STUDY ON PROBABILITY DISTRIBUTION OF PHOTOVOLTAIC POWER FLUCTUATIONS AT MULTI-TIME SCALES

Ling LUO
EPRI, State Grid Shanghai Municipal Electric Power Company-China
luol@sh.sgcc.com.cn

Chen FANG
EPRI, State Grid Shanghai Municipal EPRI, SGSMEPC-China
fangc02@139.com

Yong YANG
Shanghai University of Electric Power-China
1530843443@qq.com

Chunyang LI
Shanghai University of Electric Power-China
851439317@qq.com

Fen LI
Shanghai University of Electric Power-China
beckyhust@163.com

Peng ZHANG
EPRI, State Grid Shanghai Municipal Electric Power Company-China
zhang_peng@sh.sgcc.com.cn

ABSTRACT

By analyzing the probability distribution function (p.d.f) of photovoltaic (PV) power, the influence of randomness and uncertainty of PV power fluctuation on the operation of power system can be effectively reduced. It will be beneficial to improve the PV grid-connected penetration. In this paper, the generalized Gaussian distribution, the finite student t-mixture model and other models are utilized to analyze the power fluctuation characteristics under multiple time scales, based on a large amount of data gathered from a distributed PV station. Firstly, the study shows that generalized Gaussian distribution performs best to describe the probability distribution of PV power variation and the average power variation under 10~15-min time scales while Gaussian mixture model performs best under 30~60-min time scales. Further, a model based on the relationship between the variability of hourly PV energy production and global solar radiation is proposed, which can be used for quantitative analysis of the energy production variability of PV stations.

INTRODUCTION

PV power generation is a multivariate coupled highly nonlinear stochastic process. The uncertainty and volatility of photovoltaic output power will bringing enormous impact and challenges to the power grid's scheduling, plan and control and other aspects.

Study on the relationship between the fluctuation of PV output and the fluctuation of radiation, is not only conforms to the reality of meteorological science and photovoltaic power generation, but also meaningful to the power system peaking and power grid planning. Probability is a way of expressing uncertainty, which has been increasingly concerned in the field of new energy power generation [1]. Such as Monte Carlo simulation, probability density function modeling and other methods. In general, there are two main methods for describing the wind power or PV power fluctuation characteristics based on the probability density function method: modeling the differential or rolling average of wind power or PV power [2-6] or modeling wind power or PV power directly [7-8]. The applied probabilistic models can be divided into single distribution, mixed distribution and nonparametric model.

In [3], it is considered that the fluctuations of wind power will affect the system utilization ratio of climbing capacity, using the t location-scale distribution in the probability distribution of wind power. In [4], the Beta distribution is considered to be suitable for describing the probability distribution of the power output of the concentrated wind farm, and the Laplace distribution is preferred to describe the probability distribution of the output changes of the concentrated wind farm. [5] uses three different p.d.f to fit the PV power changes at different time scales. Autoregressive spectrum analysis method is used in the paper [6] to research the fluctuation of PV power at different time and space scales, but the PV power data is based on the simulation of the meteorological data and the PV power fluctuation probability models under different weather types have not been studied and checked out.

Based on the measured data of a PV power station, this paper estimates the p.d.f of PV power variations under multiple time scales from aspects of single distribution and mixed model through statistical methods. In allusion to single distribution, the t location-scale distribution and the generalized Gaussian distribution model of the PV power fluctuations are studied, and the model parameters are estimated to utilize the maximum likelihood estimation algorithm. For mixed models, the model of PV power fluctuation characteristic of finite student t mixture and Gaussian mixture are presented. Then, parameters of these two models are estimated by the improved Expectation Maximization (EM) algorithm [9]. Finally, a variety of different probability distribution models are evaluated and verified. Ultimately, the relationship between the fluctuation of PV output and radiation is quantitatively analyzed.

PROMBLEM FORMULATION

Generalized Gaussian Distribution Parameter Solving

Generalized Gaussian Distribution [10-13] is applied in data modeling in the fields of speech signal, digital water marking, spectrum access, power system load demand and independent component analysis. The expression of probability density function is

$$f(x)=\frac{\alpha}{2\beta\Gamma(1/\alpha)}\exp\left\{-\left(\frac{|x-\mu|}{\beta}\right)^{\alpha}\right\} \quad (1)$$

, where α represents shape parameter which is greater than 0; β is scaling parameter, σ is mean square error (MSE).

There are various methods to calculate generalized distribution parameters, such as maximum likelihood estimation and distance estimation. When the shape parameter α is obtained by the maximum likelihood estimation algorithm, it is necessary to solve the height nonlinear function (including double gamma and three gamma functions) by means of iterative method (such as Newton-Raphson method).In this paper, the shape parameter is obtained by establishing a convex function.

Mixed Model and Parameter Solving

The finite probability hybrid model is one of the frequently-used and effective statistical methods for data analysis. It can be understood as the linear superposition of several identical p.d.f, which has the advantages of excellent expressiveness and flexibility compared with the single distribution. The finite probability hybrid model can always fit the local data feature by increasing the component number. The two-component and three-component Gaussian mixture model used in this paper are linear superposition of two and three normal distribution. The probability density function is

$$\varphi(x)=\sum_{i=1}^{s} m_i \frac{1}{\sigma_i\sqrt{2\pi}}\exp\left(-\frac{(x-\mu_i)^2}{2\sigma_i}\right) \quad (2)$$

, where x represents observed data; s is the number of components. m_i , μ_i and σ_i are respectively the weighting coefficient, mean and variance of the i^{th} normal distribution respectively.

The two-component finite student t-mixing model[11] is a linear superposition of two t location-scale distributions. The probability density function is

$$\varphi(x)=\sum_{i=1}^{2} m_i f(x;\mu_i,\sigma_i,\alpha_i) \quad (3)$$

, where *f(x)* represents t location-scale distribution; m_i , μ_i , σ_i and α_i are weighting coefficient, mean, mean square and shape parameters of the i^{th} location distribution.

The fitting effect of the sample data is determined by the parameters of mixed model. In this paper, EM algorithm based on K-means++ clustering[12] is used to obtain the mixed model parameters. EM algorithm has high demand on the initial value, because the inappropriate initial value will make EM algorithm fall into the local optimal solution. So the K-mean ++algorithm is used to divide the sample data into k classes. Then the mean value μ_i, the MSE σ_i, and the proportion of samples per class m_i (weighting factor) of each class are taken as the initial values of the EM algorithm, to reduce the probability of the algorithm falling into the local optimal solution.

The complete sample data consists of the label of the observed value and the category of the observed value, and the EM algorithm can be used to obtain the probability distribution parameter for incomplete sample data (missing category tag data). EM algorithm is generally divided into E-step and M-step two steps, E-step for the category tag, M-step for each class to obtain the mean value, MSE, weighting factor and degree of freedom.

CASE ANALYSIS

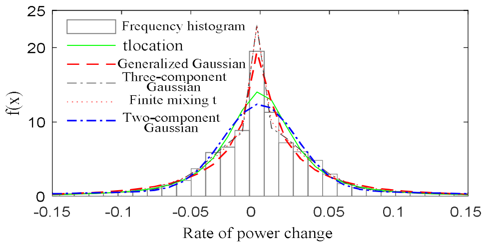
Data Sources

The measured power data per 5min of a 5.1 kW distributed PV power station from 2010 to 2011 is collected as well as the hourly solar radiation data during the same period. The hour angle of sunrise and sunset is also calculated according to the local latitude, as well as the PV power data from sunrise to sunset.

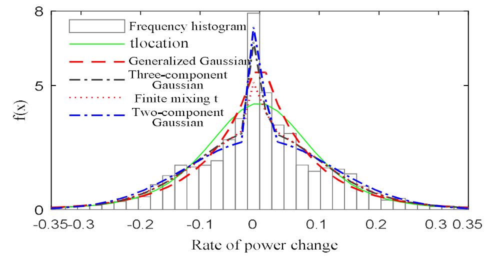
Modeling and Analysis of PV Power Fluctuation Characteristics at Different Time Scales

Moore formula [13] is chosen as the selection criteria of histogram class number. In order to judge whether the theoretical p.d.f is consistent with the observed data, we select two factors, the coefficient of determination (R^2) and root mean squared error (RMSE) [7], for judging standard.

PV power first-order difference is used to characterize photovoltaic power fluctuation [5] and Fig.1 shows the results of the fitting rate of PV power change rate in some time scales. It is observed that the histogram of change rate of PV power gradually becomes short and wide with the increase of time scale. Meanwhile, the degree of fluctuation also increases gradually. For 5-min and 60-min time scales, the PV power fluctuates 95% and confidence level is installed capacity of 8% and 23%. Besides, the fluctuation amplitude increased significantly



(a) 15min time-scale



(b) 60 min time-scale

Fig.1 PDF fitting results in different time scale

Table1 PDF fitting errors in different time scales

Time-scale/min	Fitting index		t location	Gaussian distribution	Three-component Gaussian	Two-component Gaussian	Finite student t
5	R^2		0.9945	0.9811	0.9395	0.9683	0.9938
	RMSE		0.3289	0.6100	1.0905	0.7889	0.3482
10	R^2		0.9555	0.9907	0.9844	0.9240	0.9556
	RMSE		0.7000	0.3196	0.4145	0.9153	0.6999
15	R^2		0.9307	0.9865	0.9734	0.9023	0.9775
	RMSE		0.7306	0.3230	0.4527	0.8677	0.4161
30	R^2		0.8940	0.9712	0.9793	0.8761	0.9700
	RMSE		0.6704	0.3496	0.2959	0.7246	0.3564
60	R^2		0.8305	0.9117	0.9649	0.9700	0.9106
	RMSE		0.5972	0.4311	0.2717	0.2511	0.4337

Table 1 shows the fitting indices of different distributions of different time scales in 2011. In the single distribution, the generalized Gaussian distribution fitting index R^2 is greater than the t location-scale distributed R^2 and RMSE are much smaller than the RMSE of t location-scale distribution for 10~60-min time scale, indicating that the generalized Gaussian distribution describes the p.d.f of the PV fluctuation characteristics better. In the mixed distribution, the weighted coefficient of the finite student t mixture model is less than 0.05, which can make this model into the local optimal solution at the 60-min time scale. Generally speaking, through the comparison of the fitting indexes, we can find that the t location-scale distribution is the best at the time scale of 5-min, the generalized Gaussian distribution is the best at 10~15-min, and GMM has best fitting effect at 30~60-min time scales. When it comes to large time scales (such as 30~60-min), their histograms have a longer 'tail'. The fitting effect of corresponding single distribution is slightly worse than mixed model.

Quantitative Analysis of PV Power Fluctuation and Modeling of Energy Fluctuation

The generalized Gaussian distribution parameters are used to quantitatively analyze the fluctuation of PV at different time scales.

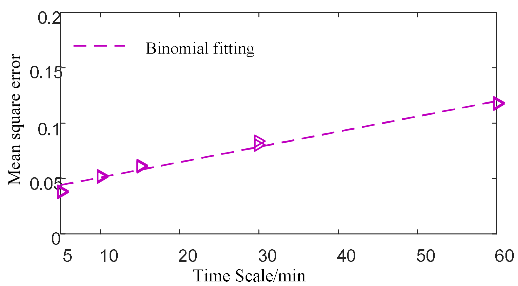


Fig.2 The relationship between standard deviation and time scales

Fig.2 shows the scatter plot and the fitting straight line of generalized Gaussian distribution at different time scales. The correlation coefficients of the fitted straight and mean square values are 0.982 and 0.991. It can be seen that the variance values of the two regions are almost coincident for the same time scale, and the variance values increase when the time scale becomes wider. That indicates that the PV power fluctuations are gradually increased.

The solar radiation and PV power data (normalized) from 2010 to 2011 is chosen to analyses the relationship between the fluctuation of the PV power generation and the fluctuation of the radiation quantitatively. The 5 min PV power data is summed up to hourly values and then normalized. In [14], the exponential distribution of the fluctuation of the radiation and the fluctuation of the PV power is established. And the exponential distribution parameters of the PV power fluctuation are considered related to the meteorological and environmental factors, such as radiometer and sharpness index. [15] analyzes the fluctuation of the cluster PV power plant with the MSE of PV power fluctuation. In this paper, the relationship between the MSE belonged to the hourly power fluctuation of the PV power plant and the solar radiation is established. And based on the relationship, the fluctuation of the PV power.is studied quantitatively.

$$\sigma_{\Delta P} = \delta \sigma_{\Delta H} + b \tag{4}$$

, where $\sigma_{\Delta P}$ represents the hourly PV power generation fluctuation, $\sigma_{\Delta H}$ represents the MSE of the radiation corresponding to the power; δ represents regression coefficients; b represents intercept, which approximately equals to the result of [15].

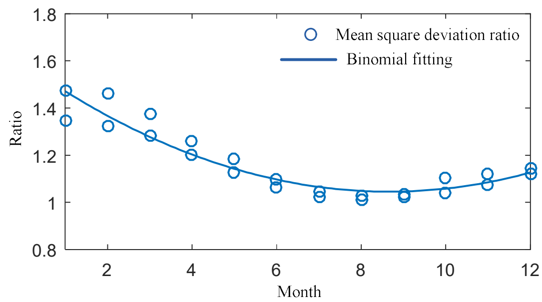


Fig.3 The rate of energy production and radiation variability standard deviation

Fig.3 shows the result what is the ratio of the MSE of the PV power generation in the different two months and corresponding time radiation. It can be seen that the MSE ratio of two different months is different, and submits to the binomial expression. The correlation coefficient of binomial fitting is 0.825. From January to April, the error of binomial fitting is larger than that of other months due to the lack of data. Fig.4 shows the MSE of the power daily variation of PV power generation and the MSE of the daily radiation quantity daily variation. The fitting function is:

$$\sigma_{\Delta P} = 0.7116\sigma_{\Delta H} + 0.0104 \quad (5)$$

The MSE and the correlation coefficient is 0.7424. According to the variation of daily radiation variance, we can approximate the MSE of daily generating capacity, and then determine the fluctuation of energy output of PV power plant.

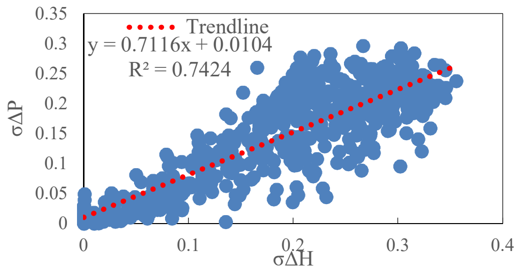


Fig.4 The relationship between daily energy production and radiation variability standard deviation

CONCLUSION

In this paper, the probability distribution modeling of PV power fluctuation is obtained from two methods. One is single distribution and the other is mixed model. The fitting results are evaluated by the coefficient of determination and RMSE. Conclusions are as follows:

(1) Under the time scale of 10~60-min, the generalized Gaussian distribution performs better than t location-scale distribution in a single distribution. In the small time scale of 5-min, the t location-scale distribution performs better.

(2) Considering the mixed model and single distribution, the Gaussian mixture model is better than the single distribution at 30 ~ 60min time scale, and the result is opposite at 5~15-min time scale.

(3) The root mean squared error of the generalized Gaussian distribution increases linearly when the time scale becomes larger. The standard deviation ratio between the radiation and the power generation submits to the binomial expression with the variations of months. According to the change of the daily radiation, the energy fluctuation of PV power plant can be estimated approximately.

ACKNOWLEDGEMENT

This work is supported by State Grid Shanghai Municipal Electric Power Company Technology Program (No. 520940160026).

REFERENCES

[1] Xue Y, Xing L, Feng X, 2014, "A Review on Impacts of Wind Power Uncertainties on Power Systems", *proceedings of the CSEE*. vol.34, 5029-5040.

[2] Ma X Y, Sun Y Z, 2013, "Scenario Generation of Wind Power Based on Statistical Uncertainty and Variability", *IEEE T SUSTAIN ENERG*. vol.4, 894-904.

[3] Lin W, Wen J, Cheng S, et al., 2013, "An Investigation on the Active-Power Variations of Wind Farms ", *IEEE T IND APPL*. vol.48, 1087-1094.

[4] Louie H, 2010, "Evaluation of probabilistic models of wind plant power output characteristics", *IEEE International Conference on Probabilistic Methods Applied to Power Systems*, Singapore, vol.2010, 442-447.

[5] Wu Z W, Jiang X P, et al., 2014, "Study on Fluctuations Characteristics of Photovoltaic Power Output in Different Time Scales", *Mod Electr Power*. vol.31, 58-61.

[6] Chen X, 2015, "A Method for AGC Reserve Capacity Analysis Considering Photovoltaic Power Fluctuation Characteristics ", *Automation of Electric Power Systems*. vol.39, 16-21.

[7] Hong L, 2014, "A Probabilistic Model to Simulate Wind Power Output Fluctuation of a Certain Wind Farm Cluster in Xinjiang Region", *Power Syst Technol*. vol.38, 1616-1620.

[8] Cui I Y, 2015, "Probability Density Distribution Function of Wind Power Fluctuation of A Wind Farm Group Based on the Gaussian Mixture Model", *Power Syst Technol*. vol.40, 1107-1112.

[9] Li F, Li C Y, et al., 2017, "Photovoltaic output fluctuation characteristics research based on variational Bayesian learning", *Electr Power Autom Eq*. vol.37, 99-104+122.

[10] Song K S, 2006, "A globally convergent and consistent method for estimating the shape parameter of a generalized Gaussian distribution", *IEEE T INFORM THEORY*. vol.52, 510-527.

[11] Peel D, Mclachlan G J, 2000, "Robust mixture modelling using the t distribution ", *Statistics & Computing*. vol.10, 339-348.

[12] Arthur D, Vassilvitskii S, 2007, "K-means++: the advantages of careful seeding", *Acm-Siam Symposium on Discrete Algorithms,* New Orleans, Louisiana, USA , vol.11, 1027-1035.

[13] Luo L, et. al., 2015, "Heterogeneous Cognitive Radio Sensor Networks for Smart Grid: Markov Analysis and Applications", *International Journal of Distributed Sensor Networks*, vol.2015, 528346.

[14] Marcos J, Marroyo L, Lorenzo E, et al., 2011, "Power output fluctuations in large scale PV plants: One year observations with one second resolution and a derived analytic model", *Progress in photovoltaics*. vol.19, 218-227.

[15] Hoff T E, Perez R, 2010, "Quantifying PV power Output Variability", *Solar Energy*. vol.84, 1782-1793.