A Novel Generation Method for the PV Power Time Series Combining the Decomposition Technique and Markov Chain Theory

Xu, Shenzhi; Ai, Xiaomeng; Fang, Jiakun; Wen, Jinyu; Li, Pai; Huang, Yuehui

Published in:
Journal of Engineering

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

The 6th International Conference on Renewable Power Generation (RPG)

19–20 October 2017

Generation method for the PV power time series combining the decomposition technique and Markov chain theory

Shenzhi Xu[1], *Xiaomeng Ai*[1], *Jiakun Fang*[2], *Jinyu Wen*[1], *Pai Li*[3], *Yuehui Huang*[3]

[1]*State Key Laboratory of Advanced Electromagnetic Engineering and Technology, School of Electrical and Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, Hubei Province, People's Republic of China*
[2]*Department of Energy Technology, Aalborg University, Aalborg DK-9220, Denmark*
[3]*State Key Laboratory of Operation and Control of Renewable Energy and Storage Systems, China Electric Power Research Institute, Beijing 100192, People's Republic of China*
E-mail: xushenzh@qq.com

Published in *The Journal of Engineering*; Received on 11th October 2017; Accepted on 3rd November 2017

Abstract: Photovoltaic (PV) power generation has made considerable developments in recent years. However, its intermittent and volatility of its output have seriously affected the security operation of the power system. In order to better understand the PV generation and provide sufficient data support for analysis the impacts, a novel generation method for PV power time series combining decomposition technique and Markov chain theory is presented here. It digs important factors from historical data from existing PV plants and then reproduce new data with similar patterns. In detail, the proposed method first decomposes the PV power time series into ideal output curve, amplitude parameter series, and random fluctuating component three parts. Then generating daily ideal output curve by the extraction of typical daily data, amplitude parameter series based on the Markov chain Monte Carlo (MCMC) method, and random component based on random sampling, respectively. Finally, the generated three parts are recombined into new PV power time series by the decomposition formula. Data obtained from real-world PV plants in Gansu, China, validates the effectiveness of the proposed method. The generated series can simulate the basic statistical, distribution, and fluctuation characteristics of the measured series.

1 Introduction

The world has been seeing dramatic development of solar energy in recent years due to its sustainability. The global new photovoltaic (PV) installed capacity further increased in 2016, reached 70 GW, an increase of ~30% over 2015. The intermittent and volatility of PV power output have become an important factor affecting the stability of power system [1–3]. To analyse the impacts of large amount of PV integration, bulk data of PV generation is necessarily required. However, for newly installed PV plants, there is insufficient data for system-level analysis. PV power time series generation is an effective method to solve the insufficient data problem. Time series generation refers extract the internal patterns from the measured power series and then use the extracted patterns to generate new power series which is well consistent with the measured series in statistical and fluctuating features.

The magnitude of the PV output depends on how much solar radiation is received. The change of solar radiant energy has both obvious regularity and unpredictable randomness. In [4], the sun–earth movement model is established, and the daily solar radiation energy curve can be calculated according to latitude, longitude, and altitude. However, this model ignores temperature, climate, and other factors; thus, there is a certain cap between the calculated results and the measured PV output. In [5, 6], the short-term and mid-long-term stochastic properties of solar power generation are analysed and provides some reference for the generation of PV power time series.

In general, studies on the PV power time series generation can be divided into two categories, namely solar radiation method [7, 8] and solar power method [9–11]. The solar radiation method first generates solar radiation intensity series and then uses the radiation-electric power conversion function to estimate power output. This method requires high accuracy of the radiation-electric power transfer function. There could be significant difference between the generated PV power series and the measured ones. The solar power method refers to generating new PV power time series directly using the actual measured data. In [9], the whole day is divided into multiple periods to model the PV output separately. Luo *et al*. [10] use the Gibbs sampling technique to construct time series model of PV output. The empirical model of PV output based on the measured data is established in [11]. The solar power method showed above cannot describe the regularity of daily PV output accurately. In summary, the existing methods of generating PV power time series are still inadequate in the description of regularity and randomness of PV output. Combining the advantages of the two kinds of methods is urgently needed.

In this paper, a novel generation method for PV power series combining the decomposition technology and Markov chain theory is proposed. Also, the effectiveness of the method is verified by the time series from real-world PV plants.

2 Decomposition of PV power time series components

In general, the output of PV plants is mainly affected by three aspects. (i) Earth rotation movement and the movement between the sun and the earth. They make regular changes in solar radiation. (ii) Atmospheric attenuation, which affects the solar radiation intensity received by the solar panels. (iii) Cloud disturbance. The shadowing effect of clouds will bring random component to the PV output. Thus, the PV output power is decomposed into three parts according to the following formula

$$P(t) = k \cdot P_{\text{norm}}(t) + P_{\text{rand}}(t) \quad (1)$$

where $P(t)$ is the PV output power, k the amplitude parameter, $P_{\text{norm}}(t)$ the ideal PV output, and $P_{\text{rand}}(t)$ the random fluctuating component.

2.1 Ideal output curve extraction

Ideal PV output is the daytime output of PV plants without regard to atmospheric attenuation and cloud disturbances. The ideal output curve calculated using the theoretical model [4, 12] always maintains the sinusoidal characteristic. However, the actual output in the morning and in the afternoon is often not symmetrical. Therefore, this paper considers the use of the ideal output curve extraction method instead of the theoretical model calculation method. The extraction method consists the following steps:

(i) Select typical days. Using the absolute value of the second-order difference of PV output to determine whether the day is a typical day, as shown in the following formula

$$\max\left|\left(x_{t+2}-x_{t+1}\right)-\left(x_{t+1}-x_t\right)\right| < D \tag{2}$$

In the formula, x_t represents the PV output at time t of 1 day and D is the critical threshold. If the inequality is satisfied, the day is a typical day. In this paper, D is 10% of the installed capacity of PV plants.

(ii) Normalise the PV output. Owing to the difference of maximum daily PV output, sunrise and sunset moments, it is needed to normalise PV output to extract the shape of an ideal output curve. Using the maximum PV output value of the day for the standard unit and normalising the time span. The PV output curve of typical day after the standardisation is shown in Fig. 1.

(iii) Formulate analytical equation of PV output curve of typical days. As the different lengths of daily output, the standard unit time of the sampling points is also different after normalisation. Thus, it is necessary to obtain the PV output analytical equations of typical days. We use the fast Fourier transform and keep the first five harmonics to achieve the resolution.

2.2 Generating ideal output curve of atypical days

The ideal output curve of atypical days needs to be generated by the linear interpolation method of the ideal output curve of typical days near its timing. Using the following formula to calculate the ideal output of atypical days

$$P_{i,\text{norm}}\left(t^*\right)=\frac{n-i}{n-m}P_{m,\text{norm}}\left(t^*\right)+\frac{i-m}{n-m}P_{n,\text{norm}}\left(t^*\right) \tag{3}$$

where i represents an atypical day; t^* is the normalised time; $P_{i,\text{norm}}(t^*)$ the normalised ideal PV output of the ith day at normalised time t^*; m and n are the nearest typical days before and after the ith day.

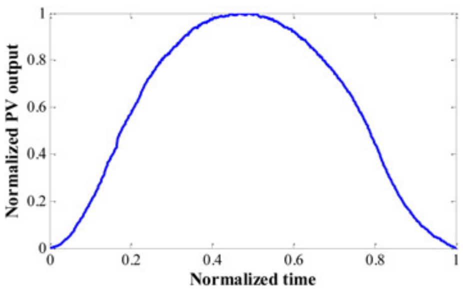

Fig. 1 *Normalised PV output of typical day*

After obtaining the normalised ideal PV output of typical and atypical days, it is necessary to convert the normalised time into the actual time. Using the geographical information of PV plants and date order, we can calculate sunrise and sunset moments of every day in the location of the PV plants. The daytime of PV output can be calculated by

$$T_i^{\text{day}}=T_i^{\text{ss}}-T_i^{\text{sr}} \tag{4}$$

T_i^{day} is the daytime of the ith day; T_i^{ss} and T_i^{sr} represent the sunset and sunrise moments of the ith day, respectively. So t^* in the normalised ideal PV output of the ith day can be converted into t through the following formula

$$t=t^*\cdot T_i^{\text{day}} \tag{5}$$

According to the above steps, we can get the ideal PV output $P_{\text{norm}}(t)$ of all days during the study period.

2.3 Calculation of amplitude parameters

$P_{\text{norm}}(t)$ reflects the shape of PV output curve when there is no cloud disturbance. Its amplitude range is [0, 1]. In practice, the peak value of daily PV output is affected by many factors, including solar radiation peak of atmospheric upper bound, atmospheric attenuation, and so on. However, these factors basically do not affect the shape of the ideal PV output curve. In this paper, the amplitude parameter is used to characterise these factors. The amplitude parameter is calculated using the least square method, as shown below

$$\min_{k_i}\left\{\sum_{t=1}^{N}\left[P_i(t)-k_iP_{i,\text{norm}}(t)\right]^2\right\} \tag{6}$$

In the equation, i represents the date; k_i is the corresponding amplitude parameter; N represents the number of PV data sampling points in 1 day. Making weekly timing curve of an actual PV power plant and the corresponding amplitude parameter series, as shown in Fig. 2.

2.4 Random component of PV output

The random component mainly reflects cloud disturbance. According to formula (1), random component can be obtained by the measured PV output series subtracting the ideal PV output which is amplified by amplitude parameter. The random component is shown in Fig. 3.

In [7], it is pointed out that the stochastic fluctuation component of the PV output satisfies the t-location scale (TLS) distribution. The fluctuation of PV output is intermittent and has a certain degree of continuity, as shown in Fig. 3.

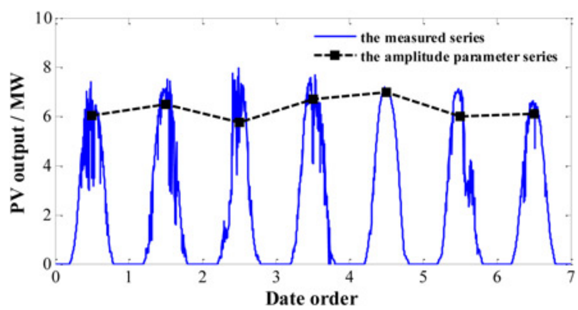

Fig. 2 *Weekly timing curve and amplitude parameter series*

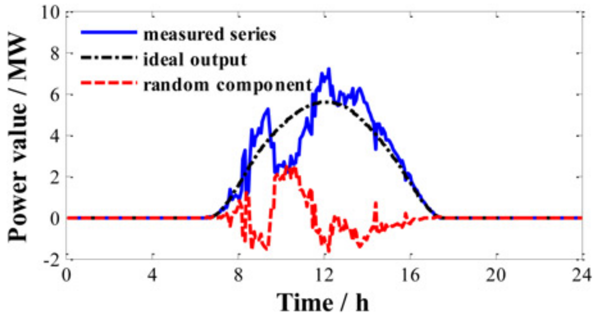

Fig. 3 *Schematic diagram of the decomposition of the PV output*

The feature is described using duration distribution in this paper. Duration is the time which the PV output remains smooth or fluctuating [13]. Its distribution refers to the probability distribution of the duration of different lengths. When the value of a sampling point of random component is >0.1 p.u., it can be assumed that PV output enters the fluctuating state and record the duration of its continuous fluctuation. When the value is <0.1 p.u. and maintains two or more sampling points long, PV output can be considered to enter the smooth state. Recording the length of time that it continues to be smooth. The probability distribution and parameters can be determined according to the duration of the fluctuating and smooth state.

Using the measured data and Matlab toolbox dfittool, we found that the inverse Gaussian distribution (IGD) is suitable for describing the duration distribution of the fluctuating part and smooth part.

3 Generation method for PV power time series

Using the above-mentioned ideal output normalisation curve extraction method of typical days and generation method for ideal output normalisation curve of atypical days. We can get the ideal output curve for every day in the study period. The ideal output normalisation curve can be considered only related to the date, so it is fixed and unique. In the following, we generate the random series of the amplitude and the random component and then combine them with the ideal output normalisation curve to obtain the generated PV power time series.

3.1 Markov Chain Monte Carlo-based generation method for amplitude parameter series

Markov Chain Monte Carlo (MCMC) method is a stochastic simulation method which takes the interaction between the various states of the system into consideration. Assuming that the value of the discrete random variable x_t is time-dependent, and t belongs to a discrete time set T. The whole possible value of x_t is a discrete state set S, and $S=\{s_1, s_2, s_3, \ldots\}$. If the conditional probability of x_t is satisfied

$$\begin{aligned} \boldsymbol{P}\{x_{t+1} = s_{t+1} | x_1 = s_1, x_2 = s_2, \ldots, x_t = s_t\} \\ = \boldsymbol{P}\{x_{t+1} = s_{t+1} | x_t = s_t\} \end{aligned} \quad (7)$$

It is said that the random variable x_t with Markov quality [14].

The transition probability matrix $\boldsymbol{P}$ is of size $N \times N$, and N is the number of states that the random variable may achieve. The value of each element p_{ij} in the matrix $\boldsymbol{P}$ represents the conditional probability

$$p_{ij} = \boldsymbol{P}(x_{t+1} = j | x_t = i) \quad (8)$$

This paper considers MCMC method for generating amplitude parameter series. Firstly, obtaining the magnitude of the amplitude parameter range corresponding to each state according to the below equation

$$k_0 = \frac{k_{\max}}{N} \quad (9)$$

In the equation, $k_{\max}$ is the maximum value in the amplitude series, N the state division number, and k_0 the size of the range of amplitude parameters represented by each state. In this paper, the division number N is 4, representing rainy days, partly cloudy days, cloudy days, and sunny days four weather conditions. If an amplitude parameter k_t satisfies the range constraint

$$k_t \in \left((i-1) \cdot k_0, i \cdot k_0\right) \quad i = 1, 2, \ldots, N \quad (10)$$

It is assumed that k_t corresponds to state i. Performing state transition for each value in the amplitude parameter series to get the corresponding amplitude parameter state series.

We can generate matrix $\boldsymbol{P}$ based on the state series according to (8). The cumulative transition probability matrix $\boldsymbol{P}_{\text{cum}}$ is further calculated according to the below equation

$$\boldsymbol{P}_{\text{cum}}(i, j) = \sum_{m=1}^{j} \boldsymbol{P}(i, m) \quad (11)$$

Then generating initial state randomly and using the $\boldsymbol{P}_{\text{cum}}$ matrix and Monte Carlo method to generate a new state series with length t_r. t_r is the length of the amplitude parameter series to be generated. The flowchart of generating new state series is shown in Fig. 4.

Finally, generating random variables which satisfy the distribution within the corresponding range of each state to convert the discrete state series into a series of consecutive random variables. The steps of generating the amplitude parameter value from the state value are described in a concrete example. Assuming that the range of amplitude parameter represented by a state is $(0.2k_{\max}, 0.3k_{\max}]$, the specific generating steps are as follows:

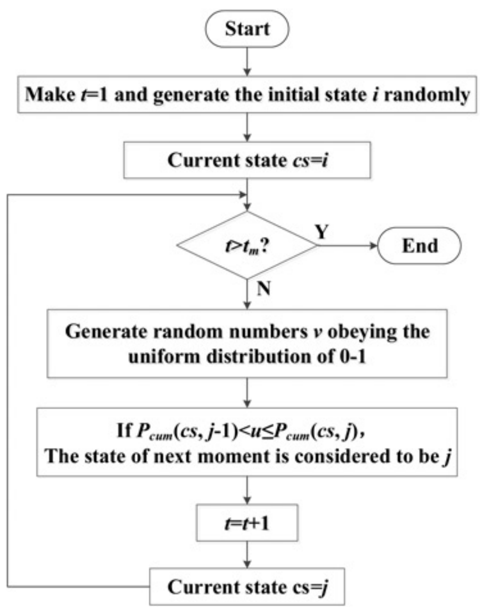

Fig. 4 *Flowchart of generating new state series*

(i) Calculating the CDF values for each amplitude parameter point of the original series in the range of $(0.2k_{\max}, 0.3k_{\max}]$, as shown below

$$F(x) = \boldsymbol{P}(a \le x) = \frac{n(a \le x)}{N\left(0.2k_{\max}, 0.3k_{\max}\right)} \qquad (12)$$

In the equation, a represents the sample points in the range of $(0.2k_{\max}, 0.3k_{\max}]$; x represents a sample value in the range of $(0.2k_{\max}, 0.3k_{\max}]$; $n(a \le x)$ is the number of sample points less than or equal to x in the range of $(0.2k_{\max}, 0.3k_{\max}]$; $N(0.2k_{\max}, 0.3k_{\max}]$ is the total number of sample points in the range of $(0.2k_{\max}, 0.3k_{\max}]$.

(ii) Generating a random number u which is distributed in [0, 1] uniformly and compares it with the CDF value obtained in step (i).

(iii) If the random number u is equal to the value of a certain $F(x_i)$, x_i is the generated value. If u is not equal to any values of $F(x)$, then u must belong to a certain interval $[F(x_i), F(x_{i+1})]$. $F(x_i)$ is the value of $F(x)$ which is less than u and closet to u. $F(x_{i+1})$ is the value of $F(x)$ which is greater than u and closest to u. Taking x_{i+1} as the amplitude parameter generation value at that time.

For the other amplitude parameter range, follow the steps above to generate specific values.

3.2 Random component generation with considering the duration

The random component is randomly generated according to the result of the parameter fitting. Therefore, it is necessary to calculate the TLS fitting parameters of its probability distribution, the IGD fitting parameters of the duration of the smooth and fluctuating parts based on the random component separated from the measured data. Then following the flowchart (Fig. 5) to generate the random component series of specified length. On this basis, the output before sunrise and after sunset is replaced with zero. We can get a complete random component series.

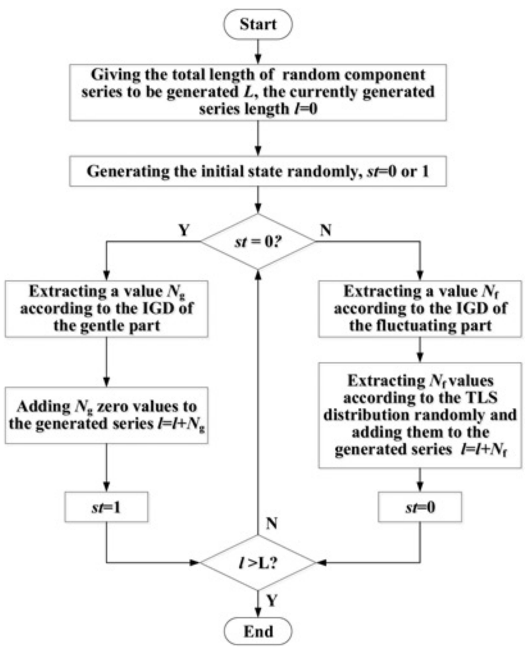

Fig. 5 *Flowchart of generating random component*

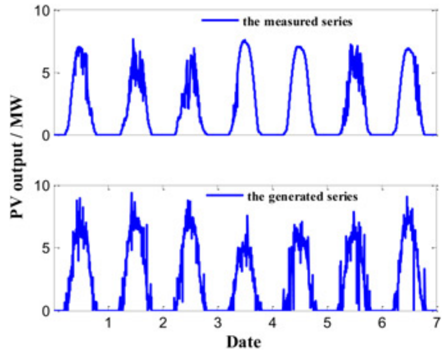

Fig. 6 *Comparison of the output curves*

3.3 Method of generating PV output series

Based on the above method, we can generate the required length of the ideal output normalisation curve, amplitude parameter series, and random component series, respectively. Then we can combine them into complete PV output series according to formula (1). This method is used to generate PV output for 1st PV power plant based on the measured data. The output of measured series and generated series of a week are shown in Fig. 6.

It can be seen that the generated series curve is well shaped to preserve the characteristics of the original series curve, inherit the exact sunrise and sunset moments, while reflecting the intermittent characteristics of PV output.

4 Simulation and verification

4.1 Data source

This paper uses the output data of six PV power plants in Gansu, China, to carry out simulation test. The basic information of the PV power plants is shown in Table 1. The length of generated series is equal to the original series length.

4.2 Comparison of statistical characteristics

4.2.1 Basic statistical characteristics: The basic statistical characteristics of PV output series include average and variance. Fig. 7 shows the basic statistical characteristics of measured and

Table 1 Basic information of PV power plants

No.	Power plants name	Cap/MW	Length	Interval, min
1	CECEPChangMa	10	52,128	5
2	CECEPSDaTan	20	52,128	5
3	CPInterJingTai	50	52,128	5
4	CPInvestWuwei	50	52,128	5
5	CPInvestJingTai	10	52,128	5
6	QSJinTa	3	52,128	5

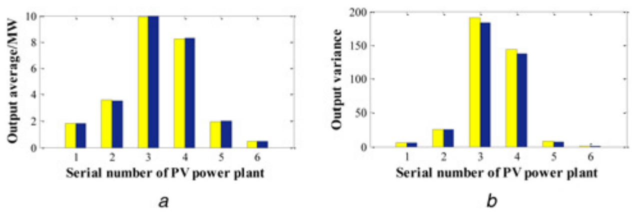

Fig. 7 *Comparison of the basic statistical properties*
a Average
b Variance

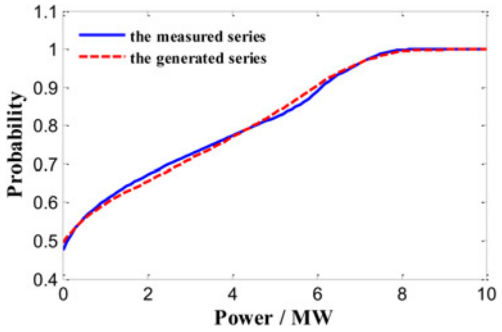

Fig. 8 *PDF curves of the measured and generated series*

generated series. In the figure, the light column represents the measured series and the dark column represents the generated ones. It can be seen that the generated series well inherits the basic statistical properties of the measured series.

4.2.2 PV output distribution: PV output distribution is one of the most basic statistical properties to measure the quality of PV output series generation. Introducing the Kolmogorov–Smirnov test (KS test) [15] to compare whether the two sets of random variables are subject to the same distribution. The KS test actually calculates the maximum vertical distance D between the two sets of CDF curves and compares it with the critical values at a given significant level. If the D value is less than the critical value, the two sets of random variables can be considered to be subject to the same distribution.

This paper considers significant levels of $1-\alpha = 0.999$. The comparison of the CDF curves of the measured series and the generated series of 1st PV power plant is shown in Fig. 8.

It can be seen from Fig. 8 that the CDF curve of the measured series and the generated series are almost coincident, indicating that the fitting effect of generated series for PV output distribution is excellent. It can be seen from the results in Table 2 that all test PV plants pass the KS test. The validity of generated series fitting distribution of the measured series is showed further.

4.2.3 Fluctuation characteristics of different time scales: The fluctuation characteristics of different time scales are the distribution of active power change in the PV output series at different time steps. For example, the 30 min level fluctuation characteristics is the amount of active power change between two sampling points separated by 30 min. In this paper, 30 min, 1, 2, and 4 h are selected as different time scales for analysing. The KS test is used to determine the effect of the generated series fitting the measured series. The four time-scale fluctuation characteristics of the 1st PV power plant are shown in Fig. 9. In Fig. 9, the blue solid line represents the fluctuation series of the measured data and the red dotted line represents the fluctuation series of the generated data. It can be seen from the figure that the fitting degree of fluctuation characteristics at different time scales are relatively high. The generated series can well inherit the fluctuation characteristics of the measured series. The validity of the method fitting fluctuation characteristics of the measured series is fully demonstrated by the results of KS test in Table 3.

Table 2 KS values of PV output distribution curves

No.	KS values	Critical values
1	0.0168	0.0210
2	0.0194	0.0210
3	0.0204	0.0210
4	0.0158	0.0210
5	0.0210	0.0210
6	0.0171	0.0210

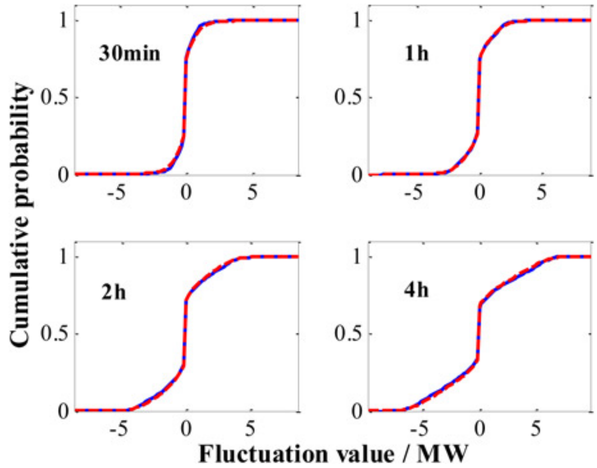

Fig. 9 *Comparison of CDF curves of fluctuating series*

Table 3 KS values of different time scales fluctuation characteristics curves

No.	30 min	1 h	2 h	4 h	Critical values
1	0.0205	0.0068	0.0123	0.0131	0.0210
2	0.0189	0.0160	0.0160	0.0190	0.0210
3	0.0166	0.0142	0.0173	0.0176	0.0210
4	0.0191	0.0104	0.0132	0.0123	0.0210
5	0.0208	0.0210	0.0203	0.0153	0.0210
6	0.0204	0.0198	0.0197	0.0155	0.0210

5 Conclusion

In order to better understand the PV generation and provide sufficient data support for system operators, a novel generation model for PV power time series combing decomposition technology and Markov chain is proposed in this paper. The following conclusions can be drawn from this work.

(i) The PV power time series can be decomposed into ideal output normalisation curve, amplitude parameter series, and random fluctuating component. These three parts correspond to regular changes in solar radiation, atmospheric attenuation, and cloud disturbance, respectively.
(ii) The generation method for amplitude parameter series based on MCMC method can simulation the distribution of the original amplitude parameter series, thus reflecting the weather situation.
(iii) The power time series generated by the method presented in this paper can well inherit the characteristics of the measured series, including average, variance, cumulative distribution, and fluctuation characteristics.

6 Acknowledgment

This work is sponsored by China Postdoctoral Science Foundation (2016M590693).

7 References

[1] Eftekharnejad S., Vittal V., Heydt G.T., *et al.*: 'Impact of increased penetration of photovoltaic generation on power systems', *IEEE Trans. Power Syst.*, 2013, **28**, (2), pp. 893–901

[2] Thomson M., Infield D.G.: 'Impact of widespread photovoltaics generation on distribution systems', *IET Renew. Power Gener.*, 2007, **1**, (1), pp. 33–40

[3] Liu S., Liu P.X., Wang X.: 'Stochastic small-signal stability analysis of grid-connected photovoltaic systems', *IEEE Trans. Ind. Electron.*, 2016, **63**, (2), pp. 1027–1038

[4] Duffie J.A., Bechman W.A.: 'Solar engineering of thermal processes' (Wiley, New York, 1980), pp. 3–133

[5] Zhang X., Kang C., Zhang N., *et al.*: 'Analysis of mid/long term random characteristics of PV power generation', *Autom. Electr. Power Syst.*, 2014, **38**, (6), pp. 6–13

[6] Zhao L., Li J., Ai X., *et al.*: 'Analysis on random component extraction and statistical characteristics of photovoltaic power', *Autom. Electr. Power Syst.*, 2017, **41**, (1), pp. 48–56

[7] Dhople S.V., Dominguez-Garcia A.D.: 'Estimation photovoltaic system reliability and performance metrics', *IEEE Trans. Power Syst.*, 2012, **27**, (1), pp. 554–564

[8] Wang M., Zong X., Yuan Y., *et al.*: 'Reliability analysis of generation systems with photovoltaic stations', *Proc. CSEE*, 2013, **33**, (34), pp. 42–49

[9] Klöckl B., Papaefthymiou G.: 'Multivariate time series models for studies on stochastic generators in power systems', *Electr. Power Syst. Res.*, 2010, **80**, (3), pp. 265–276

[10] Luo G., Shi D., Chen J., *et al.*: 'A Markov chain Monte carlo method for simulation of wind and solar power time series', *Power Syst. Technol.*, 2014, **38**, (2), pp. 321–327

[11] Ding M., Xu Z.: 'Empirical model for capacity credit evaluation of utility-scale PV plant', *IEEE Trans. Sustain. Energy*, 2016, **8**, (1), pp. 94–103

[12] Li S.: 'Solar physics' (Capital Normal University Press, Beijing, 1996), pp. 28–111

[13] Yu P., Li J., Wen J., *et al.*: 'A wind power time series modeling method based on its time domain characteristics', *Proc. CSEE*, 2014, **22**, pp. 3715–3723

[14] Liu C: 'Stochastic process' (Huazhong University of Science and Technology Press, Wuhan, 2014), pp. 41–42

[15] Massey F.J.Jr.: 'The Kolmogorov–Smirnov test for goodness of fit', *J. Am. Stat. Assoc.*, 1951, **46**, (253), pp. 68–78