Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach

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

Published in:
IEEE Transactions on Power Systems

Link to article, DOI:
10.1109/TPWRS.2016.2608740

Publication date:
2017

Document Version
Peer reviewed version

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
# Probabilistic Forecasting of Photovoltaic Generation: An Efficient Statistical Approach

<table>
<thead>
<tr>
<th>Journal:</th>
<th>IEEE Power Engineering Letters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manuscript ID</td>
<td>PESL-00003-2016.R1</td>
</tr>
<tr>
<td>Manuscript Type:</td>
<td>Letters</td>
</tr>
<tr>
<td>Date Submitted by the Author:</td>
<td>29-Apr-2016</td>
</tr>
<tr>
<td>Complete List of Authors:</td>
<td>Wan, Can; Tsinghua University, Department of Electrical Engineering; Hong Kong Polytechnic University, Department of Electrical Engineering Lin, Jin; Tsinghua University, Electrical Engineering Song, Yonghua; Smart Grid Optimization and Operation Laboratory, Department of Electrical Engineering, Tsinghua University Xu, Zhao; Hong Kong Polytechnic University, Department of Electrical Engineering Yang, Guangya; Technical University of Denmark, Department of Electrical Engineering;</td>
</tr>
<tr>
<td>Key Words:</td>
<td>PV power, Forecasting, prediction intervals, extreme learning machine, quantile regression</td>
</tr>
</tbody>
</table>
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 $q^\beta$ with probability $\beta \in [0, 1]$ of the target PV $y_t$ is defined as,
\[
\Pr(y_t \leq q^\beta_t) = \beta
\]

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

where $q^\alpha$ and $q^\beta$ 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,
\[
\alpha = 1-\bar{\alpha} = \alpha/2
\]
The future prediction target of PV power production $y_t$ is expected to be enclosed by the PI with the coverage rate,
\[
\Pr(Y_t \in I^\alpha_t) = 1-\alpha
\]

B. Optimization Problem Formulation

Based on the definition of quantile in (1), the quantile $q^\beta$ of PV prediction uncertainty can be uniquely approximated through minimizing properly formulated cost function [5], expressed as,
\[
\min \sum_{t=1}^{T} \ell_\beta(y_t - q^\beta_t)
\]

where $T$ is the size of training samples, $\ell_\beta(\cdot)$ is the asymmetrical absolute function defined by,
\[
\ell_\beta(y_t - q^\beta_t) = \begin{cases} 
\beta (y_t - q^\beta_t) & \text{if } y_t - q^\beta_t \geq 0 \\
(\beta-1)(y_t - q^\beta_t) & \text{if } y_t - q^\beta_t < 0 
\end{cases}
\]
and is a binary variable equivalent to 1 if the observation
represents the parameters of the
\( \hat{\text{E}} \) represents the linear systems of
ELM for the lower and upper bounds of PIs respectively, the
decision variables \( w_{\text{u}} \) and \( w_{\text{p}} \) represent the parameters of the
ELM, and the constraint (8) ensures that the upper bounds are
above the lower bounds. By introducing auxiliary
variables \( \gamma^u, \gamma^p, \gamma^u, \gamma^p, \gamma^u, \gamma^p \), the problem in (7) and (8) can be
transformed to an equivalent linear optimization problem, given
as,

\[
\min \sum_{q=1}^T y_q \gamma_q^u + (1-\alpha) \gamma_q^p + \tilde{\alpha} \gamma_q^p + (1-\tilde{\alpha}) \gamma_q^p \\
\text{s.t.} \quad y_q - f(x_q, w_{\text{u}}) = \gamma_q^u - \gamma_q^p \\
\quad y_q - f(x_q, w_{\text{p}}) = \gamma_q^p - \gamma_q^p \\
\quad \gamma_q^u, \gamma_q^p, \gamma_q^u, \gamma_q^p, \gamma_q^u, \gamma_q^p \geq 0 \\
\quad f(x_q, w_{\text{u}}) - f(x_q, w_{\text{p}}) \geq 0
\]

(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,

\[
ECR = \frac{1}{T} \sum_{i=1}^T I\{y_i \in I^n_i\}
\]

(14)

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

\[
CRD = ECR - (1-\alpha)
\]

(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,

\[
S^q_i = \frac{1}{T} \sum_{t=1}^T q^n_t - q^u_t + 2/\alpha [q^n_t - y_t] I\{y_t < q^u_t\} + 2/\alpha [y_t - q^u_t] I\{y_t > q^u_t\}
\]

(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 \( P_n = 10 \text{ kW} \) 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>
<thead>
<tr>
<th>Table I</th>
<th>PERFORMANCE OF PIS WITH 5-MIN LOOK-AHEAD TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCP Method</td>
</tr>
<tr>
<td>80%</td>
<td>Climatology</td>
</tr>
<tr>
<td></td>
<td>Persistence</td>
</tr>
<tr>
<td></td>
<td>BNN</td>
</tr>
<tr>
<td>90%</td>
<td>Proposed Method</td>
</tr>
<tr>
<td></td>
<td>Climatology</td>
</tr>
<tr>
<td></td>
<td>Persistence</td>
</tr>
<tr>
<td></td>
<td>BNN</td>
</tr>
<tr>
<td></td>
<td>Proposed Method</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table II</th>
<th>PERFORMANCE OF PIS WITH 10-MIN LOOK-AHEAD TIME</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>NCP Method</td>
</tr>
<tr>
<td>80%</td>
<td>Climatology</td>
</tr>
<tr>
<td></td>
<td>Persistence</td>
</tr>
<tr>
<td></td>
<td>BNN</td>
</tr>
<tr>
<td>90%</td>
<td>Proposed Method</td>
</tr>
<tr>
<td></td>
<td>Climatology</td>
</tr>
<tr>
<td></td>
<td>Persistence</td>
</tr>
<tr>
<td></td>
<td>BNN</td>
</tr>
<tr>
<td></td>
<td>Proposed Method</td>
</tr>
</tbody>
</table>

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
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