

Enhancing Solar Power Forecasting with Regularized Constrained Quantile Regression Averaging and Bootstrapping Techniques

Tawsif Ahmad

Pacific Northwest National Laboratory
Richland WA, USA
E-mail: tawsif.ahmad@pnnl.gov

Ning Zhou

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

Abstract— Probabilistic solar power forecasting (SPF) plays an essential role in optimizing power-grid operations by quantifying the forecast uncertainty. To improve the accuracy and robustness of PSPF, this paper introduces the regularized constrained quantile regression averaging (rCQRA) method to combine outputs from multiple PSPF models. In addition, a bootstrapping method was used to quantify model uncertainty, providing insights into the reliability and significance of each ensemble component. To evaluate its efficacy, the proposed rCQRA method is used to integrate four PSPF methods. The resulting SPF models are trained and validated using a real-world six-year dataset from a rooftop solar plant in the USA. The performance of the proposed rCQRA method is evaluated and compared with two benchmark methods under three categories of weather conditions. It is shown that the rCQRA method has superior performance in its forecast reliability, sharpness, and accuracy.

Keywords— Probabilistic solar power forecasting, quantile regression averaging, uncertainty quantification.

I. INTRODUCTION

Integrating large amounts of PV generation brings new challenges in maintaining efficient and reliable operations of the power grid. The challenges primarily stem from the intermittent nature of PV generation caused by ever-changing and unpredictable weather conditions [1], [2]. To address this issue, solar power forecasting (SPF) has emerged as an essential tool in power grid operations, which has been widely used by utility companies and plant owners in the day-ahead energy markets [3]. Therefore, developing accurate and reliable SPF methods has captured a tremendous amount of attention from power system researchers and engineers.

Data-driven SPF methods such as statistical, machine learning, and deep learning algorithms have been widely adopted because they offer enhanced adaptability and scalability, bypassing the complexities and limitations inherent in modeling the actual physical processes of PV panels. Historically, tremendous research efforts have been made to develop deterministic SPF methods that produce a point value forecast on a forecast horizon [4]–[7]. However, these methods fall short of quantifying the uncertainty inherent in the forecasts, which is vital for scheduling generation reserves. To address this gap, probabilistic SPF (PSPF) methods have

emerged to provide forecast distributions. In the realm of PSPF methods, the persistence ensemble (PerEn) method stands out as a fundamental approach. It produces empirical distributions based on the most recent solar power data and often serves as a benchmark to evaluate the effectiveness of new methods [8]. The analog ensemble (AnEn) method, on the other hand, ranks historical solar generation data based on the similarity between the historical weather conditions and the predicted weather conditions on the forecast day and estimates the distribution by fitting the forecast errors to a pre-defined distribution [9]. Quantile regression (QR) and its derivatives are implemented in the PSPF domain, exploiting the non-parametric estimation capabilities of QR to bypass the distribution assumptions inherent in other methods [10], [11]. These methods, while producing probability distributions, operate on a single model, which potentially limit their applicability due to their confining assumptions in the model.

To overcome the limitation, ensemble PSPF methods combine multiple models to enhance forecast accuracy and reliability [12]. The ensemble learning method (ELM), one of the pioneering multi-model ensemble methods, uses multiple deterministic SPF outputs to fit a pre-defined distribution, such as the Gaussian distribution [13]. The lower-upper bound estimate (LUBE) ensemble method [14] constructs robust prediction intervals (PIs) by training a neural network that features a two-neuron output layer. Specifically, the neurons of the output layer represent the upper and lower quantile levels of the forecast. Importantly, these methods combine the forecast of deterministic SPF methods without estimating the distribution of the individual forecast outputs.

While there has been considerable research in combining deterministic SPF outputs, the realm of combining multiple PSPF outputs remains largely uncharted. A notable exception is the method proposed in [15], which combines estimated distributions from multiple deterministic SPF methods by weighted averaging using the Bayesian model averaging (BMA) technique. These individual deterministic methods utilize predefined distributions (like the beta and truncated normal distributions) to quantify forecasting uncertainty. However, the actual forecasts may not always follow the presumed distributions in real-world applications.

This paper is based on work supported by the National Science Foundation under grant no. #1845523. The opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the funding agency. The Pacific Northwest National Laboratory is operated by Battelle for DOE under Contract DE-AC05-76RL01830.

To address this challenge, this paper introduces a new approach to combine multiple PSPF outputs using quantile regression averaging (QRA). The QRA method combines quantile outputs from multiple probabilistic methods using QR, eliminating the need for pre-defined distributions of the individual methods. Whereas QRA has shown notable success in other applications such as electricity price, load, and irradiance forecasting [16]–[19], past implementations have not delved into the statistical performance of the estimated QRA coefficients. To overcome this limitation, this paper adopts a bootstrapping approach to calculate confidence intervals (CIs) of the estimated QR coefficients [20] along with the regularized constrained QRA (rCQRA) method.

The rest of the paper is organized as follows: Section II introduces the rCQRA methodology and some variants of QRA for model combination. Section III presents a case study, evaluating the efficacy of the proposed rCQRA method using real-world data and comparing its performance with other combination techniques. Conclusions are drawn in section IV.

II. QUANTILE REGRESSION AVERAGING FOR MODEL COMBINATION

In this section, the mathematical foundation of the rCQRA method is presented and formulated to combine multiple individual PSPF models.

A. Quantile Regression as an Individual SPF Model

The first step of model combination is to generate multiple individual PSPF models. The QR method is used in this subsection as an example to illustrate the procedure of building an individual model.

A linear QR model is formulated as (1). Here, y is the response variable, which is solar power generation. X is the matrix of predictors, which often includes some meteorological variables for SPF. β is the vector of QR coefficients to be estimated. e is the noise accounting for the impact of other variables not included in the model. $h \in [1, 2, \dots, 24]$ denotes the forecast hour. $q_n \in (0, 1)$ for $n \in \{1, 2, \dots, N\}$ is the n^{th} quantile level from $N+1$ divisions. Thus, the model in (1) represents the linear QR model for the n^{th} quantile level of the solar generation forecast at hour h .

$$y_h^{q_n} = X_h^{q_n} \beta_h^{q_n} + e_h. \quad (1)$$

The QR coefficient estimate $\hat{\beta}_h^q$ can be determined by solving the optimization problem defined by (2.a) and (2.b). Here, y_h^i is the solar power measurement at hour h for the i^{th} instance of the training data. $L(\beta_h^q)$ is the pinball loss function. The optimization problem in (2) can be solved using linear programming to find $\hat{\beta}_h^q$. Then, the distribution of the forecast from the QR method at current time t can be described by (3), where $\hat{Q}_{h,QR}^t(q_n) = X_h^{q_n,t} \hat{\beta}_h^{q_n}$. Note that symbols h and t in the above formulation will be the same for all forecasting methods. Thus, notations h and t will be omitted from here forth for simplicity.

$$L(\beta_h^q) = q \sum_{i: y_h^i > X_h^{q,i} \beta_h^q} (y_h^i - X_h^{q,i} \beta_h^q) + (1 - q) \sum_{i: y_h^i < X_h^{q,i} \beta_h^q} (y_h^i - X_h^{q,i} \beta_h^q) \quad (2.a)$$

$$\hat{\beta}_h^q = \arg \min_{\beta_h^q} L(\beta_h^q) \quad (2.b)$$

$$\hat{Q}_{h,QR}^t = [\hat{Q}_{h,QR}^t(q_1), \hat{Q}_{h,QR}^t(q_2), \dots, \hat{Q}_{h,QR}^t(q_N)] \quad (3)$$

B. Problem Statement for Model Combination

This paper focuses on combination methods instead of individual SPF methods. As such, it is assumed that K individual forecasting models have been established using some state-of-the-art PSPF methods, such as the QR models discussed in II.A. Each of the individual models produces a forecast distribution, which can be described by $\hat{Q}_k = [\hat{Q}_k(q_1), \hat{Q}_k(q_2), \dots, \hat{Q}_k(q_N)]$. Here, $k \in \{1, 2, \dots, K\}$ is the index of an individual model. $\hat{Q}_k(q_n)$ is the forecast from the k^{th} individual model at the n^{th} quantile level. The objective of a model combination method is to combine all the $\hat{Q}_k$ s to form an improved probability forecast at each quantile level q_n .

C. Constrained Quantile Regression Averaging

In the QRA method, forecasts from individual models are combined using the linear model represented in (4) [17], [21]. In (4), $\hat{y}_{q_n}$ is the combined probability forecast for the n^{th} quantile, while $\hat{\beta}_k(q_n)$ is the estimated QRA coefficient for the n^{th} quantile of the k^{th} model. More details on this formulation can be found in [21].

$$\hat{y}_{q_n} = \sum_{k=1}^K \hat{\beta}_k(q_n) \hat{Q}_k(q_n). \quad (4)$$

The coefficients $\hat{\beta}_k(q_n)$ in (4) can be determined by reformulating the optimization problem in (2) into (5.a). Here, y^i is the solar power measurement at the i^{th} instance of the training data. The constraints in (5.b) ensure that $\beta_k(q_n)$ adheres to the properties of a cumulative distribution function (CDF). Because of the constraints, this model is named the constrained QRA (CQRA) [17].

$$\hat{\beta}_k(q_n) = \arg \min_{\beta_k(q_n)} L(\hat{y}_{q_n}^i, y^i) = \arg \min_{\beta_k(q_n)} \left\{ \begin{array}{l} q_n \sum_{i: y^i > \hat{y}_{q_n}^i} (y^i - \hat{y}_{q_n}^i) + \\ (1 - q_n) \sum_{i: y^i < \hat{y}_{q_n}^i} (y^i - \hat{y}_{q_n}^i) \end{array} \right\} \quad (5.a)$$

$$\text{subject to: } \sum_{k=1}^K \beta_k(q_n) = 1 \text{ and } \beta_k(q_n) \geq 0 \quad (5.b)$$

D. Regularized CQRA

A regularization factor can be imposed on the coefficient $\beta_k(q_n)$ to increase the influence of the most informative PSPF

method and reduce the coefficients of less impactful methods towards zero. The regularization factor is added to the objective function in (5.a) as follows:

$$\hat{\beta}_k(q_n) = \arg \min_{\beta_k(q_n)} L(\hat{y}_{q_n}^i, y^i) + \lambda \sum_{k=1}^K |\beta_k(q_n)|. \quad (6)$$

Here, λ is the penalty factor, selected to minimize the objective function in (6). The regularization term in (6) is also known as the least absolute shrinkage and selection operator (LASSO) [22]. Hence, the proposed rCQRA includes the objective function presented in (6) and the constraints defined in (5.b).

E. Bootstrapping for Model Uncertainty

To assess the reliability and uncertainty of the estimated coefficients, a bootstrap-based CI estimation procedure is adopted in this paper. A residual bootstrap procedure is selected because of its notable efficacy for non-independent and identically distributed (non-i.i.d.) data. Conversely, for i.i.d. data, the xy -pairs bootstrap is more suitable. The residual bootstrap works by resampling from the residual vector with replacement and then re-estimating the QRA coefficients. This process is repeated multiple times (say, n_B times), generating n_B estimates of rCQRA coefficients $\hat{\beta}_k(q_n)$. This procedure allows users to estimate the sample covariance matrix of $\hat{\beta}_k(q_n)$. Subsequently, a 95% CI can be estimated from the sample covariance matrix using the percentile method [20]. The implementation of the bootstrap procedure is inspired by [23].

F. Performance Evaluation

In this paper, the continuous rank probability score (CRPS) is used to evaluate the performance of the proposed method. CRPS is a widely used metric in probabilistic forecasting evaluation and is a good measure of forecast reliability, sharpness, and accuracy [24]. To assess the reliability performance of the proposed method, the reliability diagram is used in this paper [24]. To obtain the reliability diagram, the prediction interval coverage probability (PICP) is calculated and plotted against the nominal coverage of the forecast. However, a reliable forecast does not always mean a better forecast. To assess the sharpness of the forecast, the prediction interval normalized average width (PINAW) is calculated. Taking both PICP and PINAW into account, the coverage width-based criterion (CWC) is considered, which can be used to evaluate both the reliability and sharpness of the forecast. More details on CRPS, PICP, PINAW, and CWC can be found in [24], [25]. It is important to note that lower values of CRPS and CWC signify superior method performance.

G. Competing Methods

Two other combining methods are used as benchmarks in this paper [17]. The first one, known as mean QRA (mQRA), averages the quantile forecasts from each individual SPF method as in (7).

$$\hat{y}_{q_n} = \frac{1}{K} \sum_{k=1}^K \hat{Q}_k(q_n). \quad (7)$$

The second combining technique, known as weighted QRA (wQRA), assigns weights to the individual SPF outputs according to their CRPS values. The weights of the wQRA method can be defined in (8), where $CRPS_k$ is the CRPS value of the k^{th} SPF method. Note that methods with lower CRPS values are assigned higher weights.

$$\hat{\beta}_k(q_n) = \frac{1/CRPS_k}{\sum_k 1/CRPS_k} \quad (8)$$

III. CASE STUDY

The proposed rCQRA method is applied to a real-world dataset and is compared with the traditional mQRA and wQRA methods for day-ahead forecasting. Specifically, the forecast horizon ranges from 1 hour to 24 hours ahead.

A. Dataset Information

The 450 kW rooftop solar PV plant located at Vestal, NY, USA (*lat.* 42°05'37.0"N, *long.* 76°00'06.0"W) is chosen for this study. Funded by the New York State Energy Research and Development Authority (NYSERDA), the hourly solar power observations from this plant during the years 2016-2022 were made available on NYSERDA's website [26]. The weather data for this location are collected from Visual Crossing [27]. Standard data pre-processing techniques are used to identify and remove outliers [28]. Weather variables with the highest correlation to solar generation are used as predictors of the individual SPF methods. Readers are referred to [29] for a more comprehensive discussion of these methods.

B. Construction of Individual Models

This study uses the QR, AnEn, ELM, and PerEn methods to construct individual PSPF models. In practice, any available models producing quantile forecasts can be integrated using the proposed method. As the primary focus of this study is not the performance of the individual PSPF methods, the selection and evaluation of the individual methods are not discussed in detail. The data from 2016 to 2020 are used to train and validate these models. 10-fold cross-validation is utilized in this study to reduce overfitting. The data for 2021 are used to train and validate the QRA models, i.e., the rCQRA, mQRA, and wQRA models. The data for 2022 are used for testing these models.

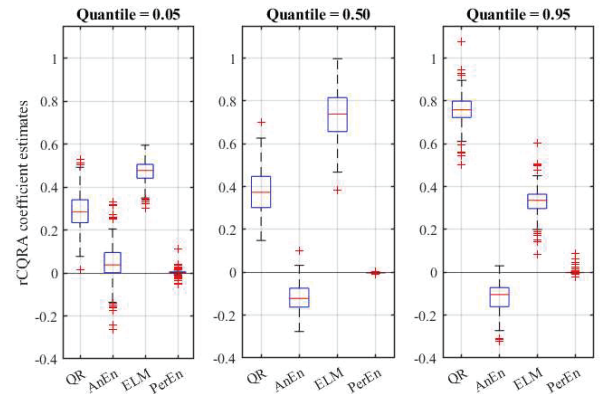


Fig. 1. Box-whisker plots showing the bootstrap estimates of model coefficients at three quantile levels (5%, 50%, and 95%) for hour 12 on the forecast day.

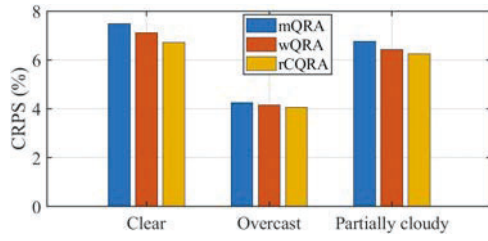


Fig. 2. CRPS values under three different weather conditions.

The performances of these models are evaluated under three weather conditions, i.e., *clear*, *overcast*, and *partially cloudy*.

C. Model Uncertainty

The bootstrap procedure is used to quantify model uncertainty. The bootstrap estimates of the rCQRA coefficients for three quantiles for a certain hour of the day are shown in the box-whisker plots in Fig. 1. It is evident that the QR and ELM models' coefficients for the rCQRA model are positive across all quantiles. The boxplots do not include '0', which means that these models' outputs have significant interpretability for the rCQRA performance. The AnEn method has a mixed contribution to the rCQRA model. For the 5% quantile level, the AnEn coefficient's boxplot contains '0' well within its boundary, showing that the AnEn forecast might not have a significant impact on the rCQRA estimates. PerEn coefficients estimates do not participate in the rCQRA estimates for all three quantiles, as evident from the boxplot in Fig. 1. The coefficients are close to '0'. This is because the regularization penalty introduced in the rCQRA objective function shrinks the PerEn coefficients to '0'. The bootstrap procedure can be helpful for users in selecting individual models to combine when they have a large number of models at their disposal.

D. rCQRA Performance under Different Weather Conditions

The proposed rCQRA method is evaluated under the three weather conditions. The CRPS and CWC metrics, along with the reliability diagram, are used to quantify the performance of the proposed method. The CRPS and CWC values for the mQRA, wQRA, and rCQRA methods under the three weather conditions are shown in Fig. 2 and Fig. 3, respectively. The reliability diagrams for the three methods are shown in Fig. 4. (a)-(c) for the three weather conditions. Under all the weather conditions, the rCQRA method consistently records lower CRPS values than the mQRA and wQRA methods, indicating

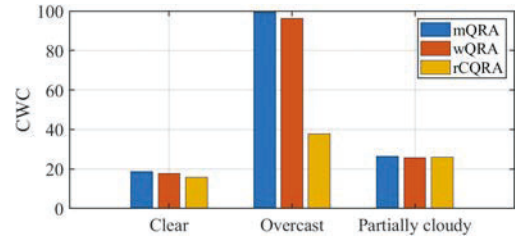


Fig. 3. CWC values under three different weather conditions.

lower deviations from the actual solar power measurements, i.e., improved accuracy and reliability. This reliability performance is further illustrated by the reliability diagrams in Fig. 4. During *clear* conditions (as depicted in Fig. 4.(a)), the rCQRA exhibits near-perfect PICP. In contrast, the other two methods deviate significantly from the nominal curve. However, a near-perfect reliability diagram does not necessarily equate to a superior forecast. A forecast with a wider PI can exhibit perfect reliability but with a compromise in the sharpness of the forecast. The CWC metric takes both the reliability (PICP) and sharpness (PINAW) into account, which makes it suitable for further assessment on top of the reliability diagram. The CWC values for the *clear* weather in Fig. 3 indicate that the improved reliability of the rCQRA method comes with an improvement in the sharpness of the forecast.

Under the *overcast* condition, rCQRA has larger deviations from the nominal curve for lower nominal coverages compared to the mQRA and wQRA methods (Fig. 4.(b)). However, for higher nominal coverages, the rCQRA shows better reliability. This statement can also be verified by the significantly lower CWC value for the rCQRA method under *overcast* conditions, as shown in Fig. 3. Finally, Fig. 4.(c) demonstrates the reliability diagrams for the three methods under *partially cloudy* weather conditions. The PICP for the rCQRA method exhibits lower deviations from the nominal curve, indicating better reliability. However, the CWC values in Fig. 3. for *partially cloudy* conditions show that the rCQRA improvement over other methods is not that significant. Overall, the rCQRA method shows notable advantages over the benchmark combination methods across all weather conditions.

Fig. 5 compares the estimated 95% PIs for a single day from the test set with the actual measurements. The rCQRA method produces narrower PIs than the other methods and

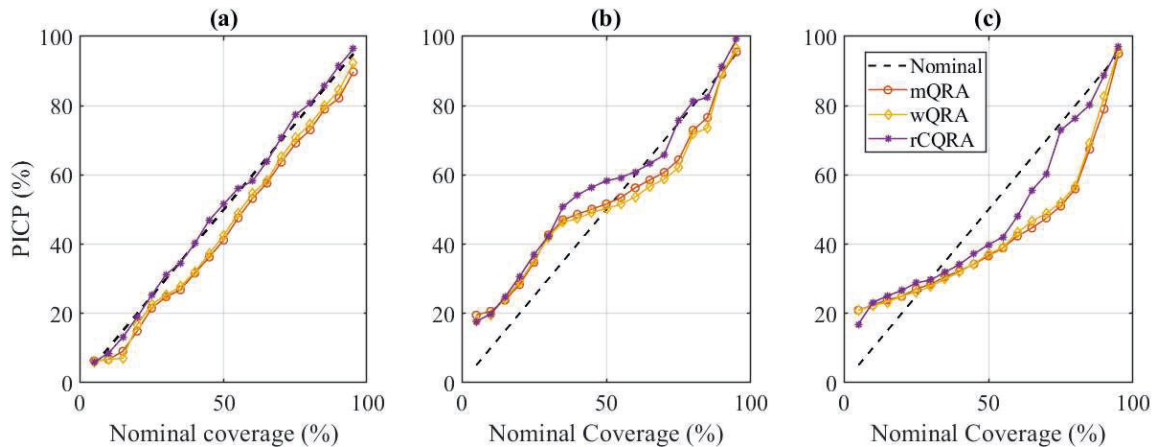


Fig. 4. Reliability diagrams for the three methods under three weather conditions: (a) *clear*, (b) *overcast*, and (c) *partially cloudy*.

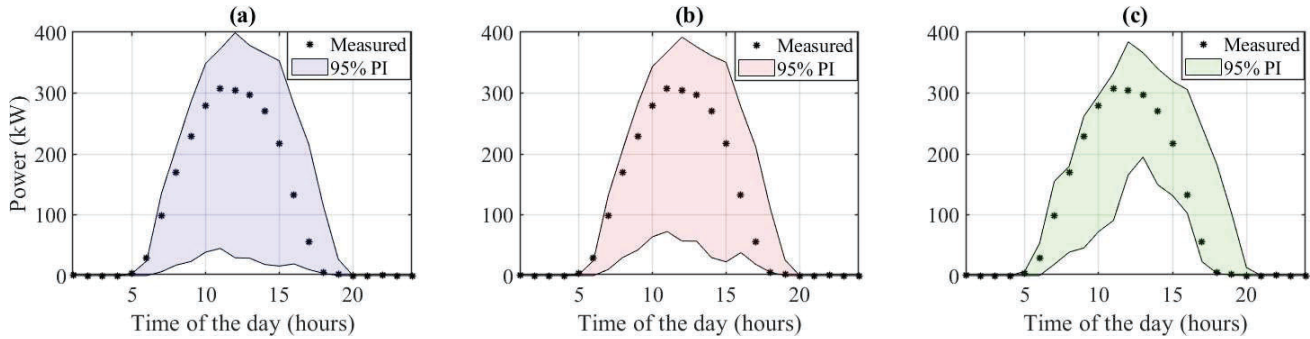


Fig. 5. Estimated 95% PIs compared with actual measurements for (a) mQRA, (b) wQRA, and (c) rCQRA.

maintains good reliability. This observation is consistent with the results shown in Fig. 3 and Fig. 4.

IV. CONCLUSIONS

In this paper, the rCQRA method is proposed to combine multiple PSPF methods. Through the integration of a regularization penalty, the rCQRA method allows users to identify redundant models and models that have insignificant contributions to the combined forecast. Thus, the rCQRA is particularly useful when the number of models to be combined is large. In addition, a bootstrap procedure is adopted to assess the uncertainty in the participating models. When simultaneously implemented, the rCQRA with the bootstrapping can produce robust, reliable, and sharp forecasts across diverse weather conditions. Future studies will involve working with datasets from different geographic locations and climate conditions so that the results in this study 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 J. Power Energy Syst.*, vol. 1, no. 4, pp. 38–46, Dec. 2015.
- [2] B. Li and J. Zhang, "A review on the integration of probabilistic solar forecasting in power systems," *Sol. Energy*, vol. 210, pp. 68–86, Nov. 2020.
- [3] J. M. Morales González, A. J. Conejo, H. Madsen, P. Pinson, and M. Zugno, *Integrating Renewables in Electricity Markets: Operational Problems*. in International Series in Operations Research and Management Science. Springer, 2014.
- [4] P. Singla, M. Duhan, and S. Saroha, "A comprehensive review and analysis of solar forecasting techniques," *Front. Energy*, vol. 16, no. 2, pp. 187–223, Apr. 2022.
- [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.
- [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.
- [7] B. Yang *et al.*, "Classification and Summarization of Solar Irradiance and Power Forecasting Methods: A Thorough Review," *CSEE J. Power Energy Syst.*, vol. 9, no. 3, pp. 978–995, May 2023.
- [8] D. Yang, D. van der Meer, and J. Munkhammar, "Probabilistic solar forecasting benchmarks on a standardized dataset at Folsom, California," *Sol. Energy*, vol. 206, pp. 628–639, Aug. 2020.
- [9] 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.
- [10] P. Lauret, M. David, and H. T. C. Pedro, "Probabilistic Solar Forecasting Using Quantile Regression Models," *Energies*, vol. 10, no. 10, Art. no. 10, Oct. 2017.
- [11] H. Verbois, A. Rusydi, and A. Thiery, "Probabilistic forecasting of day-ahead solar irradiance using quantile gradient boosting," *Sol. Energy*, vol. 173, pp. 313–327, Oct. 2018.
- [12] T. Ahmad and N. Zhou, "Ensemble Methods for Probabilistic Solar Power Forecasting: A Comparative Study," in *2023 IEEE Power & Energy Society General Meeting (PESGM)*, Jul. 2023, pp. 1–5.
- [13] A. Ahmed Mohammed and Z. Aung, "Ensemble Learning Approach for Probabilistic Forecasting of Solar Power Generation," *Energies*, vol. 9, no. 12, Art. no. 12, Dec. 2016.
- [14] Q. Ni, S. Zhuang, H. Sheng, G. Kang, and J. Xiao, "An ensemble prediction intervals approach for short-term PV power forecasting," *Sol. Energy*, vol. 155, pp. 1072–1083, Oct. 2017.
- [15] K. Doubleday, S. Jascourt, W. Kleiber, and B.-M. Hodge, "Probabilistic Solar Power Forecasting Using Bayesian Model Averaging," *IEEE Trans. Sustain. Energy*, vol. 12, pp. 325–337, Jan. 2021.
- [16] J. Nowotarski and R. Weron, "Computing electricity spot price prediction intervals using quantile regression and forecast averaging," *Comput. Stat.*, vol. 30, no. 3, pp. 791–803, Sep. 2015.
- [17] Y. Wang, N. Zhang, Y. Tan, T. Hong, D. S. Kirschen, and C. Kang, "Combining Probabilistic Load Forecasts," *IEEE Trans. Smart Grid*, vol. 10, no. 4, pp. 3664–3674, Jul. 2019.
- [18] B. Uniejewski and R. Weron, "Regularized quantile regression averaging for probabilistic electricity price forecasting," *Energy Econ.*, vol. 95, p. 105121, Mar. 2021.
- [19] D. Yang, G. Yang, and B. Liu, "Combining quantiles of calibrated solar forecasts from ensemble numerical weather prediction," *Renew. Energy*, vol. 215, p. 118993, Oct. 2023.
- [20] J. Hahn, "Bootstrapping Quantile Regression Estimators," *Econom. Theory*, vol. 11, no. 1, pp. 105–121, 1995.
- [21] C. W. J. Granger, "Invited review combining forecasts—twenty years later," *J. Forecast.*, vol. 8, no. 3, pp. 167–173, 1989.
- [22] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
- [23] T. Ahmad *et al.*, "Estimation of the Correlation Between Oscillation Modes and Operating Conditions using Quantile Regression: A Measurement-Based Approach," in *2020 52nd North American Power Symposium (NAPS)*, Apr. 2021, pp. 1–6.
- [24] P. Lauret, M. David, and P. Pinson, "Verification of solar irradiance probabilistic forecasts," *Sol. Energy*, vol. 194, pp. 254–271, Dec. 2019.
- [25] X. He, Z. Lei, H. Jing, and R. Zhong, "Short-Term Probabilistic Forecasting Method for Wind Speed Combining Long Short-Term Memory and Gaussian Mixture Model," *Atmosphere*, vol. 14, no. 4, Art. no. 4, Apr. 2023.
- [26] "NYSDERDA DER Integrated Data System." Accessed: Oct. 16, 2023. [Online]. Available: <https://der.nysderda.ny.gov/reports/view/performance/?project=318>
- [27] "Weather Data Services | Visual Crossing." Accessed: Oct. 16, 2023. [Online]. Available: <https://www.visualcrossing.com/weather/weather-data-services#/editDataDefinition>
- [28] A. Blázquez-García, A. Conde, U. Mori, and J. A. Lozano, "A Review on Outlier/Anomaly Detection in Time Series Data," *ACM Comput. Surv.*, vol. 54, no. 3, p. 56:1–56:33, Apr. 2021.
- [29] R. H. Inman, H. T. C. Pedro, and C. F. M. Coimbra, "Solar forecasting methods for renewable energy integration," *Prog. Energy Combust. Sci.*, vol. 39, no. 6, pp. 535–576, Dec. 2013.