

Short term wind power scenarios forecast based on multivariate normal distribution

Shuai Liu, Yongli Zhu Jiacheng Gao and Ke Zhang*

North China Electric Power University, Baoding, China

*Corresponding author e-mail: liushuai_0517@163.com

Abstract. According to the fact that the increase of wind power penetration rate causes the influence of wind power's randomness and fluctuation on the power grid, the single short-term wind power point prediction often cannot meet the needs of power grid risk assessment and decision-making. In this paper, we first calculate the theoretical probability model of each wind power forecast box by the end function in MATLAB, and then use the exponential covariance function expression to determine the best covariance matrix corresponding to the dynamic scenarios, and determine the multivariate normal distribution model of wind farm output obedience at multiple connected moments; For each predicted moment of the wind power point prediction value of the wind belongs to the prediction box, we direct sample random vector which obey multivariate normal distribution to form the wind power dynamic scenario. After a simulation experiment on a real wind farm, the results show that the scenario set considering wind power fluctuation at different time scales can cover the measured wind power curve and the reliability of the method is proved.

1. Introduction

With the increasingly prominent problem of global environmental destruction and energy shortage, the growth and utilization of renewable energy sources have attracted widespread attention. Based on the rapid rise of wind power generation and its high degree of randomness and intermittency[1], the impact of large-scale wind power grid integration on the power grid has become more and more obvious, causing the situation of “disposal of wind” in China to be very tense. Wind power forecasting is the most essential and critical technology for improving the operation of large-scale wind power access systems, which we all agree with today. Therefore, accurately predicting the output of wind farms has important value for stable and economical operation of power grids with large wind farms.

The research and application of wind power forecast in China in the past 10 years have achieved a breakthrough from scratch [2]. Currently, there are two major types of prediction methods for wind farm output: physical methods and statistical methods. The forecast forms mainly include point prediction, probability prediction, and interval prediction. The literature [3] uses the improved particle swarm optimization to optimize the parameters affecting the regression function of the least squares support vector machine, and forms a novel wind farm output point prediction model. The literature [4] proposed a method to predict the short-term output probability of wind farms based on sparse Bayesian learning that can actively prevent over-learning and it is not necessary to set the penalty factors for the equilibrium experience risk and confidence intervals in SVM. The literature [5] carries



out statistics on the probability density function of the prediction deviation in each power interval, and proposes a confidence interval prediction strategy with a certain level of significance based on the distribution of the predicted deviation distribution. Although the discussion on forecasting methods of wind farm output is more adequate, each single forecasting model has more or less problems [6]. The wind power point prediction is only a conclusive figure of the wind farm output that meets certain accuracy requirements at a certain moment, but it cannot give the possibility that the number may occur and the possible fluctuation bounds; Although probabilistic prediction gives information about the measured probability distribution that approximates the distribution of a certain probability model, it cannot reveal the possible change areas of the prediction result; Interval prediction although given the upper and lower limits of wind power to meet a certain confidence level, it cannot describe the wind power fluctuations at different times. Moreover, they all share a common problem. They do not consider that there is a correlation between wind powers in successive periods. This correlation shows that wind power does not arbitrarily fluctuate, and the amplitude and range of wind power fluctuations have a certain probability.

In view of the above research status, this paper adequately exploits this correlation and adopts a short-term wind power scenario prediction method based on multivariate normal distribution. Under the premise of known wind power point prediction, a multivariate normal distribution forecasting model was established, and using the data from Ireland wind farms in 2016, the daily power scenario prediction was performed, demonstrating the effectiveness of the method.

2. Wind Power Probability Distribution

2.1. Wind Power ECDF Model

When the probability distribution of the wind power is mostly directed at the specific wind farms. Due to the differences in the geographical environment, station conditions and meteorological conditions of wind farms, their probability distribution models are quite different. Therefore, the probability distribution model of wind power does not have a universal analytical distribution function.

For the power system, the probability distribution of wind power is a statistical inference process of its external changes. The wind power can be regarded as a random variable p , assuming that the theoretical distribution of the random variable is $F(p)$, and $F_L(p)$ is the empirical distribution function. Glivenko proved the following theory in 1953[6]:

$$\Pr \left\{ \lim_{l \rightarrow \infty} \sup_{|x| < +\infty} |F_l(p) - F(p)| = 0 \right\} = 1 \quad (1)$$

The theorem illustrates that when the experimental data l is large enough, $F_L(p)$ is very close to $F(p)$. On the one hand, it can be shown that the population can be inferred from the sample, and on the other hand, when the empirical distribution function of the sample size is large enough, it is very close to the actual theoretical distribution. This is the theoretical basis for estimating the probability distribution of wind power using non-parametric empirical distribution.

When the theoretical distribution of wind power is not known, assume that the order statistics x_1, \dots, x_L are a sample of the wind power random variable p , then the cumulative empirical distribution function of wind power is as follows:

$$F_l(p) = \frac{1}{l} \sum_{i=1}^l \theta(p - x_i) \quad (2)$$

$$\theta(p - x_i) = \begin{cases} 1, & p \geq x_i \\ 0, & p < x_i \end{cases} \quad (3)$$

In order to verify the effectiveness of the empirical distribution, we randomly generate one thousand and ten thousand standard normal distribution random numbers in the MATLAB environment, and obtain the cumulative empirical distribution of the two sets of normal distribution random numbers through formula (2). The normal distribution function is compared as shown in Figures 1 and 2. The result is obvious.

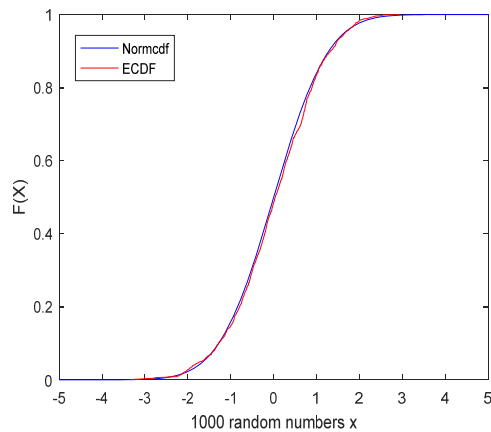


Figure 1. Theoretical distribution and ECDF

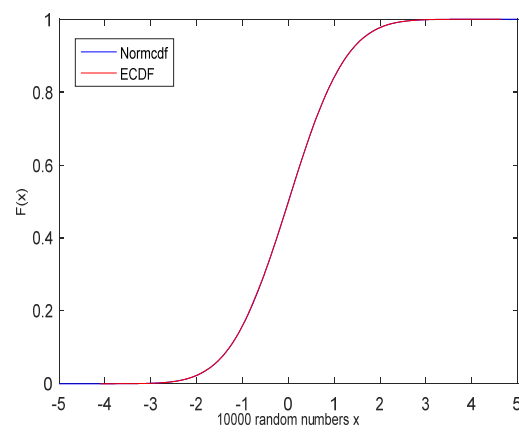


Figure 2. Theoretical distribution and ECDF

2.2. Wind Power Forecast Box

The data used in this paper was provided by the Ireland Power Company and was used to standardize the data. This article draws on the modeling idea in the literature[7], first sorts the point prediction values of wind power from the largest to the smallest, and divides the predicted value into several “value intervals”, according to the size of the predicted value, the corresponding data set [measured Value Predicted Values] is placed in the corresponding numerical range. For the data used in this paper, the numerical interval length is 0.05p.u. There are 19 numerical intervals, and all data sets in each numerical interval are called a “forecast box”.

In a forecast box for wind power, the prediction values of all data sets are very close, but the corresponding measured values are quite different. By observing the measured wind power in the forecast box to observe whether it obeys a known analytical distribution function, Figure 3 shows that the distribution of measured wind power in the 7th forecast box is closer to the normal distribution. Figure 4 shows that The distribution of measured wind power in the 18th forecast box is more inclined to the Beta distribution, so it is impossible to simply infer the theoretical distribution of its obey. In this paper, non-parametric empirical distribution will be used to approximate the theoretical distribution of wind power measured values.

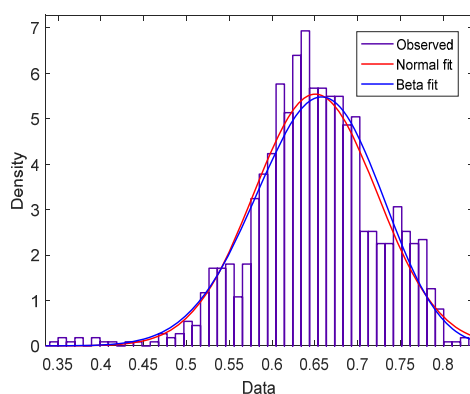


Figure 3. 7th forecast box

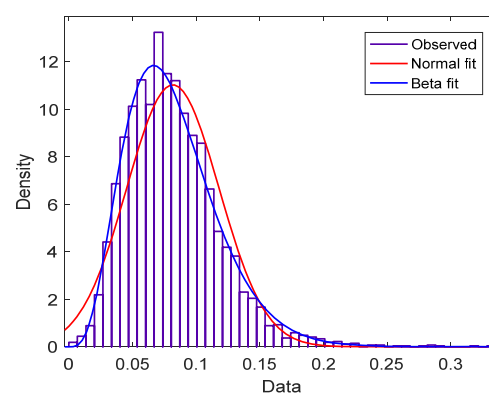


Figure 4. 18th forecast box

Direct sampling is the simplest and most common measure of producing random variables. Its theoretical basis is the principle of probabilistic integral transformation, which generates random variables by inverting the cumulative distribution function.

Let the cumulative distribution function of the random variable x be

$$F(x) = p(X \leq x) \quad (4)$$

Then the range of $F(x)$ is $[0, 1]$. For the sampling value that can generate a random variable x , a number of random numbers that conform to the uniform distribution $U[0, 1]$ are first formed, and these random numbers are used as the values of the distribution function $F(x)$, and then obtain the sample value of the random variable x by solving the inverse function of the distribution function $F(x)$, as shown in Figure 5.

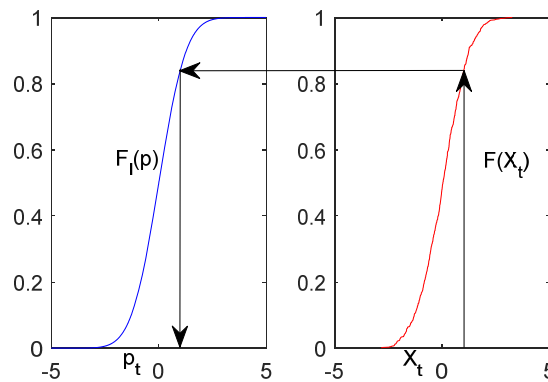


Figure 5. Direct sampling principle diagram

In view of the large number of random variables X_t that obey the multivariate normal distribution random numbers in this paper, since the set of cumulative distribution function values $F(X_t)$ of multivariate normal distributions of these random numbers is obey uniform distribution $U[0,1]$, we can use the following two formulas obtain the sampled value of the random variable Ext .

$$F(\mathbf{X}_t) = \int_{-\infty}^{\mathbf{X}_t} \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T \Sigma^{-1}(\mathbf{x}-\mu)} d\mathbf{X}_t \quad (5)$$

$$p_t = F_t^{-1}(F(\mathbf{X}_t)) \quad (6)$$

Where d is the dimension of the random variable; Σ represents the covariance matrix of the multivariate normal distribution; $|\Sigma|$ represents the determinant of the covariance matrix; Σ^{-1} represents the inverse of the covariance matrix.

3. Multivariate Normal Distribution of Wind Power

The output power of the wind farm can be regarded as a random variation process. Let the random vector $\mathbf{X}=(X_1, X_2, W_k)^T$, k is the length of time, which is 96 in this paper. Based on the Central Limit Theorem [5], we assume that the random vector \mathbf{X} conforms to the multivariate normal distribution $\mathbf{X} \sim N(\mu, \Sigma)$, and the mean vector μ can be obtained from the historical data of wind power. The covariance matrix Σ is:

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix} \quad (7)$$

Where it represents the covariance of the random vector Zip and the random vector so, imp=1, 2, k.

3.1. Determination of Covariance Matrix

This article defines the wind power fluctuations per unit time interval as

$$\Delta p = p_{t+1} - p_t \quad (8)$$

The wind power fluctuation Up can also be regarded as a random variable. The literature [8] clarifies that the student t distribution can properly describe the change of wind power. This paper uses the distribution fitting tool in the MATLAB statistical toolbox to do a variety of probability density fits and their corresponding probability plots for the theoretical distribution of wind power measured fluctuation values, as shown in Figure 6 and Figure 7.

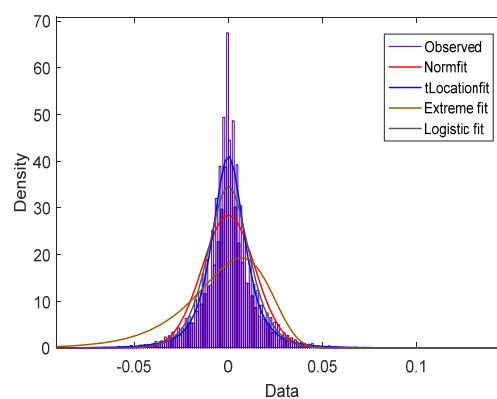


Figure 6. Measured value fluctuations fitted

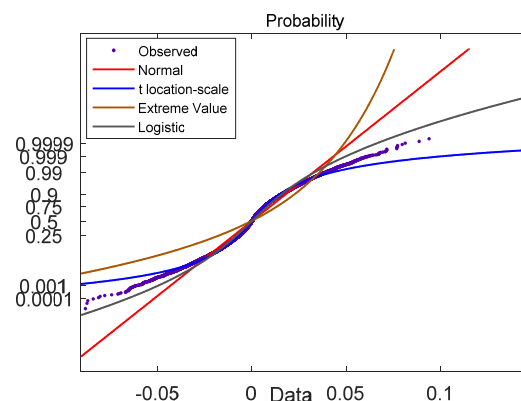


Figure 7. Measured value fluctuations probability

In this paper, the statistical diagnosis effect of Kolmogorov–Smirnov distance [10] is shown in the following table.

Table 1. Goodness of fit results.

Distribution	K-S dist.
Normal	0.0794
t location-scale	0.0293
Extreme value	0.1654
Logistic	0.0515

The results show that the distribution of wind power measured value shows a clear kurtosis and thick tail. Whether it is from the fitting of probability density function plot, probability plot or Table 1, it can be shown that the fitting effect of t Location-Scale distribution is superior to other distribution and more accurately describe the measured wind power fluctuations.

This article uses the thinking of literature [11] to perform exponential modeling of covariance:

$$\sigma_{ij} = e^{-\frac{|i-j|}{\varepsilon}}, i, j = 1, 2, \dots, k \quad (9)$$

In the formula, ε is the time scale parameter used to control the correlation between the random variables X_t . In order to make the fluctuations of wind power dynamic scenarios closer to the fluctuations statistical rule of the measured value of wind power, the best time scale parameter ε should be estimated to achieve this goal. When the time scale parameter ε is determined, the covariance matrix Σ is uniquely determined. At this time, a random vector with multiple normal distribution functions $X \sim N(\mu, \Sigma)$ can be generated by using the mvnrnd function in MATLAB to generate the wind power dynamic scenarios vector; Then exploring the fluctuation value vector Δp of the dynamic scenarios of the wind power and the historical fluctuation value vector Δp of the wind power history. When the Δp is regarded as a simulated distribution and Δp is taken as an actual distribution, the KL divergence is calculated using the following formula:

$$\min_{\varepsilon} D_{\varepsilon}(P||Q) = \sum_{x \in X} P(x) \log \frac{P(x)}{Q(x)} \quad (10)$$

Where P is the actual distribution of data, Q is the approximate distribution of data, in order to judge the approximate level between their probability models. When the two random distributions are consistent, their KL divergence value is zero; when the difference between the two random distributions increases, their KL divergence value also increases. Therefore, when the KL divergence is minimized, the time scale parameter ε at this time is the best estimate time scale parameter. From the exponential modeling of the covariance matrix, it can be seen that when the best time scale parameters are determined, the covariance matrix Σ is uniquely determined.

3.2. Dynamic Scenarios Generation Steps

In the case that the short-term wind power point prediction value set and the empirical distribution function of the measured wind power value in each forecast box are known, the dynamic scenarios generation of wind power is as follows:

Step 1: First, identify the best covariance time-scale parameter ε according to the objective function in equation 9, then determine the optimal covariance matrix of the exponential model, and then determine the multivariate normal distribution of wind power $X \sim N(\mu, \Sigma)$.

Step 2: Use the mvncdf function in the MATLAB Statistics Toolbox to generate d samples of a random vector X that obeys a multivariate normal distribution of wind power.

Step 3: Determine the forecast box to which the wind power point predict value for each leading time section belongs. In this forecast box, use the formulas (4) and (5) to directly sample for the d multivariate normal random vector, normal distribution random numbers are converted into d wind power dynamic scenarios

4. Analysis of Examples

In order to verify the effectiveness of the proposed method, the wind farm in Northern Ireland was used as a case study. The measured values and point predictions of wind power at every 15 minutes during the 2016 year were experimental data, and a total of 19 forecast boxes have been established in descending order of wind power point prediction, $[0.90, 1]$, $[0.85, 0.90]$, ..., $[0, 0.05]$. Then, the empirical distribution function was counted in turn using the measured value of wind power in each forecast box.

As shown in Fig. 8, since the time scale parameter ε will affect the correlation of multivariate normal distribution random numbers at different time slices, the rationality of the value of the time scale parameter ε will directly affect the fluctuation of randomly generated wind power dynamic scenarios. Experiments show that when $\varepsilon = 234$, The KL divergence between the probability distribution of wind power history data and the probability distribution of dynamic scenarios reaches a minimum of 8.0564, which means that the two probability distributions reach the maximum closeness.

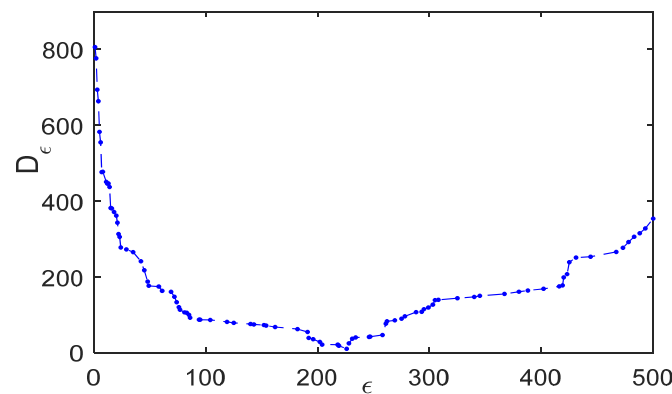


Figure 8. The connection of ε and $D\varepsilon$

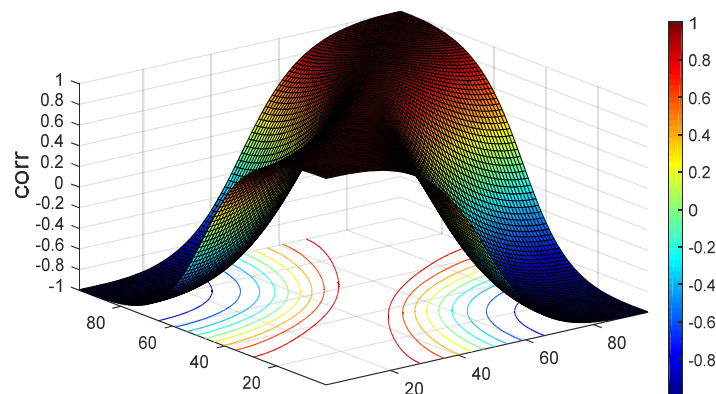


Figure 9. Correlation coefficient matrix

Second, when the optimal time scale parameters are determined, our best covariance matrix is uniquely determined. Then we use the function `corrcoef` in MATLAB to calculate the correlation coefficient matrix and plot it as shown in Figure 9. From the figure, it is not difficult to find that there are differences in the correlation strengths at different time levels, and when the distances at different time levels increase, the strong correlation decreases. It also clarifies and complies with the correlation of wind power at different time intervals.

Based on the best time scale parameter ε and the best covariance matrix Σ , a multivariate normal distribution statistical model obeyed by wind power was determined, and a number of 50 wind power dynamic scenarios set were formed according to the point prediction values of wind power on a certain day as shown in Figure 10. Each of the thin solid lines represents a possible scenario of wind power generated by the method described in this paper; the blue thick solid line represents the measured values of a certain day of wind power, and the red thick solid line represents the point prediction values corresponding to the wind power measured values on a certain day. From the figure, it is not difficult to see that the set of 50 wind power dynamic scenarios completely contains the measured values of wind power. What is more prominent is that the overall change trend of the wind power dynamic scenarios set is very similar to the change trend of the wind power measured value curve. It explains the effectiveness of this article approach.

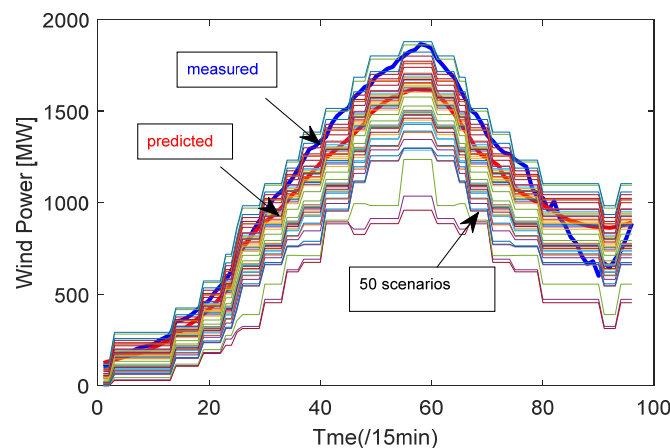


Figure 10. Scenarios forecast result

5. Conclusion

In this paper, under the premise of point prediction of wind power, we fully explore the relevance of wind power as a random variable at different time intervals and the inherent properties of wind power, and establish an empirical distribution of wind power and a multivariate normal distribution model. It shows a forecasting strategy of wind power dynamic scenarios and illustrates the effect of the method more profoundly through calculation examples. At the same time, the wind power scenario prediction can not only reduce the wind abandonment phenomenon, increase the utilization of wind energy and reduce the system operation risk, but also can flexibly plan the output of the hydro-thermal power unit in the system and optimize the system's spinning reserve capacity, thus improving the economic efficiency of the system operation.

References

- [1] Huang Guodong, Li Weigang, Li Zhenbin, et al. A review on generation dispatching of power system with large-scale wind power[J]. *Electrical Measurement & Instrumentation*, 2015, 52(9): 1-4.
- [2] Gu Xingkai, Fan Gaofeng, Wang Xiaorong, et al. Summarization of wind power prediction

- technology[J]. Power System Technology, 2007, 31(2): 335-338.
- [3] Wang He, Hu Zhijian, et al. Short-term prediction of wind power based on IPSO-LSSVM[J]. Power System Protection and Control 2012, 40(24): 108-111.
- [4] Yang Ming, Fan Shu, et al. wind farm generation forecast based on componential sparse bayesian learning[J]. Automation of Electric Power Systems, 2012, 36(14): 125-130.
- [5] Vapnik V.N. Statistical learning theory[M]. 1998, New York: Wiley.
- [6] Wang Linhong, Yang Yuhong. A short-term wind power predication method based on improved probability weights theory[J]. Electrical Measurement & Instrumentation, 2015, 52(8): 45-48.
- [7] Bludszuweit H, Dominguez-Navarro J. A., Llombart A. Statistical analysis of wind power forecast error[J]. IEEE, Transactions on Power Systems, 2008, 23(3), 983-991.
- [8] Pinson P., Madsen H., Nielsen H. A., et al. From probabilistic forecasts to statistical scenarios of short-Term wind power production[J]. Wind Energy, 2009, 12(1): 51-62.
- [9] Lin W., Wen J., Cheng S., Lee W.-J. An investigation on the active-power variations of wind farms[J]. IEEE, Transactions on Industry Application, 2012, 48(3): 1087-1094.
- [10] Zhu Jianguo, Feng Cuilian. Statistical diagnosis based on Kolmogorov-Smirnov distance[J]. Statistics & Decision, 2010, 24(13): 37-38.
- [11] Pinson P., Girard R. Evaluating the quality of scenarios of short-term wind power generation[J]. Applied Energy, 2012, 96(1): 12-20.