 doi:10.20944/preprints202110.0037.v1

Article

Improving Forecast Reliability for Geographically Distributed Photovoltaic Generations

Daisuke Kodaira [1,*], Kazuki Tsukazaki [1], Taiki Kure [1], Junji Kondoh [1]

[1] Tokyo University of Science; daisuke.kodaira03@gmail.com; 7317090@ed.tus.ac.jp; brightness10.4tkt@gmail.com; j.kondoh@rs.tus.ac.jp
* Correspondence: daisuke.kodaira03@gmail.com

Abstract: Photovoltaic (PV) generation is potentially uncertain. Probabilistic PV generation forecasting methods have been proposed with prediction intervals (PIs). However, several studies have dealt with geographically distributed PVs in a certain area. In this study, a two-step probabilistic forecast scheme is proposed for geographically distributed PV generation forecasting. Each step of the proposed scheme adopts ensemble forecasting based on three different machine-learning methods. In this case study, the proposed scheme was compared with conventional non-multistep forecasting. The proposed scheme improved the reliability of the PIs and deterministic PV forecasting results through 30 days of continuous operation with real data in Japan.

Keywords: photovoltaic generation forecast; probabilistic forecast; prediction interval; ensemble forecast; day ahead forecasting; multiple PV forecasting

1. Introduction

Photovoltaic (PV) generation in a distribution network plays a key role in promoting clean energy production. One of the well-recognized problems of PV generation is the increased power flow at the substation and in the distribution line under the substation [1]. The peak time and amount of power flow depend on the demand and PV generation in the network. The peak is mitigated by operations of energy storage systems (ESSs), such as fixed batteries reported in [2][3]. The peak time and amount of energy generated from the PVs must be forecasted to operate the ESS with the best efficiency. In [3], the proposed peak-shaving algorithm is performed based on prediction intervals (PIs), which indicate the probability of peak demand at the substation. The PIs are evaluated using two contradictory fundamental ideas: the coverage rate and the width of the intervals [4]. If the PIs cover all observations, the coverage rate is the best at 100%. By contrast, the PI widths are preferred to be narrower. As the PIs have a high coverage rate of observations and become narrower, the performance of the peak mitigation improves [3].

PVs are distributed within a certain area connected to the same distribution network. Thus, a spatiotemporal model is required to extract and use spatial and temporal data from multiple PVs to improve PI reliability [5] [6] [7]. Ref [5] proposed a deep learning framework that can generate PV forecasts for multiple regions and horizons with 56 locations in the US. Ref. [6] proposed a model to forecast six hours based on 136 PV installations in France. Irradiance forecasting for 11 PVs distributed in a certain region is performed as accumulated generations [8]. The cloud motion vector-based method [9][10] is an established approach for covering distributed PVs in a certain area. Numerical weather predictions are used for forecasting hours to days ahead [11]. Satellite images, ground measurements, and sky imaging were combined to improve deterministic and probabilistic forecast reliability [12]. Optical flow [13] deals with non-uniform cloud motion and is originally a technique for image processing [15]. The optical flow is a distribution of the apparent velocities of the movement of brightness patterns in an image. An optical flow

 doi:10.20944/preprints202110.0037.v1

that tracks the amount of distributed PV generation was developed in our laboratory [14]. Consequently, the mean absolute percentage error is 4.23% in the case of forecasting 30 min [14]. However, the error increases when the prediction time is later.

Another promising approach for forecasting PV generation is the historical data-driven approach. Data-driven approaches require a large amount of measured past-generation data for deep learning [16]. However, once the correct dataset with a small number of missing records is arranged, the forecasting ability is excellent, especially in day-ahead forecasting [17]. Developing PIs with data-driven approaches for various objectives, such as not PV forecasting, is proposed based on Delta [18] [19], Bayesian [20], mean-variance estimation [21], and bootstrap techniques [22], which are comprehensively compared in [23]. Quantile regression was adopted in [7]. The bootstrap technique was proposed in [17] to quantify the uncertainty with PIs for PV forecasting. In addition, the performance of the bootstrap technique has been proven for wind farm power generation forecasts [24].

Sky or cloud image-based methods and data-driven approaches have been developed as aforementioned; however, as the intensive reviews of PV forecasting, as reported in [11], studies on regional models for multiple PVs are limited. In addition, the existing forecast models are too specific to circumscribe to a particular region [25].

In this study, we introduce the idea of optical flow to data-driven methods, such as machine-learning-based methods to improve existing probabilistic PV generation forecasting methods. Existing machine learning utilizes past data, including generations, temperature, humidity, and precipitation, and the most important predictor is radiation. The forecasting model was mainly developed for each PV system. Conversely, the original idea of PV generation forecasting with the optical flow developed in our laboratory is that the generation of geographically distributed PVs moves as the sun and clouds move [14]. Herein, we propose a PV forecasting method for geographically distributed PVs in a certain area. The PVs are geographically close. Therefore, the past-generation data of one PV can be a meaningful predictor of another PV generation forecasting, which is proven in Section 3 as a case study. Ensemble forecasting comprising three machine-learning methods is proposed in this study as an example of probabilistic forecasting. The proposed ultimate forecasting scheme comprises a single PV forecast model and multiple PV forecast models. The ensemble forecast was adopted for both single and multiple PV forecast models. The proposed ensemble method is enhanced by utilizing the past-generation data of multiple PVs. The simulation shows that the reliability of the forecasting is improved by both deterministic and probabilistic forecasting. The contributions of this study are as follows:

1) In this study, we propose a method to develop boundaries for PIs based on past forecast errors. The case study shows that the boundaries are stable and useful for multiple PVs based on real PV generation data.
2) A multi-step PV forecasting scheme for geographically distributed PVs in a certain area is proposed. The case study shows that the proposed scheme improves the forecasting reliability with real PV generation data.
3) The performance of the proposed multi-step PV forecasting scheme was evaluated with a long-term simulation case as continuous 30 days. The statistical analysis indicates that the proposed scheme improves the root mean square error (RMSE) and mean average percentage error (MAPE) for deterministic forecasting. In addition, the PI cover rate and the width of PI for probabilistic forecasting are improved compared to conventional single PV forecast methods.

The rest of the paper is organized as follows: Section 2 introduces the methodology of the ensemble forecasting model and way to generate the PIs. Section 3 introduces the case study to prove that the proposed forecasting algorithms can improve the reliability of probabilistic forecasting in terms of the PI cover rate and PI width. Finally, Section 4 concludes the study.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

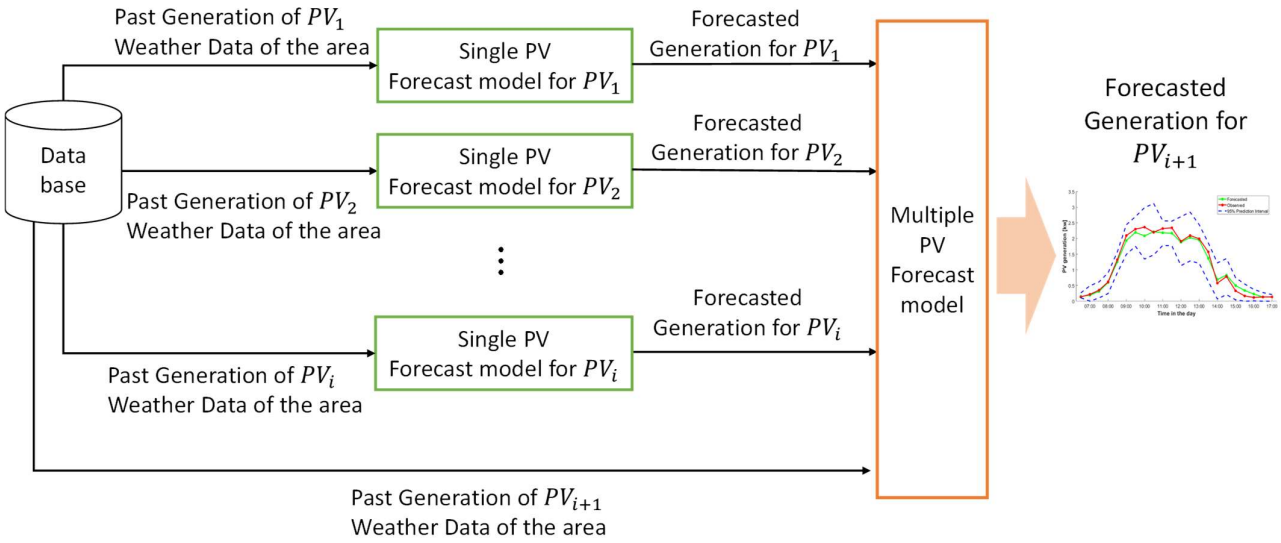


Figure 1. Configuration of the single and multiple PV forecast model

2. Forecast Methodology

The proposed forecast model consists of two steps: a single-forecast model and a multiple forecast model, as shown in Figure 1. The single-forecast model is composed for each PV, indicated as PV_1, PV_2, … PV_i in Figure 1. Past-generation data and weather data are inputs for the ensemble forecast model, as explained in Section 2.1. The forecasted PV generation by the single-forecast model for each PV was utilized as inputs to the multiple forecast model. In Figure 1, PV_{i+1} is forecasted based on the forecasted generation from PV_1 to PV_i, which were chosen based on the Euclidean distance calculated by the latitude and longitude of each PV location. In the case study, the five nearest PVs were chosen to compose the multiple forecast models. The multiple forecast model is performed based on the past data of the target PV, weather data, and the results of other PV forecasts by the single-forecast models.

2.1. Ensemble forecasting with prediction intervals

Both the single and multiple PV forecast models were designed for ensemble forecasting. Three data-driven regressions—k-means, neural network (NN), and long short-term memory (LSTM)—are utilized for ensemble forecasting. In this study, we will handle multiple PVs that are geographically distributed in a certain area. The ensemble model is

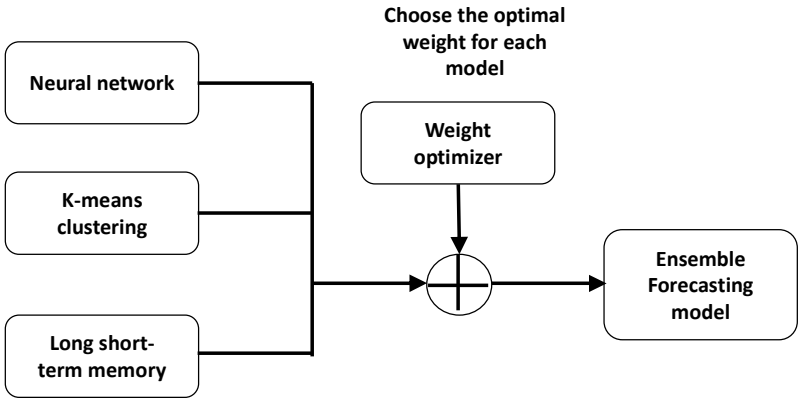


Figure 2. Configuration of an ensemble forecasting model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

arranged for each PV in the proposed method. If we need to forecast five PVs at once, we need to build five individual models for each PV based on different training data. The configuration of the ensemble model is shown in Figure 2. Each data-driven regression model was individually trained to configure the best parameters based on past data. All individual models were added with different weights and one ensemble model. The weight optimizer in Figure 2 calculates the optimal weight for addition based on the past performance of each model. The k-means-based prediction was reported in [26]. The NN model was designed using the function-fitting neural network available in MATLAB [27]. LSTM was also implemented using the function in MATLAB [28].

The PV forecast process is assumed to be performed once a day using continuously updated observed data. The forecast result is provided as day-ahead forecasting; hence, the forecasted PV generation can be utilized to determine the operations of the ESS charge and/or discharge, such as [3]. The process of ensemble forecasting and the development of the prediction intervals are shown in Figure 3. The PV forecasting process comprises two parts: the training process with past data and the forecasting process with test data. The steps from (i) to (vi) in Figure 3 are explained as follows.

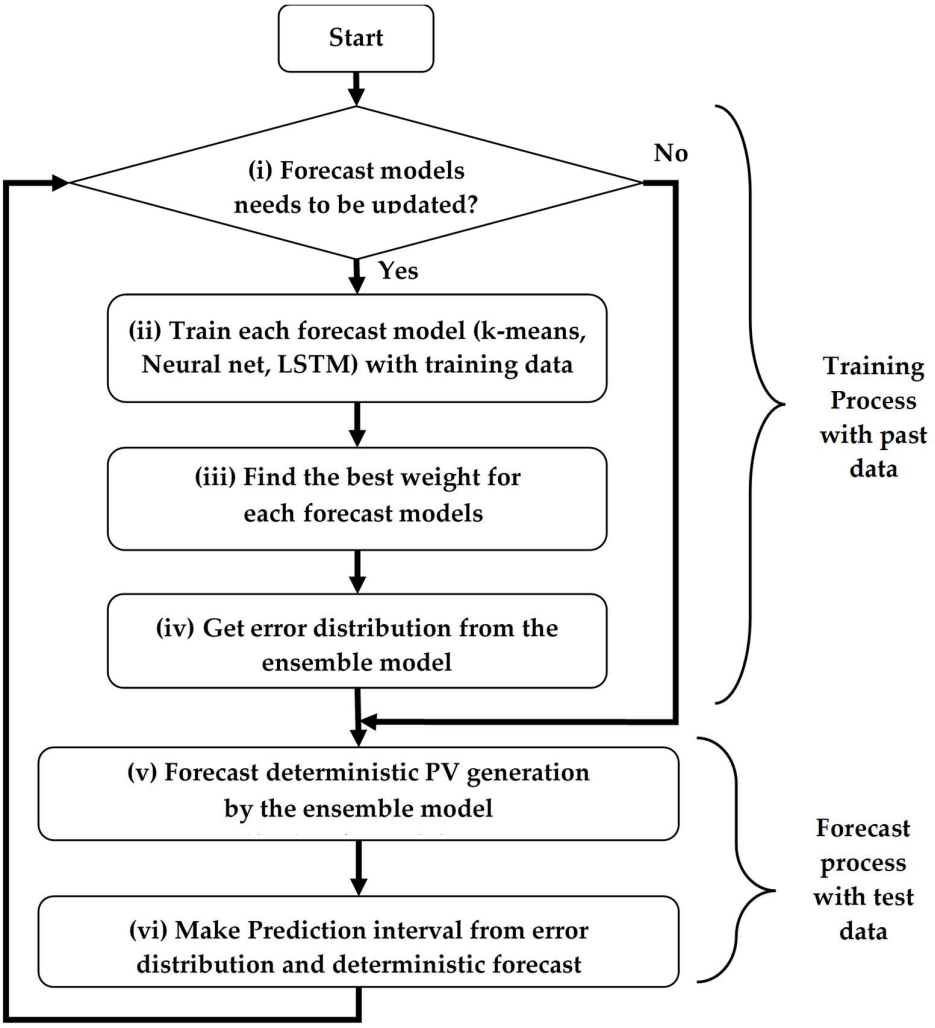


Figure 3. Configuration of an ensemble forecasting model.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

(i) Check if forecast models need to be updated

In step (i), the ensemble forecast model is inspected to check whether the trained parameters are latest. If the trained model with the determined parameters does not include the latest observed data, the model is re-trained to update the parameters of every forecasting method. In the training process, the parameters for the ensemble forecasting methods are determined using past data. Once the training process is completed, the parameters for the forecasting methods remain fixed until a new training process is performed. Therefore, the model parameters must be updated periodically to catch up with the latest observed data. In the case study, the model was updated every 30 days.

(ii) Train each forecast model with training data

In step (ii), the forecasting models k-means, NN, and LSTM are individually trained. The data configuration is shown in Figure 4. Two groups are arranged for model building and forecasting: long-term past data (training and validation data) and forecast data. Long-term data contain predictors (timestamps, temperature, and weather conditions) and target (PV generation); forecast data contain only predictors. In the forecast data, weather information is obtained from weather forecasts available to the public via the web. Training data were utilized as a training dataset to construct the k-means, NN, and LSTM models. Long-term past data preferably contain at least one year of collection to capture seasonal features. Validation data in the long-term past data were selected in sets of arbitrary length from long-term past data. The validation data were utilized to determine the optimal weight for the ensemble forecast model, as shown in Figure 2. In addition, the validation data are utilized to compose the error distribution, leading to PIs. Based on the validation data, the error distribution can reveal bias errors caused by recent facility changes, such as installing new PV farms [29]. This bias error can also be reflected in model training with long-term past data, including validation data. However, the significance of the error takes more than several weeks to show up because the biased new data records are significantly smaller than the existing long-term past data.

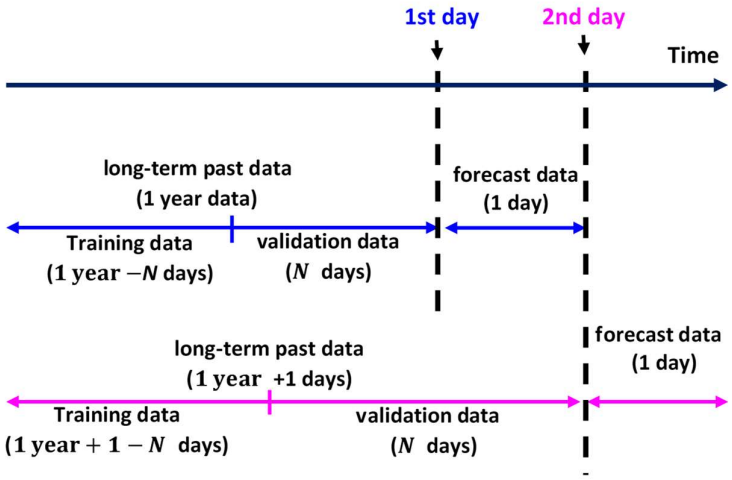


Figure 4. Data configuration; long-term past data for training and error validation. Forecast data for forecasting unknown PV generations.

(iii) Find the best weight for each forecast model

The optimal coefficients for an ensemble model composed of these two trained models were determined. For the k-means model, an optimal k is determined, which indicates how many groups need to be generated. The NN model learns the weights of each neuron. An ensemble prediction model was built by combining these two prediction models with weights, as shown in Eq. (1):

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

$$\hat{y}_t^i = \sum_{i=1}^{N} c_t \hat{F}_t^i, i \in N, t \in T \quad (1)$$

Here, $\hat{y}_t^i$ is the ultimate deterministic forecasted value of PV generation for time instance t on the i-th day. T is the time instances in a day, which is 48 times in the case study. N is defined as the number of days for error validation indicated by N in Figure 4. $\hat{F}_t^i$ is the deterministically forecasted PV generation using the individual forecast methods at time t. In this case, three methods (N = 3) were adopted: k-means, NN, and LSTM. The coefficients c_t are the weights of each forecasting method. c_t is common for all days N. The weights are time-consistent, as determined by the particle swarm optimization (PSO) algorithm that minimizes the error between the observed and predicted loads, as shown in Eq. (2):

$$\boldsymbol{arg}\ \min_{c_t} \left\| \mathbb{Y}_i - \hat{\mathbb{Y}}_i \right\|_2 \quad (2)$$

Here,

$$\begin{aligned} \mathbb{Y}_i &:= \{y_1^i, y_2^i, \ldots y_t^i \ldots y_T^i\} \\ \hat{\mathbb{Y}}_i &:= \{\hat{y}_1^i, \hat{y}_2^i, \ldots \hat{y}_t^i \ldots \hat{y}_T^i\} \end{aligned} \quad (3)$$

$\mathbb{Y}_i$ is the set of observed data y_t^i corresponding to the predicted PV generation $\hat{y}_t^i$ at time t on the i-th day, and $\hat{\mathbb{Y}}_i$ is the set of predicted PV generation $\hat{y}_t^i$. For instance, as shown in Figure 5, if the observed data comprise 30 min-intervals, t comprises 48 instances a day. The deterministic prediction by the ensemble model is performed for past data for a specific time duration, such as the period of one year (i=1, 2, 3, …, 365).

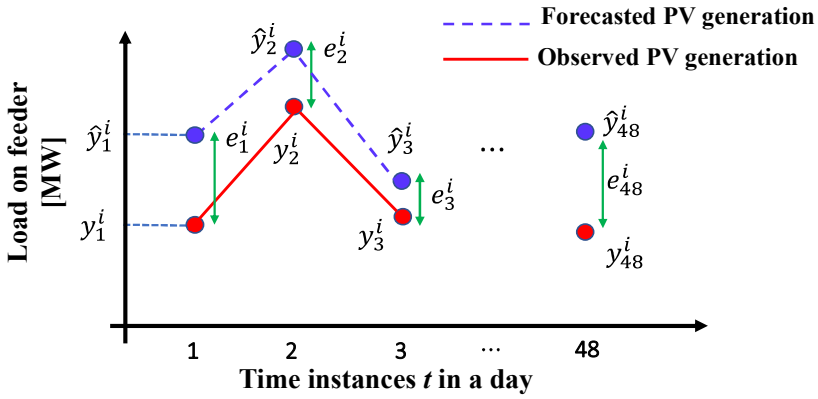


Figure 5. Forecasted and observed PV generation on the i-th day.

(iv) Get error distribution from the ensemble model

Once the optimal coefficients are obtained, future PV generation is forecasted using the trained ensemble model. The boundaries of the PIs were subsequently calculated. The absolute error set $\mathbb{E}_t$ for a specific time t is derived by comparing the predicted and observed data throughout the short-term past data. A series of errors are indicated in Figure 5, and the error set is expressed as follows:

$$e_t^i = y_t^i - \hat{y}_t^i \quad (4)$$

$$\mathbb{E}_t := \{e_t^1, e_t^2, e_t^3 \ldots e_t^i \ldots e_t^N\}, i \in N, t \in T \quad (5)$$

Here, e_t^i is the forecasting error for the *ith* day at time t. N indicates the number of days included in the validation dataset. Each time t has an error record for several days.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

The set $\mathbb{E}_t$ forms the histogram for each time t and is referred to as the error distribution in this document.

(v) Forecast deterministic PV generation by the ensemble model

After the error distribution is formulated in the model training process, deterministic PV generation is forecasted for the next 24 h. The deterministic forecast and error distribution were added into a set $\mathbb{D}_t$. The set $\mathbb{D}_t$ for time t is defined as follows:

$$\mathbb{D}_t := \{\hat{y}_t + e_t^1, \hat{y}_t + e_t^2, \hat{y}_t + e_t^3 \ldots \hat{y}_t + e_t^N\} \quad (6)$$

(vi) Make prediction interval from error distribution and deterministic forecasting

The PIs comprise upper and lower boundaries. In this study, these boundaries are obtained by taking confidence intervals from set $\mathbb{D}_t$ in (6). The set $\mathbb{D}_t$ is not guaranteed to be distributed as a normal distribution; thus, how to make PIs should be investigated further in future work. In the following case study, the confidence interval level is 95% as an example, which can change as the application requires.

2.1. Multiple forecast model

The multiple forecast models have the same ensemble model as the single-forecast models. The operation flow of the multiple forecast model is also similar to that of the single PV forecast model, as shown in Figure 3. The difference between the multiple and single-forecast models is the input data into the ensemble models, as shown in Figure 1. First, the target PV was chosen as the output of the multiple PV forecast model. Second, the PVs forecasted by the single PV forecast models were selected based on the geographical distance from the target PV. In the case study, four PVs were selected for the single-forecast model as an example. The criteria that choose PVs for the single-forecast model are still open to discussion, which can consider the ground form, the direction of the PV panels, obstacle conditions (sometimes the building makes shade for PV panels at a specific time), and others. The correlation coefficient between the generation data from the target PV and PVs for the single PV forecast model is a promising candidate for choosing PVs for a single PV forecast model.

3. Case Study

3.1. Given data set and premises

The data were collected around the Kanto region in Japan. The observed points were distributed as shown in Figure 6. PV generation is forecasted for five PVs named PV (i), (ii), (iii), (iv), and (v) to validate that the proposed multiple PV forecast model improves the reliability of the forecasting for each PV generation. The rated power of PVs are (i) 4.80 kW, (ii) 2.88 kW, (iii) 3.42 kW, (iv) 20.09 kW, and (v) 3.00 kW, respectively. PV generation

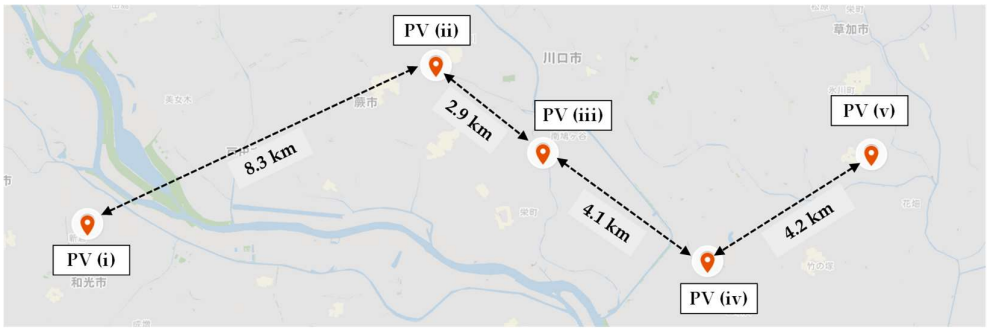


Figure 6. Location of the PV systems to be forecasted in Japan

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

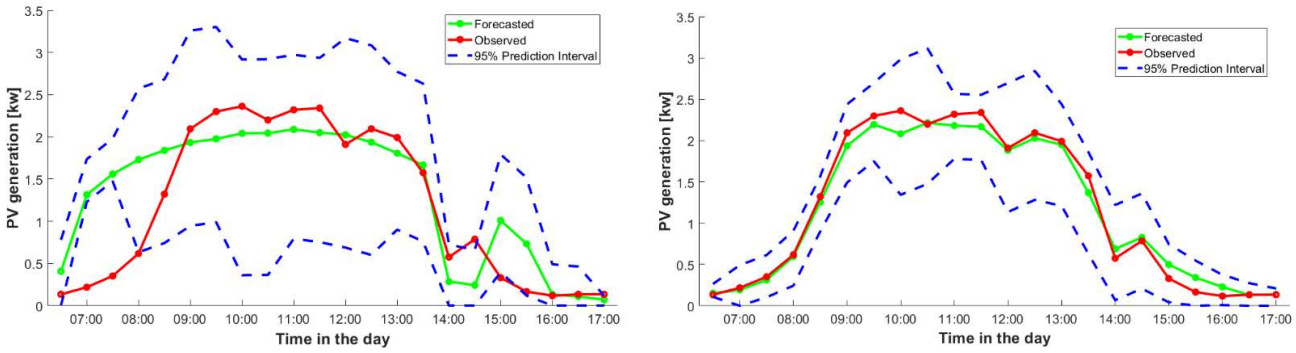
was observed every 30 min from 6:30 AM to 5:00 PM daily. The observed generation data are recorded with year, month, day, hour, minutes, temperature, precipitation, and weather (sunny, cloud, or rain). The location of each PV is also given by latitude and longitude. These five PVs are chosen to be close to each other in terms of distance. The distances between each PV are shown in Figure 6. The area PVs are almost flat, therefore, the altitude is assumed to be the same in the case study.

The observed data for the case study are from August 15, 2013 to July 31, 2014. The observed data from 10:30 PM to 6:00 AM are not available; therefore, the data from 5:30 PM to 10:00 PM are not utilized for model training and forecasting. Some missing records from 10:30 PM to 6:00 AM were interpolated by taking the average before and after the missing records. The PV generation for each PV is forecasted using a single PV forecast model and a multiple PV forecast model. The one-day-ahead forecasting is continuously performed for 30 days from July 2, 2014 to July 31, 2014.

3.2. Simulation results

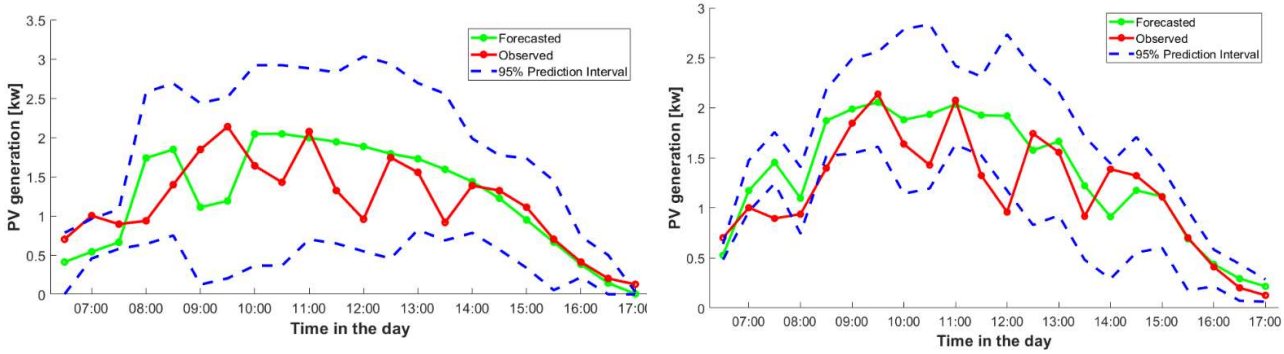
The simulation result is evaluated based on four criteria: the cover rate of the prediction interval, the width of the prediction interval, MAPE, and RMSE. The forecasted result is calculated daily with a 30-min interval because the ESSs in distribution networks are assumed to operate following the predetermined schedule daily.

3.2.1 Forecast result on the best and worst day



(a) Single PV forecast model for PV (iii) (Cover rate = 72%) (b) Multiple PV forecast model for PV (iii) (Cover rate = 100%)

Figure 7. Improvement of forecasted PV generation for PV (iii) on 11th July, 2014. The day has the worst PI coverage rate by the single PV forecast model among 30 days.



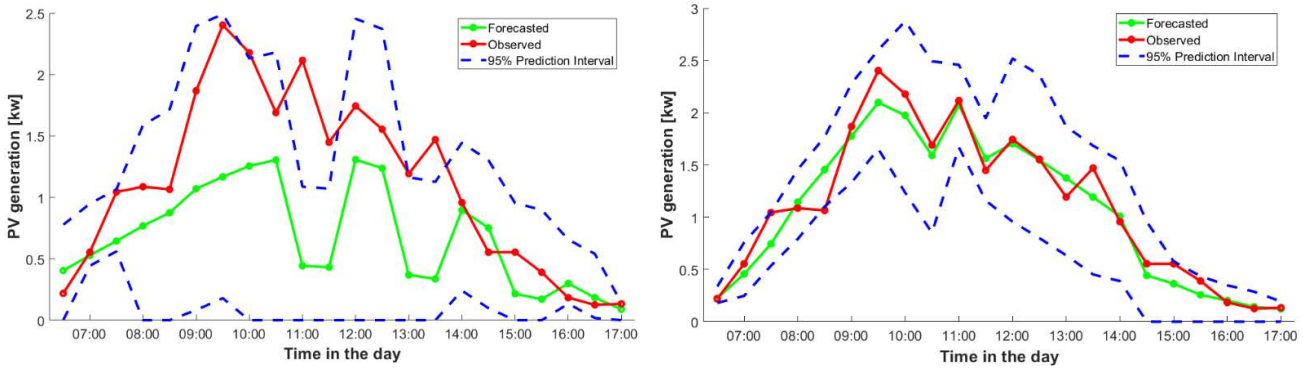
(a) Single PV forecast model for PV (iii) (Cover rate = 90%) (b) Multiple PV forecast model for PV (iii) (Cover rate = 77%)

Figure 8. Deteriorating of forecasted PV generation for PV (iii) on 15th July 2014. The day has the worst PI coverage rate by the multiple PV forecast model among 30 days.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

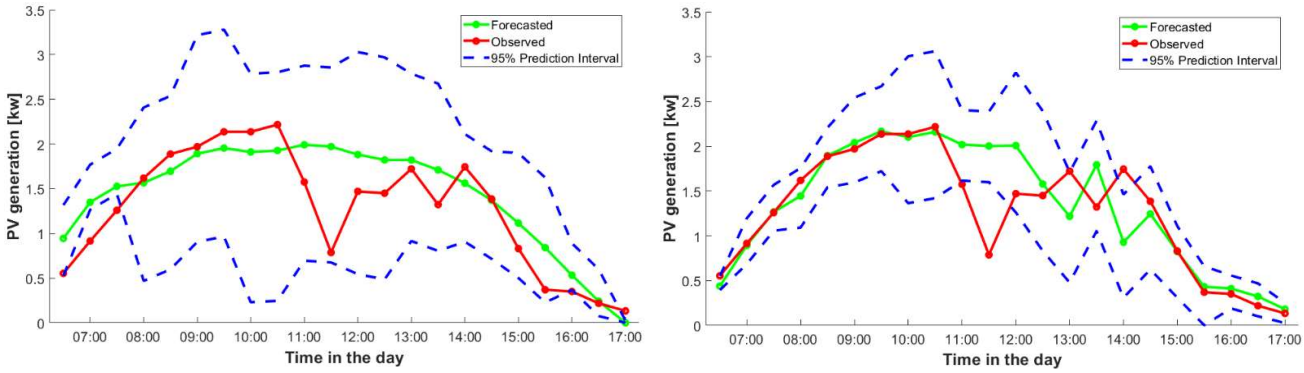
Figure 7 shows the improvement in the forecasted results for PV (iii) on July 11, 2014. The rated power of PV (iii) is 3.42 kW. Figure 7 (a) presents the day with the worst coverage rate among the 30 days using the single PV forecast model. Figure 7 (b) was obtained using the multiple PV forecast model. The cover rate was improved from 72% to 100% using the multiple PV forecast model. In addition, the RMSE was reduced from 0.517 to 0.117 kW. However, in some cases, the PI cover rate is deteriorated by the multiple PV forecast model. Figure 8 shows the deteriorating of the forecasted results for PV (iii) on July 15, 2014. Figure 8 (b) presents the day with the worst coverage rate among the 30 days using the multiple PV forecast model. Figure 8 (a) was obtained using the single PV forecast model. Using the multiple PV forecast model, the cover rate deteriorated from 90 to 77%. However, the RMSE was reduced (improved) from 0.475 kW to 0.347 kW. The reason for the PI cover rate is that the PI width generated by the multiple PV forecast model is not narrower than that of the single PV forecast model.

As with the PI cover rate above, the RMSE calculated based on the single PV forecast model is also improved by the multiple PV forecast model. Figure 9 shows the improvement in the forecasted results for PV (iii) on July 10, 2014. Figure 9 (a) shows the day with the worst RMSE among the 30 days using the single PV forecast model. Figure 9 (b) is obtained using the multiple PV forecast model for the same day. The average of the RMSE in a day was improved from 0.667 to 0.165 kW by the multiple PV forecast model. Figure 10 shows the deteriorating of the forecasted results for PV (iii) on July 25, 2014. Figure 8



(a) Single PV forecast model for PV (iii) (RMSE = 0.667 kw) (b) Multiple PV forecast model for PV (iii) (RMSE = 0.165 kw)

Figure 9. Improvement of forecasted PV generation for PV (iii) on 10th July, 2014; the worst RMSE by the single PV forecast model among 30 days performances.



(a) Single PV forecast model for PV (iii) (RMSE = 0.372 kw) (b) Multiple PV forecast model for PV (iii) (RMSE = 0.382 kw)

Figure 10. Deteriorating of forecasted PV generation for PV (iii) on 25th July, 2014; the worst RMSE by the multiple PV forecast model among 30 days performances.

 doi:10.20944/preprints202110.0037.v1

(b) presents the day with the worst RMSE among the 30 days using the multiple PV forecast model. Figure 10 (a) was obtained using the single PV forecast model. The RMSE slightly worsened from 0.372 to 0.382 kW using the multiple PV forecast model.

According to Figure 8 and Figure 10, the multiple PV forecast model is not always superior to the single PV forecast model for any case. Therefore, in the next subsection, we statistically analyze the forecast results to verify if the multiple PV forecast model is superior to the single PV forecast model for most cases.

3.2.1 Statistical analysis of forecast result in the whole forecast duration

Figure 11 shows a box plot of the cover rate of the prediction interval. In each box plot in Figure 11, the median of 30 days forecasted result is represented by a red line. The edge of the box represents the 75th and 25th percentiles. Notches display the variability of the median between samples as confidence intervals. The width of a notch is computed such that boxes whose notches do not overlap have different medians at the 5% significance level. The significance level is based on a normal distribution assumption, but comparisons of medians are reasonably robust for other distributions [30]. Table I shows a summary of the boxplots from Figure 11 to Figure 14. In Table I, the indicators that have improved compared to the single PV forecast model are highlighted.

Regarding the mean value of the PI cover rate, which is represented by the red line in Figure 11, the multiple PV forecast model outperforms the single PV forecast model in cases PV (ii), (iii), and (iv). PV (i) does not show significant differences between single and multiple PV forecast models. The weather in this season changes from west to east. The PV (i) does not change the result because the PV (i) is located on the west side among the five PVs and cannot obtain any information from the other PVs to improve the forecasting accuracy. PV (v) shows that the median of the PI cover rate decreases from 90.9 to 86.4% in the multiple PV forecast model, as shown in Table I. Nevertheless, the minimum cover rate of the multiple PV forecast model is improved from the single PV forecast model, as indicated in the edge of the boxplot in Figure 11. PV (v) is located on the west side among the five PVs; therefore, the information from the far PVs such as PV (i) and PV (ii) is not variable to improve the forecast accuracy. Figure 12 shows the box plots of the PI width for the five PVs forecasted by the single PV and multiple PV forecast models. The PI width generated by multiple PV models was narrower in all PVs than in the single PV forecast model. As the PI width becomes narrower, the scheduling of the energy management systems becomes easier, and the scheduled operation can be realized with more probability. As shown in Figure 11, the PI cover rate was also improved or remained by the multiple PV forecast model. Regarding PV (ii), PV (iii), and PV (iv), Figure 11 and Figure 12 lead to an ideal result that the multiple PV forecast model simultaneously improves both the cover rate and PI width. The PV (ii), PV (iii), and PV (iv) are located in the middle of the five PVs; therefore, they retrieve variable information from the surrounding PVs to improve forecast accuracy. With respect to PV (i) and PV (v), the PI width was improved. By contrast, the PI cover rate remains, which is also a useful result for energy management using ESSs or electric vehicle scheduling.

Figure 13 shows the boxplot of mean MAPE for five PV generations forecasted by the single and multiple PV forecast models, respectively. In all cases, the multiple PV forecast model shows a smaller MAPE than the single PV forecast model in terms of the mean. Figure 14 shows the boxplot of RMSE for five PV generations forecasted by the single and multiple PV forecast models, respectively. In all cases, the multiple PV forecast model showed a smaller RMSE than the single PV forecast model with respect to the mean. The results are shown in Figure 13 and **Error! Reference source not found.** The multiple PV forecast model improved the deterministic forecast accuracy in both MAPE and RMSE.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 4 October 2021 doi:10.20944/preprints202110.0037.v1

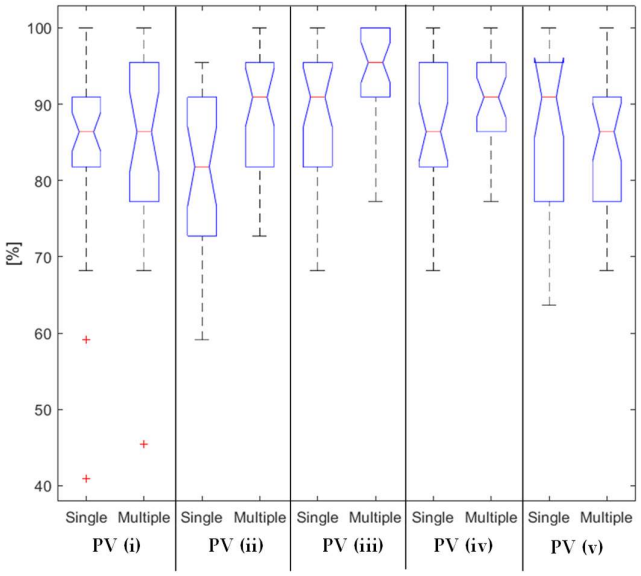


Figure 11. Box plot for prediction interval cover rate; Single PV vs Multiple PV forecast model at five locations

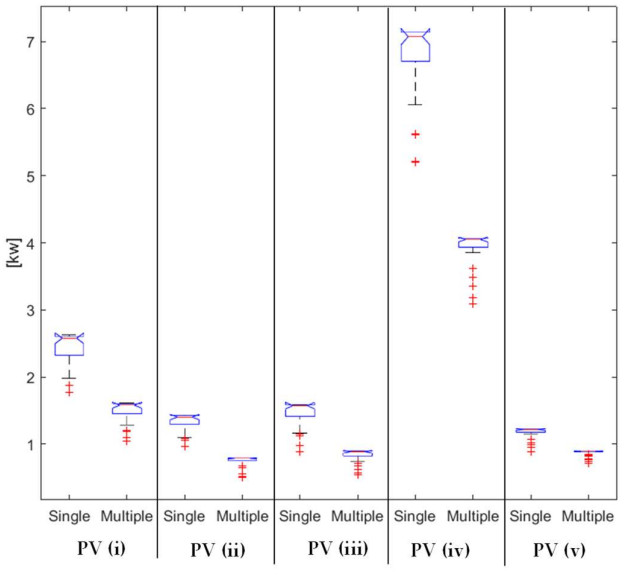


Figure 12. Box plot of prediction interval width; Single PV vs Multiple PV forecast model at five locations

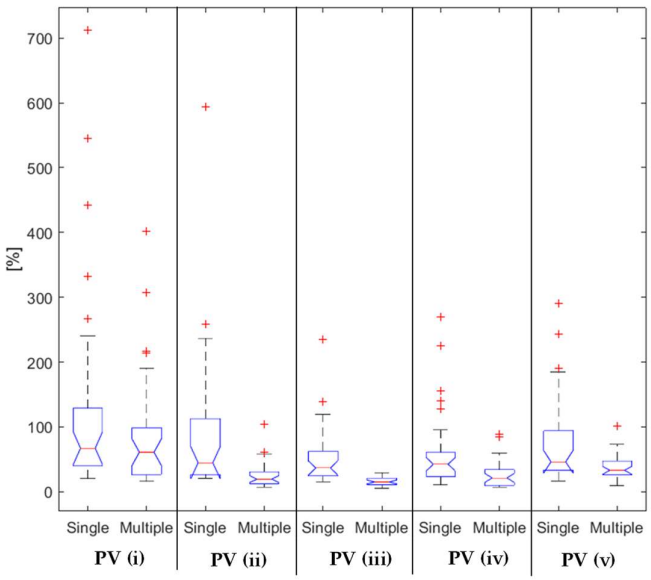


Figure 13. Box plot for mean absolute percentage error (MAPE); Single PV vs Multiple PV forecast model at five locations

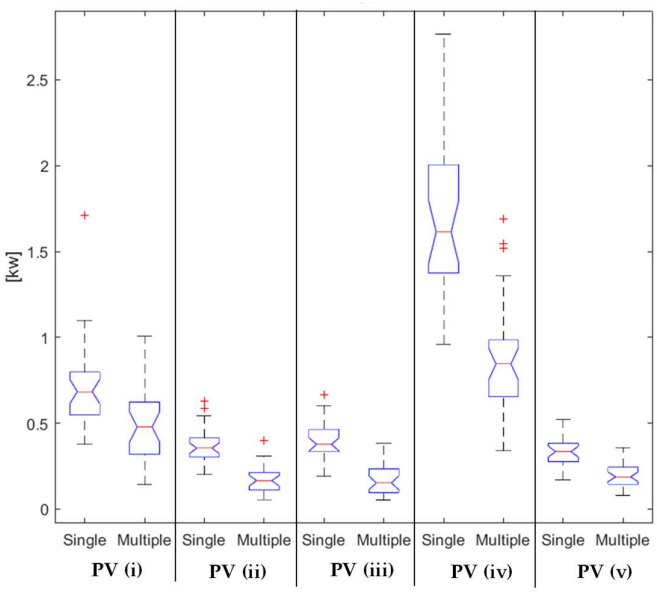


Figure 14. Box plot for root mean square error (RMSE); Single PV vs Multiple PV forecast model at five locations

Table I Summary of the multiple PVs forecast for 30 days

		Cover Rate [%]		PI width [kW]		MAPE [%]		RMSE [kW]	
		Single	Multi	Single	Multi	Single	Multi	Single	Multi
PV (i)	M+2σ	89.0	91.6	2.659	1.624	92.7	81.6	0.755	0.566
	Median (M)	86.4	86.4	2.578	1.581	67.1	60.9	0.684	0.479
	M-2σ	83.8	81.2	2.497	1.537	41.5	40.3	0.614	0.392
PV (ii)	M+2σ	87.0	94.8	1.442	0.797	69.5	24.8	0.386	0.196
	Median (M)	81.8	90.9	1.403	0.788	44.7	19.4	0.355	0.166
	M-2σ	76.6	87.0	1.364	0.780	20.0	14.1	0.324	0.136
PV (iii)	M+2σ	94.8	98.1	1.622	0.906	47.7	18.0	0.415	0.192
	Median (M)	90.9	95.5	1.574	0.887	36.6	15.1	0.377	0.153
	M-2σ	87.0	92.8	1.525	0.867	25.6	12.1	0.339	0.113
PV (iv)	M+2σ	90.3	93.5	7.196	4.086	53.1	28.2	1.794	0.939
	Median (M)	86.4	90.9	7.073	4.051	42.4	21.3	1.613	0.845
	M-2σ	82.5	88.3	6.950	4.015	31.7	14.4	1.432	0.751
PV (v)	M+2σ	96.1	90.3	1.232	0.897	63.6	39.3	0.365	0.216
	Median (M)	90.9	86.4	1.218	0.894	46.2	33.0	0.335	0.188
	M-2σ	85.7	82.5	1.205	0.890	28.7	26.8	0.304	0.160

4. Conclusions

In this study, we proposed a multiple PV forecast model based on ensemble forecasting for distributed PV in a certain area. The ensemble forecasting comprises k-means, NN, and LSTM with optimized weights using the PSO algorithm. In addition, error-based PI construction has also been proposed to convert deterministic forecasting into probabilistic forecasting. The proposed multiple PV forecast model utilizes the result of the neighboring PV forecast result based on the proposed ensemble forecast method. The proposed multiple PV forecast model is designed to provide more reliable PIs for probabilistic forecasting and smaller errors for deterministic forecasting. The proposed multiple PV forecasting model is verified using five real PV generation data and climate data in the case study. As a result of continuous simulations with 30 days data, the RMSE, MAPE, PI cover rate, and PI width were improved by the multiple PV forecast model compared with the conventional single PV forecast model for all five PV cases. In future work, the multiple PV forecast model outperforms the single PV forecast model, but the effective way of selecting neighboring PVs should be investigated theoretically.

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

Acknowledgments: The authors thank S. Ogata and M. Ohashi at OMRON Social Solutions Corp. for their invaluable comments.

References

[1] N. Haghdadi, A. Bruce, I. MaCgill, and R. Passey, "Impact of Distributed Photovoltaic Systems on Zone Substation Peak Demand," IEEE Trans. Sustain. Energy, vol. 9, no. 2, pp. 621–629, 2018.

[2] A. Nagarajan and R. Ayyanar, "Design and Strategy for the Deployment of Energy Storage Systems in a Distribution Feeder With Penetration of Renewable Resources," IEEE Trans. Sustain. Energy, vol. 6, no. 3, pp. 1085–1092, 2015.

[3] D. Kodaira, W. Jung, and S. Han, "Optimal Energy Storage System Operation for Peak Reduction in a Distribution Network Using a Prediction Interval," IEEE Trans. Smart Grid, vol. 11, no. 3, pp. 2208–2217, 2020.

[4] A. Khosravi, S. Nahavandi, and D. Creighton, "Construction of optimal prediction intervals for load forecasting problems," IEEE Trans. Power Syst., vol. 25, no. 3, pp. 1496–1503, 2010.

[5] S. Chai, Z. Xu, Y. Jia, and W. K. Wong, "A Robust Spatiotemporal Forecasting Framework for Photovoltaic Generation," IEEE Trans. Smart Grid, vol. 11, no. 6, pp. 5370–5382, 2020.

[6] X. G. Agoua, R. Girard, and G. Kariniotakis, "Photovoltaic power forecasting: Assessment of the impact of multiple sources of spatio-temporal data on forecast accuracy," Energies, vol. 14, no. 5, 2021.

[7] X. G. Agoua, R. Girard, and G. Kariniotakis, "Probabilistic Models for Spatio-Temporal Photovoltaic Power Forecasting," IEEE Trans. Sustain. Energy, vol. 10, no. 2, pp. 780–789, 2019.

[8] E. Lorenz, J. Hurka, D. Heinemann, and H. G. Beyer, "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems," IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens., vol. 2, no. 1, pp. 2–10, 2009.

[9] T. Carrière, R. Amaro e Silva, F. Zhuang, Y.-M. Saint-Drenan, and P. Blanc, "A New Approach for Satellite-Based Probabilistic Solar Forecasting with Cloud Motion Vectors," Energies, vol. 14, no. 16, p. 4951, 2021.

[10] H. T. C. Pedro, D. P. Larson, and C. F. M. Coimbra, "A comprehensive dataset for the accelerated development and benchmarking of solar forecasting methods," J. Renew. Sustain. Energy, vol. 11, no. 3, 2019.

[11] A. Mellit, A. M. Pavan, E. Ogliari, S. Leva, and V. Lughi, "Advanced methods for photovoltaic output power forecasting: A review," Appl. Sci., vol. 10, no. 2, 2020.

[12] M. Pierro et al., "Multi-Model Ensemble for day ahead prediction of photovoltaic power generation," Sol. Energy, vol. 134, pp. 132–146, 2016.

[13] C. W. Chow, S. Belongie, and J. Kleissl, "Cloud motion and stability estimation for intra-hour solar forecasting," Sol. Energy, vol. 115, pp. 645–655, 2015.

[14] Y. Miyazaki, Y. Kameda, and J. Kondoh, "A Power-Forecasting Method for Geographically Distributed PV Power Systems using Their Previous Datasets," Energies, vol. 12, no. 24, 2019.

[15] B. K. P. Horn and B. G. Schunck, "Determining optical flow," Comput. Vis., vol. 17, no. 1–3, pp. 185–203, Aug. 1981.

[16] H. Wen et al., "Deep Learning Based Multistep Solar Forecasting for PV Ramp-Rate Control Using Sky Images," IEEE Trans. Ind. Informatics, vol. 17, no. 2, pp. 1397–1406, 2021.

[17] S. Al-Dahidi, O. Ayadi, M. Alrbai, and J. Adeeb, "Ensemble approach of optimized artificial neural networks for solar photovoltaic power prediction," IEEE Access, vol. 7, pp. 81741–81758, 2019.

[18] J. T. G. Hwang and A. A. Ding, "Prediction Intervals for Artificial Neural Networks," J. Am. Stat. Assoc., vol. 92, no. 438, pp. 748–757, Jun. 1997.

[19] R. D. De Veaux, J. Schumi, J. Schweinsberg, and L. H. Ungar, "Prediction intervals for neural networks via nonlinear regression," Technometrics, vol. 40, no. 4, pp. 273–282, 1998.

[20] D. J. C. MacKay, "The Evidence Framework Applied to Classification Networks," Neural Comput., vol. 4, no. 5, pp. 720–736, 1992.

[21] D. A. Nix and A. S. Weigend, "Estimating the mean and variance of the target probability distribution," IEEE Int. Conf. Neural Networks - Conf. Proc., vol. 1, pp. 55–60, 1994.

[22] T. Heskes, "Practical confidence and prediction intervals," in Advances in Neural Information Processing Systems, 1997, pp. 176–182.

[23] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," IEEE Trans. Neural Networks, vol. 22, no. 9, pp. 1341–1356, 2011.

[24] A. Khosravi, S. Nahavandi, and D. Creighton, "Prediction Intervals for Short-Term Wind Farm Power Generation Forecasts," IEEE Trans. Sustain. Energy, vol. 4, no. 3, pp. 602–610, Jul. 2013.

[25] R. Ahmed, V. Sreeram, Y. Mishra, and M. D. Arif, "A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization," Renew. Sustain. Energy Rev., vol. 124, no. February, 2020.

[26] S. Park, S. Han, and Y. Son, "Demand power forecasting with data mining method in smart grid," 2017 IEEE Innov. Smart Grid Technol. - Asia Smart Grid Smart Community, ISGT-Asia 2017, pp. 1–6, 2018.

[27] "fitnet, Matlab." .

[28] "Long Short-Term Memory Networks - MATLAB & Simulink - MathWorks." https://jp.mathworks.com/help/deeplearning/ug/long-short-term-memory-networks.html?lang=en (accessed Sep. 02, 2021).

[29] K. A. Agyeman, G. Kim, H. Jo, S. Park, and S. Han, "An Ensemble Stochastic Forecasting Framework for Variable Distributed Demand Loads," Energies, vol. 13, no. 10, p. 2658, May 2020.

[30] MathWorks, "Visualize summary statistics with box plot," 2021. https://jp.mathworks.com/help/stats/boxplot.html?lang=en (accessed Aug. 23, 2021).