

# Research on the Relationship between Wind Power Correlation and Volatility

Mingshu Song<sup>1,\*</sup>, Jingjie Bai<sup>2</sup>, Qing Zhu<sup>2</sup>, Qi Liu<sup>2</sup> and Yongchao Yu<sup>2</sup>

<sup>1</sup>State Grid Xinjiang Electric Power Company, Urumqi, Xinjiang, 830000

<sup>2</sup>NARI Group Beijing Branch East Company, Beijing 100192

\*Corresponding author

**Abstract**—Based on the wind farm operating data of acquisition and analysis in real-time, analyzing the regional relativity and output volatility of wind power situation. Through the cooperation of the regional relativity and output volatility, we can know the relationship of the regional relativity and output volatility. The relationship of the regional relativity and output volatility could be used in dispatching management to support the professional management of new energy grid scheduling. That can support useful information for wind priority scheduling and strengthen the advancement of wind power net and priority consumptive.

**Keywords**—wind power; the regional relativity; the output volatility

## I. INTRODUCTION

Wind power is clean and renewable green energy with abundant reserves, which makes wind power the most suitable renewable energy to be large-scale developed and commercial operated<sup>[1]</sup>. As wind power installed capacity grows and wind farms expand their scale, wind power causes more and more important influence on power system<sup>[2]</sup>. The randomness and volatility of wind power bring load and challenges for power system peak shaving. However, though thermal power is relatively stable, the pollutant brought by thermal power seriously affects environment, which violates the principle of priority consume clean energy. Therefore, how to improve the utilization of wind power is still one of the most important research priorities.

Wind resource distribution would make the output volatility of large-scale wind power less than the summation of corresponding numbers of individual wind turbine, which is called Smoothing Effect<sup>[4]</sup>. The geography adjacent wind farms would have similar valley-peak output change trend, and this phenomenon is called correlation. Those are two opposition but unity aspects of wind power fluctuation characteristics. The stronger the correlation the worse the complementarity and the Smoothing Effect would not be satisfying. Otherwise the Smoothing Effect would be better. Therefore smoothing effect and correlation are the main exemplifications of wind power volatility that has been widely studied. On aspect of time domain, literature [3-4] analyzed the statistical data of wind power output and described wind power volatility and change trend in different spatial and temporal scales of Inner Mongolia in detail based on the. Literature [5] respectively did qualitative analysis and quantitative calculation of wind power output features in Northeast China, Gansu and Coastal Areas in Jiangsu. Only

few literatures include corresponding models. Literature [6] built output power curve model based on measured data of wind power. Literature [7] uses normal distribution model of wind power output change rate. Literature [8] built an exponential model between wind power output correlation coefficient and the distance between wind farms. Literature [9] used Fourier decomposition method to analyze how the volatility of an individual wind turbine output influences overall output of the wind farm. Though researches on frequency domain could analyze the impact of wind turbine output volatility on overall output by building models, the research scales are generally limited in single wind farm and involve complicated deduction and calculation, which makes them not practical enough for engineering application.

This paper analyzes large amounts of measured data and combines Hypothesis Testing methods to build a parametric fitting model that could describe wind power output Smoothing Effect and correlation. First, use data testing to provide that the normal distribution model is not reasonable for wind power output changing rate. Then build a two-component one-dimensional Gaussian mixture model for wind power output changing rate probability distributions and analyzed the law that the parameters of the model influenced by wind power scale. Based on this model, this paper defined the confidence interval, smoothing coefficient and relevant coefficient for wind power changing. At last this paper deduced the relationship between large-scale wind power smoothing coefficient and relevant coefficient, and build corresponding models based on measured data. This research has guiding significance for generating wind power cluster test data and provide statistical model reference for large-scale wind power grid connection analysis.

## II. WIND POWER OPERATION ANALYSIS AND EVALUATION MODEL

Wind power grid connection power grows with the increasing of wind power installed capacity. However, the abandoned wind power is also increasing, which mean big waste of clean energy and corresponding investment. Therefore, environment conflicts are exacerbated and good energy-saving effect is hard to achieve. This paper researches on evaluation model for wind power operation status and analyze the volatility, correlation, output rate, units' effective utilization hour and forecasting hours. The framework of this paper is shown in Figure 1.

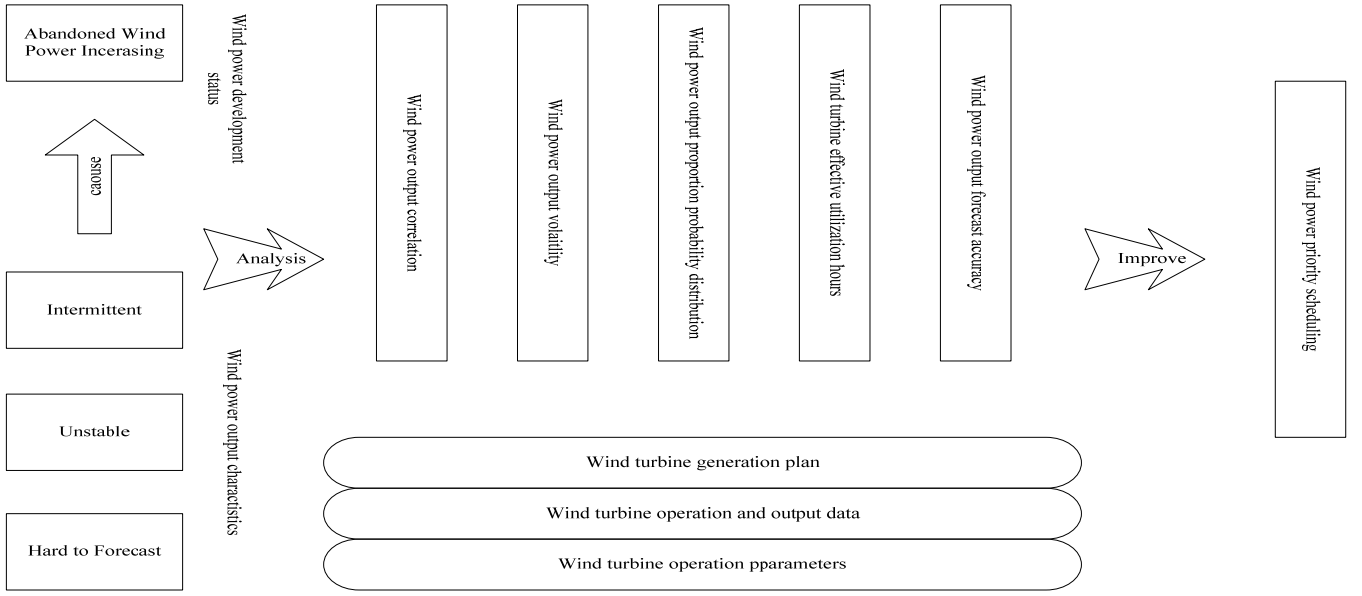


FIGURE 1. FRAMEWORK OF WIND POWER OPERATION ANALYZE AND EVALUATION MODEL

#### A. Probability Distribution Model of Wind Power Rapidly Volatility

Building probability distribution model of wind power changing rate is the foundation of analyzing wind power Smoothing Effect and correlation. Define wind power changing rate  $P_R$  as equation (1).

$$P_R = \frac{P(t-1) - P(t)}{P_N} \quad (1)$$

Wherein,  $P(t)$  is the total output of all wind turbines at time  $t$ ;  $P_N$  is the total rated installed capacity.

Many literatures assume the probability distribution of the changing rate  $P_R$  follows a normal distribution. However, by data simulation testing it could be found that the accuracy of the model is not satisfying under some conditions. Therefore, this paper introduces Gaussian model to improve the accuracy of the probability distribution model.

The probability distribution model of Gaussian mixture model [17] is the weighted summation of several Gaussian probability distribution models. This paper chooses one-diamonded Gaussian mixture model, as equation (2) and equation (3).

$$f(x) = \sum_{j=1}^n \alpha_j N(\mu_j, \sigma_j^2) \quad (2)$$

$$N(\mu_j, \sigma_j^2) = \frac{1}{(2\pi)^{1/2} \sigma_j} e^{-\frac{1}{2\sigma_j^2}(x-\mu_j)^2} \quad (3)$$

The fitting function is equation (4) and  $\alpha_i$  is the weight of component  $i$ .

$$f(x) = \alpha_1 \times \frac{1}{2\sqrt{2\pi}\sigma_1} \exp\left(-\frac{x^2}{2\sigma_1^2}\right) + \alpha_2 \times \frac{1}{2\sqrt{2\pi}\sigma_2} \exp\left(-\frac{x^2}{2\sigma_2^2}\right) \quad (4)$$

When the number of wind turbines changes, the above model could also get satisfying fitting results. The fitted parameters are list in Table I.

TABLE I. FITTING PARAMETERS OF GAUSSIAN MIXTURE MODEL

Wind turbine number	5	10	15	20	25	28
$\sigma_1$	0.0027	0.0020	0.0016	0.0014	0.0012	0.0011
$\sigma_2$	0.0085	0.0058	0.0045	0.0038	0.0033	0.0031
$\alpha_1$	0.6525	0.6352	0.5842	0.5667	0.549	0.5373
$\alpha_2$	0.3475	0.3648	0.4158	0.4333	0.4502	0.4627

According to Table 1, for the two components, one has large variance and another has small variance. This phenomenon shows that the probability distribution would generate two components while the output power changing. One of the components shows rapid decay and the other shows slow decay, this also provides the probability is different from normal distribution. Logarithmic the fitted standard deviations and wind turbine number to get the relationship Figure, as Figure II. According to Figure II, the

two standard deviations show power function decays when wind turbine number increase, which reflects that second level wind power output volatility would decrease as wind farm scale expand. It can be supposed that when wind farm scale reaches a certain level, the probability distribution with smaller standard deviation would attenuate rapidly and approaches normal distribution. In this condition, if the number of wind turbine is large, normal distribution is reasonable to be used to describe the probability distribution.

### B. Large-scale Wind Power Output Volatility Modeling and Analyzing

Large-scale cluster grid connection is the main developing mode of wind power in China. Large-scale wind power cluster generally crosses 100 kilometers to 500 kilometers and controlled by the secondary power FM of the corresponding regional power grid. Therefore, regional power grid puts more attention on minute level power volatility law of large-scale wind power cluster. This paper chooses the measured data of five-minute-interval wind power output in serval provinces of north China to do simulation. According to simulation results, it could be provided that Gaussian mixture model is also fit for second level wind power cluster output volatility analysis, and the specific parameters would be different for different time and space scale. Using normal distribution and two-component Gaussian mixture model to simulate single or multiple wind farms output power changing rate probability distribution, and the fitting deviations are list in Table II and Table III.

TABLE II. FITTING DEVIATION OF SINGLE-FARM OUTPUT VOLATILITY PROBABILITY DISTRIBUTION

Single wind farm output	5 min	10 min	15 min
Normal distribution fitting deviation/%	34.35	31.51	31.98
Gaussian mixture model fitting deviation/%	8.82	9.85	12.69

TABLE III. FITTING DEVIATION OF MULTI-FARM OUTPUT VOLATILITY PROBABILITY DISTRIBUTION

Cluster wind farm output	5 min	10 min	15 min
Normal distribution fitting deviation/%	22.06	24.42	25.91
Gaussian mixture model fitting deviation/%	4.68	5.44	8.25

According to Table 2 and Table 3, fitting deviation of Gaussian mixture model is much less than that of normal distribution, which verified that the model put forward by this paper has better accuracy. On the other hand, for a same time scale, when using normal distribution model, multiple wind farm output fitting deviation is obviously less than single wind farm output fitting deviation. The overall output fitting deviation is less than the fitting deviation summation of corresponding number of single wind farm, which means multiple wind farm output changing rate is more close to normal distribution.

Obviously, when Gaussian mixture model variances of the two components gradually getting close or the weight of one component is approaching zero, two-component one dimensional Gaussian model would degenerate into normal

distribution. Combining this law, and comparing the fitting parameters in Table IV and Table V it could be found that for a certain time scale, both of the standard deviations  $\sigma_1$  and  $\sigma_2$  would decrease, which reflects the Smoothing Effect of cluster output. On the other hand,  $\sigma_1 / \sigma_2$  would approaching 1, this provide the probability distribution of cluster output changing rate would approaching normal distribution.

TABLE IV. FITTING PARAMETERS OF SINGLE-FARM OUTPUT VOLATILITY PROBABILITY DISTRIBUTION

Wind turbine number	5	10	15
$\sigma_1$	0.587	0.4505	0.4513
$\sigma_2$	0.413	0.5495	0.5487
$\alpha_1$	0.04458	0.04878	0.05985
$\alpha_2$	0.1337	0.1578	0.1975
$\sigma_1 / \sigma_2$	3.00	3.23	3.30

TABLE V. FITTING PARAMETERS OF MULTI-FARM OUTPUT VOLATILITY PROBABILITY DISTRIBUTION

Wind turbine number	5	10	15
$\sigma_1$	0.5947	0.6712	0.625
$\sigma_2$	0.4053	0.3288	0.375
$\alpha_1$	0.02505	0.04040	0.05115
$\alpha_2$	0.05715	0.09947	0.12966
$\sigma_1 / \sigma_2$	2.28	2.46	2.53

In summary, when wind farm is in small scale, two-component one dimensional Gaussian model I needed to accurately describe wind power output changing rate. But with wind farm scale expanding, the probability distribution would gradually approach normal distribution.

## III. WIND POWER OUTPUT CORRELATION STATISTICAL ANALYSIS

### A. Correlation Analysis

Correlation analysis analyzes two of serval variables to measure the degree of correlation between the variables. Correlation analysis requires the variables has certain correlation or probability. Correlation coefficient can reflect the degree of correlation between the variables<sup>[3]</sup>. The correlation coefficient  $|r| = 0$  means the variables do not have correlation;  $0 < |r| \leq 0.3$  means the correlation is weak;  $0.3 < |r| \leq 0.5$  means the correlation is not significant;  $0.5 < |r| \leq 0.8$  means the correlation is significant;  $0.8 < |r| < 1$  means the variables are highly correlated;  $|r|=1$  means the variable are fully correlated. Correlation is not causation ore simple personalize. The scope and domain of correlation almost covers all aspects of human's life. The definitions of correlation vary greatly in different subjects. The definition of correlation could be equation (5) or (6).

$$r = \frac{\overline{xy} - \bar{x} * \bar{y}}{\sqrt{(\overline{x^2} - \bar{x}^2) * (\overline{y^2} - \bar{y}^2)}} \quad (5)$$

$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{n \sum x^2 - (\sum x)^2} \sqrt{n \sum y^2 - (\sum y)^2}} \quad (6)$$

### B. Regional Correlation Analysis of Wind Farm Output

Regional correlation analysis of direct transfer wind farms in Northern Hebei could be deduced based on equation (6), as equation (7).

$$r = \frac{\sum (x_i - \bar{x}) * (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} * \sqrt{\sum (y_i - \bar{y})^2}} \quad (7)$$

Wherein  $i$  is the output time index;  $x$  and  $y$  is wind farm.

This paper chooses measured data of direct transfer wind farms in Northern Hebei as basic data. The statistical analysis shows the main factor that influence regional correlation in a certain time period is geographical location.

### C. Wind Farm Output Volatility Statistic

Wind resource is uncontrollable, and wind power output statue is determined by wind velocity. The unstable and intermittent make wind power output has the same characteristics, which would influence power system stable operation. Voltage fluctuations and flash edge are the main reasons that bring negative impact to power system<sup>[4]</sup>.

The volatility of wind power output is mainly influenced by climate, weather, terrain and other uncontrollable natural factors. Wind power has low credit density and distributed in a wide range of space. Wind turbine has its particular operation and control characteristic. Currently, those two aspects make most of the wind turbines do not have the capability of inhibit power fluctuations<sup>[5]</sup>. Wind power output volatility mainly influences on power system planning. The volatility would make flow distribution more complex, the unstable factors would bring attack to the original power system operation mode. As wind power scale expands, wind power would influence on power system, and also influence wind farm location. For example, if wind power installed capacity counts 20% or more of the grid capacity, the peaking shaving capability and safe operation would face great challenge. Therefore, from the aspect of system operation, wind farm output volatility would influence the injecting trend and scheduling scheme.

Based on the measured daily operation data of the direct transfer wind farms in Northern Hebei, this paper uses the following equations to do statistical analysis.

Fifty-minute volatility:

$$F(n) = f(n+15) - f(n) \quad (8)$$

$$(0 \leq n \leq 1424)$$

Five-minute volatility:

$$F(n) = f(n+5) - f(n) \quad (9)$$

$$(0 \leq n \leq 1434)$$

One-minute volatility:

$$F(n) = f(n+1) - f(n) \quad (10)$$

$$(0 \leq n \leq 1439)$$

## IV. LARGE-SCALE WIND POWER OUTPUT SMOOTHING EFFECT AND CORRELATION ANALYSIS

### A. Wind Power Smoothing Effect Analysis

To measure the Smoothing Effect of large-scale wind power, this paper defines the smoothing coefficient  $S$ , which takes the global and local volatility characteristics into consideration.

Definition 1, using the relative change value of sequence standard deviation of non-dimensional individual wind turbine output<sup>[12]</sup> and overall standard deviation to measure  $S$ , as equation (11).

$$S_1 = \frac{\sigma_{signal, P.u.} - \sigma_{cluster, P.u.}}{\sigma_{signal, P.u.}} \quad (11)$$

The above definition is simple, intuitional and easy to be understood, which is suit for a fitted time scale. It can be used to compare smoothing effect under different time scale, but not suit for evaluate smoothing effects. It could be provide mathematically that using the average value of wind power output to gain data of long time scale cannot avoid new sequence's standard deviation increasing, which would cover the data's own law. To compare the changing laws of the smoothing effect under different time scale, this paper introduces definition 2.

Definition 2, using the relative change value of the volatility confidence interval to measure  $S$ .

$$S_2 = \frac{R_{signal} - R_{cluster}}{R_{signal}} \quad (12)$$

Definition 2 describes the relative difference between the volatilities of single wind farm and cluster wind farm under the same cumulated probability, which is an indictor based on probability distribution. Definition 2 has good consistency for different time scale smoothing effect. The smaller the smoothing coefficient, the worse the improving effect.

This paper chooses measured data of 7 wind farms in north China, which are nearby to each other, as the basic data, and the simulation result is list in Table VI. As time scale expand, the correlation between increases, the smoothing effect of cluster output decreases and the

smoothing coefficient decreases.

TABLE VI. SMOOTHING PARAMETERS OF DIFFERENT TIME SCALES CALCULATED ON MEASURED DATA

Time scale	5min	10min	15min	1h
$S_1$	0.2427	0.2576	0.270	0.3467
$S_2$	0.5692	0.5096	0.4758	0.3006

#### B. Wind Power Smoothing Effect and Correlation Analysis

Based on the measured data of a cluster wind farm in north China, this paper calculated the relationship coefficients according to equation (9). Wherein, the output sequences of two wind farms are regarded as random variables X and Y. Calculating the smoothing coefficient S, defined by standard deviation, and gain the scatterplot as Figure III.

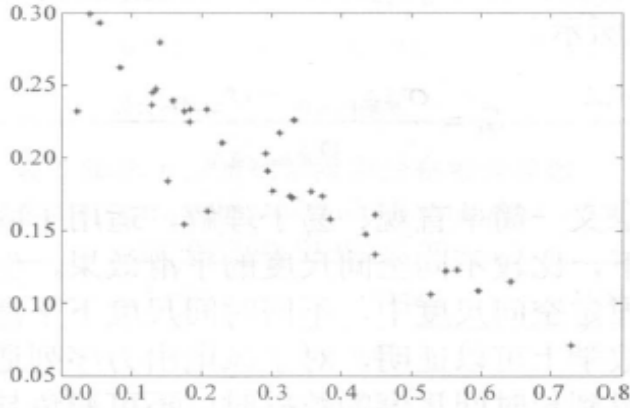


FIGURE II. SCATTER PLOTS OF SMOOTHING EFFECTS AND CORRELATION COEFFICIENTS

According to Figure III, as correlation coefficient increases, the smoothing effect coefficient decreases, this meets the related experience. Fitting function  $S = -0.2872r_{x,y} + 0.2775$  could be gain based on linear function. The negative sign before the correlation coefficient means smoothing effect and correlation have negative correlation. According to the fitting function, when the smoothing coefficient equals 0, the corresponding correlation is 1, which means the two wind farms have positive correlation.

In summary, when wind farm number increases, the overall volatility would be weaken in vary degrees. When the correlation of wind farm output is big, this weaken effect would be small.

#### V.CONCLUSIONS

Based on the measured wind power output data, this paper combined statistical analysis and fitting methods to study wind power output volatility characteristics under different time sale and space scale. According to the simulation result, the following conclusions could be obtained.

1) The probability distribution of single or cluster wind

farm output volatility changing rate do not strictly obey normal distribution. The simulation results show two-component one-dimensional Gaussian mixture model could improve the fitting accuracy. But with wind farm scale expanding, the probability distribution would approach normal distribution. Therefore, when accuracy requirement is not high, large scale wind farm output volatility could also use normal distribution to describe its changing probability distribution.

2) This paper defined the smoothing coefficient that determined by the relative change value of the volatility confidence interval, which can be used to measure wind power output volatility changing law. As the number of wind farm increases, the smoothing coefficient shows an exponential decay, but till determined by wind velocity, wind turbine type, wind turbine control strategy and other factors. The power function could describe the quantitative relationship between wind power output and the number of wind turbine or wind farm, which can be a good prosthesis for simple qualitative analysis in the existing researches.

3) According to correlation data and wind power output volatility data, when regional correlation less than 0.1 the output curves of the two farms do not have obvious trend relationship. When the regional correlation come to 0.6, the output volatility curve of the two farms would almost keep the same, and when it achieves 1 the curve almost keep full consistent trend. Therefore, the regional correlation and volatility show a positive relationship.

#### REFERENCES

- [1] WANG Xu, ZHANG Yuan. Wind Power[M]. BeiJing: Publishing China Electric Power.
- [2] SUN Rongfu, ZHANG Tao, LIANG Ji. Assessment and Application of Grid acceptance of wind power capacity[J]. Automation of Electric Power Systems, 2011,04:70-76.
- [3] HU Yuanyuan, WANG Xiaolong, PANG Erjun. The Relativity, Randomness and Volatility of Wind Power Operation[J]. Instrumentation Analysis Monitoring, 2012, 04: 9-13.
- [4] QU Keding, YU Jun, JIN Fuguo. Analysis on wind Power Operation in Northeast China power system[J]. Northeast Electric Power Technology, 2008, 11: 22-24.
- [5] TIAN Ru, ZHANG Dongying. Wind Power Output Volatility Overview,2012,07:12.
- [6] XIAO Chuangying, WANG Ningbo, DING Kun. Power characteristics of Jiuquan wind power base[J]. Automation of Electric Power System, 2010,17:64-67.
- [7] LI Feng, LU Yichuan. The Impact of Large-scale Wind Power on the Power System[J]. China Electric Power, 2006,11:80-84.
- [8] PANG Aili. The Impact on the System After the Wind Power accessed[J]. ShangHai Electric Power,2007,01:33-35.