Photovoltaic power interval forecasting method based on kernel density estimation algorithm

Min Shi1, Rui Yin1, Yifeng Wang1, Dengxuan Li2*, Yutong Han3, Wansi Yin3
1 State Grid Hebei Electric Power Company, Shijiazhuang 050022, China
2 China Electric Power Research Institute, Nanjing 210003, Jiangsu, China
3 School of New Energy, North China Electric Power University, Beijing 102206, China
*Corresponding author’s e-mail: lidengxuan@epri.sgcc.com.cn

Abstract. Photovoltaic (PV) power forecasting is of great significance to improve the access level of photovoltaic power generation. Deterministic forecasting methods often fail to meet the needs of grid risk analysis and decision-making, and a single model is also difficult to adapt to the changes of PV objects under different meteorological conditions. In order to solve the above problems, this paper describes the seasonal distribution characteristics of PV output fluctuations, and a method of PV power uncertainty forecasting based on seasonal classification is proposed. Compared with the parameter estimation method, the nonparametric kernel density estimation method does not need to make assumptions about the distribution of the prediction error, and can obtain more information about the actual distribution of the error, and has no fixed requirements for the various distribution characteristics of the error, it has strong adaptability.

1. Introduction

The research on uncertainty forecasting of PV output power has just started. On the one hand, the parameter estimation method assumes that the form of probability density function is known in advance. However, affected by a variety of physical processes, its output is difficult to meet a specific distribution. Sometimes the distribution of the assumption of PV power may be unreasonable, the parameter estimation method is difficult to apply. On the other hand, for the parameter estimation method, a large amount of historical operation data requires to be analysed, but due to the nonlinear characteristics of the PV power curve, the shape of the predicted probability density function will change over time, which has a certain impact on the effectiveness of the parameter method. An effective way to solve the above problems is to use the non-parametric estimation method to realize the uncertain forecasting of output power. The function form and parameters of non-parametric estimation method are unknown, it does not make any assumption about the basic distribution. Compared with the parameter estimation method with assumption, the non-parametric estimation method is one of the typical methods of model estimation which is more in line with the real distribution of random variables [1]. The KDE is a very effective non-parametric density estimation method. Zhang et al. [2] proposed to use the K-nearest neighbor algorithm to find weather types...
similar to the day to be forecasted, according to the historical similar days, the wind power probability density distribution was constructed by using the optimized KDE based on log-based. Jiang et al. [3] optimized the conditional KDE of bandwidth by using normal reference criterion, and the forecasting probability density function of each subsequence was obtained, then the expectation and variances were derived. Finally, the deterministic forecasting was obtained by summarizing these expectations, and the probability forecasting interval was obtained according to the covariance and Gaussian distribution assumptions on target wind speed. He et al. [4] proposed a hybrid wind power probability density forecasting method based on quantile regression neural network and kernel function with unbiased cross validation. The forecasting results of wind power under different quantiles are used as the input of KDE, and the comprehensive wind power probability density forecasting information can be estimated at any time in the future. The main idea of this paper is to use the non-parametric estimation method to achieve the uncertainty forecasting of PV power considering the law of PV output characteristics.

Different from wind power generation, PV output has obvious periodicity and non-linearity [5-8], so it is very difficult to forecast the power generation accurately under various meteorological conditions by using a single model. If the change of PV output caused by the differences of meteorological is ignored, the applicability of the forecasting model under different meteorological conditions will be affected. Therefore, the main research content of this paper is how to combine the multi-model method with the non-parametric estimation interval forecasting method.

2. Basic theory and forecasting method

The theory and configuration method of ELM and non-parametric estimation used in the paper were given in this chapter, and finally explains the thesis method.

In power forecasting, there is a great impact on the forecasting results of any kind of forecasting model due to the difference of model input variables. For example, in general statistical forecasting, the influence factors of PV power include irradiance, temperature, humidity and wind speed, etc. If there are too few input variables are selected for power forecasting, part of the information may be missing and it is difficult to fully reflect the variation characteristics of PV power, but too many selected variables may lead to information redundancy and reduce the generalization ability of the model. In this paper, solar irradiance, ambient temperature, humidity and wind speed are selected as the input variables of the forecasting model.

PV power forecasting error is defined as the deviation between the forecasting value $P_f$ and the actual measured value $P_a$ at a certain time point.

$$e = P_a - P_f$$

(1)

The fluctuation of forecasting error is quite different for different PV power values. Therefore, in this paper, the power forecasting values are divided into multiple power levels at equal intervals $\Delta P = 1000W$, and PV power forecasting error distributions of different power forecasted levels are established for four seasons. At the same time, it is necessary to ensure that each power level has enough samples for probability statistics, because too few sample points cannot better reflect the actual distribution of forecasting error. Therefore, in order to make the number of sample points in the interval meet the requirements, it is necessary to combine the intervals with fewer sample points in the adjacent power level distribution. The spring samples are divided into 8 power levels and the remaining seasons are 7 power levels according to the number of samples of deterministic forecasting power.

After the probability density function of forecasting error is obtained by KDE method, the cumulative probability distribution function can be obtained by integral. For a given confidence level $1 - \alpha$, it can be determined that the boundary values of forecasting error $e^*_{\min}$ and $e^*_{\max}$ satisfy $P(e^*_{\min} < e < e^*_{\max}) = 1 - \alpha$, and the forecasting interval is shown as follows

$$[P_f + e^*_{\min}, P_f + e^*_{\max}]$$

(2)

The estimated interval under a certain confidence level can be obtained combined with the
deterministic forecasting value of PV power, and the fluctuation of PV power can be described, so as
to improve the accuracy and reliability of power forecasting. The uncertainty forecasting process of
PV power mainly includes the following steps.

Step 1: Data preparation. Dividing the original data according to seasons.

Step 2: Deterministic forecasting. The deterministic forecasting models of different seasons are
established respectively based on the ELM to obtain deterministic forecasting power, then the
deterministic forecasting error is calculated.

Step 3: Forecasting model training. The deterministic forecasting power data is divided into
appropriate power levels, and the forecasting error is fitted by KDE, then the boundary value of the
forecasting error intervals of each power level can be calculated according to the confidence level.

Step 4: Interval forecasting. The validation data is substituted into the deterministic forecasting
model to obtain the deterministic forecasting power. The PV power forecasting interval can be
obtained by substituting the forecasting power of the corresponding level into formula (2). Finally, the
performance of the forecasting results can be evaluated.

Figure 1. Flow chart of uncertainty forecasting.

3. Example verification
In order to verify the effectiveness of the uncertainty forecasting method proposed in this paper, the
performance of the proposed method is evaluated from the error fitting effect and power forecasting
effect respectively. In this paper, RMSE and MAE are used to evaluate the deterministic forecasting
effect of seasonal multi-model and annual model.

Taking into account the seasonal characteristics of PV output fluctuation, the deterministic seasonal
forecasting models are constructed respectively based on the data samples are classified according to
the seasons, comparing it with the annual single model power forecasting method.

The error and accuracy of deterministic forecasting under the two forecasting methods are shown in
table 1. In terms of RMSE and MAE, the seasonal multi-model power forecasting method considering
the output fluctuation characteristics is smaller than the single annual model. The forecasting accuracy of seasonal multi-model considering the characteristics of output fluctuations is higher than 94%, which is better than the single annual model.

Table 1. Error comparison of deterministic forecasting

| Indicator | Classification | Seasonal model | Annual model |
|-----------|----------------|----------------|--------------|
|           | Spring         | 2.31           | 2.40         |
|           | Summer         | 2.89           | 4.74         |
|           | Autumn         | 3.10           | 5.12         |
|           | Winter         | 2.57           | 4.41         |
| MAE       | Spring         | 4.27           | 4.30         |
|           | Summer         | 5.37           | 8.55         |
|           | Autumn         | 5.90           | 9.20         |
|           | Winter         | 5.72           | 9.07         |
| RMSE      | Spring         | 95.73          | 95.70        |
|           | Summer         | 94.63          | 91.45        |
|           | Autumn         | 94.10          | 90.80        |
| Accuracy/%| Winter         | 94.28          | 90.93        |

The deterministic power forecasting results obtained are classified according to the level, and the confidence interval under a certain confidence level can be obtained by using the KDE to fit the forecasting error accordingly. Figure 2. shows the forecasting intervals of PV power at the confidence levels of 95%, 90% and 80% for the uncertainty forecasting samples of seasonal multi-model and annual model respectively. It is obvious that under the same confidence level, the forecasting results of the seasonal multi-model can ensure that the PV time series changes are tracked. At the same time, it has a narrower upper and lower limits, and no more points fall outside the forecasting interval. In addition, the wider the confidence interval, the greater the probability of inclusion of the actual value, so the width of the confidence interval should also increase correspondingly once the confidence level increases. When selecting the confidence level, if the estimation interval is too large, the reference significance of the estimation results will be reduced, while the estimation interval is too small, which will lead to the missing of some larger PV power estimation error data. Therefore, in order to achieve the required reliability and accuracy, an appropriate confidence level should be selected in practical application.
4. Conclusion
In this paper, an uncertainty forecasting method of seasonal multi-model PV power considering output fluctuation characteristics is proposed. The seasonal distribution characteristics of PV output are analyzed, and it is proved that the seasonal model is helpful to improve the forecasting accuracy. Non-parametric KDE method does not need to make assumptions about the distribution of forecasting error compared with parameter estimation method. It is flexible in shape, and more information about the actual error distribution can be obtained, since it has no fixed requirements for the various distribution characteristics of the error, it has a strong adaptability. The proposed method can effectively describe the possible fluctuation interval of PV power and evaluate the reliability of the forecasting interval. The possible fluctuation interval of PV power under different confidence levels is given, which provides important reference information for scientific dispatching management of PV power generation.

Acknowledgments
Supported by: Science and Technology Project of State Grid Hebei Electric Power Company “Research and application of medium and long-term forecasting technology for regional wind and photovoltaic resources and generation capacity” (5204BB170007); Special Fund Project of Hebei Provincial Government (19214310D)

Reference
[1] Almeida M P, Muoz M, Parra I D L, et al. Comparative study of PV power forecast using parametric and nonparametric PV models[J]. Solar Energy, 2017, 155:854-866.
[2] Zhang Y, Wang J. K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting[J]. International Journal of Forecasting, 2016, 32(3):1074-1080.
[3] Jiang Y, Huang G, Yang Q, et al. A novel probabilistic wind speed prediction approach using real-time refined variational model decomposition and conditional kernel density estimation[J]. Energy Conversion and Management, 2019, 185(APR.):758-773.
[4] He, Yaoyao, Li, et al. Probability density forecasting of wind power using quantile regression neural network and kernel density estimation[J]. Energy Conversion and Management, 2018.
[5] Huang Y, Cao C, Gu H. Short-term photovoltaic power generation forecasting scheme based on IKFCM and multi-mode social spider optimization SVR [J]. Power System Protection and Control, 2018, 46(24):96-103.
[6] Wang F, Mi Z, Zhen Z, et al. A Classified Forecasting Approach of Power Generation for Photovoltaic Plants Based on Weather Condition Pattern Recognition [J]. proceedings of the CSEE, 2013, 33(34):75-82.
[7] Van d M D W, Widén, J, Munkhammar J. Review on probabilistic forecasting of photovoltaic power production and electricity consumption[J]. Renewable and Sustainable Energy Reviews, 2017:S1364032117308523.

[8] Tewari S, Geyer C J, Mohan N. A Statistical Model for Wind Power Forecast Error and its Application to the Estimation of Penalties in Liberalized Markets[J]. IEEE Transactions on Power Systems, 2011, 26(4):2031-2039.