DTU Library

Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach

Wan, Can; Lin, Jin; Song, Yonghua; Xu, Zhao; Yang, Guangya

Link back to DTU Orbit

Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach

Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach

Can Wan, *Member, IEEE,* Jin Lin, *Member, IEEE,* Yonghua Song, *Fellow, IEEE,*
Zhao Xu, *Senior Member, IEEE,* Guangya Yang, *Member, IEEE*

Abstract—**This letter proposes a novel efficient probabilistic forecasting approach to accurately quantify the variability and uncertainty of the power production from photovoltaic (PV) systems. Distinguished from most existing models, a linear programming based prediction interval construction model for PV power generation is proposed based on extreme learning machine and quantile regression, featuring high reliability and computational efficiency. The proposed approach is validated through the numerical studies on PV data from Denmark.**

Index Terms—**PV power, forecasting, prediction intervals, extreme learning machine, quantile regression.**

I. INTRODUCTION

Due to large-scale integration of PV systems with significant intermittency and uncertainty, PV generation forecasting is crucial to assist optimal operation and control of power systems [1]. Considering the chaotic nature of weather systems, prediction errors of PV power can be unavoidable. In previous studies, PV forecasting uncertainty is roughly modeled as normal distribution [2], [3], which can contradict the actual conditions. Probabilistic forecasting would be very important to quantify the uncertainty of PV systems. To the best knowledge of the authors, there are few studies on this topic in the literature. In this letter, a unique efficient statistical approach is developed to generate the quality prediction intervals (PIs). This approach uses extreme learning machine (ELM) as the predictor [4] that is a novel algorithm for training a single hidden-layer feedforward neural network. The objective of PI formulation is established

via quantile regression [5], then the PIs construction model is transformed to a simple linear programming problem, which can be resolved with high accuracy and computational efficiency. In general, the proposed approach can efficiently produce optimal PIs of PV generation and have high potential for online decision-making activities of the integration of PV in modern power systems, including voltage control [2], probabilistic load flow [6], power management of PV plants [7], storage system operation [8], etc.

II. METHODOLOGY

A. Basics of Nonparametric PIs

Denote the PV power production measured (to be forecasted) at time t by y_t, and let x_t be the input vector of the prediction model. The quantile q_t^β with probability $\beta \in [0, 1]$ of the target PV y_t is defined as,

$$\Pr(y_t \le q_t^\beta) = \beta \tag{1}$$

Prediction intervals can give pairs of bounds within which the realized value is expected to be covered with a certain probability. A PI I_t^α of PV power at time t with nominal coverage rate (NCR) $1-\alpha$, $\alpha \in [0, 1]$, can be defined as,

$$I_t^\alpha = \left[q_t^\alpha, \ q_t^{\bar{\alpha}} \right] \tag{2}$$

where q_t^α and $q_t^{\bar{\alpha}}$ denote the lower and upper bounds of the PI respectively. Generally, the PI is set to be centered on the probability density function, which can be represented as,

$$\underline{\alpha} = 1 - \bar{\alpha} = \alpha/2 \tag{3}$$

The future prediction target of PV power production y_t is expected to be enclosed by the PI with the coverage rate,

$$\Pr(Y_t \in I_t^\alpha) = 1 - \alpha \tag{4}$$

B. Optimization Problem Formulation

Based on the definition of quantile in (1), the quantile q_t^β of PV prediction uncertainty can be uniquely approximated through minimizing properly formulated cost function [5], expressed as,

$$\min \sum_{t=1}^{T} \ell_\beta \left(y_t - q_t^\beta \right) \tag{5}$$

where T is the size of training samples, $\ell_\beta(\cdot)$ is the asymmetrical absolute function defined by,

$$\ell_\beta \left(y_t - q_t^\beta \right) = \begin{cases} \beta(y_t - q_t^\beta) & \text{if } y_t - q_t^\beta \ge 0 \\ (\beta - 1)(y_t - q_t^\beta) & \text{if } y_t - q_t^\beta < 0 \end{cases} \tag{6}$$

Because of ELM's input weights and hidden biases are

randomly produced, training a single hidden-layer feedforward neural network just becomes searching a least-square solution of a linear system [4], which motivates its application in this work. According to the definition of PIs in (2) and the approximation model of quantiles in (5) and (6), an innovative PIs construction model can be established through the minimization of the cost function formulated as,

$$\min_{w_{\underline{\alpha}},w_{\bar{\alpha}}} \sum_{t=1}^{T} \ell_{\underline{\alpha}}\left(y_t - f(x_t,w_{\underline{\alpha}})\right) + \ell_{\bar{\alpha}}\left(y_t - f(x_t,w_{\bar{\alpha}})\right) \qquad (7)$$

$$\text{s.t.} \qquad f(x_t,w_{\underline{\alpha}}) \le f(x_t,w_{\bar{\alpha}}) \qquad (8)$$

where $f(x_t,w_{\underline{\alpha}})$ and $f(x_t,w_{\bar{\alpha}})$ represent the linear systems of ELM for the lower and upper bounds of PIs respectively, the decision variables $w_{\underline{\alpha}}$ and $w_{\bar{\alpha}}$ represent the parameters of the ELM, and the constraint (8) ensures that the upper bounds are above the lower bounds. By introducing auxiliary variables $\gamma_t^{\underline{\alpha}}, \hat{\gamma}_t^{\underline{\alpha}}, \gamma_t^{\bar{\alpha}}, \hat{\gamma}_t^{\bar{\alpha}}$, the problem in (7) and (8) can be transformed to an equivalent linear optimization problem, given as,

$$\min_{\substack{w_{\underline{\alpha}},w_{\bar{\alpha}} \\ \gamma_t^{\underline{\alpha}},\hat{\gamma}_t^{\underline{\alpha}},\gamma_t^{\bar{\alpha}},\hat{\gamma}_t^{\bar{\alpha}}}} \sum_{t=1}^{T} \underline{\alpha}\gamma_t^{\underline{\alpha}} + (1-\underline{\alpha})\hat{\gamma}_t^{\underline{\alpha}} + \bar{\alpha}\gamma_t^{\bar{\alpha}} + (1-\bar{\alpha})\hat{\gamma}_t^{\bar{\alpha}} \qquad (9)$$

$$\text{s.t.} \quad y_t - f(x_t,w_{\underline{\alpha}}) = \gamma_t^{\underline{\alpha}} - \hat{\gamma}_t^{\underline{\alpha}} \qquad (10)$$

$$y_t - f(x_t,w_{\bar{\alpha}}) = \gamma_t^{\bar{\alpha}} - \hat{\gamma}_t^{\bar{\alpha}} \qquad (11)$$

$$\gamma_t^{\underline{\alpha}},\hat{\gamma}_t^{\underline{\alpha}},\gamma_t^{\bar{\alpha}},\hat{\gamma}_t^{\bar{\alpha}} \ge 0 \qquad (12)$$

$$f(x_t,w_{\bar{\alpha}}) - f(x_t,w_{\underline{\alpha}}) \ge 0 \qquad (13)$$

It can be found that the formulated optimization problem (9)-(13) can be efficiently solved by linear programming algorithms. This can guarantee the significantly high optimization quality and computational efficiency of the proposed approach.

C. Assessment of PIs

1) Reliability: Reliability is a primary index indicating the probabilistic validity of PIs. The empirical coverage rate (ECR) of PIs is a key measure, defined by,

$$ECR = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{y_t \in I_t^{\alpha}\} \qquad (14)$$

where $\mathbf{1}\{\cdot\}$ is a binary variable equivalent to 1 if the observation lies within the PIs, and to 0 otherwise. The coverage rate deviation (CRD) gives the deviation between ECR and NCR,

$$CRD = ECR - (1-\alpha) \qquad (15)$$

The smaller absolute CRD means the higher reliability.

2) Overall Skill: The interval score can be used to evaluate the overall skill of PIs to involve all aspects of PIs properties. The interval score of obtained PIs based on the entire training dataset can be expressed as,

$$S_t^{\alpha} = \frac{1}{T}\sum_{t=1}^{T}\left[q_t^{\bar{\alpha}} - q_t^{\underline{\alpha}}\right] + 2/\alpha\left[q_t^{\underline{\alpha}} - y_t\right]\mathbf{1}\{y_t < q_t^{\underline{\alpha}}\}$$
$$+ 2/\alpha\left[y_t - q_t^{\bar{\alpha}}\right]\mathbf{1}\{y_t > q_t^{\bar{\alpha}}\} \qquad (16)$$

The score is negatively oriented, i.e., the smaller score, the better PIs.

III. NUMERICAL STUDY

In the study, a realistic PV generation system with nominal capacity P_n=10 kW from Denmark is utilized for the numerical analysis to validate the effectiveness and efficiency of the proposed approach. The 5-min resolution PV generation data covering June, September and October 2014 are collected for the study. Due to the absence of solar irradiance, the PV system has constant zero power production in night, which are not needed to forecast. Therefore, PV generation data during the night are disregarded in the study.

Climatology approach that can be considered as an unconditional quantile regression method is applied as naïve benchmark in the study. Its nonparametric predictive distribution can be obtained on basis of observed PV generation. The well-established persistence approach for very short-term forecasting of PV is used as an essential benchmark [1], of which the prediction error is normally distributed. In addition, the bootstrap based neural network (BNN) model is applied as an advanced benchmark [9]. About 60% of the PV data are used for training the prediction model, and the rest data are used for test. Different look-ahead times ranging from seconds to weeks correspond to different practical applications in power systems [1]. For instance, the regulation operational reserve is responsible for balancing the difference between the scheduled generation and actual load in the intra-hour market ranging from 5 min to 30 min [3]. The case studies in this letter focus on very short-term probabilistic PV power forecasting with 5-min and 10-min look-ahead times. The numerical results are listed in Tables I and II.

TABLE I
PERFORMANCE OF PIS WITH 5-MIN LOOK-AHEAD TIME

NCP	Method	ECR	CRD	Score
80%	Climatology	80.85%	0.85%	0.5960
	Persistence	92.29%	12.29%	0.0985
	BNN	93.08%	13.08%	0.0931
	Proposed Method	80.35%	**0.35%**	**0.0690**
90%	Climatology	90.75%	0.75%	0.6452
	Persistence	93.58%	3.58%	0.1392
	BNN	94.66%	4.66%	0.1280
	Proposed Method	90.28%	**0.28%**	**0.1109**

TABLE II
PERFORMANCE OF PIS WITH 10-MIN LOOK-AHEAD TIME

NCP	Method	ECR	CRD	Score
80%	Climatology	80.85%	0.85%	0.5960
	Persistence	88.20%	8.20%	0.1405
	BNN	90.14%	10.14%	0.1285
	Proposed Method	80.39%	**0.39%**	**0.1151**
90%	Climatology	90.75%	0.75%	0.6452
	Persistence	90.86%	0.86%	0.1931
	BNN	92.97%	2.97%	0.1693
	Proposed Method	89.78%	**-0.22%**	**0.1626**

From Tables I and II, the proposed method demonstrates much better performance than the three benchmarks. PIs obtained by the proposed approach have pretty good reliability, with CRDs less than 0.5%. Particularly, the persistence and BNN models have very low reliability for PIs of 80% NCR, with

CRDs larger than or close to 10%. This explicitly implies that normal distribution cannot accurately describe the stochastic characteristics of PV prediction uncertainty. In addition, the proposed method achieves the smallest interval score values, demonstrating the best overall skill. As a nonparametric approach, the climatology obtains high reliability with CRDs less than 1%, but it causes poor sharpness due to the unconditional estimation. It can be found from Tables I and II that the interval scores of PIs with look-ahead time 10 min are larger than that of look-ahead time 5 min, which should be reasonable that larger uncertainty would be involved in the prediction of longer look-ahead time.

TABLE III
COMPARISONS OF MODEL TRAINING TIME

Method	Time (s)
BNN	2504.53
Proposed Method	8.49

The computational efficiency of the proposed approach is studied through comparison with BNN, as illustrated in Table III. The numerical experiment is fulfilled on a PC with Intel Core Duo 3.6GHz CPU and 16GB RAM. From Table III, traditional BNN requires about 300 times more computation time than the proposed approach for model training, which shows the extremely high efficiency of the proposed approach. It indicates the high potential of the proposed approach for online applications in power systems.
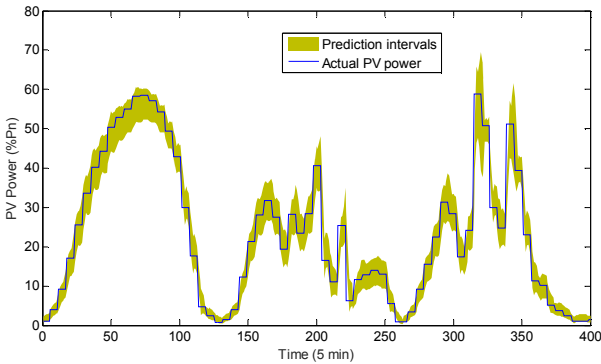

Fig. 1. PIs with NCR 90% and 5-min look-ahead time in Oct. 2014 obtained by the proposed approach.
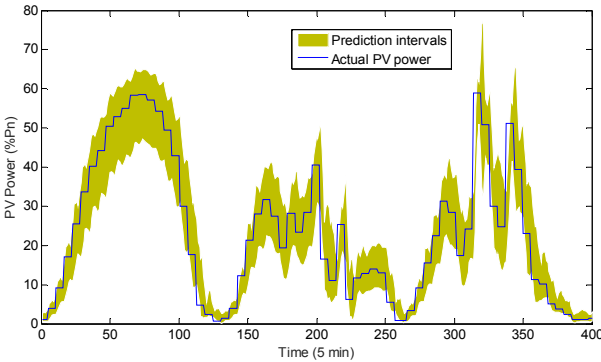

Fig. 2. PIs with NCR 90% and 10-min look-ahead time in Oct. 2014 obtained by the proposed approach.

The PIs with NCR 90% and 5-min and 10-min look-ahead times approximated by the proposed approach and observed PV power in three days of Oct. 2014 are respectively depicted in Figs. 1 and 2 to visually demonstrate the significant performance. It can be found that PV power Actual PV productions are well covered by the constructed PIs. In addition, it is apparent that the width of PIs with 5-min look-ahead time is much wider than that with 10-min look-ahead time, which indicates that longer look-ahead time can lead to larger uncertainty.

IV. CONCLUSION

This letter proposes a novel efficient statistical approach to produce prediction intervals of PV generation, which sufficiently takes the advantages of ELM and quantile estimation. The proposed approach formulates the complicated nonparametric PI construction problem as a simple linear programming model. Preliminary experiments have verified the high computational efficiency and superiority of the proposed approach for probabilistic forecasting of PV power.

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, Dec. 2015.

[2] Z. Ziadi, M. Oshiro, T. Senjyu, A. Yona, N. Urasaki, T. Funabashi, *et al.*, "Optimal voltage control using inverters interfaced with PV systems considering forecast error in a distribution system," *IEEE Trans. Sustain. Energy,* vol. 5, no. 2, pp. 682-690, Apr. 2014.

[3] M. D. Tabone and D. S. Callaway, "Modeling variability and uncertainty of photovoltaic generation: A hidden state spatial statistical approach," *IEEE Trans. Power Syst.,* vol. 30, no. 6, pp. 2965-2973, Nov. 2015.

[4] G. B. Huang, H. Zhou, X. Ding, and R. Zhang, "Extreme learning machine for regression and multiclass classification," *IEEE Trans. Syst. Man Cybern. B, Cybern.,* vol. 42, no. 2, pp. 513-529, Apr. 2012.

[5] R. Koenker, *Quantile Regression*: Cambridge University Press, 2005.

[6] M. Fan, V. Vittal, G. T. Heydt, and R. Ayyanar, "Probabilistic power flow analysis with generation dispatch including photovoltaic resources," *IEEE Trans. Power Syst.,* vol. 28, no. 2, pp. 1797-1805, May 2013.

[7] G. Wang, M. Ciobotaru, and V. G. Agelidis, "Power management for improved dispatch of utility-scale PV plants," *IEEE Trans. Power Syst.,* vol. 31, no. 3, pp. 2297-2306, May 2016.

[8] M. J. E. Alam, K. M. Muttaqi, and D. Sutanto, "Mitigation of rooftop solar PV impacts and evening peak support by managing available capacity of distributed energy storage systems," *IEEE Trans. Power Syst.,* vol. 28, no. 4, pp. 3874-3884, Nov. 2013.

[9] G. Papadopoulos, P. J. Edwards, and A. F. Murray, "Confidence estimation methods for neural networks: A practical comparison," *IEEE Trans. Neural Netw.,* vol. 12, no. 6, pp. 1278-1287, Nov. 2001.