Ensemble Methods for Probabilistic Solar Power Forecasting: A Comparative Study

Tawsif Ahmad
Electrical and Computer Engineering Department
State University of New York at Binghamton
Vestal NY, USA
E-mail: tahmad2@binghamton.edu

Ning Zhou
Electrical and Computer Engineering Department
State University of New York at Binghamton
Vestal NY, USA
E-mail: ningzhou@binghamton.edu

***Abstract*—To guide the selection of probabilistic solar power forecasting methods for day-ahead power grid operations, the performance of four methods, i.e., Bayesian model averaging (BMA), Analog ensemble (AnEn), ensemble learning method (ELM), and persistence ensemble (PerEn) is compared in this paper. A real-world hourly solar generation dataset from a rooftop solar plant is used to train and validate the methods under *clear*, *partially cloudy*, and *overcast* weather conditions. Comparisons have been made on a one-year testing set using popular performance metrics for probabilistic forecasts. It is found that the ELM method outperforms other methods by offering better reliability, higher resolution, and narrower prediction interval width under all weather conditions with a slight compromise in accuracy. The BMA method performs well under *overcast* and *partially cloudy* weather conditions, although it is outperformed by the ELM method under *clear* conditions.**

Keywords—Analog ensemble, Bayesian model averaging, Ensemble learning, probabilistic solar power forecasting.

I. INTRODUCTION

The power system has been experiencing a rapid increase in solar photovoltaic (PV) penetration as the advancement toward a sustainable power grid continues. Large-scale integration of PV generation introduces various challenges to the efficient and reliable operation of the grid because solar PV generation is often intermittent due to the random variation of weather conditions [1], [2]. Solar power forecasting (SPF) is a vital decision-making tool for utility companies participating in the day-ahead energy market [3]. Thus, SPF has been emerging as a major research topic in the last couple of decades.

Taking advantage of historical data, data-driven techniques such as statistical methods, machine learning, and deep learning algorithms have been utilized in recent years to forecast solar PV generation. Many studies have been carried out on deterministic forecasts where a single value is forecasted for each temporal point on the forecast horizon [4]–[7]. However, deterministic SPF does not quantify the forecast uncertainty, which is needed to book generation reserves. In contrast, probabilistic SPF quantifies forecasting uncertainty in the form of probability distribution function (PDF), cumulative distribution function (CDF), or prediction intervals (PI). Statistical and machine learning methods such as quantile regression (QR), quantile gradient boosting (QGB), Gaussian process regression (GPR) are some well-known probabilistic methods and have been applied in the field of SPF [8]–[10]. These probabilistic SPF methods are different from each other in terms of loss functions, model hyper-parameters, training algorithms, etc. As a result, their forecasting accuracies are often different from each other and can vary significantly under different application scenarios.

To reduce the variability of forecasting performance, ensemble methods combine outputs from multiple SPF methods to increase forecasting robustness. The persistence ensemble (PerEn) is a simple probabilistic SPF method that fits the solar generation to a normal distribution whose parameters are estimated from the most recent generation data at the same lead time [11]. The analog ensemble (AnEn) method is the first ensemble method that produces an ensemble forecast by ranking the historical power observations according to the minimum distance between the current and past weather forecast at the same lead time [12]. The ensemble learning method (ELM), proposed in [13], takes advantage of the outputs from multiple deterministic SPF methods to estimate the distribution of solar power generation. An ensemble prediction system (EPS) based on perturbations of the initial conditions of the European Centre for Medium-Range Weather Forecasts (ECMWF) prediction models and post-processing via statistical and machine learning techniques is proposed in [14] to produce probabilistic SPF. In [15], an ensemble approach called lower upper bound estimate (LUBE) is proposed to construct accurate and robust PIs via training a neural network with a two-neuron output layer. In [16], Bayesian model averaging (BMA) is used to estimate the PDF of solar generation via weighted averaging of PDFs from ensemble members. The member-specific PDFs are estimated using separate machine-learning algorithms such as logistic regression and expectation-condition maximization. These ensemble methods are shown to have achieved higher forecasting accuracy than an individual SPF model in their ensembles.

Although several probabilistic SPF ensemble methods can be found in literature, their performance has not been thoroughly studied under various weather conditions. For example, in [13], only pin-ball loss is studied, while the reliability and sharpness of the forecast are not considered. In [16], the forecast reliability and sharpness are considered, whereas the coverage probability is not considered. Both methods' performance is not studied under different weather conditions. In [15], the impact of weather conditions on the forecast is studied in terms of coverage probability and PI width, while the reliability and sharpness are not analyzed. The studies carried out so far do not provide sufficient information for users to select a proper ensemble method. To bridge the gap, this paper thoroughly evaluates the performance of four ensemble probabilistic SPF methods, i.e., the BMA, AnEn,

This paper is based on work supported by the National Science Foundation under grant no. #1845523.

ELM, and PerEn methods, under various weather conditions using the real-world data. The major contributions of the paper can be summarized as follows:

(a) The performances of four widely used ensemble SPF methods are compared using six well-defined metrics based on a real-world dataset. The study results can be directly used by the local utility in choosing a forecasting method.

(b) The proposed performance comparison method can be used by peer engineers to evaluate SPF methods at different geographic locations and under different climate conditions.

The rest of the paper is organized as follows: section II explains the ensemble methods, section III presents results from a case study, and conclusions are drawn in section IV.

II. Ensemble Methods for Probabilistic Forecasting

To lay a ground for discussion, four ensemble methods for probabilistic forecasting are overviewed in this section. These methods are chosen because they represent state-of-the-art techniques in probabilistic SPF and can serve as a benchmark for future studies.

A. Bayesian Model Averaging

BMA is a statistical technique to combine PDFs of ensemble members using the weighted averaging technique. The ensemble members are SPF outputs from different deterministic models [16]. Let K be the number of ensemble members in a BMA model and let $\hat{P}_1, \hat{P}_2, \dots, \hat{P}_K$ be the forecast outputs of each of the K members at a certain lead time. The PDF of the solar power forecast conditioned on the member forecasts can be obtained by using BMA as follows:

$$p\left(y|\hat{P}_1, \hat{P}_2, \dots, \hat{P}_K\right) = \sum_{k=1}^{K} w_k f_k(y|\hat{P}_k), \tag{1}$$

where y represents the solar power output, $f_k(y|\hat{P}_k)$ is the member-specific PDF which is obtained using the historical forecast from member k, and w_k is a non-negative weight assigned to the PDF of member k based on its relative performance during the training stage. Note that $(w_1, w_2, \dots, w_K)$ form a PDF and hence $\sum_{k=1}^{K} w_k = 1$. $f_k(y|\hat{P}_k)$ is obtained by fitting the historical forecast data of member k by selecting an appropriate kernel. Truncated normal kernel is used in this study amongst other popular kernels such as Gaussian distribution, beta distribution [17], [18]. The solar power output is limited to the interval $[0, P_{cap}]$, where P_{cap} is the generation capacity of the solar power plant. When a member forecast, $\hat{P}_k$, is very close to or over P_{cap}, clipping is necessary. A clipping threshold, λ is introduced to account for the clipping and it is set at 99.5% of P_{cap}. The probability of clipping $P(y \geq \lambda P_{cap}|\hat{P}_k)$ given that $\hat{P}_k$ is the best member forecast can be estimated using logistic regression on $\hat{P}_k$ as follows:

$$logit\left(P\left(y \geq \lambda P_{cap}\middle|\hat{P}_k\right)\right) = a_{0k} + a_{1k}\hat{P}_k \tag{2}$$

Here, member forecast $\hat{P}_k$ is the predictor and the response is a binary event with '0' and '1' representing no clipping and clipping respectively.

A truncated normal kernel can be obtained by truncating a standard normal distribution to interval $[0, P_{cap}]$. The PDF of this kernel for a specific ensemble member can be defined as:

$$g_k(y, \mu_k, \sigma_k) = \frac{\phi\left(\frac{y-\mu_k}{\sigma_k}\right)}{\sigma_k\left(\Phi\left(\frac{P_{cap}-\mu_k}{\sigma_k}\right) - \Phi\left(\frac{0-\mu_k}{\sigma_k}\right)\right)} \tag{3}$$

where, μ_k is the mean, σ_k is the standard deviation, and $\phi(\cdot)$ and $\Phi(\cdot)$ are the PDF and CDF of the standard normal distribution respectively. μ_k is estimated from member forecast k by a scaling factor b_k:

$$\mu_k = b_k \hat{P}_k \tag{4}$$

b_k is estimated using linear regression on historical data considering forecast $\hat{P}_k$ as predictor and observed power as the response. σ_k is estimated via a variance coefficient c_k and kernel mean μ_k as follows:

$$\sigma_k^2 = \left[-\frac{c_k}{0.25}\left(\frac{\mu_k}{P_{cap}} - 0.5\right)^2 + c_k\right].P_{cap} \tag{5}$$

Variance coefficient c_k and member weights w_k are estimated using maximum log-likelihood estimation via the Expectation Condition Maximization (ECM) algorithm [17].

With the estimated $a, b,$ and c coefficients, the PDF of each ensemble forecast $\hat{P}_k$ given that $\hat{P}_k$ is the best forecast at that instant can be obtained as:

$$\begin{aligned} f_k\left(y|\hat{P}_k\right) = {} & \frac{P\left(y \geq \lambda P_{cap}|\hat{P}_k\right)}{(1-\lambda)P_{cap}} \mathbb{1}\left[y \geq \lambda P_{cap}\right] \\ & + \frac{P\left(y < \lambda P_{cap}|\hat{P}_k\right)}{G_k(y, \mu_k, \sigma_k)|_\lambda} g_k(y, \mu_k, \sigma_k)\, \mathbb{1}\left[y < \lambda P_{cap}\right] \end{aligned} \tag{6}$$

where $G_k(y, \mu_k, \sigma_k)|_\lambda$ is the corresponding CDF evaluated at λ. With these estimates of $f_k\left(y|\hat{P}_k\right)$ for each ensemble member k and their corresponding weights w_k, the BMA output PDF can be obtained using (1). Finally, the CDF can be estimated by integrating the estimated PDF.

B. Analog Ensemble

AnEn develops ensemble forecasts by utilizing historical solar power observations and historical predictions of solar power using a deterministic model and associated weather variables [12]. For a lead time h, the distance between the current forecast and past forecasts is used to rank the current forecast's similarity with the past forecasts. This distance is calculated as follows:

$$d = \sum_{i=1}^{N_v} \frac{w_i}{\sigma_i} \sqrt{\sum_{j=-h'}^{h'} \left(x_{i,h-j}^{current} - x_{i,h+j}^{past}\right)^2} \tag{7}$$

Here, N_v and w_i are the number of weather variables and their weights, respectively. σ_i is the standard deviation of the historical data of the i^{th} weather variable. h' is the half-width of the time window over which the distance is calculated. $x_{i,h}^{current}$ and $x_{i,h}^{past}$ are the current and past forecasts of the i^{th} weather variable at lead time h, respectively.

The past forecasts of weather variables are ranked according to their distance from the current weather forecast. From this set, n forecasts with the smallest distance are

selected. The solar power observations, concurrent with the n forecasts, form the AnEn ensemble. The statistical properties of the ensemble can be estimated by fitting the n forecasts to a standard normal distribution. The underlying assumption here is that the forecast error of the current forecast will likely be similar to the errors of the selected past forecasts. The AnEn method is computationally inexpensive compared to other ensemble methods as it requires only a single run of the deterministic model.

C. Ensemble Learning Method

ELM generates probabilistic forecasts by exploiting an ensemble of various statistics and machine learning models [13]. Different deterministic models perform differently with different types of data. Some models perform well only when a large number of training data are available, whereas others perform well with smaller sample sizes. ELM takes advantage of the strengths of different algorithms in different situations by creating an ensemble of these models.

First, the training data are grouped based on hours of the day. Then, for each hour of the day, the following nine models are trained to produce hourly deterministic forecasts: multiple linear regression (MLR), decision tree regressor, gradient boosting regressor, k-nearest neighbors (kNN) with uniform weights, kNN with distance-based weights, lasso regression, random forest regression, ridge regression, and persistence model. The training strategy of these models, such as model hyperparameters, loss functions and optimizers, are developed as suggested in [13].

The deterministic forecasts from the nine models are then combined to generate a probabilistic forecast using the following three methods: linear method, normal distribution method, and normal distribution with additional features. Detailed discussion about these methods can be found in [13]. In this paper, only the normal distribution method is considered for generating a probabilistic forecast.

D. Persistence Ensemble

PerEn is commonly used as the benchmark forecast to evaluate new forecasting algorithms. A deterministic persistence model simply forecasts today's hourly power generation remains the same tomorrow which can be mathematically written as:

$$\hat{P}_d^h = P_{d-1}^h \tag{8}$$

where, $\hat{P}_d^h$ is the persistence forecast for hour h of day d, and P_{d-1}^h is the observation at the same hour of day $d-1$.

In contrast, the PerEn method uses the most recent 20 solar power observations at the same hour to generate probabilistic forecasts by ranking them to obtain quantiles. The methods described in the previous section can be used, with the normal distribution method being the most popular choice. An example of the implementation of the PerEn method for solar power forecasting can be found in [12].

E. Evaluation of Probabilistic SPF

Performance metrics are widely used to evaluate how well a probabilistic forecasting model fits the observation data [19]. Evaluation metrics used in this paper are- *continuous rank probability score* (CRPS) and its decompositions- *reliability* (REL), *resolution* (RES), and *uncertainty* (UNC) [20], *Brier score* (BS), *prediction interval coverage probability* (PICP), *prediction interval normalized width* (PINAW). The CRPS and BS compare the CDF of the forecast, whereas the PICP and PINAW evaluate the PIs. When a probabilistic forecast method produces smaller CRPS, REL, BS, PINAW, and larger RES, and PICP than other methods, the method is often preferred by users because all the metrics unanimously agree on its higher performance. When these metrics do not agree with each other, users may choose a method based on performance metrics that are applicable to their applications.

III. Case Study

The ensemble SPF methods discussed in section II are implemented to produce day-ahead hourly probabilistic SPF. The term 'day-ahead' refers to 1 hour – 24 hours ahead forecast. A 450-kW rooftop solar PV plant located at Vestal, NY, USA (*lat. 42˚05′37.0″N, long. 76˚00′06.0″W*) is selected to perform this study. Hourly solar power observations from the year 2016 to 2021 are downloaded from New York State Energy Research and Development Authority (NYSERDA)'s website [21]. The hourly weather forecast data for this site are collected from the Visual Crossing weather data services [22]. Weather variables such as *temperature, relative humidity, visibility,* and *cloud cover* are selected as predictors as they proved to have significant correlation with solar radiation [23]. Moreover, solar power observation from the previous day is also considered as a predictor based on autocorrelation. Standard data pre-processing techniques are utilized to identify and remove bad data [24].

Because the BMA and ELM require deterministic forecasts from multiple models, the nine models discussed in section II are selected. The AnEn requires a single model to determine the predictor weights. The MLR model is selected in this case. The PerEn doesn't require additional deterministic models. The data from the year 2016 to 2020 is used for training the models, and the data for the year 2021 is used for testing. All models are cross-validated using the 10-fold cross-validation technique [25]. The CDFs of the hourly solar generation are estimated using the cross-validated models as well as the ensemble methods. Hourly observations of weather conditions in the year 2021 are collected from the weather data services to assess the performance of the models under different weather conditions.

Three weather conditions are considered, namely- *'Clear', 'Partially cloudy',* and *'Overcast'*. Out of 365 days in the test set, 124 *clear* days, 112 *partially cloudy* days, and 129 *overcast* days are identified. Performance metrics are computed for each day of the test set, and then averaged over all the days under a particular weather condition. The average performance metrics of the ensemble methods are shown in Tables I, II, and III. Three days under three different weather conditions (i.e., clear day on 05/19/2021, partially cloudy day on 07/09/2021, and overcast day on 12/28/2021) are selected to visualize the 95% PIs estimated by each of the ensemble methods in Fig. 1.

During the *clear* days shown in Table I, the ELM method achieves the lowest CRPS value. Moreover, the ELM method has the lowest REL and highest RES values. The ELM method also exhibits the lowest PINAW, which indicates that the width of the 95% PI is narrower than the other methods. These results indicate that the ELM estimates are the most reliable, sharpest, and closest to the measurement distribution. These results can be further verified from Fig. 1. It can be observed that the PI width of the ELM is narrower than the other methods. However, some observations fall outside the estimated PI of the ELM. This explains the lower values of BS and PICP for the ELM method. The BMA and AnEn methods also exhibit lower BS and PICP because of their narrower width of estimated PIs. The PerEn method achieves better BS and PICP with a cost of the lower resolution of the estimated distribution.

During the *partially cloudy* days shown in Table II, the BMA and ELM methods' performance on the are similar, as evident from their CRPS, REL, and RES values. However, the BMA achieves higher accuracy than the ELM, as evident from the BMA's lower BS value and higher PICP value. The ELM method, on the other hand, offers better resolution, which is indicated by the lower PINAW value. The PerEn method shows low BS and high PICP, indicating better coverage. However, the resolution of its forecasted CDF is significantly poorer than the other methods. These results can be further validated from the 95% PI estimates on a *partially cloudy* day in Fig. 1.

TABLE I. PERFORMANCE METRICS OF THE ENSEMBLE METHODS ON THE *124 CLEAR* DAYS

Method	CRPS (%)	CRPS Decomposition			BS (%)	PICP (%)	PINAW
		REL	RES	UNC			
BMA	9.0	0.89	49.4	57.5	7.1	78.6	11.92
AnEn	15.4	1.50	43.7	57.5	7.4	61.7	14.67
ELM	7.2	0.63	50.9	57.5	7.5	64.1	7.26
PerEn	10.5	0.90	47.9	57.5	4.0	95.7	41.5

TABLE II. PERFORMANCE METRICS OF THE ENSEMBLE METHODS ON THE *112 PARTIALLY CLOUDY* DAYS

Method	CRPS (%)	CRPS Decomposition			BS (%)	PICP (%)	PINAW
		REL	RES	UNC			
BMA	5.6	0.40	34.0	39.2	4.5	86.9	20.95
AnEn	9.9	0.64	29.9	39.2	6.8	73.5	28.60
ELM	5.6	0.38	33.9	39.2	6.7	73.6	15.49
PerEn	8.2	0.62	31.7	39.2	4.1	96.2	63.56

TABLE III. PERFORMANCE METRICS OF THE ENSEMBLE METHODS ON THE *129 OVERCAST* DAYS

Method	CRPS (%)	CRPS Decomposition			BS (%)	PICP (%)	PINAW
		REL	RES	UNC			
BMA	3.2	0.08	7.8	10.9	2.5	97.2	817.8
AnEn	6.2	0.54	5.2	10.9	5.6	89.5	1353.9
ELM	2.9	0.23	8.2	10.9	5.6	90.1	628.8
PerEn	4.9	0.57	6.5	10.9	5.4	93.1	440.8

During the *overcast* days shown in Table III, the BMA method shows better reliability and sharpness than the other methods. Although the ELM shows poorer reliability (higher REL) than the BMA, it offers slightly better resolution, which results in an improved overall CRPS. The BMA method shows better accuracy than the AnEn and ELM methods. However, the 95% PI width (shown in Fig. 1.) of the ELM estimates is significantly narrower, indicating better resolution.

The CDFs of the hourly solar generation estimated by the methods at 10:00 AM on the three selected days (under three different weather conditions) are presented in Fig. 2. On the *clear* day, the CDF estimated from the ELM shows the lowest deviation from the CDF of the measurement, which indicates it provides the most reliable and sharpest forecast. This is also the case on the *partially cloudy* days. However, on the *overcast* days, its deviation is higher. The BMA estimates are the next closest CDF under all the weather conditions. The CDF estimates from the AnEn and PerEn methods show the most deviations from the measurement CDF under all the weather conditions.

Finally, the computation time for these methods is counted using MATLAB® 2022a environment running on an Intel® Core™ i5-8400 CPU @ 2.80 GHz with 12 GB RAM and 64-bit Windows operating system. The computation time of all the methods is less than one minute, which indicates that they can be implemented in real time for day-ahead applications.

IV. CONCLUSIONS

In this paper, the performance of four ensemble SPF methods, namely BMA, AnEn, ELM, and PerEn, are evaluated using real-world data from a rooftop solar PV plant. It is found that the ELM performs consistently well under different weather conditions. The BMA method performs similarly to the ELM during *partially cloudy* and *overcast* days. However, it is outperformed by the ELM during *clear* days. Although these two methods show lower accuracy and

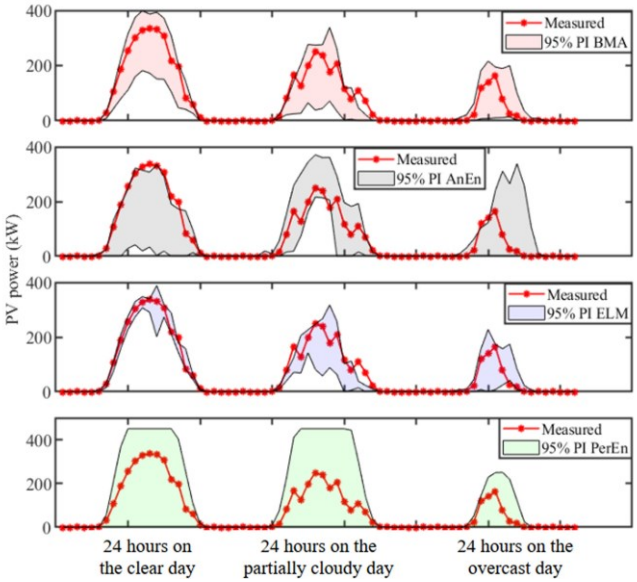

Fig. 1. Estimated 95% PIs from the BMA, AnEn, ELM and PerEn ensemble methods on the clear day (05/19/2021), partially cloudy day (07/09/2021), and overcast day (12/28/2021).

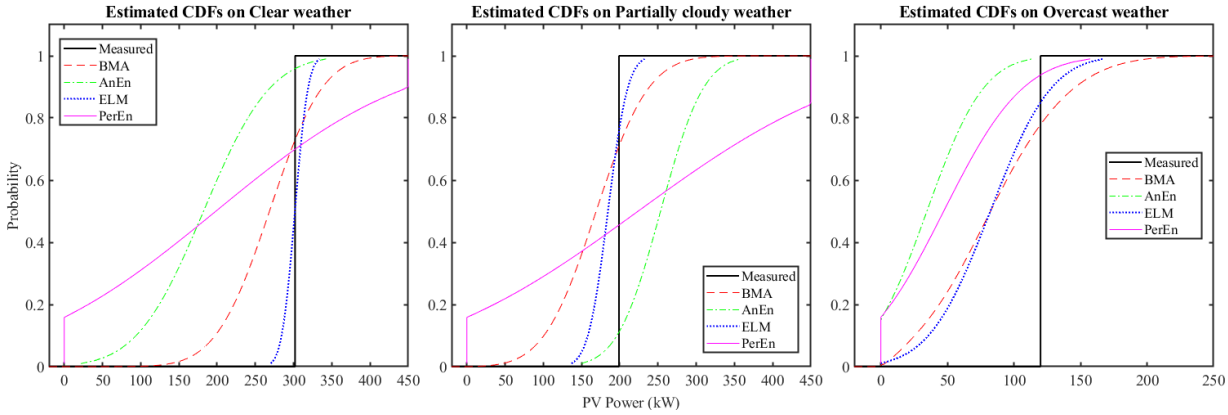

Fig. 2. Estimated CDFs from the BMA, AnEn, ELM, PerEn methods compared with the measured power at 10:00 AM on the clear day (05/19/2021), partially cloudy day (07/09/2021), and overcast day (12/28/2021).

coverage probability than the PerEn, they produce more reliable and sharper distributions. The AnEn method performs poorly compared to the other three methods under all weather conditions. The BMA and ELM methods require deterministic forecasts, whereas the AnEn method only needs historical observations. Thus, the AnEn method will be helpful when probabilistic SPF needs to be produced only from past observations. Considering forecast reliability, sharpness, resolution as well as computation time, the ELM is the best-performing method under all weather conditions. As continuous work, the authors are applying the proposed performance-comparison method on some datasets from different geographic locations and climate conditions to identify the conditions under which the study results can be generalized.

REFERENCES

[1] C. Wan, J. Zhao, Y. Song, Z. Xu, J. Lin, and Z. Hu, "Photovoltaic and solar power forecasting for smart grid energy management," *CSEE Journal of Power and Energy Systems*, vol. 1, no. 4, pp. 38–46, Jan. 2016, doi: 10.17775/CSEEJPES.2015.00046.

[2] B. Li and J. Zhang, "A review on the integration of probabilistic solar forecasting in power systems," 2020, doi: 10.1016/j.solener.2020.06.083.

[3] J. M. M. González, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, "Integrating Renewables in Electricity Markets: Operational Problems," *Springer*, vol. 205, p. 429, 2014, doi: 10.1007/978-1-4614-9411-9.

[4] P. Singla, M. Duhan, and S. Saroha, "A comprehensive review and analysis of solar forecasting techniques," *Frontiers in Energy 2021 16:2*, vol. 16, no. 2, pp. 187–223, Mar. 2021, doi: 10.1007/S11708-021-0722-7.

[5] M. Guermoui, F. Melgani, K. Gairaa, and M. L. Mekhalfi, "A comprehensive review of hybrid models for solar radiation forecasting," *J Clean Prod*, vol. 258, p. 120357, Jun. 2020, doi: 10.1016/J.JCLEPRO.2020.120357.

[6] P. Kumari and D. Toshniwal, "Deep learning models for solar irradiance forecasting: A comprehensive review," *J Clean Prod*, vol. 318, p. 128566, Oct. 2021, doi: 10.1016/J.JCLEPRO.2021.128566.

[7] B. Yang *et al.*, "Classification and summarization of solar irradiance and power forecasting methods: A thorough review," *CSEE Journal of Power and Energy Systems*, 2021, doi: 10.17775/CSEEJPES.2020.04930.

[8] P. Lauret, M. David, and H. T. C. Pedro, "Probabilistic solar forecasting using quantile regression models," *Energies 2017, Vol. 10, Page 1591*, vol. 10, no. 10, p. 1591, Oct. 2017, doi: 10.3390/EN10101591.

[9] H. Verbois, A. Rusydi, and A. Thiery, "Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting," *Solar Energy*, vol. 173, pp. 313–327, Oct. 2018, doi: 10.1016/J.SOLENER.2018.07.071.

[10] H. Sheng, J. Xiao, Y. Cheng, Q. Ni, and S. Wang, "Short-term solar power forecasting based on weighted Gaussian process regression," *IEEE Transactions on Industrial Electronics*, vol. 65, no. 1, pp. 300–308, Jan. 2018, doi: 10.1109/TIE.2017.2714127.

[11] D. Yang, D. van der Meer, and J. Munkhammar, "Probabilistic solar forecasting benchmarks on a standardized dataset at Folsom, California," *Solar Energy*, vol. 206, pp. 628–639, Aug. 2020, doi: 10.1016/J.SOLENER.2020.05.020.

[12] S. Alessandrini, L. Delle Monache, S. Sperati, and G. Cervone, "An analog ensemble for short-term probabilistic solar power forecast," *Appl Energy*, vol. 157, pp. 95–110, Nov. 2015, doi: 10.1016/J.APENERGY.2015.08.011.

[13] A. A. Mohammed and Z. Aung, "Ensemble learning approach for probabilistic forecasting of solar power generation," *Energies 2016, Vol. 9, Page 1017*, vol. 9, no. 12, p. 1017, Dec. 2016, doi: 10.3390/EN9121017.

[14] S. Sperati, S. Alessandrini, and L. Delle Monache, "An application of the ECMWF Ensemble Prediction System for short-term solar power forecasting," *Solar Energy*, vol. 133, pp. 437–450, Aug. 2016, doi: 10.1016/J.SOLENER.2016.04.016.

[15] Q. Ni, S. Zhuang, H. Sheng, G. Kang, and J. Xiao, "An ensemble prediction intervals approach for short-term PV power forecasting," *Solar Energy*, vol. 155, pp. 1072–1083, Oct. 2017, doi: 10.1016/J.SOLENER.2017.07.052.

[16] K. Doubleday, S. Jascourt, W. Kleiber, and B. M. Hodge, "Probabilistic solar power forecasting using bayesian model averaging," *IEEE Trans Sustain Energy*, vol. 12, no. 1, pp. 325–337, Jan. 2021, doi: 10.1109/TSTE.2020.2993524.

[17] J. M. L. Sloughter, T. Gneiting, and A. E. Raftery, "Probabilistic wind speed forecasting using ensembles and Bayesian model averaging," *J Am Stat Assoc*, vol. 105, no. 489, pp. 25–35, Mar. 2010, doi: 10.1198/JASA.2009.AP08615.

[18] S. Baran, "Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components," *Comput Stat Data Anal*, vol. 75, pp. 227–238, Jul. 2014, doi: 10.1016/J.CSDA.2014.02.013.

[19] P. Lauret, M. David, and P. Pinson, "Verification of solar irradiance probabilistic forecasts," *Solar Energy*, vol. 194, pp. 254–271, Dec. 2019, doi: 10.1016/J.SOLENER.2019.10.041.

[20] H. Hersbach, "Decomposition of the continuous ranked probability score for ensemble prediction systems," *Weather Forecast*, vol. 15, no. 5, pp. 559–570, Oct. 2000, doi: 10.1175/1520-0434(2000)015<0559:DOTCRP>2.0.CO;2.

[21] "NYSERDA DER Integrated Data System." https://der.nyserda.ny.gov/reports/view/performance/?project=318 (accessed Apr. 26, 2022).

[22] "Weather Data Services | Visual Crossing." https://www.visualcrossing.com/weather/weather-data-services#/editDataDefinition (accessed Apr. 26, 2022).

[23] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," 2013, doi: 10.1016/j.pecs.2013.06.002.

[24] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A review on outlier/anomaly detection in time series data," *ACM Computing Surveys (CSUR)*, vol. 54, no. 3, Apr. 2021, doi: 10.1145/3444690.

[25] T. Fushiki, "Estimation of prediction error by using K-fold cross-validation," *Statistics and Computing 2009 21:2*, vol. 21, no. 2, pp. 137–146, Oct. 2009, doi: 10.1007/S11222-009-9153-8.