Applied Energy 228 (2018) 700–711

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Prediction of short-term PV power output and uncertainty analysis

Luyao Liu[a], Yi Zhao[a], Dongliang Chang[b], Jiyang Xie[b], Zhanyu Ma[b,*], Qie Sun[a,*], Hongyi Yin[a,*], Ronald Wennersten[a]

[a] Institute of Thermal Science and Technology, Shandong University, Jinan, China
[b] Pattern Recognition and Intelligent System Lab., Beijing University of Posts and Telecommunications, China

HIGHLIGHTS

- A two-stage model is developed to quantify the uncertainty of PV power forecasts.
- An integrated artificial neural network model is built to perform point forecast.
- The prediction intervals (PI) associated to the point forecasts are evaluated.
- The PIs cover more comprehensive information than the point forecasts.

ARTICLE INFO

Keywords:
PV power prediction
Neural network
Point forecast
Interval forecast
Prediction interval estimation
Probability distribution

ABSTRACT

Due to the intermittency and uncertainty in photovoltaic (PV) power outputs, not only deterministic point predictions (DPPs), but also associated prediction Intervals (PIs) are important information for promoting the application of PV in practice, especially when grid connection continues to grow. While there are few studies focused on quantifying the uncertainty of forecasting PV power outputs, this paper developed a novel two-stage model to quantify the PIs of PV power outputs. In the first stage, three different neural networks, namely Generalized Regression Neural Network (GRNN), Extreme Learning Machine Neural Network (ELMNN) and Elman Neural Network (ElmanNN), were integrated using the Genetic Algorithms optimized Back Propagation (GA-BP) method to develop a Weight-Varying Combination Forecast Mode (WVCFM) model. The WVCFM model was applied to generate DPPs. In the second stage, a nonparametric kernel density estimation (NKDE) method was adopted to estimate the PIs regarding the statistical distribution of the errors of the DPPs derived in the first stage. The proposed method was tested using four types of PV output and weather data measured from a 15 kW grid-connected PV system. The cover percentage of prediction intervals (PICPs) were computed under the confidence level of 95%, 90%, 85% and 80%, respectively. The results imply that the two-stage model proposed in the paper outperforms conventional forecast methods in terms of prediction of short-term PV power outputs and associated uncertainties.

1. Introduction

Due to the restriction by resources reserves and environmental problems of fossil energy, the development and utilization of renewable energy has become the inevitable trend of the energy transition worldwide [1,2]. As a clean energy application with broad prospects, photovoltaic (PV) power is becoming a major direction of the energy transition and has been rapidly developing in recent years [3]. The positive energy price policy and electricity price mechanism have also promoted the fast growth of PV [4]. In China, the installed capacity of PV power has reached a total of 125.8 GW by the end of 2017 [5]. Evolving technologies about smart energy network open up new opportunities for integrating solar power systems into grid [6]. However, the fluctuations in PV power output significantly hinder the grid from taking it in [7]. The fluctuation, or intermittency, can be caused by many factors such as sunlight intensity and changes in environmental conditions [8]. Therefore, a useful solution to this problem is to establish an accurate prediction model for PV power generation.

Conventionally, a large number of studies were devoted to improving the accuracy of PV deterministic point predictions (DPPs). Emanuele et al. [9] has presented a comparison of the PV output power day-ahead forecasts performed by physical models, based on three and five parameters electric equivalent circuit and statistical models on the basis of artificial neural network (ANN), finding out that the ANN,

* Corresponding authors.
E-mail addresses: mazhanyu@bupt.edu.cn (Z. Ma), qie@sdu.edu.cn (Q. Sun), yhy2011@sdu.edu.cn (H. Yin).

https://doi.org/10.1016/j.apenergy.2018.06.112
Received 29 March 2018; Received in revised form 12 June 2018; Accepted 21 June 2018

combined with clear sky solar radiation, generally achieves the best forecasting performance with the normalized mean absolute error (NMAE) of 5.6%. With a strong generality and excellent nonlinear mapping ability, ANNs are being used increasingly for solar energy predictions [10]. Various models based on ANN method such as generalized regression neural network (GRNN) [11], extreme learning machine neural network (ELMNN) [12], support vector machine (SVM) [13] and other models that have different characteristics and their respective advantages have been developed [14–16]. Recent researches that are adopting ANN as a main prediction model are furthering their work with pre-processing steps, i.e. weather classifications [17], feature extraction or selection [18,19], or key parameter analysis [20], and decorrelation of neutral vector variables [21]. Apart from the pre-processing steps, some works applies integrated or hybrid method to improve the forecasting precision. Cococcioni et al. [22] presented a flexible approach to forecast PV power one day-ahead by combining a time-series analysis and neural networks. Chu et al. [23] developed a reforecasting method based on genetic algorithm (GA) optimized artificial neural networks (ANNs). Hussain et al. [24] presented a hybrid technique to improve the performance of solar forecasting by using four different architectures of ANN models, namely: multilayer perceptron (MLP), Adaptive neuro-fuzzy inference system (ANFIS), Nonlinear autoregressive recurrent exogenous neural network (NARX), and generalized regression neural networks (GRNN). However, the deviations of short-term predictions from the actual PV power output, whatever method was employed, remained between 15% and 40%, unsatisfying for real-time control of grid-connection or security analysis [25].

Based on the traditional DPPs, the prediction interval (PI) estimation can provide the possible values of PV power generation and their related occurrence probability at the next moment [26]. With more comprehensive information, a PI can help decision-makers to understand not only the possible outputs but also the associated changing trend of the outputs, which can greatly facilitate grid planning, risk analysis and reliability evaluation. The PI estimations have been applied in some forecasting applications, e.g. wind power [27], load [28], price [29], net demand [30]. For solar power forecasting, existing studies have adopted different methods to implement estimation of PIs. The approaches can be distinguished into two categories where one assumes a probability density function (PDF) beforehand, i.e., parametric, and where no such assumptions are made, i.e., nonparametric [31]. The parametric approaches often assume that the random variables involved have known parametric distributions, e.g., normal distribution, gamma distribution, Laplacian distribution [32]. Junior et al. [33] regarded the error distribution of PV power output as known normal distribution and Laplacian distribution and estimate the parameters of such distribution via maximum likelihood estimation, then confidence interval under a certain level is obtained as a PI. In essence, more recent papers generally don't make assumptions regarding the distribution of the observations and take a nonparametric approach since the assumption of a distribution is generally speculative, far from realistic and not appropriate [34]. In order to avoid restrictive assumptions on the shape of predictive distributions, most recent papers utilize nonparametric to generate PIs of solar power [35]. Lu et al. [36] established a quantile regression (QR) model based on spline estimation to construct the PI of PV power output. The used QR is one of the most common methods to construct a nonparametric PDF, which was introduced by Koenker & Bassett [37]. Another method that is utilized to construct nonparametric density functions is bootstrap. Alhakeem et al. [38] used the bootstrap method that works as a resampling technique to conduct an PI of PV power. Bootstrap was proposed by Efron [39] as a method to estimate the probability distribution of a random variables random samples, drawn with replacement from an unknown parent distribution. Furthermore, Ni et al. [40] developed an improved bootstrap technique to construct PIs for uncertainties of PV power generation. There are also some other nonparametric methods that have been proposed to construct PIs in solar radiation and solar power forecasting, e.g. gradient boosting [41], Bayesian method [42], Gaussian processes [43,44], k-nearest neighbour [45], analog ensemble [46]. In these methods, the PIs are constructed to be measured with a specific probability and these works are arguably more valuable than a single value, since it allows for uncertainty management. By providing the variation intervals of the forecasts under a confidence level, these methods can effectively improve the applicability of PV forecasts, e.g. in favor of integrating intermittent solar energy into conventional fossil fuel-based energy systems. When compared to these methods, the nonparametric kernel density estimation (NKDE) method has the advantages of simplicity and easy implementation and has a lot of successful practice applications [47–49], demonstrating that this method could achieve superior performance than others. Nevertheless, there is still room for the NKDE method to be further improved, since setting the parameters improperly will affect the quality of PI estimation, which calls for more special attention.

To the authors' knowledge, there are few studies which systematically considered the uncertainty in forecasting the outputs of PV system in terms of the integration of point forecast and interval forecast. Therefore, the paper intends to fill the knowledge gap by proposing a novel two-stage approach for quantifying the uncertainty in forecasting the outputs of PV systems. To this aim, the paper firstly made an optimized DPP by a weight-varying combination forecast mode (WVCFM) model under four weather types. Second, a series of error samples of PV power derived from the point forecasts were processed by the nonparametric kernel density estimation method (NKDE) to obtain their probability distribution functions (PDFs). Third, based on the DPPs and the PDFs, the PIs were obtained for the next moment at different confidence levels of 95%, 90%, 85% and 80%, respectively. The contribution of the study is mainly in the integrated model for quantifying the PI of PV power outputs. The integrated model helps to achieve the uncertainty estimation that will determine the robust performance of PV point forecasts. By applying the model to a PV plant, the study proved that the method is very effective for producing comprehensive information about the PIs and thus can be useful for practical decisions about e.g. grid connection, system operation and security and stability control.

2. Methodology

The study proposed an integrated two-stage model to quantify the PIs of PV power output, namely a point forecast model to obtain DPPs at the first stage and an interval forecast model to estimate the PIs at the second stage (Fig. 1).

2.1. Point forecast model

In the first stage, the central component of the point forecast model is a WVCFM model, which was developed to perform the DPP of PV power outputs. To achieve this purpose, a total of four steps were taken in the first stage. The proposed WVCFM model was tested using four types of weather data, which were categorized by Kohonen clustering network. The specific flow of the point forecast model comprises of three steps.

Step (1) The original historical dataset was split into two datasets, in which the first-half set, Dataset 1, was used in Step (1) for training and testing the point prediction models and the second-half, Dataset 2, was adopted in Step (3) for forecasting. A weather cluster analysis was implemented to further divide the first-half set of data into four types, namely Type 1-4, based on the Kohonen clustering neural network.

Step (2) The WVCFM models, i.e. WVCFM 1-4, were respectively established on the basis of the data Type 1-4.

Step (3) The trained WVCFM models, i.e. WVCFM 1-4, were used to predict five-min ahead PV power outputs, i.e. DPPs, employing