Article

A Study on the Power Generation Prediction Model Considering Environmental Characteristics of Floating Photovoltaic System

Han Sang Jeong [1,2], Jaeho Choi [2,]* , Ho Hyun Lee [3] and Hyun Sik Jo [3]

[1] Geumgang River Basin Head Office, Korea Water Resources Corporation (K-water), Jeonju, Jeonbuk 54851, Korea; han-sang@kwater.or.kr
[2] School of Electrical Engineering, Chungbuk National University, Cheongju, Chungbuk 28644, Korea
[3] K-water Research Institute (KWI), Korea Water Resources Corporation (K-water), Deajeon 34045, Korea; lhh@kwater.or.kr (H.H.L.); hsjo@kwater.or.kr (H.S.J.)
*   Correspondence: choi@cbnu.ac.kr

check for updates

Abstract: The main contents of this paper are to verify the environmental factors affecting the power generation of floating photovoltaic systems and to present the power generation prediction model considering environmental factors by using regression analysis and neural networks studied during the last decade. This study focused on a comparative analysis of which model is best suited for the power generation prediction of the floating photovoltaic (PV) system. To compare the power generation characteristics of a floating and a land-based PV system, two identical 2.5 kW PV systems were installed—one on the water surface in the Boryeong Dam, Korea, and the other nearby on dry land—and their performances were compared. The solar irradiance of the floating PV system was 1.1% lower than that of the land-based PV. Nevertheless, the floating PV module temperature was 4.9% lower than that of the land-based PV, generating approximately 3% more power. Using the correlation analysis of data mining techniques, environmental factors affecting the efficiency of the floating PV system were investigated. The correlation coefficient between the module temperature and water temperature was $r = 0.6317$ which proves that the high efficiency and low module temperature characteristics of the floating PV system, when compared with that of the land-based PV, are due to the water evaporation effect. Considering environmental factors, power-generation prediction models based on regression analysis and neural networks are presented, and their accuracies are compared. This comparison confirms that the accuracy of the power generation prediction model using neural networks was approximately 2.59% higher than that of the regression analysis method. As a result of adjusting the hidden nodes in the neural network algorithm, it was confirmed that a neural network algorithm with ten hidden nodes was most suitable for calculating the amount of power generation.

Keywords: floating photovoltaic system; neural network; prediction power generation; regression analysis

1. Introduction

Recently, the renewable energy market has been growing rapidly, owing to fossil fuel reduction and environmental problems [1,2]. With the expansion of the renewable energy market, the global photovoltaic (PV) capacity is projected to exceed 1 TW by 2022 [3]. Especially, the solar power market for large-scale PV power-generation systems with capacities in the range of MW or higher is expanding rapidly. However, it is difficult to secure land for the construction of such large-scale systems; hence, large-scale floating PV facilities have been proposed as high-efficiency alternatives [4].

As of 2018, 1314 MW of floating PV systems were installed worldwide, and it has been predicted that 5,211,086 GWh/y of electricity will be generated annually if 10% of the man-made reservoir surfaces are used for floating PV systems [5]. Various researches have accompanied the growth of the floating PV market. In Korea, as in many countries around the world, several kinds of research on floating solar systems are being conducted, such as research on the installation of floating PVs in oceans, the grounding of floating PVs, and tracing-type floating PVs that track sunlight [6,7].

The floating PV system is a new concept of a power plant, which uses buoyant bodies to float solar power plants on the surfaces of dam reservoirs. The system is designed such that the solar module is fixed to a float that is secured by a mooring device on the water surface [8].

Figure 1 shows the concept of a typical floating PV system. The benefit of a large-scale floating PV system is that it can be installed in the reservoir of a dam without the need for deforestation, which is a common problem encountered while installing land-based PV systems.
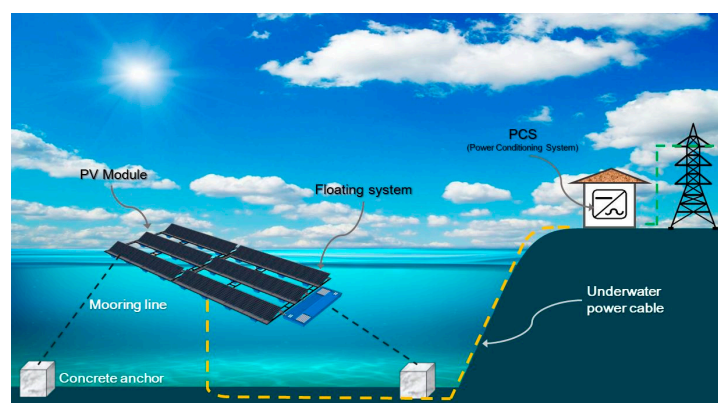
Figure 1. Concept of floating PV system.

The first floating PV system was built in 2007 in Aichi, Japan, following which several other countries, including France, Italy, Korea, Spain, and the United States, installed small-scale systems for research and development (R&D) purposes [5]. Recently, PV systems have been widely deployed in many countries, and their capacity has increased to several tens of MW or higher. A 7.5 MW system and a 6.3 MW system on the Umenoki reservoir, Japan, and the Queen Elizabeth II Lake, UK, are currently in operation. In Huainan City, China, a 40 MW PV system is in operation [5,9]. In Korea, a 100 kW demonstration floating PV system was installed, for the first time, on the water surface of the Hapcheon Dam reservoir, in October 2011. After successfully installing this PV system, a 500 kW commercial floating PV system was installed at another nearby location in 2012. As of 2020, the representative floating PV systems operating in Korea are a 2 MW system on the Boryeong Dam reservoir, a 3 MW system on the Chungju Dam reservoir, and an 18.7 MW system on the lagoon at the Gunsan Industrial Complex. Figure 2 shows the 3MW floating PV system installed on the reservoir of the Chungju Dam, Korea.

Figure 2. Three MW floating PV systems on the reservoir of Chungju Dam, Korea (source: K-water).

The power generation characteristics of PV modules are heavily influenced by the cell/module temperature [10]. Generally, the PV module output can be estimated as shown in Equation (1), and thus, the rating of a PV module may differ depending on the temperature differences from the actual site under the standard test conditions (STC), wherein the solar radiation intensity is 1000 W/m^2, the module temperature is 25 °C, and the air mass (AM) coefficient is 1.5 [11].

$$P = G_T \tau_{pv} \eta_{Tref} A [1 - 0.0045(T_c - 25)], \tag{1}$$

where P is the output power of the PV module [W], G_T is the solar irradiance on the module plane [W/m^2], τ_{pv} is the transmittance of the PV cells' outside layers, η_{Tref} is the electrical efficiency of the module at a specified temperature, A is the surface area of the PV module [m^2], and T_c is the module operating temperature [°C]. From Equation (1), it can be expected that the amount of power generation decreases by 0.45% as the module temperature increases by 1 °C from 25 °C.

Published research has concluded that the power generation of floating PV systems is 10 to 15% greater than that of the land-based PV systems because of the cooling effect [12,13]. In one study, a PV module cooling system with intermittent water spray to module was installed to get the above mentioned result, and in another, the solar-cell module temperature was presumed to be reduced by the evaporation of water. This paper investigates the PV systems without any mechanical cooling system and with the same sizes and nearby locations, which is rarely reviewed in the other papers. In addition, this paper proposes how to calculate the generated power by a machine learning algorithm, which is not done much in floating PV systems. The various factors used for comparing the land and floating PV systems were different; for example, the modules, inverters, years of installation, and power-generation capacities were different. Therefore, there was a limit to the precise comparison of the power-generation characteristics of the two types of PV systems.

To develop a large-scale floating PV power plant, the economic feasibility must be secured, and the most important factor in analyzing the economic feasibility is the power-generation prediction. The land-based PV system output forecasting problem has been thoroughly studied during the last decade using machine or neural network-based learning. However, the installation environmental of floating PV systems is different from that of land-based PV systems. Therefore, this paper focuses on selecting the most suitable power generation prediction model, considering the power generation characteristics of floating PV system.

Two identical 2.5 kW PV systems were installed on the water surface of the Boryoung Dam reservoir, Korea, and on the nearby land, respectively, to ensure precise comparison [14]. Based on the operation results over six months, the following facts were confirmed: The solar irradiance of the floating PV was 1.1% less than that of land-based PV. Nevertheless, the floating PV module temperature was 4.9% lower than that of the land-based PV, generating approximately 3% more power. The correlation coefficient between the module temperature and water temperature was $r = 0.6317$, which proved that the high efficiency and low module temperature characteristics of

the floating PV, when compared to that of the land-based PV, were due to the water evaporation effect. Through regression analysis, the power-generation prediction model of a floating PV system considering environmental factors was presented. Recently, neural network analysis is being widely used for estimating the generation of power because of its more accurate prediction when compared to regression analysis [15]. Therefore, a prediction model using neural networks, represented by a nonlinear algorithm, was developed in this study. A hyperbolic tangent was used as the activation function, and prediction models were compared using one to twenty nodes to choose the best model.

In this study, prediction models for floating PV power generation, considering environmental factors, are presented using regression analysis and neural networks. Upon comparing the accuracy, it was confirmed that the prediction model using neural networks was more accurate than the one using regression analysis. Therefore, this study presents more logical and accurate approaches to power generation prediction and economic feasibility analysis, which are essential for the development of large-scale floating PV power plants.

2. Floating PV Power Generation Characteristics

2.1. Power Generation Characteristics Comparison System (PGCCS)

To precisely compare and analyze the power generation characteristics of floating and land-based PV systems, depending on the environmental effects, both systems should have the same system specifications and installation conditions, such as PV modules, inverters, the capacity of facilities, installation orientation, and angle [16].

In this study, 2.5 kW power generation characteristics comparison systems (PGCCS) were installed, as shown in Figure 3. The configuration of the power circuit of the 2.5 kW PGCCS is shown in Figure 4. Table 1 lists the specifications of the PV module and Power Conditioning System (PCS) of the PGCCS installed, as shown in Figure 4. The installation direction of PGCCS was south, and the mounting tilt was fixed at 30 degrees.

(a)

(b)

(c)

(d)
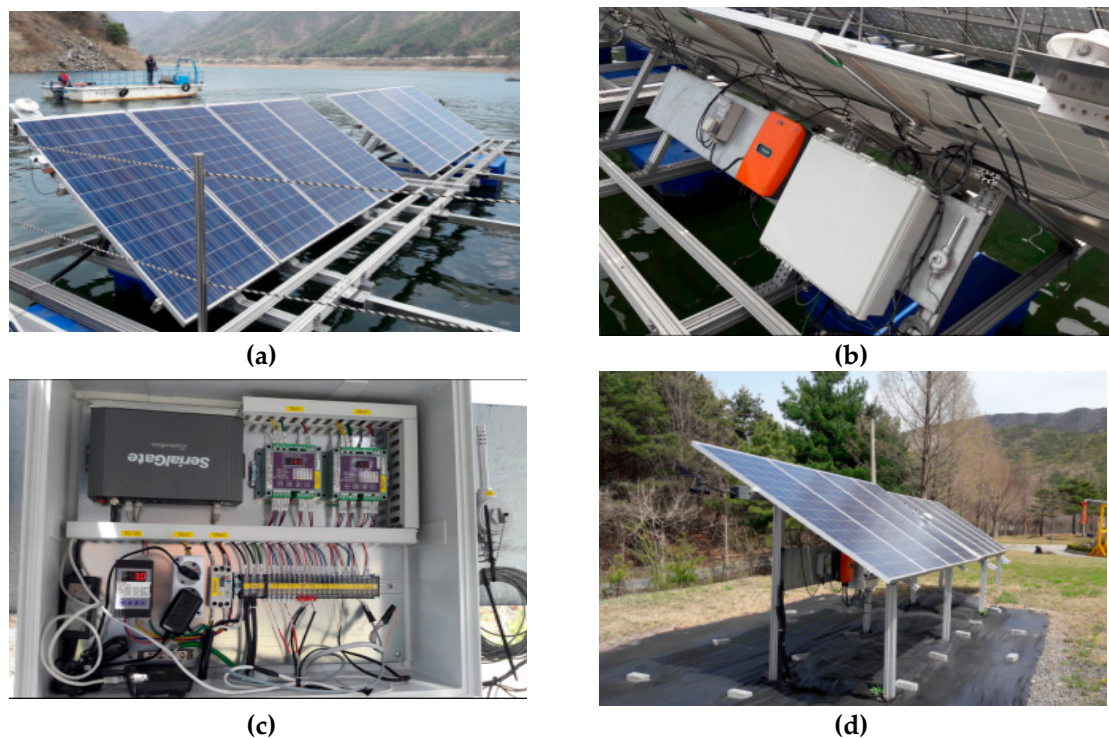
Figure 3. Pictures of PGCCS for floating and land-based PV systems: (**a**) 2.5 kW floating PV, (**b**) floating PV inverter, (**c**) measuring instrument box, (**d**) 2.5 kW land-based PV.
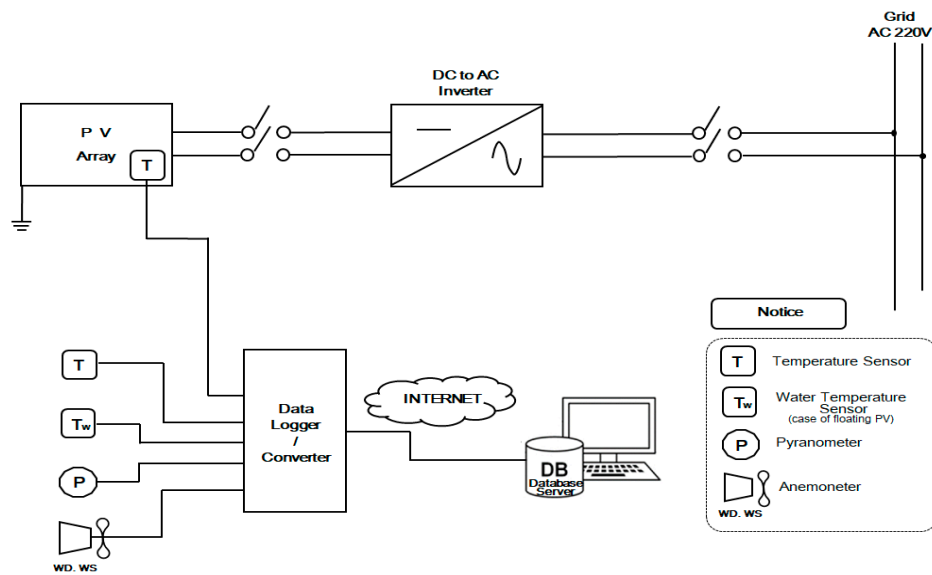
Figure 4. Configuration of PGCCS power circuit.

Table 1. Specification of Power Generation Characteristics Comparison System (PGCCS).

	PV Module and Array		PCS		
P_{max}	Cell	Array Configuration	Rated Output	Operating Voltage	Control Method
315 W	Polycrystal72	4×2	3.1 kW	V_{dc}: 100~500 V V_{ac}: 220 V, 60 Hz	Pulse–Width Modulation (PWM)

To analyze the power generation characteristics, the PV output and meteorological data, as summarized in Table 2, were collected.

Table 2. Measurement indexes of PGCCS.

Index	Array Output		Irradiation		Temperature			Wind Speed
	V_{dc}	I_{dc}	Horizontal	Slope	Air	Water	PV Module	
Unit	V	A	W/m^2	W/m^2	°C	°C	°C	m/s

The analysis period was six months from 1 July 2017 to 31 December 2017, and the analysis time was from 7:00 to 19:00 during the daytime when solar power is generated. Because the PGCCS was installed at the end of June 2017, it was analyzed based on the results of operation over the six months from 1 July 2017 to 31 December 2017. As the climate during spring and autumn exhibits the same characteristics, in Korea, all the different seasonal characteristics for these seasons could be covered during this period of analysis. The analysis was performed in the daytime because the module temperature was influenced by solar irradiance during the daytime and this directly affected the system efficiency. Night-time was excluded because the deviation in the PV module temperature characteristics reduced when it was included in the analysis time.

2.2. Operating Characteristics

Figures 5–7 show the comparative results of the floating and land-based PV systems in terms of the amount of power generated, the PV module temperatures, and irradiation, respectively.
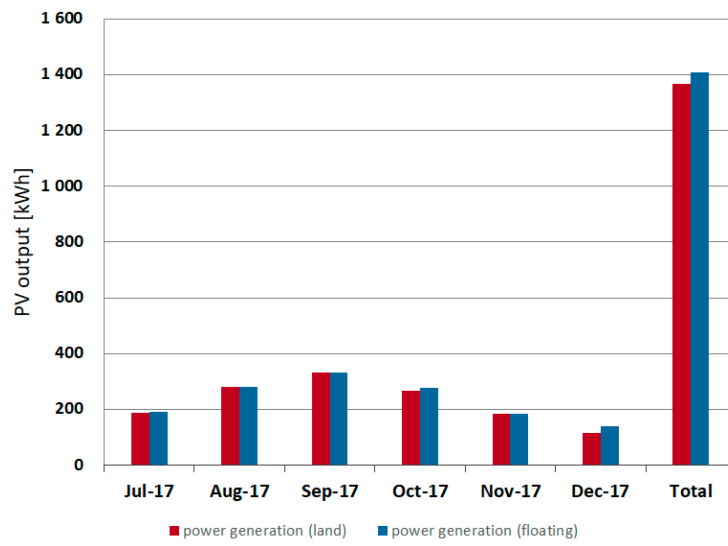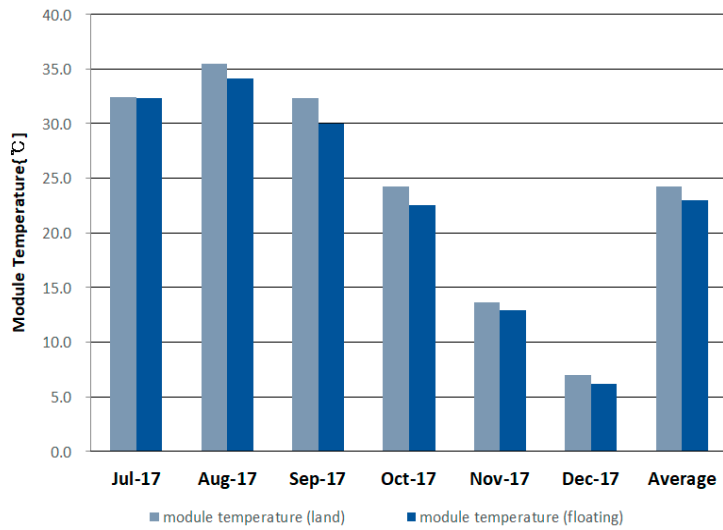
Figure 5. Comparative results of power generation.

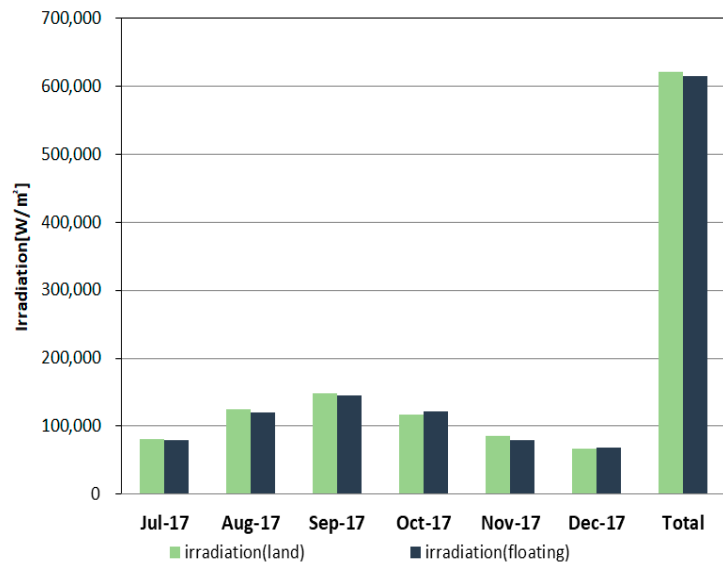**Figure 6.** Comparative results of module temperature.

Figure 7. Comparative results of irradiation.

During the survey period, the total solar irradiances were 622,262 and 615,696 W/m^2 for the land-based and floating PV systems, respectively, as shown in Figure 7. This means that the solar irradiance of the land-based PV system was 1.1% higher than that of the floating PV system. The reason for the solar irradiance of the floating PV system being slightly lower than that of the land PV system was presumed to be the mist at the water surface that occurred frequently. However, as shown in Figure 6, the daytime PV module temperature of the floating PV system was 4.9% lower than that of the land-based PV system. The outputs of the floating and land-based PV systems were 1408 and 1367 kWh, respectively, as shown in Figure 5. Therefore, the output of the floating PV system was 3% higher than that of the land-based PV system.

3. Data Analysis Algorithm

For analyzing the power-generation characteristics of the floating PV system, correlation and regression analyses were used first, following which the neural networks-based analysis was applied as an improved method for more accurate analysis.

3.1. Correlation Analysis

Correlation analysis is a method of numerically expressing the correlations of the mutual effects in the data [17]. Pearson's correlation, which is the correlation statistic most widely used to measure the degree of the relationship between linearly related variables, was used to analyze the relationship between the temperatures of the PV modules and the environmental factors, including water temperature and wind speed. Pearson's correlation coefficient, r, is described in Equation (2).

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}, \tag{2}$$

where, $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_n$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_n$, n is the number of samples, and $-1 \leq r \leq 1$ [18].

The variables x_i and y_i, which affect the module temperature of the floating PV, are defined as in Equations (3)–(5).

$$x_i = Tc_{spv} - Tc_{fpv}, \tag{3}$$

$$y_{i1} = T_a - T_w, \tag{4}$$

$$y_{i2} = v_{spv} - v_{fpv}, \tag{5}$$

where x_i is the module temperature difference between the temperatures of the land-based PV module, Tc_{spv}, and floating PV module, Tc_{fpv}. y_{i1} is the temperature difference between the PV system on land, T_a, and the water surface, T_w, and y_{i2} is the wind speed difference between land, v_{spv}, and water surface, v_{fpv}. The reason for setting the temperature differences, x_i, y_{i1}, and y_{i2}, as above is to focus on the precise influence of each variable.

3.2. Regression Analysis

Regression analysis was used to analyze the power-generation characteristics of the floating PV system, according to the change in module temperature. For describing the dependent variable Y, the multiple regression model was defined as in Equation (6) with the k number of X_k [19–21].

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \epsilon_i, \tag{6}$$

where, Y = dependent variable, PV generated power. X_k = independent variables, such as module temp, slope irradiation, water temp, and so on. $\beta_0, \beta_1, \ldots, \beta_k$ are the regression coefficients to be estimated, ϵ_i. is the residual error, and i is the data number.

The progress of estimating the regression coefficients in the multiple regression models is as follows. Assuming that the estimates of the regression coefficients, $\beta_0, \beta_1, \ldots, \beta_k$, are $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$, the least-squares method is applied to obtain $\hat{\beta}_0, \hat{\beta}_1, \cdots, \hat{\beta}_k$, that minimizes the squared sum of the residuals.

3.3. Neural Network Analysis

Neural networks that mimic the information processing functions of the brain nerves of organisms can be classified into supervised learning and unsupervised learning networks, according to the learning methods.

Figure 8 shows the supervised learning process of neural networks, which use training data consisting of input values and target values to adjust the weight, which is the strength of the formation, according to certain rules, to minimize the error between the neural network output value and target output value. Therefore, the weight is set such that the correspondence between the input and output from the given input and output data is implemented well [22].
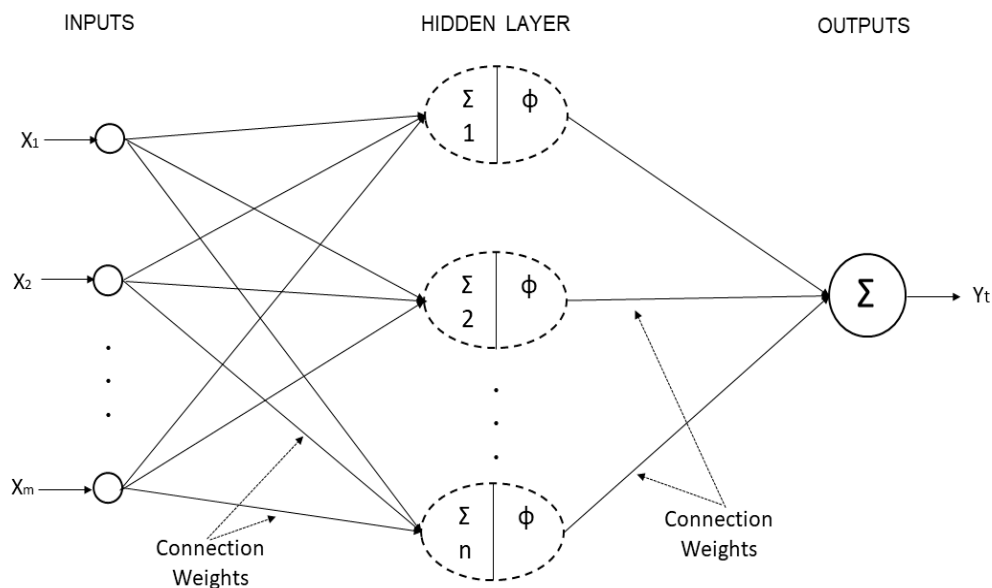

Figure 8. Structure of neural network.

In unsupervised learning, on the other hand, no target values corresponding to the input values are provided. In this case, the weights are adjusted for similar input patterns so that each node produces a similar output.

In this study, the multilayer perceptron (MLP) was used, which is the most popular model for supervised learning [23]. The MLP consists of units with many interconnected nodes. Each node function usually consists of a weighted sum and differential nonlinear activity function. Figure 9 shows the most useful activation functions of MLP.

In artificial neural networks, the activity function of a node defines the output of that node, given an input or set of inputs. A standard integrated circuit can be expressed as a digital network of activity functions, which can be "ON" (1) or "OFF" (0), depending on the input.
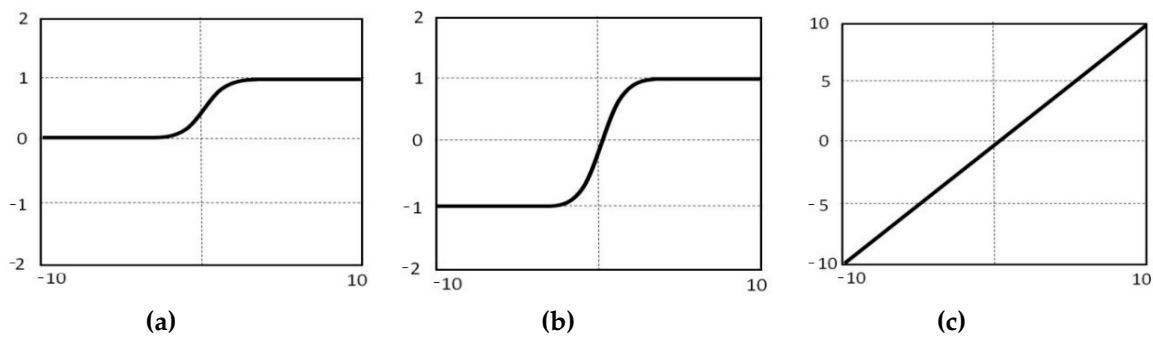
If three activation functions are to be described as equations, they can be expressed as the following equations:

$$\text{Logistic function} : f(x) = \frac{1}{1 + e^{-x}}, \tag{7}$$

$$\text{Hyperbolic tangent function} : f(x) = tanh\left(\frac{x}{2}\right) = \frac{1 - e^{-x}}{1 + e^{-x}}, \tag{8}$$

$$\text{Identity function} : f(x) = x, \tag{9}$$

Figure 9. Characteristics of activity function: (**a**) logistic function, (**b**) hyperbolic tangent, (**c**) identity function.

4. Analysis Results

4.1. Influence Factors

To analyze the environmental factors affecting the floating PV module temperature, correlations among the module temperature difference, air–water temperature difference, and wind speed difference between land and water, were analyzed.

Figure 10a shows a scatter plot that compares the module temperature difference between the land-based and floating PV modules to the temperature difference between the air over land and water [19]. It is observed that the scatter plot has a regular form and a positive correlation wherein the larger the temperature difference between the air over land and the water, the larger the module temperature difference between the land-based and floating PV modules is. The linear regression algorithm shows a high correlation coefficient of $r = 0.6317$. In general, the degree of correlation to the correlation value r is given, as in Table 3 [24,25]. Thus, significant correlations exist between the land-based and floating module temperature differences and the air and water temperature differences.

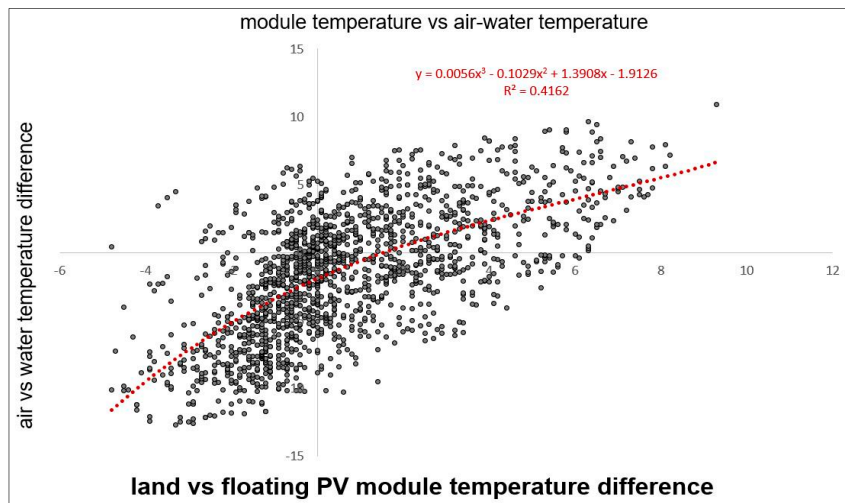**Table 3.** Correlation coefficient and relationship degree.

Correlation Coefficient r	Relationship Degree
0~0.2	Almost no correlation
0.2~0.4	Some correlation
0.4~0.7	Significant correlation
0.7~1.0	Strong correlation

Figure 10b shows the results of the nonlinear learning method applied to improve the results of the linear regression analysis. As a result of this analysis, the correlation coefficient r was improved to 0.6541, and it exhibited mutual nonlinearity.

Figure 11 shows a scatter plot showing the mutual characteristics of the module temperature difference and wind speed difference between land and water. The land-based and floating module temperature difference and wind speed difference indicate that there is no special pattern and that the correlation coefficient is very low (0.0032), which indicates that there is no specific correlation. Therefore, the module temperature that directly affects power generation has a very low correlation with wind speed; however, the change in water temperature has a significant correlation with the module temperature.
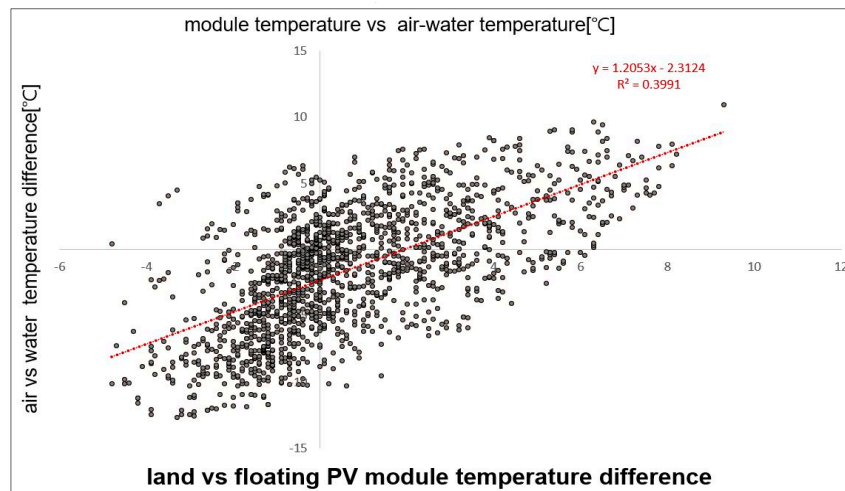
(**a**)

(**b**)

Figure 10. Scatter plot and correlation coefficient for comparing module temperature and air–water temperature: (**a**) linear, $r = 0.6317$, (**b**) nonlinear, $r = 0.6541$.
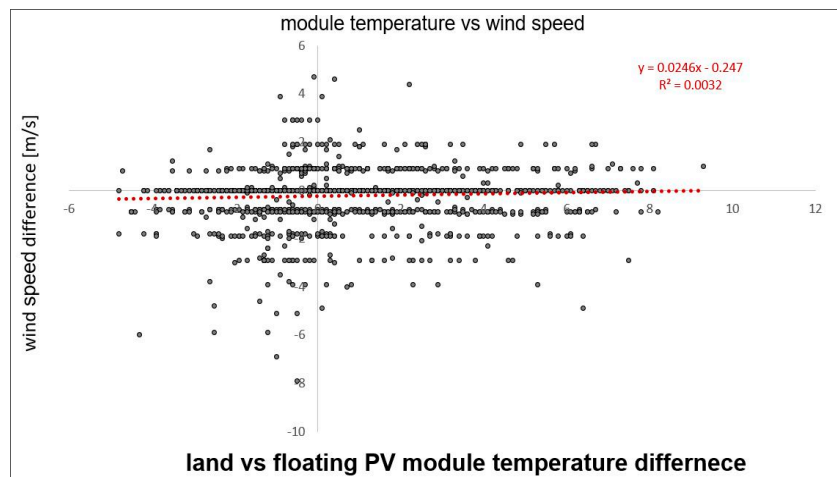

Figure 11. Scatter plot and correlation coefficient for comparing module temperature and wind speed ($r = 0.0565$).

4.2. Prediction Model Based on Linear Regression Analysis

Table 4 summarizes that solar radiation is the most influential factor in the correlation analysis for calculating the amount of solar power generated by the floating PV module, followed by the module temperature and ambient temperature, in sequence. The wind speed and direction have little effect on power generation. From Table 4, the correlation between the horizontal surface irradiance and slope irradiance, and that between the ambient temperature and module temperature, are 0.97 and 0.89, respectively, which means that the data are relatively similar. Thus, it is desirable to consider only the slope irradiance and module temperature.

Table 4. Floating PV correlation analysis.

Correlation	Power Generation	Module Temperature	Ambient Temperature	Horizontal Irradiation	Slope Irradiation	Wind Speed	Wind Direction
Power Generation	1	—	—	—	—	—	—
Module Temperature	0.64	1	—	—	—	—	—
Ambient Temperature	0.25	0.89	1	—	—	—	—
Horizontal Irradiation	0.97	0.69	0.34	1	—	—	—
Slope Irradiation	0.98	0.61	0.21	0.97	1	—	—
Wind Speed	−0.06	−0.09	−0.05	−0.07	−0.06	1	—
Wind Direction	−0.16	−0.17	−0.10	−0.15	−0.15	0.32	1

Table 5 summarizes the analysis results of the floating PV regression. From this table, the multiple correlation coefficient is 0.9859, which is more accurate than that of the land-based PV system. In detail, the t-statistic of irradiation is 178.63; therefore, the linear relationship between the independent variable and the dependent variable is very high. As the *p*-values of the module temperature and solar irradiation are 0 and the y-intercept is 0.01, whose absolute value is less than 0.05, the null hypothesis that there is no linear relationship is rejected. Therefore, the module temperature and solar irradiation coefficient can be used as the variables [26]. As a result, the formula for estimating the power generation of the floating PV module using linear regression is described in Equation (10).

$$P_{fpv} = T_c \times (0.0024851) + G_t \times (0.0019761) + (-0.0141039), \tag{10}$$

where P_{fpv} is the floating PV module output (kW), T_c is the module temperature (°C), and G is the solar irradiation amount (W/m^2).

Table 5. Floating PV regression analysis results.

Division	Coefficient	Standard Error	t-Statistics	*p*-Value
Module Temperature	0.0024851	0.00	9.88	0
Irradiation	0.0019761	0.00	178.63	0
y-Intercept	−0.0141039	0.01	−2.75	0.01
Multiple Correlation Coefficient	**0.9859**			

Upon generating random numbers from the floating PV operating results data, 50% of the data were used for training, and the remaining 50% were used for verification, to confirm whether overfitting

occurred or not. In addition, an error function, the mean absolute percentage error (MAPE) was used to compare the accuracy, as shown in Equation (11) [27].

$$MAPE = \frac{100}{n} \sum_{i=1}^{n} \left| \frac{E_i - D_i}{E_i} \right|, \tag{11}$$

where D_i is the value predicted by the predictive model, E_i is the measured value, and n is the number of datasets.

Figure 12 shows the results of the linear regression analysis with the learning data. The linearity can be confirmed as a result of plotting the true output and predicted output. Upon converting this into MAPE, the error showed 9.61%.
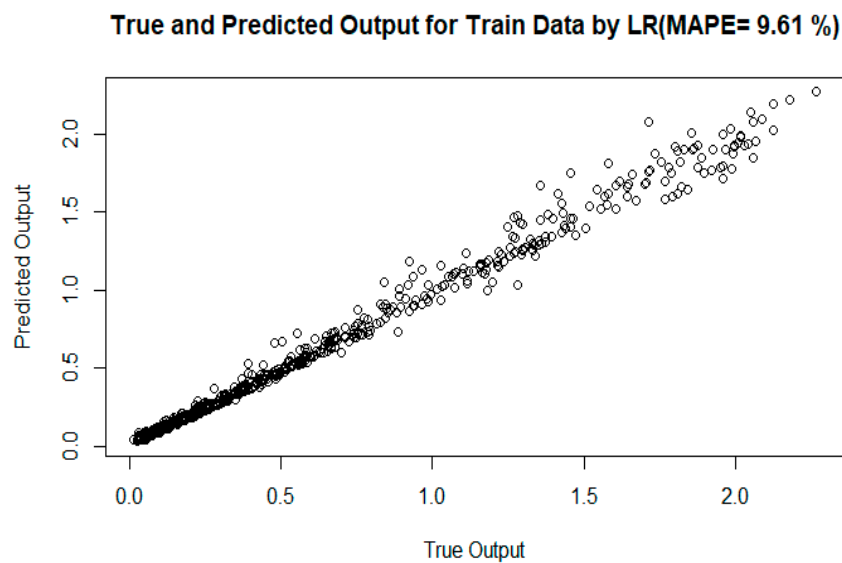
Figure 12. Linear regression (LR) analysis results (train data).

Figure 13 shows the results obtained by entering the test data into the learning model generated from the learning data. The MAPE is 11.35%, which is greater by 1.74%, but it was confirmed that overfitting did not occur.
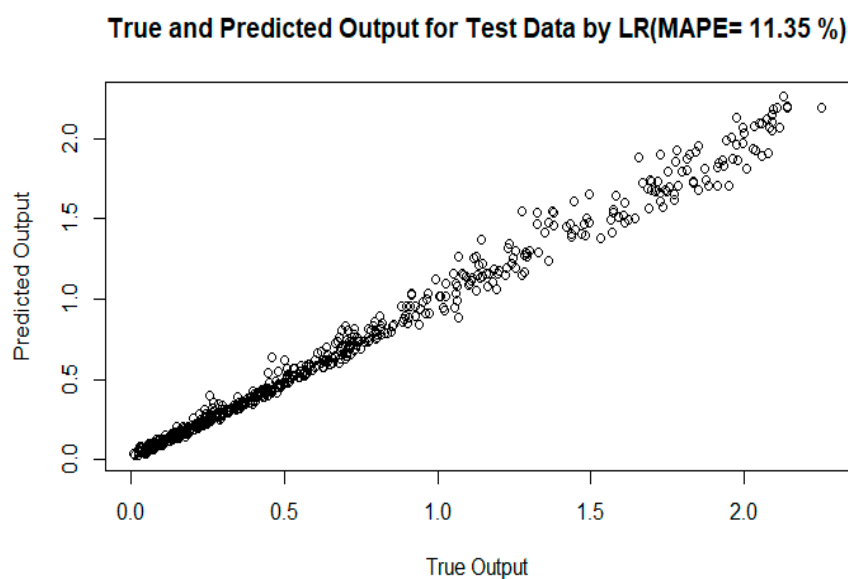
Figure 13. Linear regression (LR) analysis results (test data).

4.3. Prediction Model Based on Neural Network Analysis

Neural networks are used as representative nonlinear algorithms, which can be learned by adjusting the layer, node, and activation functions. In this work, the hyperbolic tangent was used as the activation function, and the numbers of nodes were set as 1, 5, 10, and 20. Before those models were made, the dataset was preprocessed to remove the outlier data to have more than 40% MAPE in the linear model, which was also applied in linear regression. Figure 14 shows the learning result model with one hidden layer node.
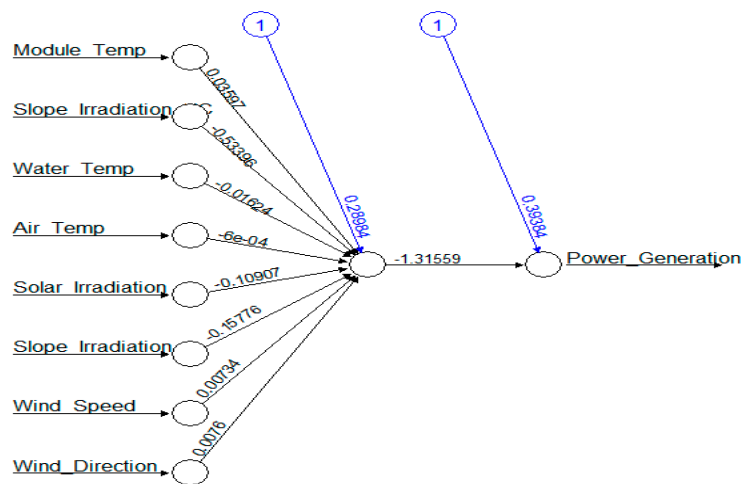
Figure 14. Neural network learning model (hidden node 1).

In Figure 15, the trained results show stable results on the graph. Although a MAPE of 10.25% occurred, this is 0.64% lower than that of the linear regression of 9.61%, but can be judged to be almost similar.
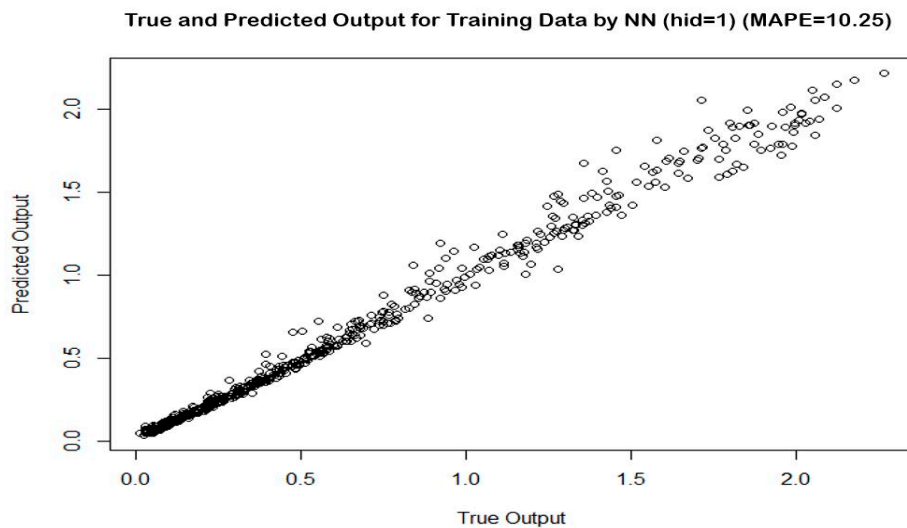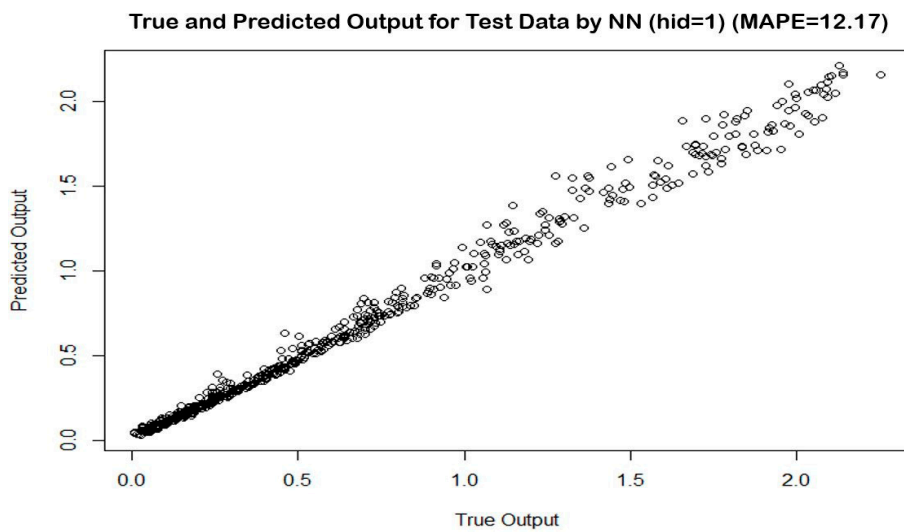

Figure 15. Neural network analysis results (training data).
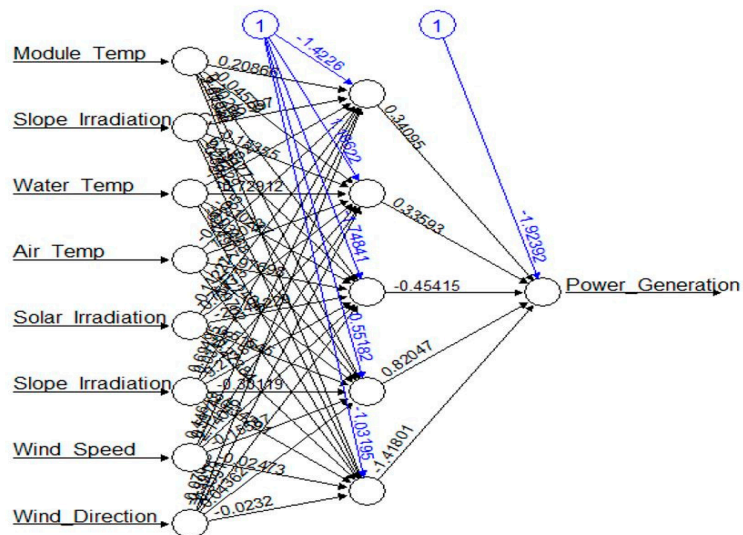
Figure 16 shows the result of verifying the accuracy by entering the test data into the trained model. The MAPE was 12.17%, which was similar to that of the learning model, without overfitting.

Figure 17 shows a model that predicts the power generation after calculating eight input variables with five hidden nodes.
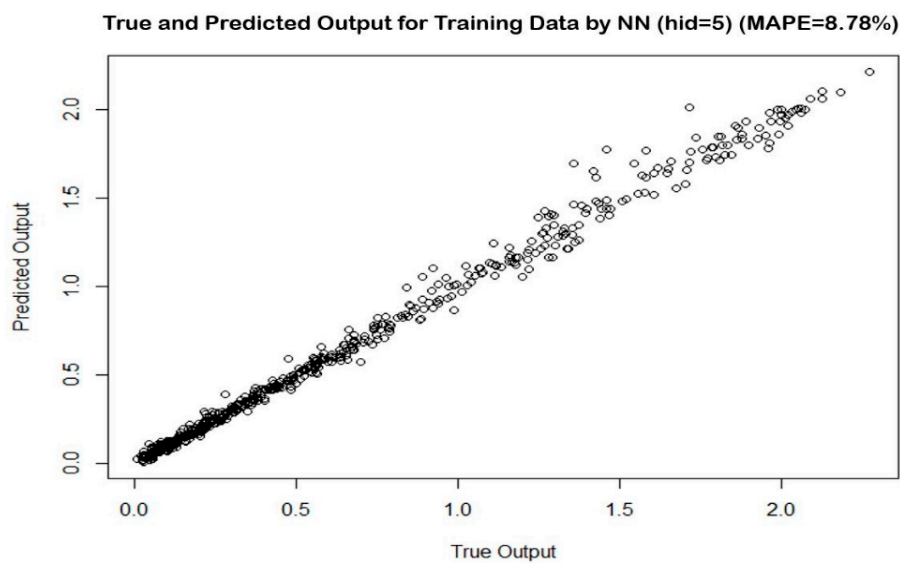
Figure 18 shows that the percentage error is reduced to 8.78% as a result of the five hidden nodes learned.

True and Predicted Output for Test Data by NN (hid=1) (MAPE=12.17)

Figure 16. Neural network analysis results (verification data).

Figure 17. Neural network learning model (hidden node 5).

True and Predicted Output for Training Data by NN (hid=5) (MAPE=8.78%)

Figure 18. Neural network analysis results (training data).

Figure 19 presents the results of verifying the accuracy of the model by entering test data into the trained model. The percentage error is 10.36% and an error increase of 1.58% occurs; however, it can be confirmed that the percentage error is lower than that of the result of the learning and verification of the linear regression analysis.
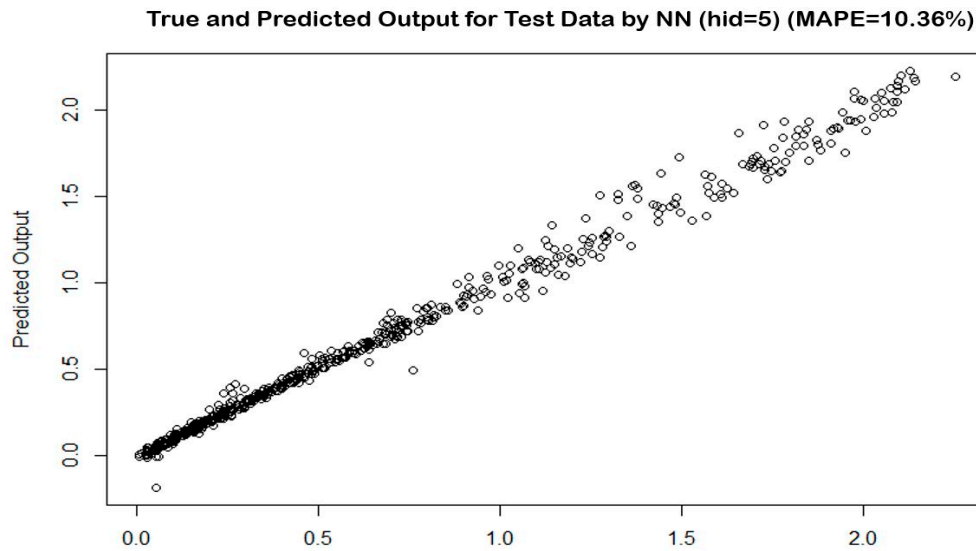

Figure 19. Neural network analysis results (Test data).

Figure 20 shows a model that predicts the power generation after calculating eight input variables with ten hidden nodes.
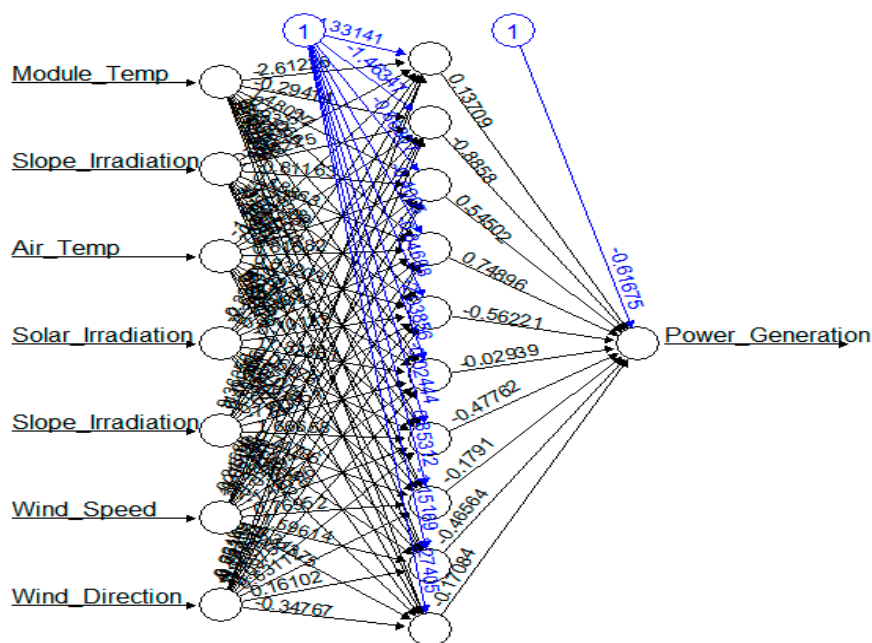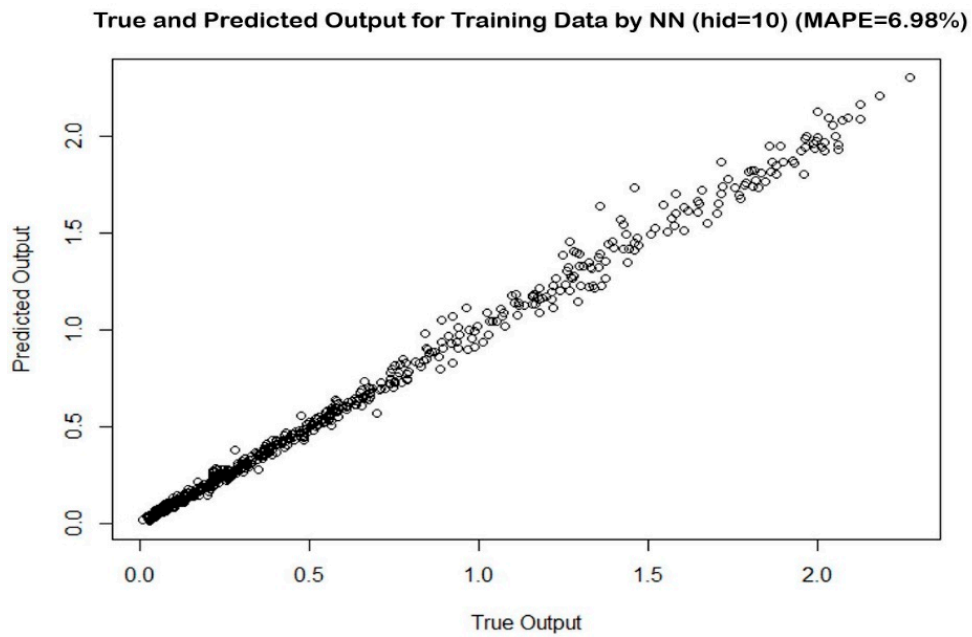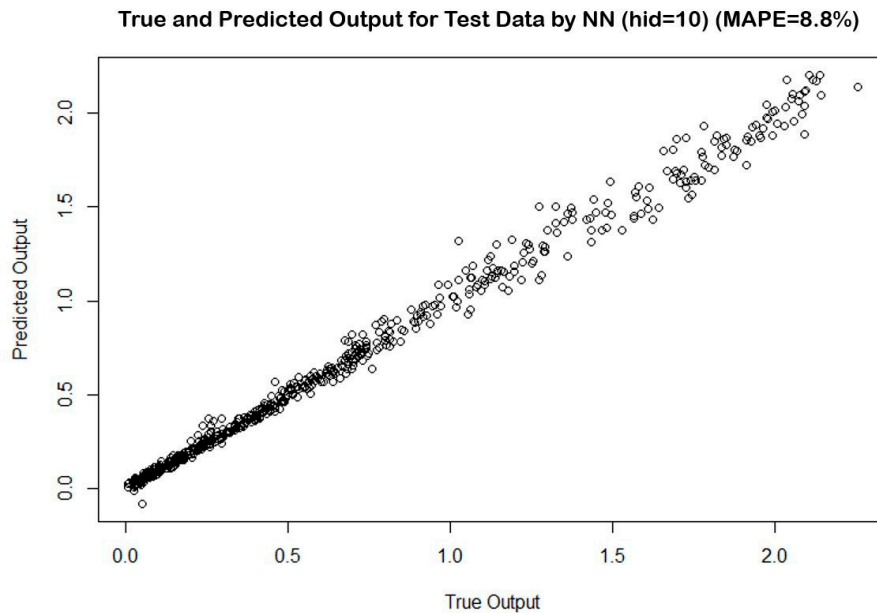

Figure 20. Neural network learning model (hidden node 10).

Figure 21 shows that the percentage error is reduced to 6.98% as a result of the ten hidden nodes learned.

True and Predicted Output for Training Data by NN (hid=10) (MAPE=6.98%)

Figure 21. Neural network analysis results (training data).

Figure 22 presents the results of verifying the accuracy of the model by entering verification data into the trained model. The percentage error is 8.8% and an error increase of 1.82% occurs; however, it can be confirmed that the percentage error is lower than that of the result of the learning and verification of the linear regression analysis and the 5 hidden nodes of the neural network.

True and Predicted Output for Test Data by NN (hid=10) (MAPE=8.8%)

Figure 22. Neural network analysis results (test data).

Figure 23 shows a model that predicts the power generation after calculating eight input variables with twenty hidden nodes.
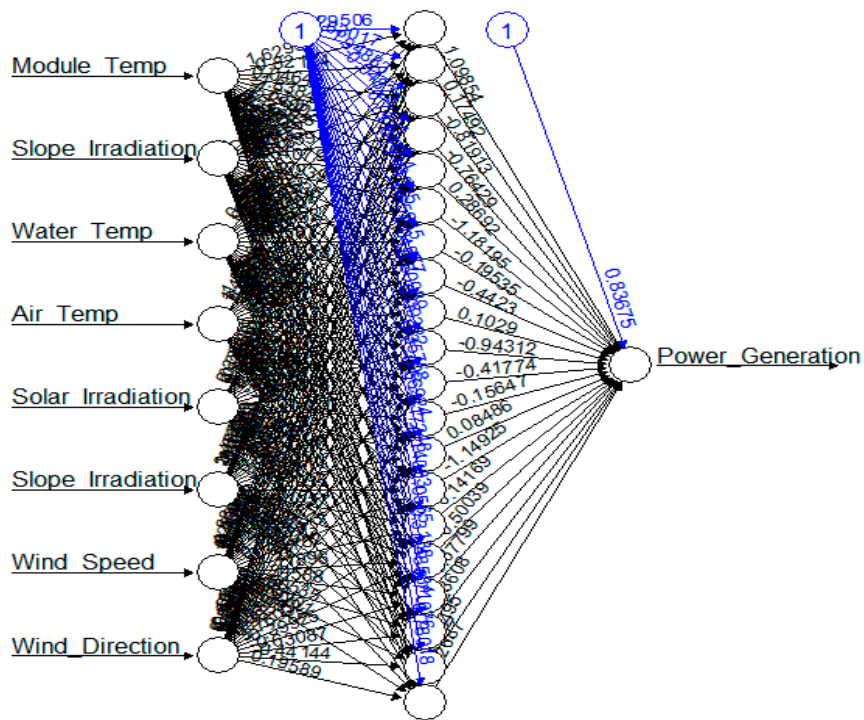
Figure 23. Neural network learning model (hidden node 20).

Figure 24 shows that the percentage error is reduced to 7.54% as a result of the five hidden nodes learned.
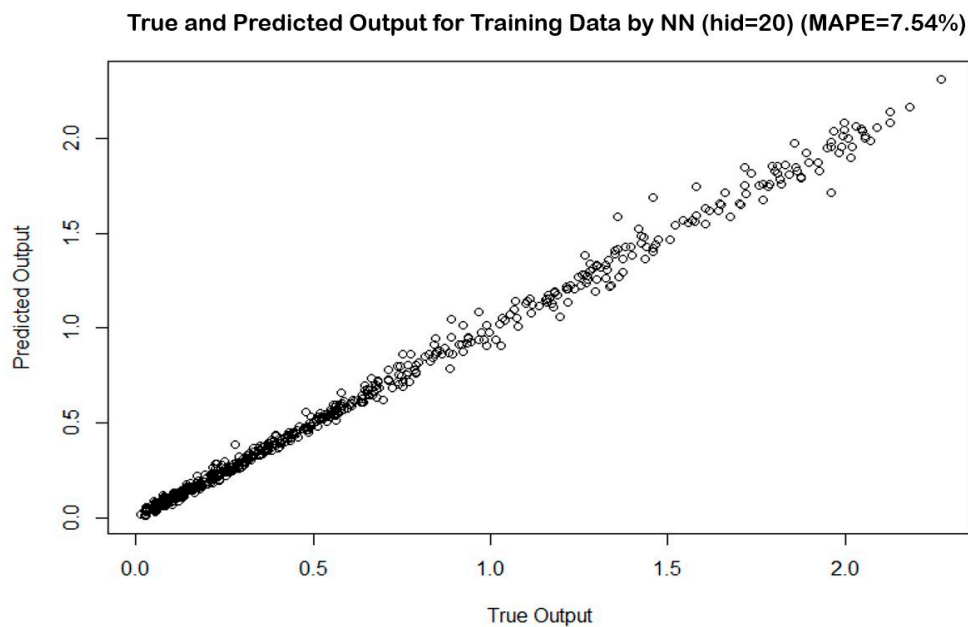

Figure 24. Neural network analysis results (training data).

Figure 25 presents the results of verifying the accuracy of the model by entering verification data into the trained model. The percentage error is 11.35% and an error increase of 3.81% occurs. It is confirmed that overfitting occurred with an error value much higher than that of the linear regression or the neural networks of the hidden nodes 1, 5, and 10.
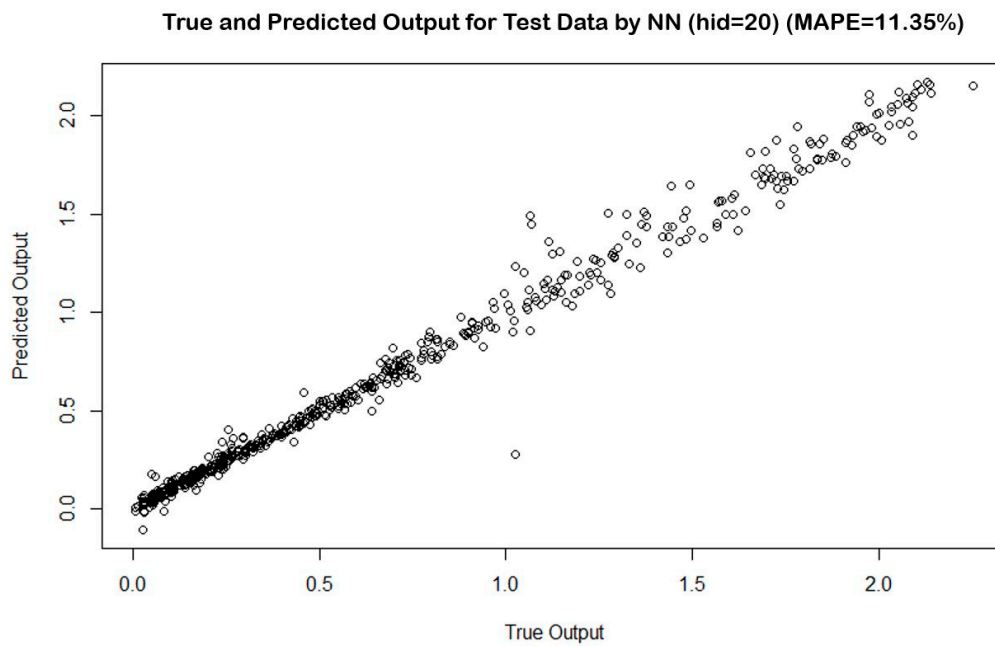
True and Predicted Output for Test Data by NN (hid=20) (MAPE=11.35%)

Figure 25. Neural network analysis results (test data).

4.4. Prediction Model Comparison

Table 6 summarizes an analysis of the linear and nonlinear algorithms reviewed to predict the amount of power generated by the floating PV system. The linear regression prediction model showed a percentage error of 9.61–11.35%, and the neural network prediction model with five hidden nodes showed an error range of 8.78–10.36%. As a result, it was confirmed that the neural network prediction model, which was a nonlinear algorithm, showed 2.59% higher accuracy than the linear regression analysis prediction model, on average. The numbers of hidden nodes were set to one to twenty in the neural network algorithm, and it was confirmed that overfitting happened in twenty nodes and ten hidden nodes produced the best accurate results among them. In conclusion, the power generation must be calculated in the non-linear model and the number of hidden nodes should be taken into consideration so as not to have overfitting, as below.

Table 6. MAPE of training and test dataset, obtained using algorithms.

MAPE	Linear Regression	Number of Hidden Nodes of Neural Network			
		1	5	10	20
Training	9.61%	10.25%	8.78%	6.98%	7.54%
Test	11.35%	12.17%	10.36%	8.8%	11.35%

5. Conclusions

In this study, the power-generation prediction model for a floating PV system using neural networks was compared and analyzed with one based on regression analysis. The summary of the characteristics of the floating solar power generation model verified in this study is as follows:

- The existing knowledge that the cooling effect and efficiency of the floating PV system increased with the evaporation of water from the surface was supplemented by more objective and concrete analyses. Based on the operation of two 2.5 kW land-based and floating PV systems installed near and on the water surface of the Boryeong Dam, Korea, respectively, the results were analyzed comparatively by applying a statistical probability method.

- During the period of analysis, the amount of solar irradiation for the land-based PV system was approximately 1.1% higher than that of the floating PV system; however, the average module temperature of the floating PV system was 4.9% lower than that of the land-based PV system.
- To analyze the reason behind the module temperature of the floating PV system being lower than that of the land-based PV system, the correlation between the module temperature difference and the air–water temperature, as well as the module temperature and wind speed, were analyzed. From the results, it was shown that there was a positive correlation between the module temperature difference and the land and water temperature difference, with a correlation coefficient of 0.6451. On the other hand, there was no correlation between the module temperature difference and the wind speed, with a correlation coefficient of 0.0032. Based on the operational data, the regression analysis method suggested a characteristic function that could predict the power generation, even considering solar radiation and module temperature.
- This study proved that the efficiency of a floating PV system was higher than that of a land-based PV system, owing to the cooling effect of water, although there are some differences, according to the environmental factors of the area. Using the water temperature characteristics, it was possible to predict the amount of electricity generated more accurately.
- In addition, the accuracy of the power generation prediction model using neural networks was approximately 2.91% higher than that of the regression analysis method. As a result of adjusting the hidden nodes in the neural network algorithm, it was confirmed that a neural network algorithm with ten hidden nodes was most suitable for calculating the amount of power generation
- This paper is the result of utilizing 2.5 kW research equipment installed in the dams and nearby areas of South Korea. Therefore, as a result of this study, it is possible to more accurately predict the power generation of the floating PV system in an environment similar to these study conditions, such as installation locations, system scales and configurations, etc.

References

1. U.S. Energy Information Administration. International Energy Outlook 2019 with Projections to 2050. Available online: https://www.eia.gov/outlooks/ieo/pdf/ieo2019.pdf (accessed on 14 April 2020).
2. International Renewable Energy Agency. Renewable Energy Market Analysis: GCC 2019. Available online: https://www.irena.org/-/media/Files/IRENA/Agency/Publication/2019/Jan/IRENA_Market_Analysis_GCC_2019.pdf (accessed on 17 April 2020).
3. Solarpowereurope. Global Market Outlook for Solar Power/2018–2022. 2018. Available online: https://www.solarpowereurope.org/global-market-outlook-2018-2022/ (accessed on 12 November 2012).
4. ArrayExpress—A Database of Functional Genomics Experiments. Available online: http://www.ebi.ac.uk/arrayexpress/ (accessed on 12 November 2012).
5. SERIS. Floating Solar Market Report. 2019. Available online: https://img.saurenergy.com/2019/06/floating-solar-market-report-executive-summary.pdf (accessed on 25 March 2020).
6. Ko, J.W.; Cha, H.L.; Kim, D.K.-S.; Lim, J.R.; Kim, G.G.; Bhang, B.G.; Won, C.S.; Jung, H.S.; Kang, D.H.; Ahn, H. Safety Analysis of Grounding Resistance with Depth of Water for Floating PVs. *Energies* **2017**, *10*, 1304. [CrossRef]
7. Choi, Y.-K.; Lee, Y.-G. A study on development of rotary structure for tracking-type floating photovoltaic system. *Int. J. Precis. Eng. Manuf.* **2014**, *15*, 2453–2460. [CrossRef]

8. Choi, Y.K.; Lee, N.H.; Kim, K.J. Empirical research on the efficiency of floating PV systems compared with overland PV systems. Advances in Information Technology and Computer Science. Available online: https://www.semanticscholar.org/paper/Empirical-Research-on-the-efficiency-of-Floating-PV-Choi-Lee/a38910b49973d099603974c24cd55f43a8b0f64c (accessed on 26 June 2020).

9. Mesbahi, M.; Minamino, S. Top 70 Floating Solar PV Plants, 2018. Available online: https://www.solarplaza.com/channels/top-10s/11761/top-70-floating-solar-PV-plants/ (accessed on 16 December 2019).

10. Tina, G.M.; Rosa-Clot, M.; Rosa-Clot, P. Electrical behavior and optimization of panels and reflector of a photovoltaic floating plant. In Proceedings of the 26th European Photovoltaic Solar Energy Conference and Exhibition, Hamburg, Germany, 5–9 September 2011; pp. 4371–4375.

11. Skoplaki, E.; Palyvos, J. On the temperature dependence of photovoltaic module electrical performance: A review of efficiency/power correlations. *Sol. Energy* **2009**, *83*, 614–624. [CrossRef]

12. Ueda, Y.; Sakurai, T.; Tatebe, S.; Itoh, A.; Kuroka, K. Performance analysis of PV system on the water. In Proceedings of the 23rd European Photovoltaic Solar Energy Conference, Valencia, Spain, 1–5 September 2008; pp. 2670–2673.

13. Choi, Y.-K. A Study on Power Generation Analysis of Floating PV System Considering Environmental Impact. *Int. J. Softw. Eng. Its Appl.* **2014**, *8*, 75–84. [CrossRef]

14. Jeong, H.; Choi, J.; Ok, Y. Characteristics of floating photovoltaic power generation based on probability statistics. In Proceedings of the 10th International Conference on Power Electronics and ECCE-Asia, Busan, Korea, 27–30 May 2019.

15. Heiat, A. Comparison of artificial neural network and regression models for estimating software development effort. *Inf. Softw. Technol.* **2002**, *44*, 911–922. [CrossRef]

16. Colli, A.; Marzoli, M.; Zaaiman, W.; Guastella, S.; Sparber, W. Comparative Performance of Various PV Technologies in Different Italian Locations. In Proceedings of the World Renewable Energy Congress, Linkoping, Sweden, 8–13 May 2011; pp. 2722–2729. [CrossRef]

17. Cohen, J.; Cohen, P.; West, S.G.; Aiken, L.S. *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed.; Lawrence Erlbaum: Newark, NJ, USA, 2003; pp. 19–31.

18. Zhong, J.; Liu, L.; Sun, Q.; Wang, X. Prediction of Photovoltaic Power Generation Based on General Regression and Back Propagation Neural Network. *Energy Procedia* **2018**, *152*, 1224–1229. [CrossRef]

19. Mindrila, D.; Balentyne, P. Scatterplots and Correlation, Khan Academy. 2014. Available online: https://www.westga.edu/academics/research/vrc/assets/docs/scatterplots_and_correlation_notes.pdf (accessed on 12 November 2012).

20. Bishop, C.M. *Pattern Recognition and Machine Learning*; Springer: Berlin/Heidelberg, Germany, 2006.

21. Nakayama, K.; Tsuji, M.; Chantana, J.; Kawano, Y.; Nishimura, T.; Hishikawa, Y.; Minemoto, T. Description of short circuit current of outdoor photovoltaic modules by multiple regression analysis under various solar irradiance levels. *Renew. Energy* **2020**, *147*, 895–902. [CrossRef]

22. Elhewy, A.H.; Mesbahi, E.; Pu, Y. Reliability analysis of structures using neural network method. *Probabilistic Eng. Mech.* **2006**, *21*, 44–53. [CrossRef]

23. Hontoria, L.; Aguilera, J.; Zufiria, P. Generation of hourly irradiation synthetic series using the neural network multilayer perceptron. *Sol. Energy* **2002**, *72*, 441–446. [CrossRef]

24. Myers, R.H. *Classical and Modern Regression with Application*, 2nd ed.; Duxbury Classic: Boston, MA, USA, 2000; pp. 82–119.

25. Kalman, R.E. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.* **1960**, *82*, 35–45. [CrossRef]

26. Frick, R.W. Accepting the null hypothesis. *Mem. Cogn.* **1995**, *23*, 132–138. [CrossRef] [PubMed]

27. Chen, S.; Gooi, H.; Wang, M. Solar radiation forecast based on fuzzy logic and neural networks. *Renew. Energy* **2013**, *60*, 195–201. [CrossRef]