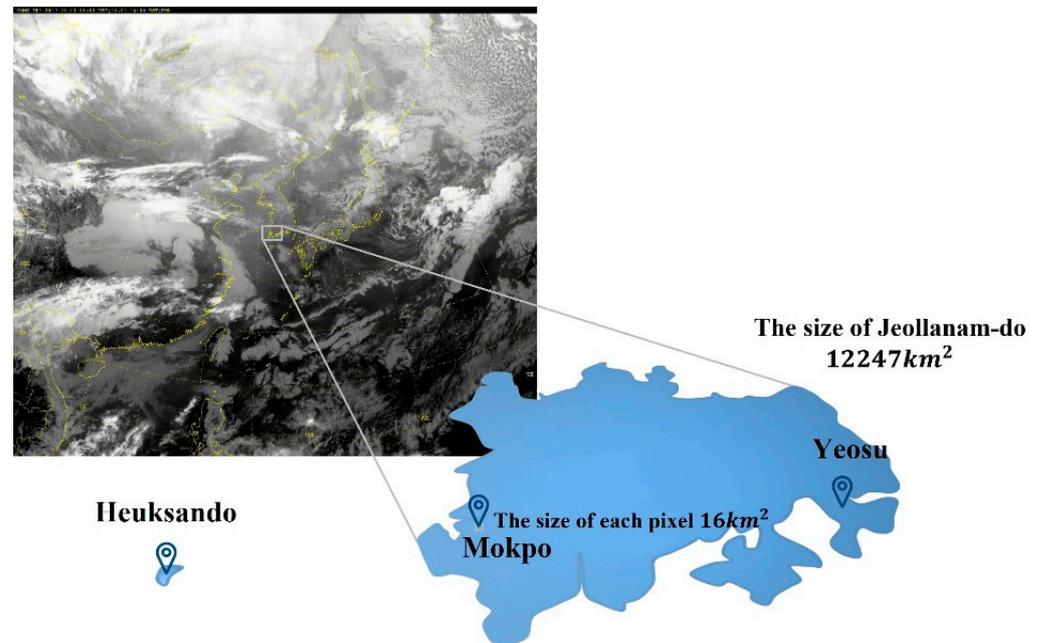


Figure 1. Location of weather station in Jeollanam-do and the size of each pixel.

The data period used in this study was from 1 January 2017 to 31 December 2019, and the period of weather data was in units of 1 h. Because observational data were obtained using a measuring device, missing data and outlier data were also included. Therefore, the error was preprocessed before the correlation analysis. Missing data were interpolated using the average of the previous and following time zones. If the outliers were not in the range of 1% to 99% of the variance of the total data distribution, the data point was considered as an outlier and was replaced with values of 1% and 99%.

2.2. Satellite Image

Satellite images for three years from 1 January 2017 to 31 December 2019 were used. The satellite image was taken by the geostationary orbit satellite “Cheonlian 1” of Korea and released to the National Meteorological Satellite Center. Images measured at wavelengths in various areas were provided, and in this study, infrared images in the $10.8 \mu\text{m}$ wavelength and visible images in $0.87 \mu\text{m}$ wavelength were collected to track the movement of clouds. The visible image can reflect the intensity of sunlight reflected by the clouds and the ground, and the thicker the cloud, the stronger the reflective intensity, making it brighter in the image. Visible images are not available because there is no solar light at night. Infrared images represent images through large and small amounts of infrared energy emitted by an object; therefore, the observation is possible for 24 h. The amount of infrared energy depends on the temperature of the object. The higher the temperature, the lower the cloud, and the darker it appears.

A visible image is useful to detect daytime cloud images, yellow dust, forest fires, fog observation, and atmospheric motion vectors, while an infrared image is useful for detecting cloud information, sea-level temperature, and yellow dust observation. Each image collected in this study is gray-scaled and has a pixel value between 0 and 255. Similarly, the images are 1500 pixels wide and 1300 pixels high, and each pixel represents a

4 km × 4 km area, thus representing an area of 16 km². An example of each image is shown in Figure 2, and the images are displayed for the same time, that is, 2 PM, when the light is strong. Although the time instant is the same for all the images, it can be seen that the observed cloud shapes in the visible image and infrared image are different. The accuracy of the data obtained through each image will be presented in Section 4.

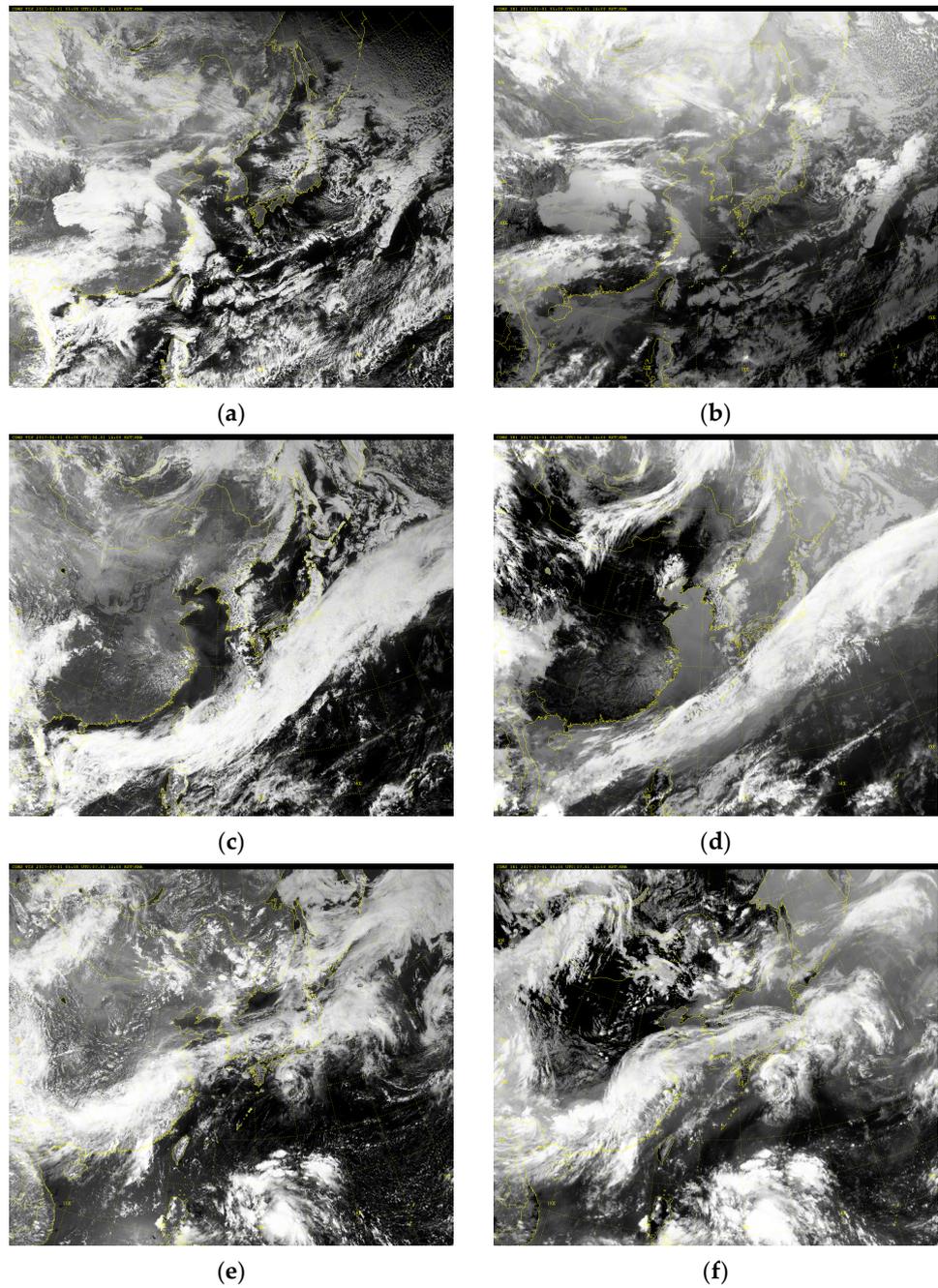


Figure 2. Cont.

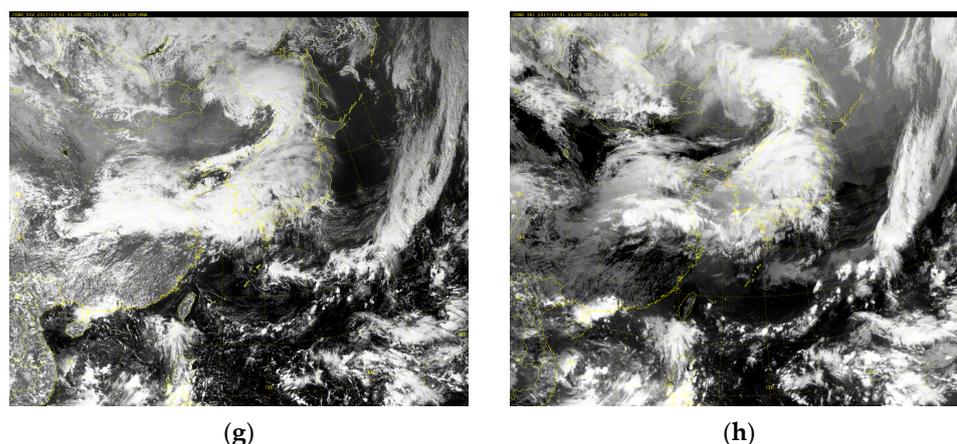


Figure 2. Examples of visible images and infrared images captured by “Cheonlian 1”. (a) Visible image at 2 PM 1 January 2017. (b) Infrared image at 2 PM 1 January 2017. (c) Visible image at 2 PM 1 April 2017. (d) Infrared image at 2 PM 1 April 2017. (e) Visible image at 2 PM 1 July 2017. (f) Infrared image at 2 PM 1 July 2017. (g) Visible image at 2 PM 1 October 2017. (h) Infrared image at 2 PM 1 October 2017.

Table 1 shows the weather and satellite data during daylight hours corresponding to Figure 2. The weather data represent the data measured at the Mokpo Weather Station, and the satellite data represent the pixel value corresponding to the Mokpo area on the pixel coordinates. The maximum value of the cloud cover is 10, and the maximum pixel value of the satellite image is 255. The larger the value of cloud cover, visible image, and infrared image, the more is the cloud volume. Note that some measurements of visible images on 1 January 2017 are missing because the data were not available owing to short daylight hours in winter. This can be confirmed from Figure 3.

Table 1. Meteorological data and satellite image data of Mokpo area for one day.

Date	Time	Solar Irradiance (Mokpo)	Cloud Cover (Mokpo)	Visible Image	Infrared Image
1 January 2017	7 AM	0.00	10	-	111
	8 AM	0.00	8	-	119
	9 AM	0.11	7	123	119
	10 AM	0.47	6	165	123
	11 AM	0.78	6	114	108
	12 PM	1.34	2	103	106
	1 PM	1.53	0	84	104
	2 PM	1.46	0	94	104
	3 PM	1.23	0	88	105
	4 PM	0.82	0	83	109
	5 PM	0.38	0	17	111
6 PM	0.00	0	-	111	
1 April 2017	7 AM	0.01	9	130	171
	8 AM	0.07	8	192	128
	9 AM	0.28	7	184	121
	10 AM	0.66	6	182	119
	11 AM	1.34	6	144	101
	12 PM	2.05	5	113	114
	1 PM	2.55	4	93	161
	2 PM	1.67	5	83	132
	3 PM	2.15	3	77	229
	4 PM	1.87	0	82	246
	5 PM	1.28	0	82	238
6 PM	0.64	0	51	190	

Table 1. Cont.

Date	Time	Solar Irradiance (Mokpo)	Cloud Cover (Mokpo)	Visible Image	Infrared Image
1 July 2017	7 AM	0.09	10	191	195
	8 AM	0.33	10	220	180
	9 AM	0.46	10	230	195
	10 AM	0.57	10	235	196
	11 AM	0.82	10	226	201
	12 PM	1.05	10	241	129
	1 PM	1.29	10	207	166
	2 PM	0.93	10	191	207
	3 PM	1.22	10	205	169
	4 PM	0.90	10	193	176
	5 PM	0.68	10	173	150
	6 PM	0.49	10	137	98
1 October 2017	7 AM	0.00	10	186	226
	8 AM	0.04	10	227	251
	9 AM	0.07	10	241	250
	10 AM	0.14	10	241	235
	11 AM	0.21	10	247	245
	12 PM	0.25	10	243	225
	1 PM	0.52	10	242	171
	2 PM	0.53	10	239	205
	3 PM	0.34	10	244	162
	4 PM	0.23	10	208	183
	5 PM	0.11	10	153	140
	6 PM	0.02	10	41	224

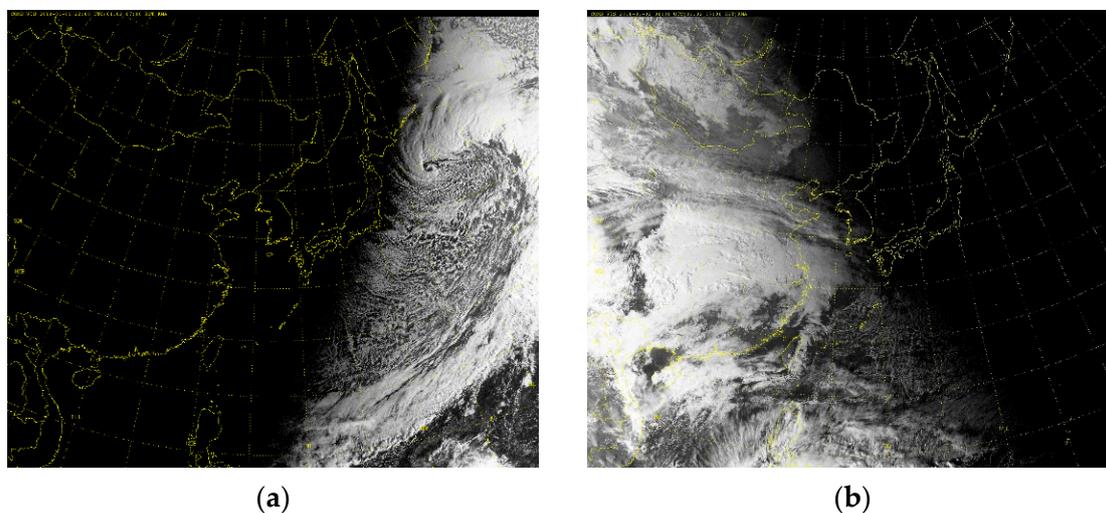


Figure 3. Comparison of visible image at 7 AM and 5 PM. (a) Visible image at 7 AM 2 January 2018. (b) Visible image at 5 PM 2 January 2018.

2.3. Photovoltaic Data

Historical power generation data are essential for analyzing the characteristics of PV power generation. In this study, the PV power generation data were acquired from power plants in Jeollanam-do, Korea. These data are publicly available. The period of the collected data was from 1 January 2017 to 31 December 2019, and the interval of data was the same as that of the weather data. The maximum PV power generation in the region was approximately 533 MW.

Figure 4 displays the distribution of data and shows the visible light-based pixel value, infrared-based pixel value, measurement cloud of the weather station, and power

generation of power plants present in Jeollanam-do. Figure 4a,b shows that the distribution of values obtained from the satellite images is diverse and detailed, and Figure 4c shows that the range of data is small, and the data are concentrated at both ends. Figure 4d shows an approximately uniform data distribution from 0 to 533 MW. Just because the data obtained from the satellite images are finely organized, it cannot be inferred that the variables and amount of power generated are always highly correlated. However, because the number of various causes can be analyzed compared to the cloud data, a more precise power generation prediction can be performed.

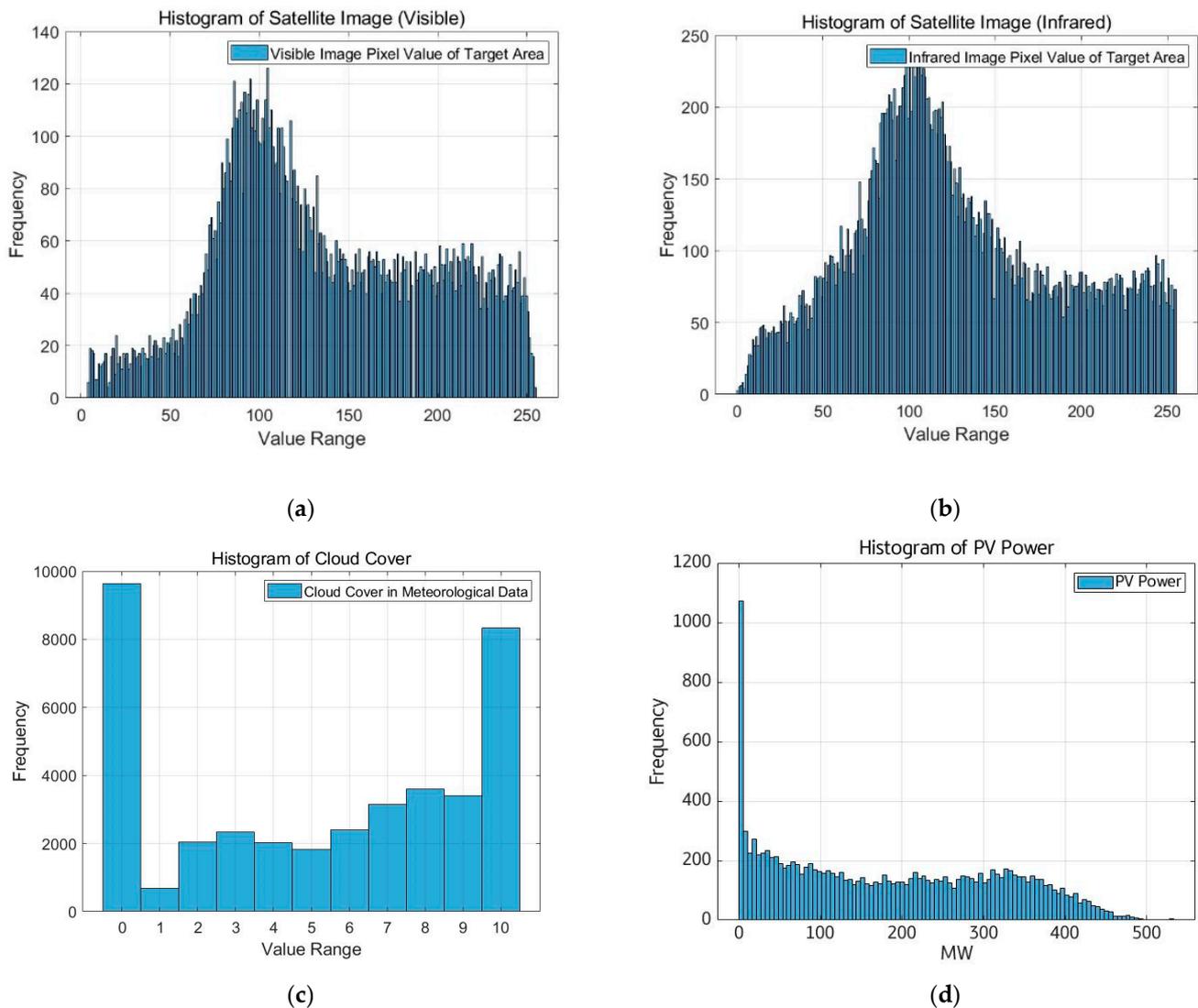


Figure 4. Histogram of the pixel value of the target region. (a) Histogram of visible image pixel value presenting target area (Mokpo); (b) histogram of infrared image pixel value presenting target area (Mokpo); (c) histogram of cloud cover measured at Mokpo Weather Station; (d) histogram of PV power at Jeollanam-do.

3. Methodology

3.1. Image Processing

To extract data from satellite images and perform correlation analysis using other data, it is necessary to convert the satellite images into numerical data and synchronize the time. This process is depicted in Figure 5. First, images were sequentially retrieved from a database that stored satellite images. The file name included 12 digits, such as “201701010100”, and consisted of 4 digits per year, 2 digits per month, 2 digits per day,

2 digits per hour, and 2 digits per minute. A procedure for checking whether the photographing date and time of the image were correct based on the corresponding file name was performed. This reduces errors when calculating the correlation between data, such as weather data and power generation. If no orthogonal data existed, an image taken 15 min later was used instead. This is because the image at the closest time is the image after 15 min because the interval of the image is 15 min, and the cloud does not change rapidly within 15 min. If there was no image after 15 min, the image after 30 min or 45 min was applied as an alternative, and if there was no such image, the data change for that time was reserved.

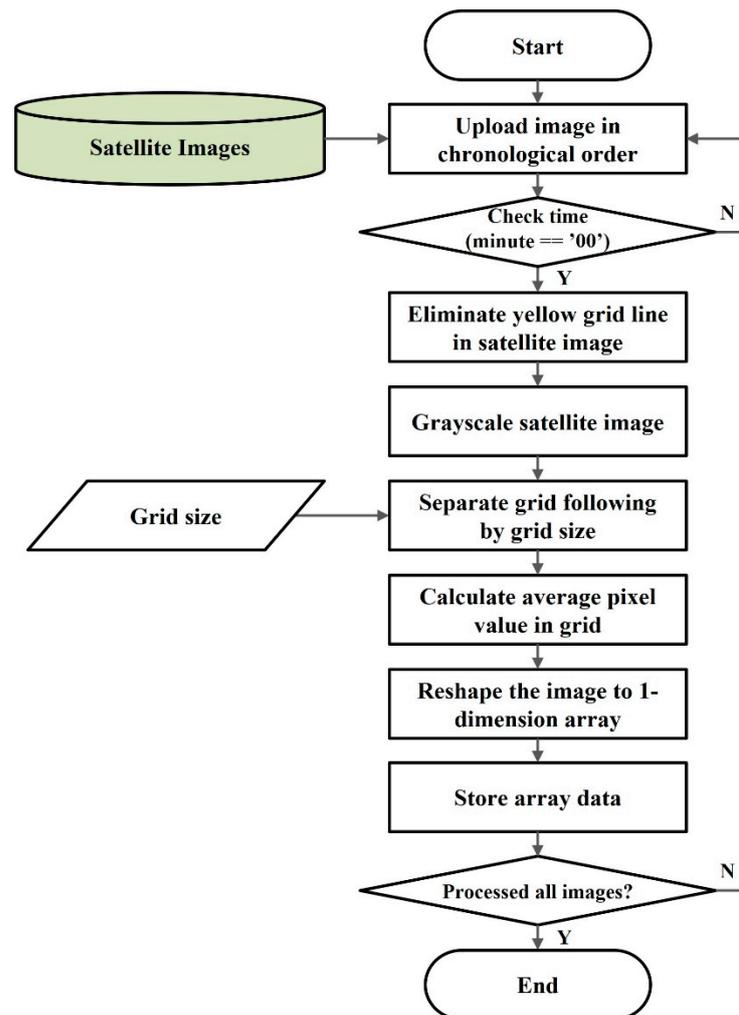


Figure 5. Process of image processing.

Table 2 shows the average change between images after 15 min, 1 h, 2 h, and 6 h at the target time based on the target area.

Table 2. Average change and average rate of change in the same pixel for different time intervals.

	15 min	1 h	2 h	6 h
Average Change	5.65	9.22	12.23	19.04
Average Rate of Change	2.2%	3.6%	4.8%	7.5%

The next step involved removing the guidelines existing in the satellite image, and the same value was input for all images to avoid affecting the calculation. As shown in Figure 6, the original image is drawn with a yellow guideline, and in this study, it was

eliminated by using 0, as shown in Figure 7. Thereafter, for convenience of calculation, the image is gray-scaled, as shown in Figure 8. In the case of infrared or visible images, even though the RGB values were the same, they were duplicated and recorded, so they were unnecessarily 3D images. Therefore, because this was likely to cause additional operations, the one dimension must be converted into a two-dimensional image by unifying it into a single matrix.

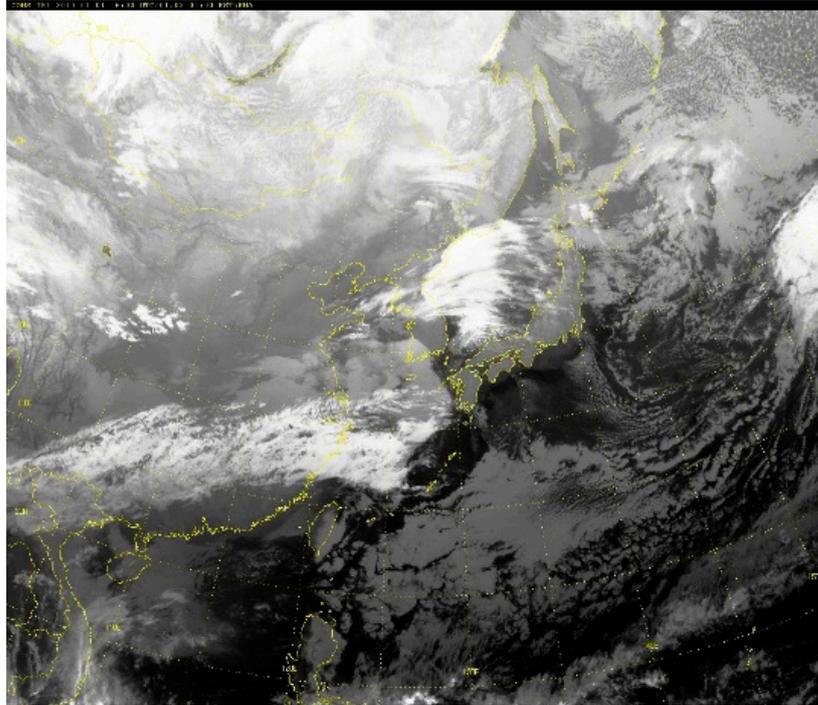


Figure 6. Original image from the database.

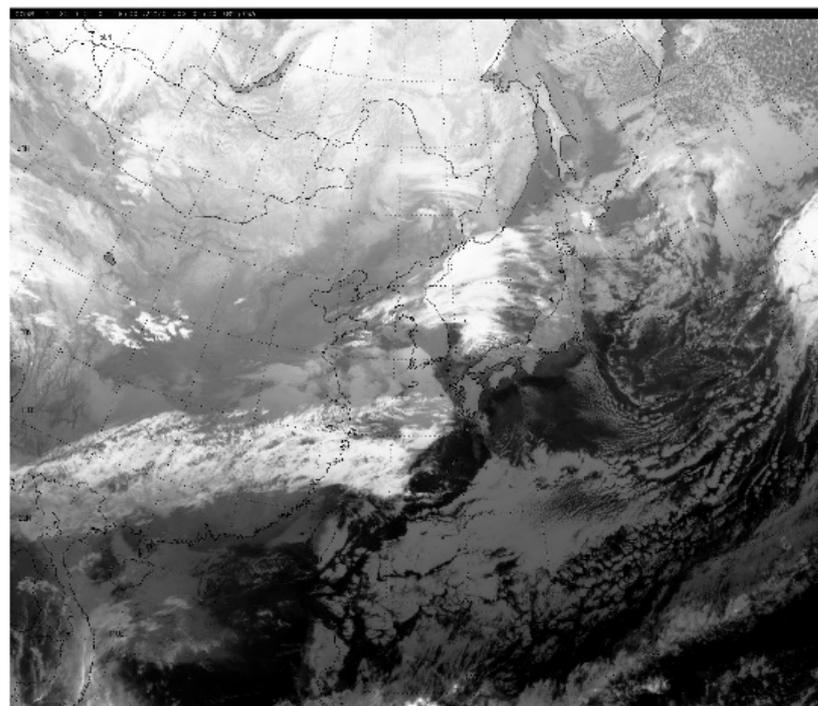


Figure 7. Eliminated guideline of image.

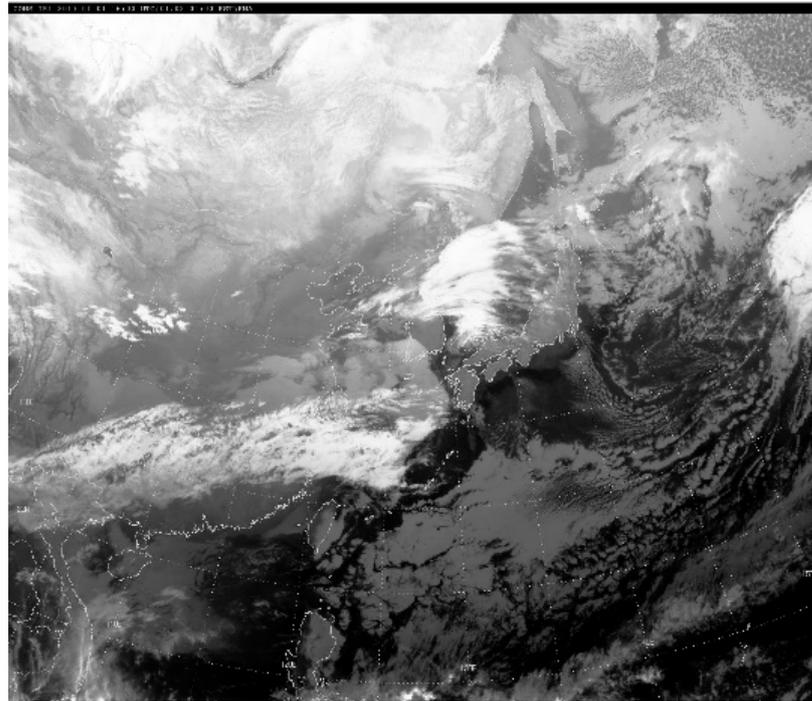


Figure 8. Gray-scaled image.

When image color conversion was completed, the correlation must be divided to track the clouds. However, analyzing all correlations for images of $1500 \text{ px} \times 1300 \text{ px}$ was difficult because of resource limitations. Therefore, after separating the image into several grids, as shown in Figure 9, we can determine whether a cloud exists using the representative value of the grid. In this study, satellite images were divided into grids in units of $10 \text{ px} \times 10 \text{ px}$ and reduced to $150 \text{ px} \times 130 \text{ px}$ images. During the reduction process, the representative value of each grid was determined using an average value of 100 pixels. The equation expressing the average value of the grid is presented as follows, and the results are shown in Figure 10.

$$\text{Average value of Grid}_n = \frac{1}{x_{\text{grid size}} \times y_{\text{grid size}}} \sum_{i=1}^{x_{\text{grid size}}} \sum_{j=1}^{y_{\text{grid size}}} \text{Grid}_{nij} \quad (1)$$

When size conversion was completed, the two-dimensional image was converted into a one-dimensional arrangement and stored in a database. This will facilitate calculations in future correlation analyses.

Table 3 shows the average value of how much the value changes after 15 min, 1 h, 2 h, and 6 h for the target area after image processing. Compared to Table 1, the change decreased by approximately 1% for 15 min and increased by 2% for 1 h intervals.

Table 3. Average change and average rate of change in the same pixel for different time intervals after image processing.

	15 min	1 h	2 h	6 h
Average Change	2.86	14.34	21.59	38.08
Average Rate of Change	1.1%	5.6%	8.5%	14.9%

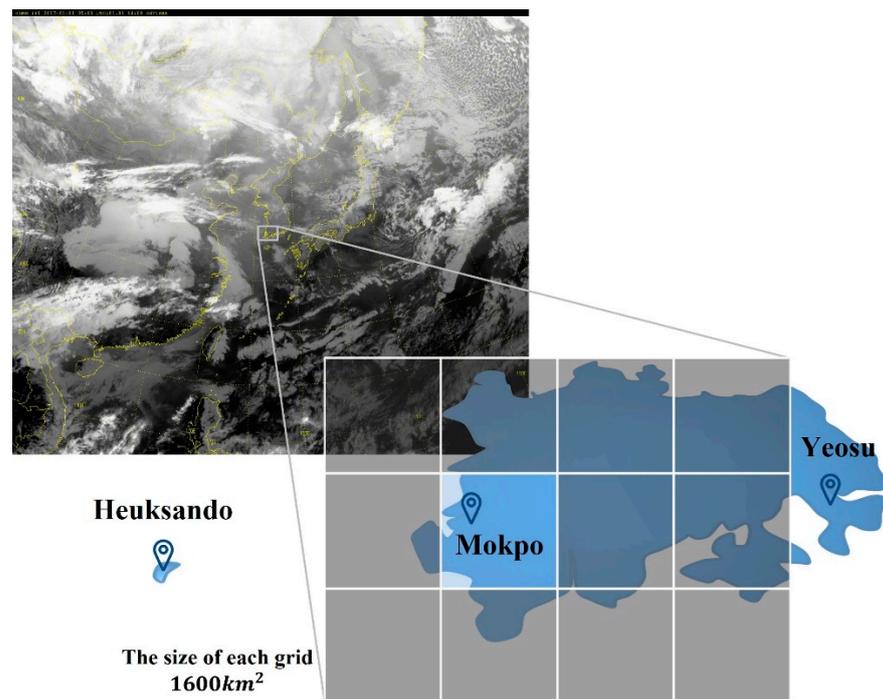


Figure 9. Location of weather station in Jeollanam-do and the size of each grid.

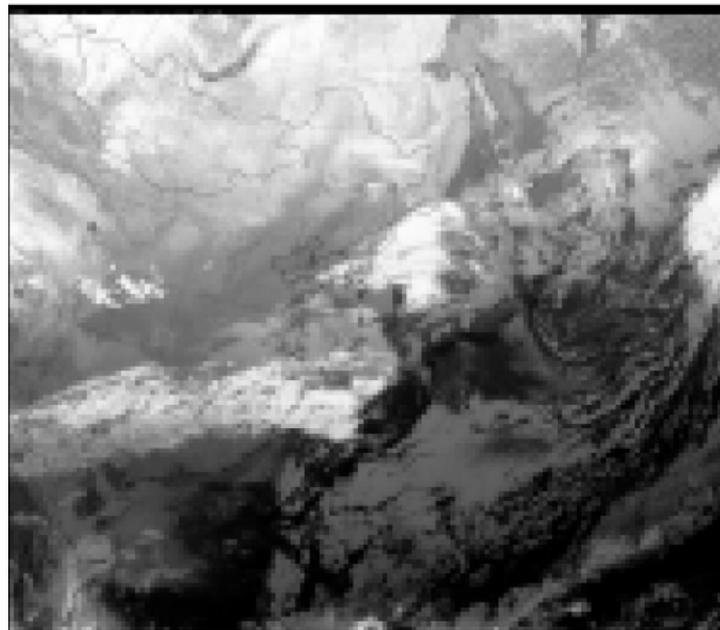


Figure 10. Image converted to grids.

3.2. Correlation Analysis

The correlation analysis was the core process of this study, and this process is shown in Figure 11. The algorithm starts with inputting the value in the target time, target area, and T. The target time denotes reference time and selects 24 h period composed of 1-h intervals. The target area denotes a reference area or pixel. The coordinates of the pixel were found by matching the targeted area in the preprocessed satellite image. The T can be 23 h maximum before the reference time, and correlation analysis was performed for 1-h interval time as necessary.

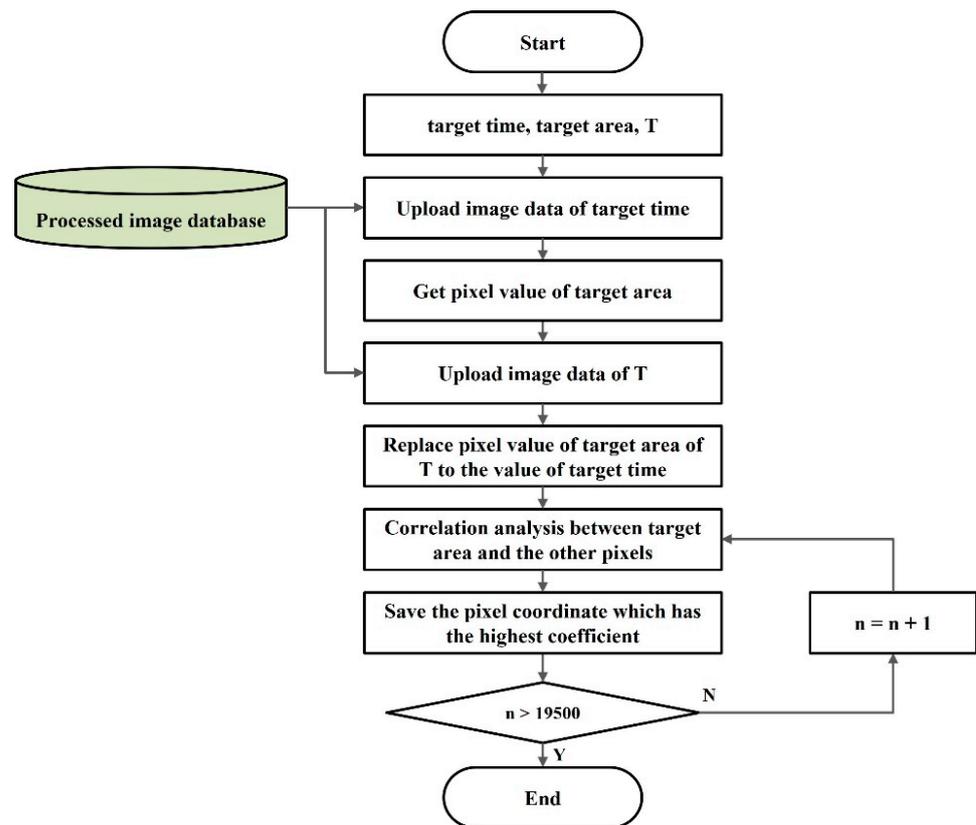


Figure 11. Process of correlation analysis.

In the paper, the Mokpo area of Jeollanam-do was selected as the reference area, and the target area would be the pixel (65, 75) matched with the targeted area in preprocessed images. To analyze how the cloud varied between 11 AM and 2 PM, we inputted 2 PM in the target time and 11 AM in T.

In the next step, uploading the image corresponding to the target time will enable the data of the pixel indicated by the target area to be extracted. Then, uploading the new image corresponding to the time T, the value of the target area of time T was replaced by the extracted value and stored in the workspace. If the same steps were applied to the same time interval for three-year satellite images, data would be collected for approximately 1000 images. The data have N scalar observations, and then the Pearson correlation coefficient is defined as:

$$\rho(\text{target area}, \text{grid}_n) = \frac{1}{N-1} \sum_{i=1}^N \left(\frac{(\text{target area})_i - \mu(\text{target area})}{\sigma(\text{target area})} \right) \left(\frac{(\text{grid}_n)_i - \mu(\text{grid}_n)}{\sigma(\text{grid}_n)} \right) \quad (2)$$

where $(\text{target area})_i$ and grid_n denote the cloud cover value of a designated area and other grids' cloud cover values, respectively. n ranges from one to the total number of pixels, which totals 19,500 pixels in the image. N is the number of data for correlation analysis. $\mu(\text{target area})$ and $\mu(\text{grid}_n)$ are the averages of the target and nth grids, respectively. $\sigma(\text{target area})$ and $\sigma(\text{grid}_n)$ are the covariances of the target and nth grids, respectively.

Assuming 2 PM as the target time, comparing the relationship with 11 AM, which was 3 h ago, the results of the correlation analysis are shown in Figure 12. In the image, there are 150 x -axis pixels and 130 y -axis pixels, which are of the same size as the existing adjusted satellite image. In other words, each point represents a correlation between the region and the target region. The black dot in the middle of Figure 12 represents the target area, the yellow color in the surrounding pixels denotes that the cloud correlation between

the target area and the area is high, and the blue color denotes that the cloud correlation between the target area and the area is low.

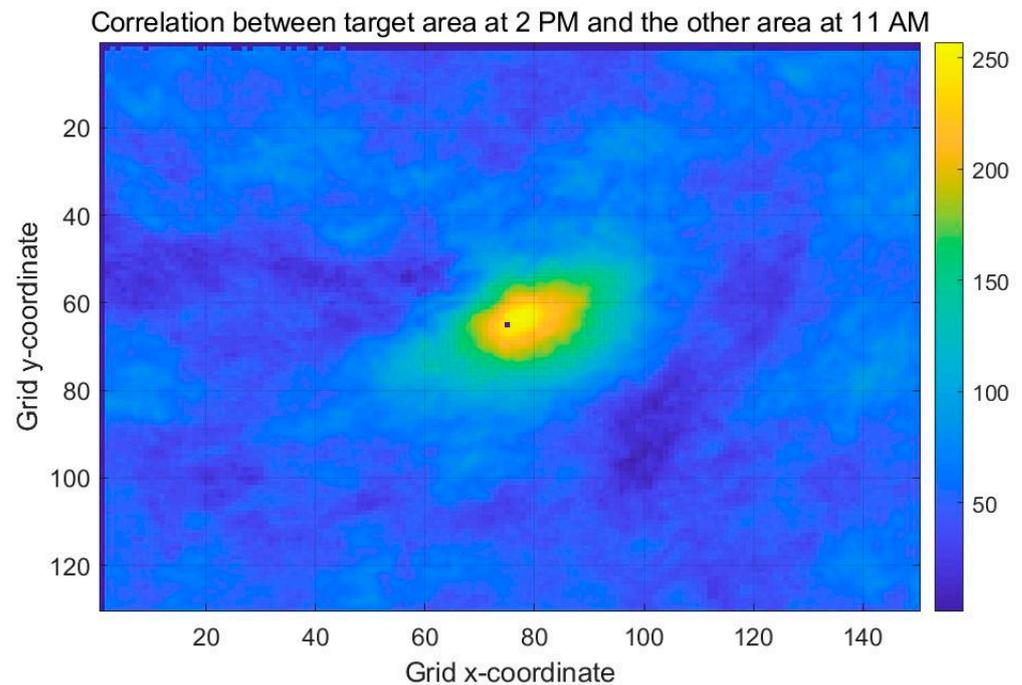


Figure 12. Correlation analysis between satellite images captured at 11 AM and 2 PM.

When the correlation analysis was finished, the coordinates of the pixel having the highest correlation between 19,499 pixels and the target area were stored in the database. Since correlation analysis was performed for up to 23 h, a correlation coefficient matrix of size 24×24 was generated when all correlation analysis was performed for 24 h.

These were classified and stored with respect to reference time and target time zone, and each stored value has a role to inform pixel coordinates for the designated time point when entering a new satellite image. For example, when the current time is 11 AM, and the cloud volume of the target area is to be predicted at 2 PM, the value of the satellite image is extracted using the pixel coordinate value corresponding to the interval of three hours from 11 AM. This was used under the assumption that if there were large amounts of clouds in the correlated area at that time, there were also many clouds in the target area a few hours later.

3.3. Prediction Process

When performing the actual prediction process, the time at which the correlation analysis was performed was reversed. If the prediction time was 9 AM, the cloud prediction was performed for 24 h using a correlation matrix with an image taken 1 h before 10 AM, 11 AM with an image taken 2 h ago, and noon with an image taken 3 h ago. Therefore, when a satellite image for the current time point was input, the pixel coordinates of the highly correlated regions were retrieved for each period, and the pixel values in the current image were extracted. Based on this, it was possible to predict the target region's cloud cover value for 24 h and to determine the solar power generation prediction value for the target region using this as an input value of the power generation prediction model. The process is shown in Figure 13.

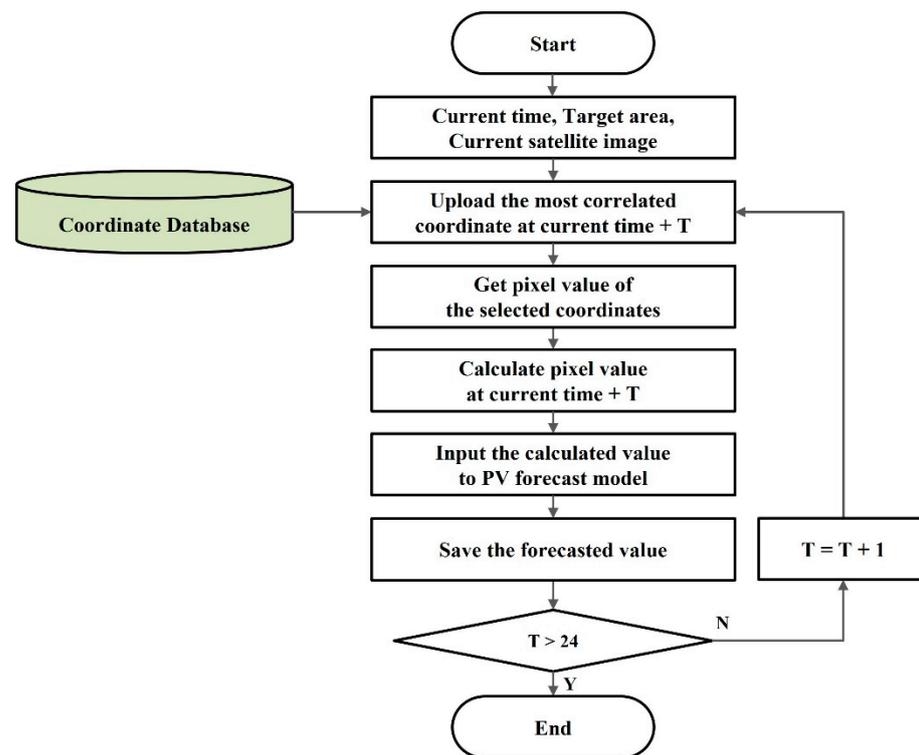


Figure 13. Process of prediction.

3.4. Forecasting Model with ANN

For evaluating the impact and performance of the data, the prediction model was fixed with ANN, and only the inputs were set differently. The defined ANN model in the paper is a fully connected structure and consists of one input layer, two hidden layers, and one output layer. At the input layer, a single variable input uses one neuron, and a multivariable input uses three neurons. Each of the hidden layers consists of 10 neurons. The mathematical expression of the first hidden layer is shown in Equation (3). $w_{ji}^{(1)}$ denotes the weight connected from the i th of the input layer to the j th neuron of the first hidden layer. $b_j^{(1)}$ denotes the bias of the j th neuron, and x_i means the i th neuron of the input layer. m means the number of neurons in the input layer. Since the number of neurons in the first hidden layer is 10, 1 to 10 are input in j . a represents the sum of weight and bias. $z_j^{(1)}$ represents transferred signals through the activation function shown in Equation (4). The calculation method of the second hidden layer is the same as the calculation method of the first hidden layer, which is shown in Equations (6) and (7). The activation function uses the same function, and the rectified linear unit (ReLU) function is used.

$$a_j^{(1)} = b_j^{(1)} + \sum_i^m w_{ji}^{(1)} x_i \quad (m = 1 \text{ or } 3, j = 1, \dots, 10) \quad (3)$$

$$h(a) = \begin{cases} a & (a > 0) \\ 0 & (a \leq 0) \end{cases} \quad (4)$$

$$z_j^{(1)} = h(a_j^{(1)}) \quad (5)$$

$$a_k^{(2)} = b_k^{(2)} + \sum_j^{10} w_{kj}^{(2)} z_j^{(1)} \quad (k = 1, \dots, 10) \quad (6)$$

$$z_k^{(2)} = h(a_k^{(2)}) \quad (7)$$

The output layer is composed of one neuron, and this model uses an identity function because it is a problem of forecasting continuous values from input data. The sum of the weights of the output layers is represented by Equation (8), and y , the predicted value output, is represented by Equation (9).

$$a_l^{(3)} = b_l^{(3)} + \sum_k w_{lk}^{(3)} z_k^{(2)} \quad (l = 1) \quad (8)$$

$$y = a^{(3)} \quad (9)$$

The ANN structure representing the above equation is shown in Figure 14. ANN works as a regression model for learning nonlinear characteristics. In the input layer, x presents a cloud cover, a pixel value of an infrared image, and a pixel value of a visible image. The signal from the input layer is varied by the weight, bias, and activation function while passing through each neuron of the hidden layers. In the output layer, each signal is calculated without an activation function. The output signal y is the amount of PV power predicted by the input variable. Since there is an error between predicted PV power and actual PV power in the early training sequence, the forecasting model reduces the error by updating the weight and the bias through backpropagation and repetitive tasks.

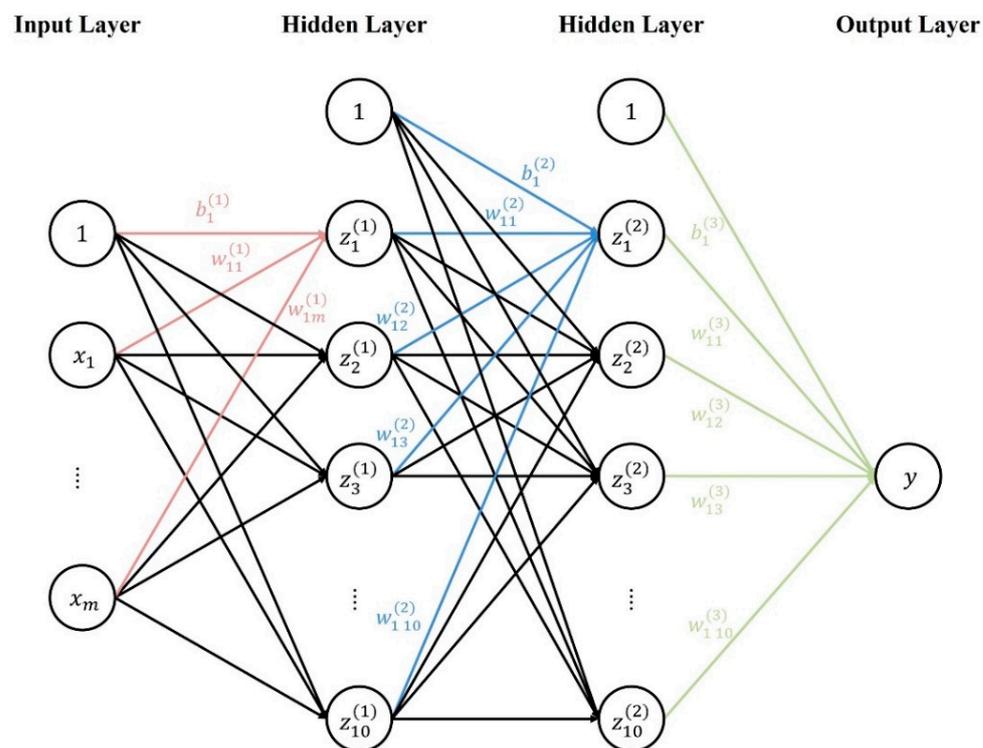


Figure 14. Structure of the proposed ANN model.

The proposed ANN model applies identically to the cloud cover-based forecasting model, visible image-based forecasting model, and infrared image-based forecasting model. The number of neurons in the input layer of the single input model was one, and the number of neurons in the input layer of the multivariable model was three.

The period of the data was from 1 January 2019 to 31 December 2019. The training period was 1 January to 30 September, and the test period was 1 October to 31 December. The data consist of the set of input values and actual PV power.

4. Simulation Result

4.1. Performance Evaluation Metric and Equipment

This section evaluates the accuracy of the PV power generation forecasting model based on satellite images. As an assessment metric to measure accuracy, the mean square error was adapted as follows:

$$MSE = \frac{1}{n} \sum_{i=1}^n \left(y_{i, real} - y_{i, predict} \right)^2 \quad (10)$$

where $y_{i, real}$ and $y_{i, predict}$ denote the real PV power generation value and the predicted value using the forecast model, respectively, and n is the total amount of data. All the operating sequences and models were realized using MATLAB 2021b. The computer was equipped with Windows 10 Pro, NVIDIA GeForce RTX 2070 Super, i7-9700k CPU, and 32 GB RAM.

4.2. Simulation Results

First, the correlation between each weather station and solar power generation was analyzed. Of the 26,279 data points collected over a total of three years, a correlation analysis was conducted on 10,950 data points based on the sunshine time of 7 AM to 6 PM. Among the existing measurement stations in Jeollanam-do, the correlation analysis targets were the Mokpo Weather Station, Yeosu Weather Station, and Heuksando Weather Station, which provide solar irradiance information.

Figure 15 displays 168 data points as an example from 1 January 2017 to 7 January 2017, and scales from 0 to 1 to compare each variable. The figure indicates that the trend of insolation at each measuring station and the trend of solar power generation in Jeollanam-do are similar. Because Mokpo Meteorological Station has the highest correlation with the amount of power generated, the performance of the meteorological site's cloud cover and the cloud cover extracted from satellite images will be determined based on the region. In addition, as can be seen from the correlation analysis results in Figure 16, the amount of solar irradiance measured at the Mokpo Meteorological Station was 0.9079, which was the highest correlation with the amount of power generation. Therefore, the results were compared based on the Mokpo area.

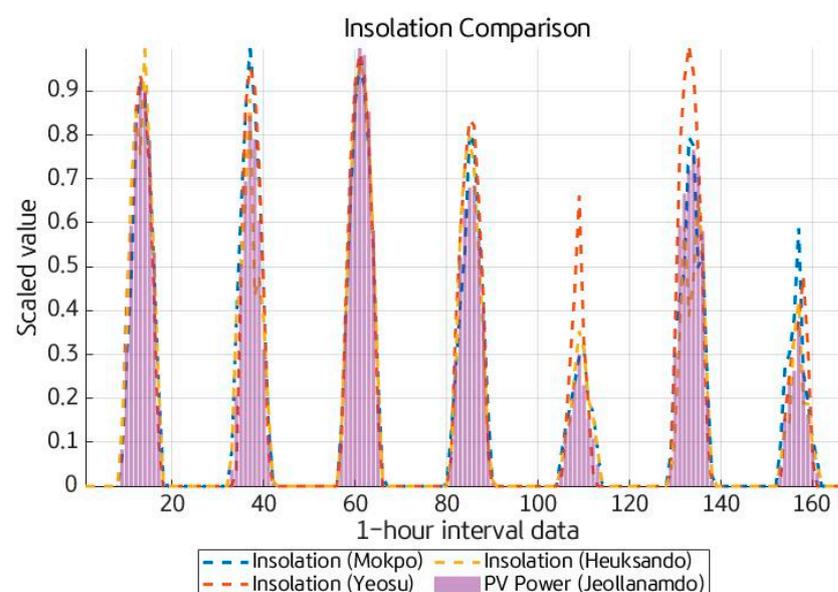


Figure 15. Comparison of the insolation of each weather station and PV power.

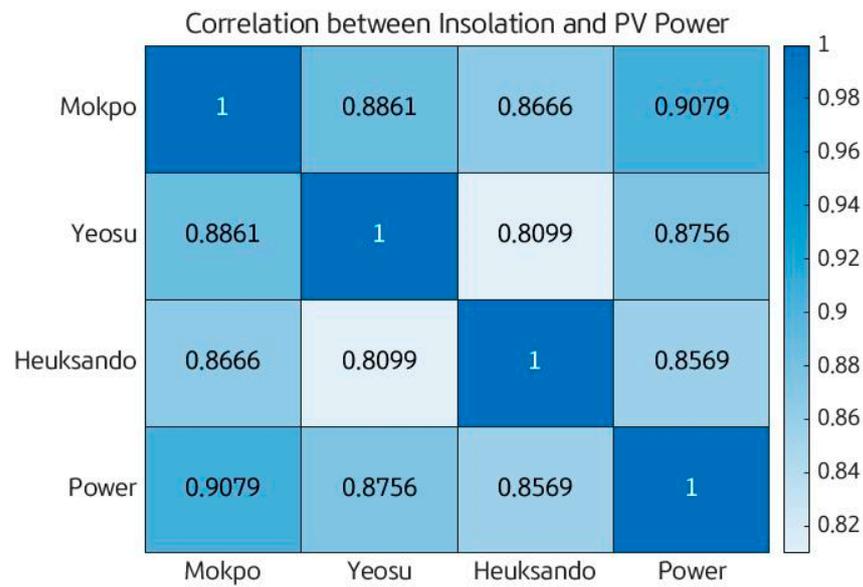


Figure 16. Correlation analysis between insolation of each weather station and PV power.

Cloud quantity divides the amount of cloud into grades from 0 to 10, and the closer the quantity is to 10, the more clouds are present. In general, because there is an inverse relationship in which the amount of power generation decreases with the increase in the number of clouds, the value was converted and used in this study by subtracting the cloud from 10. Subsequently, it was scaled from 0 to 1 for correlation analysis.

Figure 17 shows the cloud data and power generation collected for the same period and region. In the case of the cloudiness of each meteorological station, it shows a different form of the curve from the amount of solar irradiance. This is because the range of values is simple, and the fluctuation of values is large. The correlation analysis results in Figure 18 show that the correlation coefficient with the generation amount is low, unlike solar irradiance. In addition, unlike the amount of solar irradiance that directly measures the amount of light reaching the ground, it does not appear to reflect the difference in power generation due to the altitude of the sun because the focus is the amount of cloud.

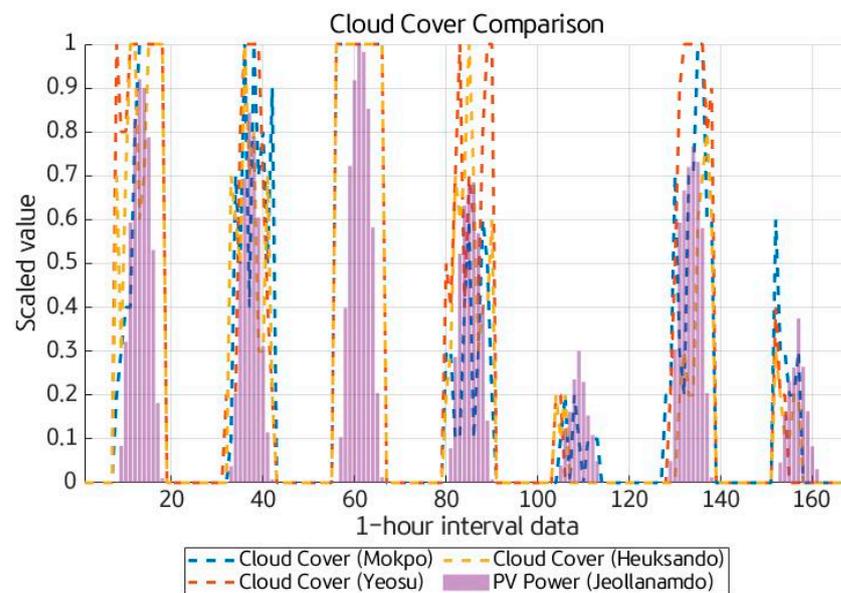


Figure 17. Comparison of the cloud cover of each weather station and PV power.

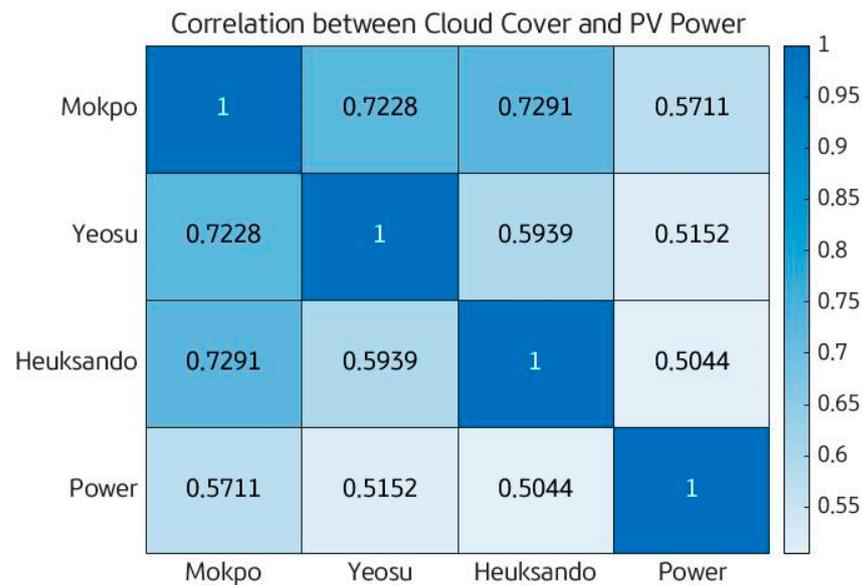


Figure 18. Correlation analysis between cloud cover of each weather station and PV power.

To evaluate the correlation of the satellite images, a comparison was conducted based on the data of the weather station in Mokpo, which has the highest correlation with the amount of solar power in Jeollanam-do. Satellite images were obtained from two types of images: visible and infrared. The values of pixels corresponding to the Mokpo area were extracted using the methodology mentioned in Section 3. Similar to the cloud provided by the weather station, the satellite image has an inversely proportional relationship with the amount of power generated. The higher the value of the cloud, the higher the value of the pixel. Therefore, the extracted pixel value was subtracted from 255, which is the maximum pixel value of the image, and applied to the correlation analysis.

It can be observed in Figure 19 that the infrared image shows a graph that is the most similar to the amount of power generation. The visible image tends to have a high value at sunrise time, which may be because it does not reflect the change in the sunrise time depending on the season. The visible image measures the intensity of sunlight reflected from the clouds and the ground; therefore, it is impossible to measure the intensity at night without sunlight, and only half of the image is photographed at sunrise and sunset; therefore, if the target area spans the area, the value is low. Consequently, it appears that the converted visible image value will always output high values. The cloud cover appears to correspond to changes in the amount of power generation; however, the correlation is unlikely to be high because the value fluctuates remarkably, and it is difficult to reflect the intensity of light.

Figure 20 presents the results of correlation analysis for the variables presented in the study. Excluding the amount of insolation, the method of extracting pixel values from infrared images was the most correlated with 0.6572, while the visible image was 0.6116, and the cloud measured at the measuring station was 0.5711. The variable with the highest correlation with insolation was infrared images, and the variables with the highest correlation between input variables were 0.8396 with infrared images and visible images.

As shown in the correlation analysis, the satellite images are similar to cloud cover, but have a higher correlation coefficient. When making actual predictions, the weather variables that the users can obtain through the Meteorological Administration or weather sites are the predictions of cloud cover and real-time satellite images. However, unlike satellite images, because the predicted value does not exist in the database, the experiment was conducted using the measured value instead of the predicted value.

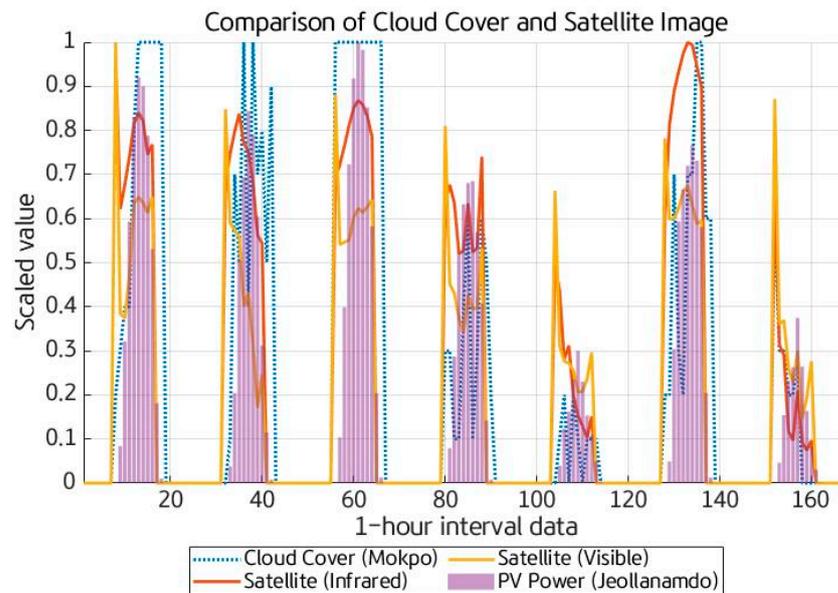


Figure 19. Comparison of the cloud cover, satellite images, and PV power.

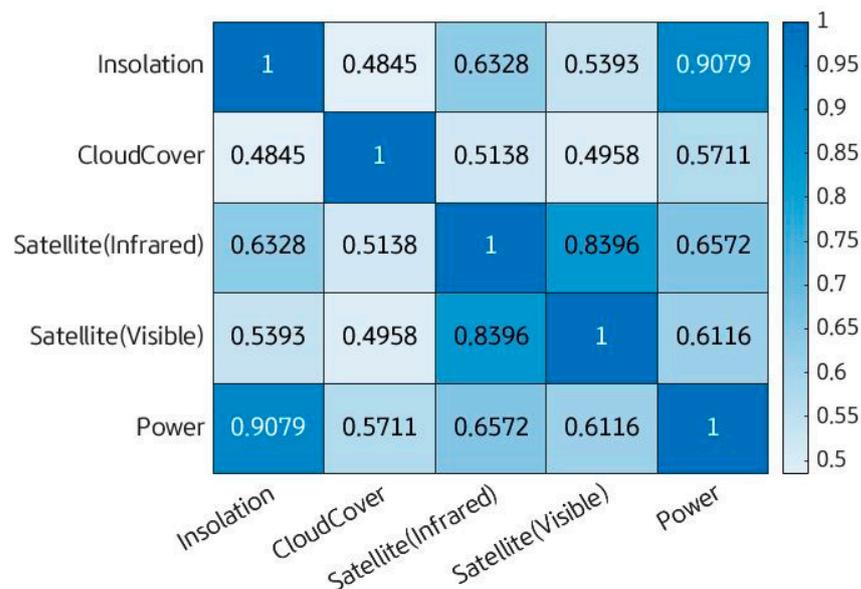


Figure 20. Correlation analysis between insolation, cloud cover, satellite images, and PV power.

First, because the predicted values extracted from the satellite images were required, correlation analysis between the target area and other areas was conducted at 1 h intervals using the method proposed in Section 3. For comparison, the correlation analysis was conducted on the infrared and visible images in the same way, and Figure 21 shows the correlation results for 10 AM, 1 PM, and 4 PM, when predictions were made for 7 AM. In Figure 21, the closer the color of each pixel to yellow, the higher the correlation, and the closer the color to blue, the lower the correlation. Figure 21a,c,e shows the results for the visible images, and Figure 21b,d,f shows the results for infrared images. In both the types of images, it can be observed that as the time interval between the current time point and predicted time point increases, the range of pixels having a high correlation increases. This can also help to infer the main direction and distance of the cloud.

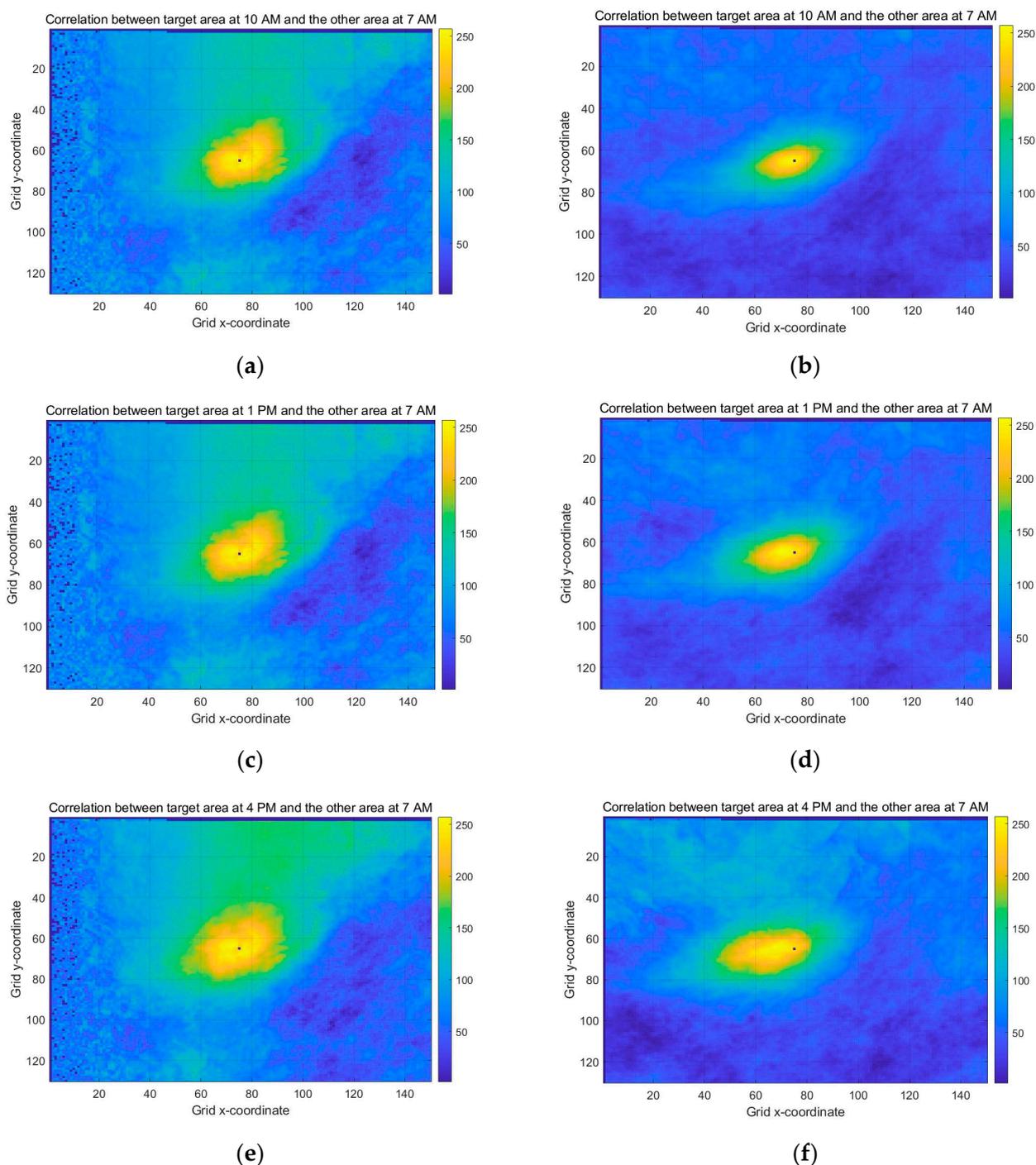


Figure 21. Comparison correlation analysis of the visible images and infrared images. (a) Visible image correlation between 10 AM and 7 AM; (b) infrared image correlation between 10 AM and 7 AM; (c) visible image correlation between 1 PM and 7 AM; (d) infrared image correlation between 1 PM and 7 AM; (e) visible image correlation between 4 PM and 7 AM; (f) infrared image correlation between 4 PM and 7 AM.

Table 4 selects pixels with the highest correlation coefficient over time for infrared and visible images and shows the corresponding values. It means the correlation between the value at the coordinates of the pixel closest to yellow in each image of Figure 21 and the target point at the current time. This result is obtained from the correlation analysis at 1 h intervals for 7 AM as the prediction start and 8 AM to 5 PM as the sunlight time. The selected maximum correlation pixel stores the coordinates and is used to predict

the flow rate later. In the case of infrared images, as shown in Figure 21, the maximum correlation coefficient decreased as the distance from the prediction time increased. In the case of visible images, in a peculiar case, if you consider the image taken at 7 AM, only half of the images are displayed because the sun has not yet completely illuminated all areas, and it seems that the exact correlation could not be inferred. In contrast, we found that the sun starts to set at 5 PM, and half of the images appear black, which results in a peculiar correlation.

Table 4. Most correlated coefficients of the infrared and visible images following time.

Time	Infrared Image Max Correlation	Visible Image Max Correlation
8 AM	0.9297	0.7609
9 AM	0.8674	0.4473
10 AM	0.7999	0.3809
11 AM	0.7223	0.3673
12 PM	0.6892	0.3793
1 PM	0.6514	0.3785
2 PM	0.6295	0.3602
3 PM	0.6025	0.3520
4 PM	0.5579	0.4192
5 PM	0.5284	0.7097

As shown in Figure 22, we compared the measured cloud cover and satellite images, predicted satellite images, and PV power for three days from 2 December to 5 December 2019. The PV power first shows a decreasing trend once during the day and then increases again, and the cloud cover measured by the Korea Meteorological Administration shows the same trend. However, owing to the lack of resolution of the values, the trend was not elaborately expressed. Green and red show cloud cover predicted using infrared and visible images, respectively. Although power generation tends to decrease during the day, errors also occur.

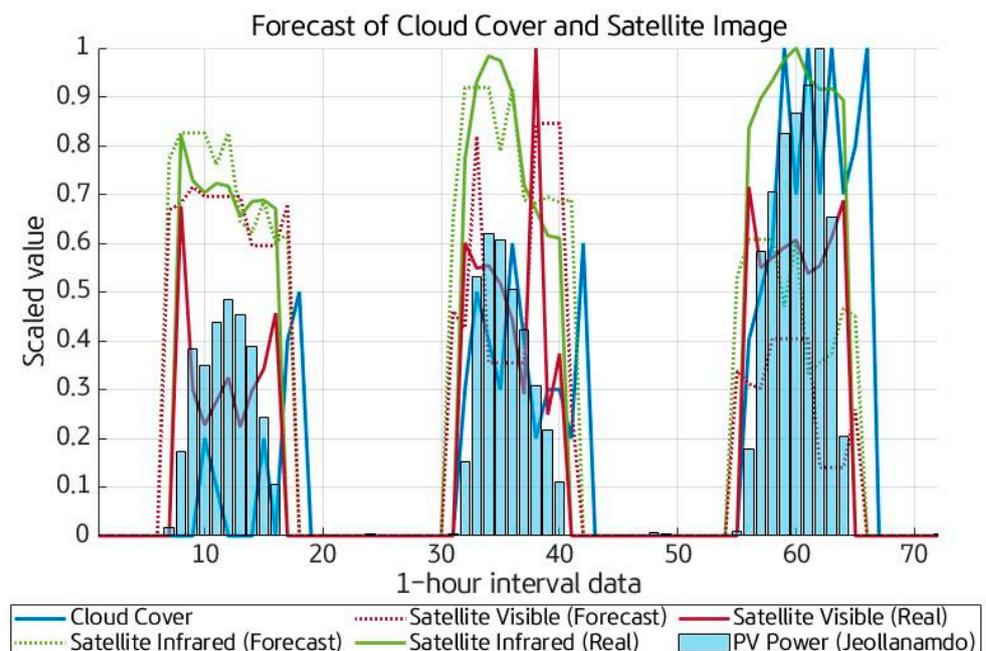


Figure 22. Comparison of the forecasted cloud cover and real values from the satellite.

Table 5 lists the results of predicting the amount of power generation after 1 h to 6 h from 7 AM. As for the prediction model, all the ANN-based models have identical

structures and hyperparameters, and the period was 1 October 2019–31 December 2019, and the verification was performed on a total of 92 data points. The results using visible image data were excellent until 2 h in a single model, but the results using cloud cover from 3 h later show better results, and the gap widens as time increases. However, the data that showed the best results are the cases where all the variables were used together. The multivariable case shows the highest accuracy from 1 h to 6 h, which means that each variable can be used complementarily.

Table 5. The prediction accuracy of 1 to 6 h ahead forecasting.

Time	Forecast Accuracy Using Cloud Cover	Forecast Accuracy Using Visible Image	Forecast Accuracy Using Infrared Image	Forecast Accuracy Using Multivariable
8 AM	41.916	19.836	38.649	16.350
9 AM	69.344	53.646	70.176	38.340
10 AM	69.288	82.349	92.286	54.447
11 AM	71.993	94.801	101.22	64.013
12 PM	69.872	106.72	110.46	66.215
1 PM	78.210	108.72	110.55	68.881
Average RMSE	66.771	77.679	87.224	51.374

4.3. Discussion

In the correlation analysis, because the correlation of the infrared images was the highest, it was expected that the best performance would be achieved when applied to the actual prediction model. However, in reality, the visible images demonstrated better performance for the ultra-short term, and the measured cloud cover demonstrated better prediction results after 3 h. In addition, to determine the possibility of data as a whole, an analysis of the images captured for three years was conducted without considering seasonality; however, further research is required on this because the movement of clouds is generally related to seasonal factors. In addition, performance evaluation is subsequently required by performing predictions on cloud volume and analyzing various prediction models for comparison.

5. Conclusions

With the increase in the proportion of solar power plants, more sophisticated prediction models than those before are required. The methods for improving forecast accuracy includes improving the quality of data or optimizing it using a suitable prediction model. This study was focused on improving the quality of the data. In particular, because the amount of solar irradiance that has the greatest correlation with the amount of PV power is not provided as a predicted value by the Meteorological Administration or weather site, the cloud cover that is correlated with PV power generation was selected as an alternative.

This study selected the weather station, the most highly correlated with the PV generation among the weather stations that record cloud cover, and compared the correlation with satellite images using data from the weather station. To this end, the pixel values of the satellite image were compared with the weather station data in the Mokpo area, and it was confirmed that the satellite image had a higher correlation with the PV power generation amount by 0.04–0.08. However, on applying cloud cover and satellite images to actual predictions, it was confirmed that the results of predictions using satellite images were better in the ultra-short term; however, the results of predictions using cloud cover were better after 3 h. Although cloud cover has an advantage of using measured values, the multivariable case showed the best predictive performance.

In conclusion, when using single input data, the satellite images were the best in the ultra-short term, and it may be better to use cloud cover when moving on to the short term. In addition, it was confirmed that when using multiple input data, the multivariable

case shows better prediction performance than that using single meteorological data or satellite data.

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

Funding: This research was supported in part by the KEPCO Research Institute under the project entitled by “A Research of Advanced Distribution Planning System for Mid-long term (R20DA16)”, in part by a Korea University Grant, and in part by the Basic Research Program through the National Research Foundation of Korea (NRF) funded by the MSIT (No. 2020R1A4A1019405).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

- IRENA. *Renewable Capacity Statistics 2021*; International Renewable Energy Agency (IRENA): Abu Dhabi, Saudi Arabia, 2021.
- Čonka, Z.; Kolcun, M.; Morva, G. Impact of Renewable Energy Sources on Power System Stability. *Power Electr. Eng.* **2014**, *32*, 25–28. [[CrossRef](#)]
- IEA. *System Integration of Renewables*; IEA: Brussels, Belgium, 2018.
- Wang, Y.; Wang, J.; Zhao, G.; Dong, Y. Application of residual modification approach in seasonal ARIMA for electricity demand forecasting: A case study of China. *Energy Policy* **2012**, *48*, 284–294. [[CrossRef](#)]
- Ahmad, T.; Zhang, H.; Yan, B. A review on renewable energy and electricity requirement forecasting models for smart grid and buildings. *Sustain. Cities Soc.* **2020**, *5*, 102052. [[CrossRef](#)]
- Wang, H.; Lei, Z.; Zhang, X.; Zhou, B.; Peng, J. A review of deep learning for renewable energy forecasting. *Energy Convers. Manag.* **2019**, *198*, 111799. [[CrossRef](#)]
- Yona, A.; Senjyu, T.; Saber, A.Y.; Funabashi, T.; Sekine, H.; Kim, C.-H. Application of neural network to 24-hour-ahead generating power forecasting for PV system. In Proceedings of the IEEE Power and Energy Society General Meeting—Conversion and Delivery of Electrical Energy in the 21st Century, Pittsburgh, PA, USA, 20–24 July 2008; pp. 1–6.
- Capizzi, G.; Napoli, C.; Bonanno, F. Innovative Second-Generation Wavelets Construction with Recurrent Neural Networks for Solar Radiation Forecasting. *IEEE Trans. Neural Netw. Learn. Syst.* **2012**, *23*, 1805–1815. [[CrossRef](#)] [[PubMed](#)]
- Cao, S.; Weng, W.; Chen, J.; Liu, W.; Yu, G.; Cao, J. Forecast of solar irradiance using chaos optimization neural networks. In Proceedings of the Asia-Pacific Power and Energy Engineering Conference, Wuhan, China, 27–31 March 2009; pp. 1–4.
- Dumitru, C.-D.; Gligor, A.; Enachescu, C. Solar Photovoltaic Energy Production Forecast Using Neural Networks. *Procedia Technol.* **2016**, *22*, 808–815. [[CrossRef](#)]
- Yan, J.; Hu, L.; Zhen, Z.; Wang, F.; Qiu, G.; Li, Y.; Yao, L.; Shafie-Khah, M.; Catalao, J.P.S. Frequency-Domain Decomposition and Deep Learning Based Solar PV Power Ultra-Short-Term Forecasting Model. *IEEE Trans. Ind. Appl.* **2021**, *57*, 3282–3295. [[CrossRef](#)]
- 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. In Proceedings of the 2011 IEEE Industry Applications Society Annual Meeting, Orlando, FL, USA, 9–13 October 2011; pp. 1–6.
- Cortes, C.; Vapnik, V. Support vector networks. *Mach. Learn.* **1995**, *20*, 273–297. [[CrossRef](#)]
- Drucker, H.; Burges, C.J.C.; Kaufman, L.; Smola, A.; Vapnik, V. Support vector regression machines. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 1997; Volume 9, 155p.
- Deng, N.Y.; Tian, Y.J. *New Data Mining Method-Support Vector Machine*; Science Press: Beijing, China, 2004; p. 75.
- Lahouar, A.; Mejri, A.; Slama, J.B.H. Importance based selection method for day-ahead photovoltaic power forecast using random forests. In Proceedings of the International Conference on Green Energy Conversion Systems (GECS), Hammamet, Tunisia, 23–25 March 2017; pp. 1–7.
- Huang, C.J.; Kuo, P.H. Multiple-Input Deep Convolutional Neural Network Model for Short-Term Photovoltaic Power Forecasting. *IEEE Access* **2019**, *7*, 74822–74834. [[CrossRef](#)]
- Yang, H.-T.; Huang, C.-M.; Huang, Y.-C.; Pai, Y.-S. A Weather-Based Hybrid Method for 1-Day Ahead Hourly Forecasting of PV Power Output. *IEEE Trans. Sustain. Energy* **2014**, *5*, 917–926. [[CrossRef](#)]
- Hossain, M.S.; Mahmood, H. Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast. *IEEE Access* **2020**, *8*, 172524–172533. [[CrossRef](#)]

20. Sangrody, H.; Zhou, N.; Zhang, Z. Similarity-Based Models for Day-Ahead Solar PV Generation Forecasting. *IEEE Access* **2020**, *8*, 104469–104478. [[CrossRef](#)]
21. Yu, Y.; Cao, J.; Zhu, J. An LSTM Short-Term Solar Irradiance Forecasting Under Complicated Weather Conditions. *IEEE Access* **2019**, *7*, 145651–145669. [[CrossRef](#)]
22. Rajagukguk, R.A.; Ramadhan, R.A.; Lee, H.-J. A Review on Deep Learning Models for Forecasting Time Series Data of Solar Irradiance and Photovoltaic Power. *Energies* **2020**, *13*, 6623. [[CrossRef](#)]
23. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M. A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renew. Sustain. Energy Rev.* **2020**, *124*, 109792. [[CrossRef](#)]
24. Escrig, H.; Batlles, F.; Alonso, J.; Baena, F.; Bosch, J.; Salbidegoitia, I.; Burgaleta, J. Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast. *Energy* **2013**, *55*, 853–859. [[CrossRef](#)]
25. Yang, D.; Jirutitijaroen, P.; Walsh, W.M. Hourly solar irradiance time series forecasting using cloud cover index. *Sol. Energy* **2012**, *86*, 3531–3543. [[CrossRef](#)]