 doi:10.20944/preprints202205.0386.v1

Article

Assessing the Impact of Features on Probabilistic Forecasting of Photovoltaic Power Generation

Hiroki Yamamoto[1]*, Junji Kondoh[1] and Daisuke Kodaira[2]

1 Department of Electrical Engineering, Graduate School of Science and Technology, Tokyo University of Science,2641 Yamazaki, Noda 278-8510, Chiba, Japan; j.kondoh@rs.tus.ac.jp (J.K.)
2 Faculty of Engineering, Information and Systems, University of Tsukuba, Tsukuba 305-8573, Japan; daisuke.kodaira03@gmail.com (D.K.)
* Correspondence: 7322604@ed.tus.ac.jp (H.Y.)

Abstract: Photovoltaic power generation has high variability and uncertainty because it is affected by uncertain factors such as weather conditions. Therefore, probabilistic forecasting is useful for optimal operation and risk hedging in power systems with large amounts of photovoltaic power generation. However, deterministic forecasting is the mainstay of photovoltaic generation forecasting; there are few studies on probabilistic forecasting and feature selection from weather or time-oriented features in such forecasting. In this study, prediction intervals were generated by the lower upper bound estimation using neural networks with two outputs to make probabilistic predictions. The objective was to improve prediction interval coverage probability (PICP), mean prediction interval width (MPIW), and loss, which is the integration of these two metrics, by removing unnecessary features through feature selection. When features with high gain were selected by random forests (RF), in the forecast of 14.7-kW PV systems, loss improved by 1.57 kW, PICP by 0.057, and MPIW by 0.12 kW on average over two weeks compared to the case where all features were used without feature selection. Therefore, the low gain features from RF act as noise in LUBE and reduce the prediction accuracy.

Keywords: Lower upper bound estimation; random forest; feature selection; probabilistic forecasting; photovoltaic generation forecasting

1. Introduction

Renewable energy sources, including photovoltaic (PV) generation, are being developed in many countries as the need for clean energy increases [1]. In particular, the number of PV installations has increased significantly in recent years due to the low cost of the modules, no carbon dioxide emissions, and the ease of installing the panels. However, PV power generation is highly variable and uncertain as it is affected by weather conditions and other uncertain factors. This variability and uncertainty significantly impact the operation of power systems in which large amounts of PV power generation is installed. As a countermeasure to this problem, it is thought that the impact on the power systems can be mitigated by conducting highly accurate PV power output forecasting, which is then considered in the operational plans of thermal power, hydroelectric power, and other power sources whose output can be adjusted.

PV forecasting models can be divided into three main categories: physical models, statistical models, and hybrid models [2]. Physical models are constructed using numerical weather prediction (NWP) and satellite imagery; Miyazaki et al. [3] used optical flow to estimate the geographical motion of PV output lump related to the cloud motion. Saint-Drenan et al. [4] probabilistically predicted PV output from reference PV output derived from NWP and meteorological data. The statistical model, including machine learning, is a data-driven forecasting model, which constructs a forecast model based on historical data. Compared to physical models, statistical models rely solely on measured data and require no prior knowledge, making them easy to implement and highly adaptable.

 doi:10.20944/preprints202205.0386.v1

Therefore, research on predicting PV power output using statistical models, mainly machine learning, is becoming increasingly popular. In PV forecasting, autoregressive moving average methods, represented by time series models [5], and machine learning-based methods such as neural network (NN) [6], support vector regression (SVR) [7,8] , and long short-term memory (LSTM) [9], are used. Hybrid models are methods that improve prediction accuracy by combining statistical models or combining physical and statistical models to complement information that cannot be compensated by single models. Kodaira et al. [10] constructed an ensemble model of NN, k-means, and LSTM and used particle swarm optimization to optimize the weights of each predictive model to predict PV output. Yang et al. [11] classified training data using self-organizing map and learning vector quantization, trained them using SVR, and used phage inference to select the model to use on the forecast date to predict PV output.

Machine learning predictions can be broadly classified into two types: predictions based on deterministic point estimates and predictions based on interval estimates that account for uncertainty in a probabilistic manner. Compared to deterministic forecasting, probabilistic forecasting has more information because it accounts for the error. Therefore, probabilistic forecasts are useful for optimal operation and risk hedging in grids with significant PV generation. However, in conventional PV forecasting, forecasts are mainly based on deterministic point estimates. In probabilistic forecasting, Gaussian process [12] and quantile regression [13], quantile regression forests[14] and others have been used to forecast PV power generation. Recently, lower upper bound estimation (LUBE) has attracted attention as a more direct method of outputting prediction intervals (PIs) without any special assumptions on the distribution [15]. LUBE has been applied in various fields of engineering. Khosravi et al. [16] used the LUBE to predict wind power generation, and Quan et al. [17] uses LUBE to forecast electrical loads. Ni et al. [18] used LUBE to forecast PV generation.

In machine learning, the features used in the prediction model have a direct impact on prediction accuracy. Incorporating unimportant features into the forecast model can lead to complex computational processing and reduced forecast accuracy due to learning unnecessary noise [19]. In general, weather variables such as temperature and solar radiation are used as features in PV forecasting. De Giorgi et al. [20] evaluated forecasting accuracy using NN for (1) a data set consisting of power generation only, (2) a data set consisting of power generation and solar radiation, and (3) a data set consisting of (2) plus module temperature and ambient temperature, and concluded that (3) was superior based on NRMSE, standard deviation, and other evaluation indicators. Zhong et al. [21] evaluated the prediction accuracy of NN, LSTM, etc. in one-hour-ahead forecasting assuming 11 sets of feature patterns from variables such as solar radiation, precipitation, temperature, water vapor mixing ratio, sea surface temperature, wind speed, etc. In NN, all meteorological variables plus accumulated solar radiation were superior. In LSTM, forecasts using features with added variables showed signs of over-learning, and forecasts that included only solar radiation and total precipitation as features were superior.

As mentioned above, in PV forecasting, there have been some studies on feature selection for deterministic forecasts in machine learning, but there are few studies on feature selection for probabilistic forecasts. In this study, we used LUBE as a probabilistic forecast to generate a day-ahead PIs for a PV power system. The PV power system is in Tokyo, Japan, and is rated at 14.7 kW. Data of 14 features, including weather variables, were measured at a nearby weather station. The training data are approximately 10 months old, and the forecast period is 2 weeks. The objective of this study is to improve prediction interval coverage probability (PICP), mean prediction interval width (MPIW), and Loss [22] , which integrates and simultaneously evaluates these two metrics, by removing unnecessary features through feature selection. The contributions of this study are as follows:

1) The effects of the 14 variables were evaluated by random forest. Features were selected according to their gains, and PIs were generated by LUBE.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

2) Features with high gain, mainly contribute to improve the accuracy of PIs. But features with low gain rarely contribute or worsen the accuracy of PIs. In particular, low gain features reduced prediction accuracy when PV output fluctuations were large.

The remainder of the paper is organized as follows: Section 2 describes feature selection with RF. Section 3 describes LUBE. Section 4 evaluates the accuracy of PIs by LUBE based on feature selection by RF. Section 5 concludes the study.

2. Feature selection using RF

Because PV power output fluctuates due to weather factors, forecast accuracy varies greatly depending on the features used in the forecast model. If features with additional variables that are effective in forecasting are used, the forecast accuracy improves; otherwise, the forecast accuracy deteriorates. It also increases the complexity of the model and computational processing. In this study, feature selection was performed by RF, which is an ensemble model that is obtained by modeling a forecasting model that repeats binary classification on a tree called a "decision tree" for each of multiple samples generated using the bootstrap method, and averaging the forecasting accuracy [23]. RF has been proved useful in forecasting renewable energy sources such as solar power and wind power, which are sensitive to environmental factors [24,25]. The conceptual diagram of feature selection by RF is shown in Figure 1 and the following procedure is used [26,27].

(i). From the training data consisting of n sets of P predictors and corresponding target variables, n sets are extracted, allowing for overlap. The extraction is repeated to generate K bootstrap samples. When generating bootstrap samples, approximately two-thirds of the training data are extracted at least once from the sample and about one-third are never extracted. The group of samples that are not extracted is referred to as out-of-bag (OOB).

(ii). A decision tree is modeled for each of the K bootstrap samples and the mean square error (MSE) is obtained using OOB as test data. $\mathrm{MSE}^{\mathrm{OOB_t}}$ represents the MSE when $\mathrm{OOB_t}$ is the test data. Y is the target variable and the hat symbol represent the predicted value.

$$MSE^{\mathrm{OOB_t}} = \frac{1}{n}\sum_{i=1}^{n}\left| Y_i - \widehat{Y_{i,t}}\right|^2 \quad (1)$$

(iii). One arbitrary feature from the OOB features is selected, randomly permuted, and the MSE is obtained again. This is repeated for all features.

$$MSE^{\mathrm{OOB_t}}\left(X_j permuted\right) = \frac{1}{n}\sum_{i=1}^{n}\left| Y_i - \widehat{Y_{i,t}}\left(X_j permuted\right)\right|^2 \quad (2)$$

where $\mathrm{X_j}$ denotes the feature and $j = 1,2,\ldots,\mathrm{P}$, $\mathrm{MSE}^{\mathrm{OOB_t}}(X_j permuted)$ denotes the MSE at $\mathrm{OOB_t}$ when the feature $\mathrm{X_j}$ is permuted.

(iv). The changes in MSE before and after permuting is obtained and the K results are averaged. Normalization is applied so that the sum of the importance of each feature is 1. If a feature is important to the accuracy of the forecast, permuting will significantly reduce the accuracy of the forecast. For unimportant features, the accuracy of the prediction is not affected.

$$MSEdif\left(X_j\right) = \frac{1}{K}\sum_{t=1}^{K}\left(MSE^{\mathrm{OOB_t}}\left(X_j permuted\right) - MSE^{\mathrm{OOB_t}}\right) \quad (3)$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

$$gainX_j = \frac{MSEdif(X_j)}{\sum_{j=1}^{P} MSEdif(X_j)} \tag{4}$$

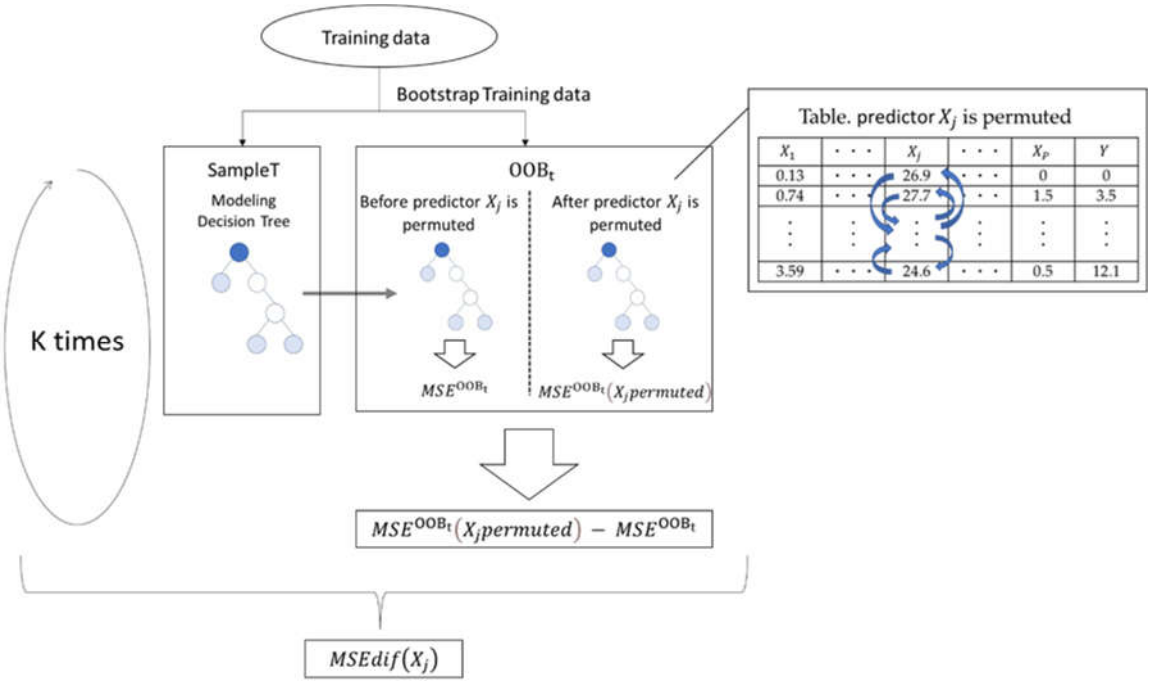

Figure 1. Feature selection by RF.

3. LUBE

LUBE is a nonparametric method that directly generates PIs using NNs with two outputs corresponding to the upper and lower bounds of the PIs [15]. Traditional methods such as the Delta and Bayesian methods are parametric methods that first perform point estimation and then generate PIs by assuming a distribution over the data [28,29]. LUBE directly generates PIs, which is simple, fast, and without special assumptions about the distribution or a large amount of computational work [17]. In LUBE, PIs are evaluated using the following three indicators

3.1. PICP

PICP is an index that evaluates the percentage of measured values that fall within the interval of PIs and is one of the important evaluation indicators. The predicted lower and upper PI bounds are $\hat{y}_{L_i}$, $\hat{y}_{U_i}$. A vector, $\boldsymbol{k}$, of length n represents whether each data point has been captured by the estimated PIs.

$$k_i = \begin{cases} 1, & if\ \hat{y}_{L_i} \leqq y_i \leqq \hat{y}_{U_i} \\ 0, & else \end{cases} \tag{5}$$

We define the total number of data points captured as c .

$$c \coloneqq \sum_{i=1}^{n} k_i \tag{6}$$

PICP is defined by equation (7).

$$PICP \coloneqq \frac{c}{n} \tag{7}$$

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

3.2. MPIW

MPIW represents the mean interval width and is an important metric for evaluating PIs; even if all measured values are within the interval and PICP is 100%, a too wide MPIW implies high uncertainty and is meaningless as a forecast.

$$MPIW := \frac{1}{c}\sum_{i=1}^{n}\left(\hat{y}_{U_i} - \hat{y}_{L_i}\right) \cdot k_i \tag{8}$$

3.3. Loss

The two key metrics, PICP and MPIW, need to be evaluated simultaneously when generating PIs using LUBE. However, if the width of PIs is narrowed, PICP is likely to decrease because some measured values will drop out of the PIs. In LUBE, the two trade-off indicators are evaluated simultaneously using Loss [22].

$$Loss = MPIW + \lambda\frac{n}{\alpha(1-\alpha)}max(0,(1-\alpha) - PICP)^2 \tag{9}$$

where n is the number of data samples and α is a measure of the confidence level of the PIs. For example, when α = 0.05, 1-α is 95%. At this time, the qualitative understanding in the equation is that if PICP does not exceed 0.95, a penalty is imposed by λ. λ is the parameter that imposes a penalty and, at the same time, a tuning parameter that relates MPIW to PICP. If it exceeds 0.95, Loss is equal to MPIW. In PIs, when PICP is much lower than the established confidence level, the PIs can be considered lacking validity [14]. Therefore, in determining λ, there is need to adjust PICP and then consider the combination with MPIW.

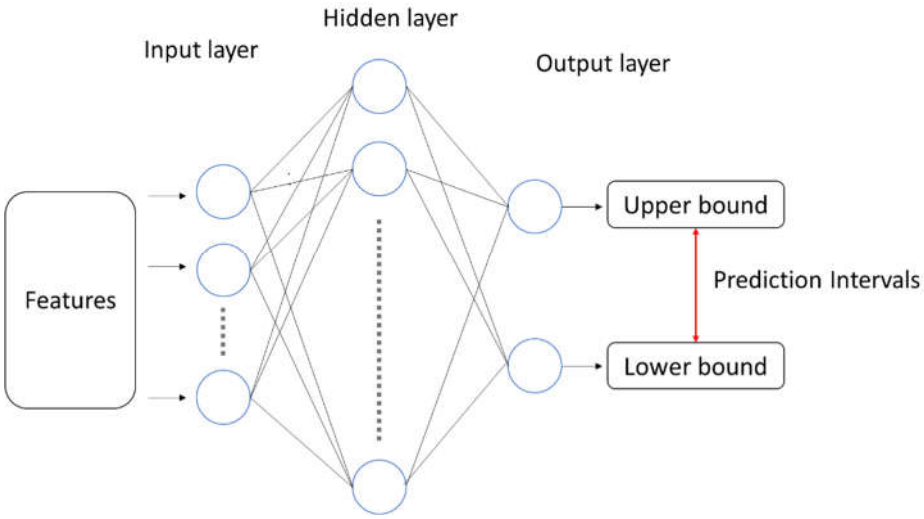

Figure 2. LUBE.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

4. Case study

For probabilistic forecasting of solar power generation using LUBE, excluding unnecessary features improves forecasting accuracy, but missing necessary features reduces forecasting accuracy. In particular, PV forecasting is difficult when output fluctuation is large, and if unnecessary features are included, the forecasting accuracy is likely to be even lower. In this case study, we evaluated the importance of each feature for the probabilistic forecasting. We first used RF to evaluate the gain of 14 features, consisting of weather and time-oriented features, for forecasting PV power output. Next, PIs were generated using LUBE with a confidence level of 95%. In LUBE, we used 14 pairs of features added in order from variables with the highest gain. The 2-week average prediction accuracy of the PIs from each feature was evaluated, and the optimal features were considered. We also evaluated the robustness of the predictions by setting the number of simulations to 85. Finally, in relation to output fluctuation, we considered each day to evaluate when predictions were successful and when they failed, with and without feature selection.

The PV power plant subject of this study is in Tokyo, Japan, and generates 14.7 kW. PV power generation is observed every 30 minutes for 24 hours. In addition to power generation data, temperature, precipitation, cloud cover, solar radiation, wind speed, and humidity were obtained from nearby meteorological observatories. The corresponding year, month, day, and hour were also added to the data set. The time data were expressed as a trigonometric function to account for periodicity. For example, the trigonometric representation of hour considers 24 hours as one cycle, and daily considers the number of days in a month as one cycle. All features were standardized for uniformity of scale. The observed data is for the period from August 15, 2013, to June 14, 2014. The feature selection by RF used data from August 15, 2013, to May 31, 2014. Training data for PIs generation by the LUBE used data from August 15, 2013, to May 16, 2014. Validation data was from May 17, 2014, to May 31, 2014, and test data was from June 1, 2014, to June 14, 2014. Add data was from June 1, 2014, to the day before the forecast day. Adding the most recent data to the training and validation data was expected to improve forecast accuracy.

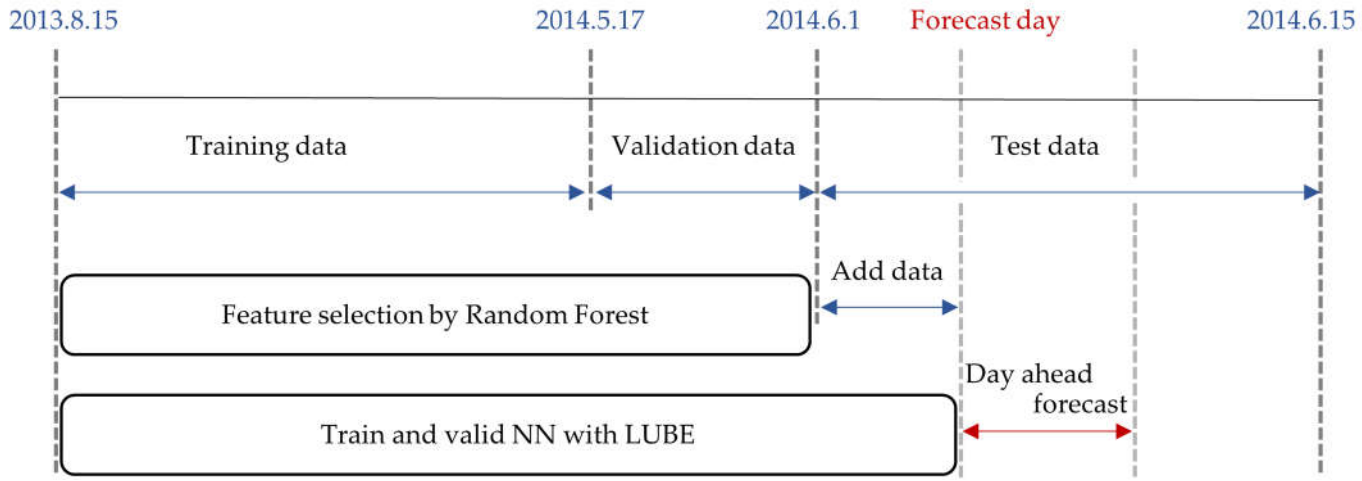

Figure 3. Lower Upper Bound Estimation Method with RF

4.1 random forest

Table 1 shows the results of feature selection using RF. The gain for solar radiation is 0.764, indicating that it is an extremely important feature. The hour sine and hour cosine have a combined gain of 0.172, and the annual cosine has a gain of 0.013. The gains for the other features are almost zero, indicating that they are of low importance in RF's predictions.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

Table 1. Evaluating the importance of features using RF

feature	gain	feature	gain
solar radiation	0.7644	atmospheric temperature	0.0065
hour sine	0.0961	monthly cosine	0.0062
hour cosine	0.0757	wind speed	0.0057
annual cosine	0.0130	daily sine	0.0038
degree of cloudiness	0.0076	daily cosine	0.0037
annual sine	0.0069	monthly sine	0.0025
humidity	0.0069	precipitation	0.0003

4.2. Prediction Intervals by LUBE

4.2.1. Evaluation of Prediction Intervals at 2-week average

Figure 4 shows the correspondence between the number of features and the two-week average of Loss and Total gain. Features (1) includes only the solar radiation which is the highest gain. Features (4) includes solar radiation, hour sine, hour cosine, and annual cosine. Features (14) includes all features. Table 2 shows the statistics of Loss for the 2-week average of Features (1), (4), and (14). Figure 5 shows a histogram of the results of 85 simulations of Loss for the 2-week average of Features (1), (4), and (14). To compare Loss between different features, we fixed α and λ used for LUBE in equation (9) to 0.05 and 5.0.

Figure 4 shows that features with a gain of nearly zero act as noise, reducing the accuracy and robustness of the forecast. Features (4) has the smallest Loss, increases thereafter, and at Features (14) has the largest. The distribution of Loss in the box-and-whisker diagram is expanding from Features (4) to Features (14). To evaluate the prediction accuracy and robustness, we compared Features (14) with Features (1), which includes only solar radiation, and Features (4), which has the lowest Loss. In Table 2, Features (14) is significantly worse than Features (1) and (4) in all statistics. At the median, Loss is 1.38 kW greater than Features (1) and 1.57 kW greater than Features (4). In addition, as shown by the standard deviations in Table 2 and histograms in Figure 5, the distribution of Features (4) is wide, and the prediction accuracy is varied. These results indicate that features that are low in gain should be removed because they reduce not only prediction accuracy but also robustness.

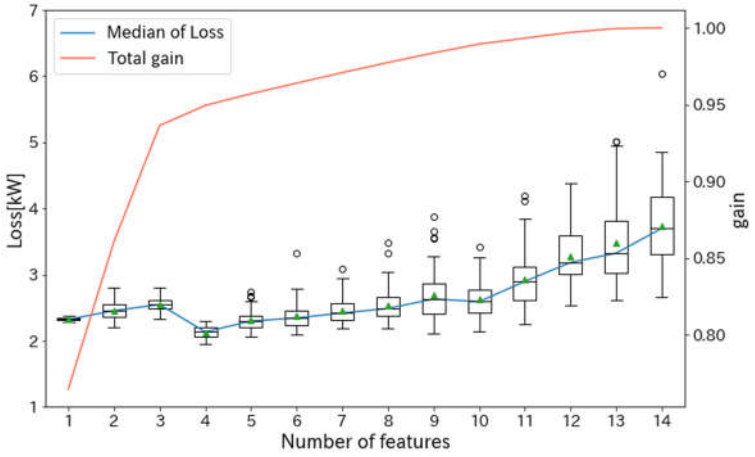

Figure 4. Number of features, 2-week average Loss, and total gain.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

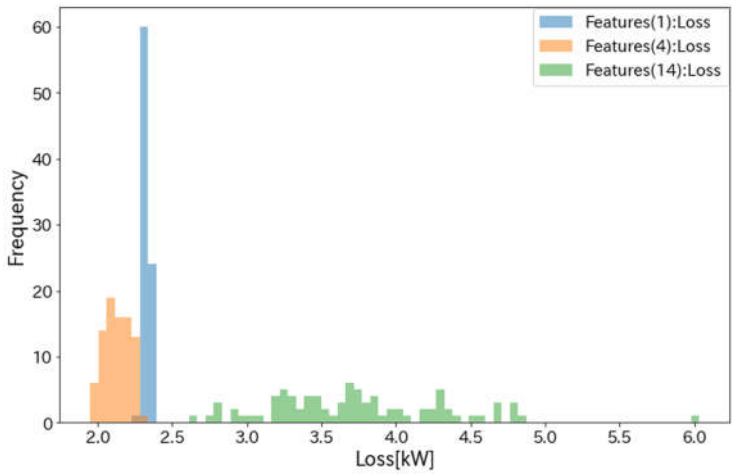

Figure 5. Histograms of 2-week average Loss for Features (1), (4), and (14).

Table 2. Statistics of 2-week average Loss.

Number of features	Min	Max	Mean	Median	Standard deviation
(1)	2.28	2.37	2.32	2.32	0.0193
(4)	1.95	2.30	2.13	2.13	0.0861
(14)	2.65	6.03	3.74	3.70	0.5967

We then similarly evaluated PICP and MPIW. Figure 6 and Figure 7 show the 2-week average PICP and MPIW, respectively; Table 3 and Table 4 show the PICP and MPIW statistics for Features (1), (4), and (14), respectively; Figure8 and Figure 9 show the histograms for each.

Figure 6 and Figure 7 show that the features after Features (4) work as noise, because the PICP tends to decrease and MPIW tends to increase after Features (4). In Table 3, comparing median values, PICP is 0.072 and 0.057 higher for Features (1), and (4), respectively, compared to Features (14). In Table 4, MPIW is 0.12 kW narrower for Features (4) than for Features (14). However, in Features (1), MPIW is 0.42 kW wider than in Features (14). This means that Features (1) is highly uncertain because it contains only one feature. Feature (4) is the narrowest MPIW because it contains enough features with high gain. Therefore, Features (4) outperforms Features (14) in MPIW and PICP as well. From the standard deviations in Table 3 and Table 4 and the histograms in Figure8 and Figure 9, PICP and MPIW, as well as Loss, features with low gain reduce the robustness of the prediction. Therefore, we can conclude that in PICP and MPIW, as in Loss, the features after Features (4) are noise.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

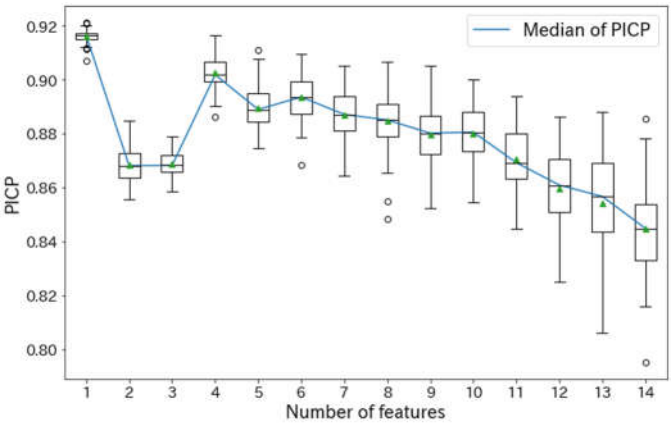

Figure 6. Number of features and 2-week average PICP

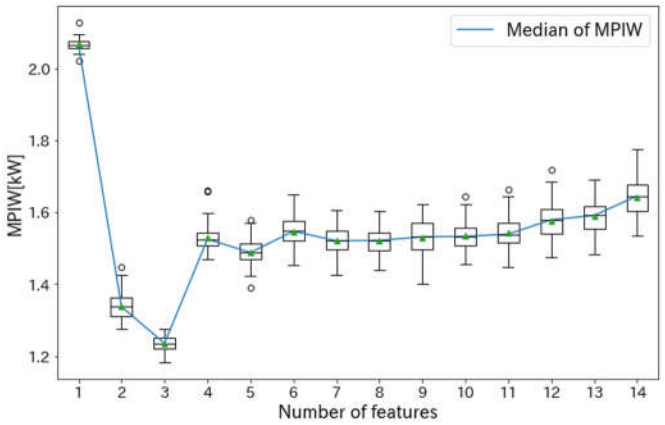

Figure 7. Number of features and 2-week average MPIW

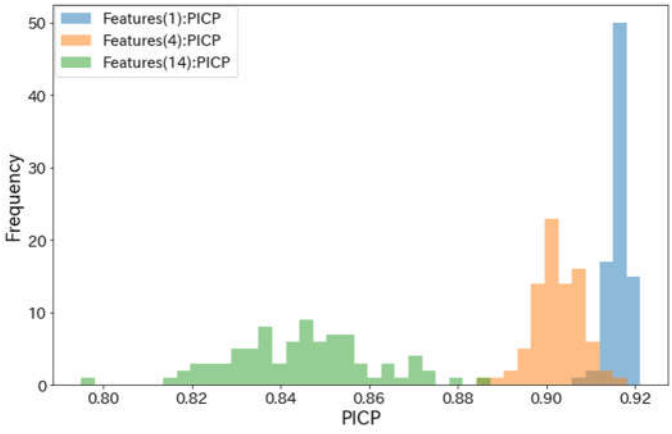

Figure8.Histograms of 2-week average PICP for Features (1), (4), and (14).

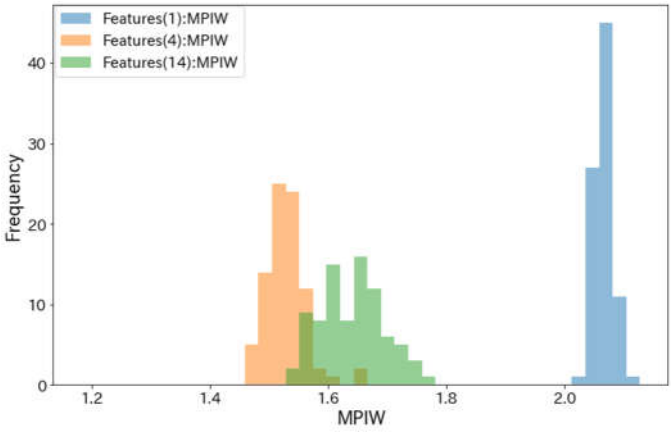

Figure 9. Histograms of 2-week average MPIW for Features (1), (4), and (14).

Table 3. Statistics of 2-week average PICP

Number of features	Min	Max.	Mean	Median	Standard deviation
(1)	0.907	0.921	0.916	0.917	0.00231
(4)	0.886	0.916	0.903	0.902	0.00537
(14)	0.795	0.886	0.845	0.845	0.0160

Table 4. Statistics of 2-week average MPIW

Number of features	Min	Max.	Mean	Median	Standard deviation
(1)	2.02	2.13	2.07	2.07	0.0151
(4)	1.47	1.66	1.53	1.53	0.0340
(14)	1.53	1.78	1.64	1.65	0.0536

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

4.2.2. Accuracy evaluation by forecast target date

To evaluate the relationship between output fluctuation and prediction accuracy according to features, we analyze Features (14) which includes all features, Features (1), and Features (4) by day, as shown in Table 5. In Table 5, out of 85 simulations, the median of Loss and corresponding MPIW and PICP for each day with each Features are shown. In Table 5, we evaluated the prediction accuracy using an index named Fluctuation, which is the average of output fluctuations at 30-minute intervals over a day. In this study, Fluctuation is defined as indicated in Equation (10).

$$Fluctuation = \frac{\sum_{i=1}^{48} |v(t_i) - v(t_{i-1})|}{48} \quad (10)$$

where $v(t_i)$ represents the power output at time t_i. Since output power is observed at 30-minute intervals, there are 48 i in a day. For example, when Fluctuation is 0 kW, the daily generation output is constant. In Table 5, the threshold for Fluctuation is 0.5 kW, which is the average of the output fluctuations for all days, and days when the output fluctuations exceed 0.5 kW are highlighted in gray. On each day, the best PICP, MPIW, and Loss are bolded among the Features (1), (4), (14). For example, on June 2, Fluctuation is 0.582 kW, Features (1) has the highest PICP of 0.896 and Features (14) has the narrowest MPIW of 1.40 kW. Features (4) has the lowest Loss at 2.30 kW. Table 6 shows the correlation coefficients between output fluctuations at 30-minute intervals and PICP, MPIW, and Loss for Features (1), (4), (14).

From Average (Fluctuation>0.5kW) and Average (Fluctuation<0.5kW) in Table 5, and strong correlation between Loss and 30-minute output fluctuation for all Features in Table 6, it is easy to predict when the output fluctuation is small and difficult to predict when it is large. When Fluctuation does not exceed 0.5 kW in Table 5, Features (4) is the best for prediction. From Average (Fluctuation<0.5kW) in Table 5, PICP, MPIW, Loss over 7 days is best for Features (4), and Features (1) is second. When Fluctuation exceeds 0.5 kW, Features (1) or Features (4) is suitable for forecasting. As indicated by the strong correlation between MPIW and output fluctuation only for Features (1) in Table 6, Features (1) has a wide MPIW and high uncertainty on days when Fluctuation is large. However, PIs in Features (1) include many real values and have the largest average PICP in Table 5. In Table 5 Average (Fluctuation> 0.5kW), Loss is smallest for Features (4), but there are several days when it is more than 0.1 below the confidence level; if the PICP is much below the confidence level, the PIs are not considered valid, so Features (1) or (4) is appropriate for forecasting. Throughout the 14 days, Loss is the smallest for Features (1) on 4 days and the smallest for Features (4) on 10 days. For Features (14), Loss is larger than Features (4) on all days, especially on days with large Fluctuation. These results show that the low gain feature by RF acts as noise and reduces the prediction accuracy, especially on days with large Fluctuations.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

Table 5. Evaluation for selected features 1, 4, and 14 from June 1 to 14.

		PICP			MPIW (kW)			Loss (kW)		
Day	Fluctuation(kW)	(1)	(4)	(14)	(1)	(4)	(14)	(1)	(4)	(14)
1	0.489	0.958	**0.979**	**0.979**	2.45	**1.55**	2.02	2.45	**1.55**	2.02
2	0.582	**0.896**	0.864	0.791	2.50	1.52	**1.40**	2.81	**2.30**	4.05
3	0.505	**0.916**	0.875	0.831	2.50	1.61	**1.58**	2.62	**2.21**	3.07
4	0.522	**0.895**	0.853	0.834	2.41	**1.49**	1.59	2.73	**2.49**	3.00
5	0.350	**0.979**	**0.979**	**0.979**	1.74	**1.66**	2.03	1.74	**1.66**	2.03
6	0.193	0.916	**0.984**	0.911	**0.85**	1.44	1.61	**0.97**	1.44	1.77
7	0.233	0.937	0.984	**1.000**	**1.25**	1.46	1.92	**1.27**	1.46	1.92
8	0.412	0.894	**0.917**	0.875	1.64	1.65	**1.50**	1.96	**1.76**	2.09
9	0.968	**0.874**	0.833	0.765	2.73	1.63	**1.47**	3.33	**3.08**	5.07
10	0.909	0.835	**0.849**	0.815	2.53	**1.63**	1.66	3.93	**2.70**	3.57
11	0.308	**0.957**	0.879	0.876	1.43	**1.10**	1.44	**1.43**	1.64	2.01
12	0.271	0.937	0.956	**0.958**	1.92	**1.58**	1.76	1.93	**1.58**	1.76
13	0.807	**0.894**	0.808	0.739	2.56	**1.36**	1.53	**2.89**	3.50	6.22
14	0.513	0.937	**0.974**	0.937	2.43	**1.60**	1.68	2.45	**1.60**	1.69
Average (Fluctuation>0.5kW)	0.687	**0.892**	0.865	0.816	2.52	**1.55**	1.56	2.97	**2.55**	3.81
Average (Fluctuation<0.5kW)	0.322	0.940	**0.954**	0.940	1.61	**1.49**	1.75	1.68	**1.58**	1.94
Average (All days)	0.504	**0.916**	0.910	0.878	2.07	**1.52**	1.66	2.32	**2.07**	2.88

Table 6. Correlation coefficient with 30-minute output fluctuation for Features (1), (4), (14).

	PICP			MPIW			Loss		
Number of features	(1)	(4)	(14)	(1)	(4)	(14)	(1)	(4)	(14)
correlation coefficient	-0.726	-0.758	-0.762	0.818	0.258	-0.355	0.932	0.879	0.821

4.2.3. Days of maximum and minimum output fluctuation

Figure 10 (a) shows the PIs for Features (1), (4), and (14) for the day with the lowest Fluctuation of 0.193kW in Table 5. Feature (1) is best for June 6. PICP is the largest at 0.984 for Features (4), and PICP exceeds 0.9 for Features (1) and (14). MPIW and Loss are the smallest in Features (1), and Loss is 0.80 kW smaller than Features (14). Therefore, on June 6, a forecast using Features (1) was appropriate.

Figure 10 (b) shows the Prediction Intervals for Features (1), (4), and (14) for the day with the largest Fluctuation of 0.968kW in Table 5. Feature (1) is the best for June 9 as well.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 30 May 2022 doi:10.20944/preprints202205.0386.v1

MPIW is the smallest for Features (14), but many real values are missing and PICP is the smallest at 0.765. Features (1) has high uncertainty and MPIW is the widest, resulting in many measured values being included in PIs, with the largest PICP of 0.874. Loss is the smallest in Features (4), about 0.25 kW less than Features (1) and 1.99 kW less than Features (14). However, PICP of Features (4) is 0.833, which is more than 0.1 below the confidence level. Therefore, the prediction in Features (1) is suitable.

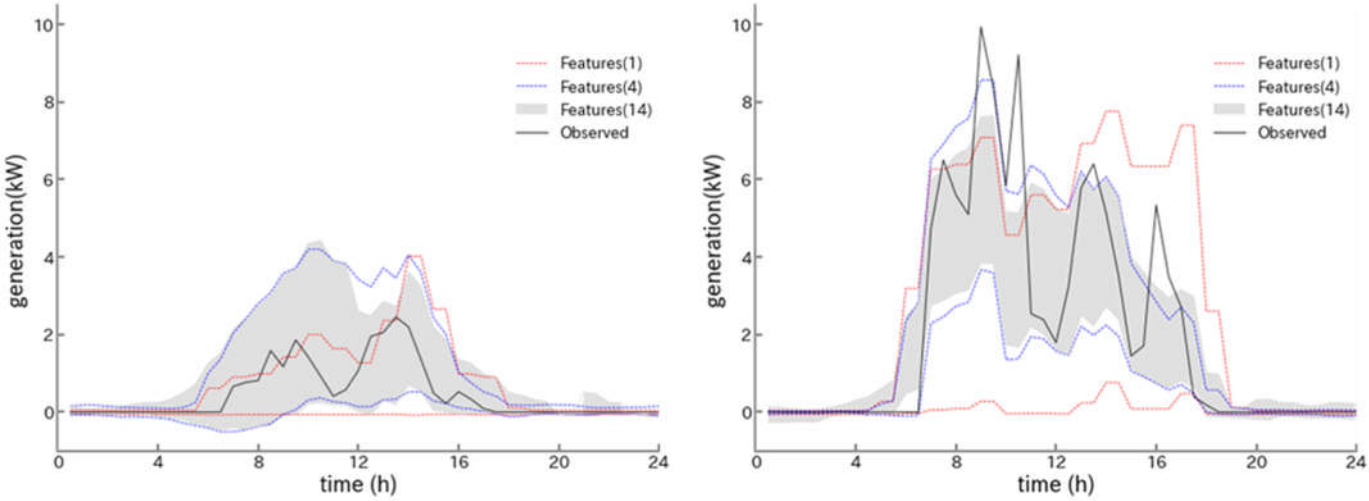

Figure 10. Prediction intervals for Features (1), (4), and (14): (**a**) June 6, the day with the smallest output fluctuation in two weeks.; (**b**) June 9, the day of the largest output fluctuation in two weeks.

5. Conclusions

The objective of this study was to improve Loss, PICP, and MPIW of PIs using LUBE by removing unnecessary features. We considered the change in prediction accuracy by incorporating, in order, the features with high gain evaluated by RF into the features used for LUBE. In number of features 1, which includes only solar radiation evaluated as the most important feature, on days with large output fluctuations, there was high uncertainty and wide MPIW, but high PICP. For number of features 14 without feature selection, features with gain nearly zero for both Loss, PICP, and MPIW became noise, resulting in lower prediction accuracy. In particular, low gain features reduced prediction accuracy when output fluctuations were large. Number of features 4, which includes solar radiation, hour sine, hour cosine, and annual cosine, includes enough features to improve Loss by 1.57 kW, PICP by 0.057, and MPIW by 0.12 kW, on average, over 2 weeks compared to number of features 14 in the output power forecast on a 14.7-kW PV system. However, for any number of features, days with large output fluctuations are found to be more than 0.1 below the confidence level. Therefore, the challenge remains to generate PIs with PICP close to the confidence level even on days with strong fluctuations.

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

Funding: This work was partly supported by JSPS KAKENHI, Grant Number JP21K14150.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors thank S. Ogata and M. Ohashi at OMRON Social Solutions Corporation for offering the output power data of the 14.7-kW PV system.

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

References

1. Ahmed, R.; Sreeram, V.; Mishra, Y.; Arif, M.D. A Review and Evaluation of the State-of-the-Art in PV Solar Power Forecasting: Techniques and Optimization. *Renew. Sustain. Energy Rev.* 2020, *124*, 109792, doi:10.1016/j.rser.2020.109792.
2. Dolara, A.; Leva, S.; Manzolini, G. Comparison of Different Physical Models for PV Power Output Prediction. *Sol. Energy* 2015, *119*, 83–99, doi:10.1016/j.solener.2015.06.017.
3. Miyazaki, Y.; Kameda, Y.; Kondoh, J. A Power-Forecasting Method for Geographically Distributed PV Power Systems Using Their Previous Datasets. *Energies* 2019, *12*, doi:10.3390/en12244815.
4. Saint-Drenan, Y.M.; Good, G.H.; Braun, M. A Probabilistic Approach to the Estimation of Regional PV Power Production. *Sol. Energy* 2017, *147*, 257–276, doi:10.1016/j.solener.2017.03.007.
5. Huang, R.; Huang, T.; Gadh, R.; Li, N. Solar Generation Prediction Using the ARMA Model in a Laboratory-Level Micro-Grid. In Proceedings of the 2012 IEEE 3rd International Conference on Smart Grid Communications, SmartGridComm 2012; 2012; pp. 528–533.
6. Liu, J.; Fang, W.; Zhang, X.; Yang, C. An Improved PV Power Forecasting Model With the Assistance of Aerosol Index Data. *IEEE Trans. Sustain. Energy* 2015, *6*, 434–442, doi:10.1109/TSTE.2014.2381224.
7. Theocharides, S.; Theristis, M.; Makrides, G.; Kynigos, M.; Spanias, C.; Georghiou, G.E. Comparative Analysis of Machine Learning Models for Day-Ahead PV Power Production Forecasting†. *Energies* 2021, *14*, doi:10.3390/en14041081.
8. Jang, H.S.; Bae, K.Y.; Park, H.S.; Sung, D.K. Solar Power Prediction Based on Satellite Images and Support Vector Machine. *IEEE Trans. Sustain. Energy* 2016, *7*, 1255–1263, doi:10.1109/TSTE.2016.2535466.
9. Tao, C.; Lu, J.; Lang, J.; Peng, X.; Cheng, K.; Duan, S. Short-Term Forecasting of PV Power Generation Based on Feature Selection and Bias Compensation–Lstm Network. *Energies* 2021, *14*, doi:10.3390/en14113086.
10. Kodaira, D.; Tsukazaki, K.; Kure, T.; Kondoh, J. Improving Forecast Reliability for Geographically Distributed PV Generations. 2021, doi:10.20944/preprints202110.0037.v1.
11. 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, doi:10.1109/TSTE.2014.2313600.
12. Sheng, H.; Xiao, J.; Cheng, Y.; Ni, Q.; Wang, S. Short-Term Solar Power Forecasting Based on Weighted Gaussian Process Regression. *IEEE Trans. Ind. Electron.* 2018, *65*, 300–308, doi:10.1109/TIE.2017.2714127.
13. Agoua, X.G.; Girard, R.; Kariniotakis, G. Probabilistic Models for Spatio-Temporal PV Power Forecasting. *IEEE Trans. Sustain. Energy* 2019, *10*, 780–789, doi:10.1109/TSTE.2018.2847558.
14. Almeida, M.P.; Muñoz, M.; de la Parra, I.; Perpiñán, O. Comparative Study of PV Power Forecast Using Parametric and Nonparametric PV Models. *Sol. Energy* 2017, *155*, 854–866, doi:10.1016/j.solener.2017.07.032.
15. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Lower Upper Bound Estimation Method for Construction of Neural Network-Based Prediction Intervals. *IEEE Trans. Neural Networks* 2011, *22*, 337–346, doi:10.1109/TNN.2010.2096824.
16. Khosravi, A.; Nahavandi, S.; Creighton, D. Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts. *IEEE Trans. Sustain. Energy* 2013, *4*, 602–610, doi:10.1109/TSTE.2012.2232944.
17. Quan, H.; Srinivasan, D.; Khosravi, A. Uncertainty Handling Using Neural Network-Based Prediction Intervals for Electrical Load Forecasting. *Energy* 2014, *73*, 916–925, doi:10.1016/j.energy.2014.06.104.
18. Ni, Q.; Zhuang, S.; Sheng, H.; Kang, G.; Xiao, J. An Ensemble Prediction Intervals Approach for Short-Term PV Power Forecasting. *Sol. Energy* 2017, *155*, 1072–1083, doi:10.1016/j.solener.2017.07.052.
19. Raza, M.Q.; Nadarajah, M.; Ekanayake, C. On Recent Advances in PV Output Power Forecast. *Sol. Energy* 2016, *136*, 125–144.

20. De Giorgi, M.G.; Congedo, P.M.; Malvoni, M. PV Power Forecasting Using Statistical Methods: Impact of Weather Data. *IET Sci. Meas. Technol.* 2014, *8*, 90–97, doi:10.1049/iet-smt.2013.0135.
21. Zhong, Y.J.; Wu, Y.K. Short-Term Solar Power Forecasts Considering Various Weather Variables. In Proceedings of the Proceedings - 2020 International Symposium on Computer, Consumer and Control, IS3C 2020; Institute of Electrical and Electronics Engineers Inc., November 1 2020; pp. 432–435.
22. Pearce, T.; Zaki, M.; Brintrup, A.; Neely, A. *High-Quality Prediction Intervals for Deep Learning: A Distribution-Free, Ensembled Approach*; 2018;
23. Breiman, L. *Random Forests*; 2001; Vol. 45;.
24. Kim, S.G.; Jung, J.Y.; Sim, M.K. A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning. *Sustain.* 2019, *11*, doi:10.3390/SU11051501.
25. Lahouar, A.; Ben Hadj Slama, J. Hour-Ahead Wind Power Forecast Based on random forests. *Renew. Energy* 2017, *109*, 529–541, doi:10.1016/j.renene.2017.03.064.
26. Grömping, U. Variable Importance Assessment in Regression: Linear Regression versus Random Forest. *Am. Stat.* 2009, *63*, 308–319, doi:10.1198/tast.2009.08199.
27. Svetnik, V.; Liaw, A.; Tong, C.; Christopher Culberson, J.; Sheridan, R.P.; Feuston, B.P. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *J. Chem. Inf. Comput. Sci.* 2003, *43*, 1947–1958, doi:10.1021/ci034160g.
28. Khosravi, A.; Nahavandi, S.; Creighton, D.; Atiya, A.F. Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances. *IEEE Trans. Neural Networks* 2011, *22*, 1341–1356, doi:10.1109/TNN.2011.2162110.
29. Panamtash, H.; Zhou, Q.; Hong, T.; Qu, Z.; Davis, K.O. A Copula-Based Bayesian Method for Probabilistic Solar Power Forecasting. *Sol. Energy* 2020, *196*, 336–345, doi:10.1016/j.solener.2019.11.079.