

Correlation Analysis of Photovoltaic Power with the Field Data of Xinjiang Korla Photovoltaic Power Plants

Kewen Yang

North China Electric Power University, Beijing, 102206, China

Abstract—Mastering the characteristics of photovoltaic power generation helps reduce the output limit of photovoltaic power generation and improve the absorption ability of photovoltaic power generation. Based on the measured data from photovoltaic power stations, the sunrise force-related characteristics of a photovoltaic power station are analyzed. Specifically, first, the correlation coefficient and the sunrise force characteristics of a photovoltaic power station are introduced. Second, the output data and time-space characteristics of two photovoltaic power stations are provided. Furthermore, based on the above actual data, the influence of weather and time on different time autocorrelation characteristics of a photovoltaic power station is analyzed. Based on the spatiotemporal characteristics of a photovoltaic power station, the correlation between time and space characteristics and photovoltaic power generation data of different substations is analyzed. Our study is conducive to further understanding the characteristics of photovoltaic power station data.

Keywords—*photovoltaic power station; autocorrelation; cross-correlation; time delay*

I. INTRODUCTION

Renewable energy has become an important solution to the world's sustainable energy development strategy, considering the problems of fossil fuel depletion, environmental pollution, and climate change. In recent years, solar photovoltaic power generation in China has developed on a rapid scale and has attracted considerable attention.

In China, solar photovoltaic capacity reached 4.9 GW [1] in 2012, up 32.4% from 2011. By 2015, the installed capacity of solar power generation will reach 21 GW, of which the installed capacity of photovoltaic power station is 10 GW, the installed capacity of solar thermal power generation is 1 GW, and the installed capacity of distributed photovoltaic power generation system connected and off-grid is 10 GW [2].

However, the randomness and intermittency of large-scale photovoltaic power generation will lead to large-scale power fluctuations, which may largely change the power flow distribution and have a significant impact on peak load regulation and frequency modulation. This would threaten the safety of power systems, leading to restrictions and cuts in photovoltaic power generation. On the contrary, the active power value of different photovoltaic power stations located relatively close has a certain relationship with time and space.

These related features can help power system operators take steps ahead of time and reduce wasted light energy. Therefore, several studies have been conducted on the correlation of photovoltaic power generation.

We analyzed random characteristics of solar radiation of previous literature [3,4,5] in the field of meteorology random, the literature [6-10] is from the randomness of solar radiation, and then study the randomness of the power. In previous studies [11,12], the influencing factors of photovoltaic power output characteristics from the viewpoint of power electronic devices were studied. In this paper, the effects of geographical location, meteorological factors, and time of sunrise on photovoltaic output will be emphasized. In general, research on the correlation of photovoltaic power generation can be divided into two categories: one is the self-correlation of the same photovoltaic power station at the same time on different days, and the other is the mutual correlation of different photovoltaic power stations at the same time. In this paper, both correlations are studied.

Although much progress has been made in the research of relevance, it cannot meet the needs of power system operators. The correlation study considers the correlation of photovoltaic power generation from the time perspective, which has an important impact on the characteristics of photovoltaic power generation, and it is difficult to take the seasonal output characteristics and peak load reverse regulation into account. Therefore, the characteristics of large-scale photovoltaic power station base need to be further studied.

On the basis of understanding the above problems, this paper further discusses the correlation between time and space and the active power generation of different photovoltaic power stations with the photovoltaic power generation data of the photovoltaic power station base in Korla, Xinjiang.

The remainders of this paper are organized as follows: the delay case is briefly introduced in section II and the concept of correlation coefficient does not consider the delay case. Section III introduces the Korla and Slek Yanqi photovoltaic power station data, In the section IV, the correlation analysis of the photovoltaic power station is carried out, and then, section V analyzes two photovoltaic power station sub-sources. Finally, the conclusion and comments are provided in section VI.

II. CORRELATION COEFFICIENT AND PHOTOVOLTAIC OUTPUT CORRELATION

A. The Correlation Coefficient

For two variations X and Y, their sample values are $X=(x_1, \dots, x_n)$, $Y=(y_1, \dots, y_n)$, the correlation coefficient calculation formula is as follows:

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

Where $COV(X, Y)$ is the covariance of variable X and Y, whose corresponding standard deviations are σ_X and σ_Y .

The correlation coefficient is the statistical indicator that reflects the degree of correlation between variables. The value of the correlation coefficient r ranges between -1 and 1. If $r = 0$, variables X and Y are not related, else if $|r| = 1$, variables X and Y are completely related; simultaneously, there is a linear relationship between X and Y. If $r < 1$, changes in variable X lead to changes in Y; the greater the absolute value, the greater the influence.

Further, for two time-series X and Y, their sample values are $X=(x_1, \dots, x_n)$, $Y=(y_1, \dots, y_n)$, The correlation coefficient R(l) at the delay l (Y delayed by X) is calculated as follows:

$$R(l) = \frac{\sum_{t=l+1}^n (x_t - \bar{x})(y_{t-1} - \bar{y})}{\sqrt{\sum_{t=l+1}^n (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^{n-1} (y_t - \bar{y})^2}} \quad (2)$$

$$\bar{x} = \frac{1}{n-1} \sum_{t=l+1}^n x_t$$

$$\bar{y} = \frac{1}{n-1} \sum_{t=1}^{n-1} y_t$$

where the maximum value of l is n/2. When the value of l changes from 0 to n/2, multiple R(l) values can be obtained, and

the delay corresponding to the maximum value of R(l) is lmax. If $R(l_{max}) > \sigma$ (an artificially defined threshold), then we say that the time-series X and Y have a delay correlation, and Y is delayed by the quantity lmax of X. The curve with l as the abscissa and R(l) as the ordinate is called the "delay correlation curve", and the point where the maximum value of R(l) is located on the curve is called the "maximum delay correlation point".

When $l=0$, (2) is equivalent to (1).

B. Qualitative analysis of the Sunrise Force Correlation in Photovoltaic Power Station

This paper mainly considers the correlation between the sunrise force of photovoltaic power station. We consider that photovoltaic output is determined by radiation intensity. In fact, the photovoltaic output is affected by weather changes, cloud cover, sampling time, geographical location, output limit, and other factors. That is, its active power output sequence starts at 0 AM and goes on till 12 PM.

Effects of climate change: There are differences in irradiation energy values in different weather conditions. On rainy days, the light intensity weakens and the cloud cover is more considerable, resulting in a serious shortage of radiation energy, which reduces the functionality of the photovoltaic system, and then the weather changes will affect the correlation between two days.

Influence of different sampling times: On a given same day, the light intensity, weather conditions, and cloud cover at different times are different. If different sampling times are selected for measurement in a day, the data will be greatly different. If different sampling times are selected across two days, the maximum correlation coefficient between two days will be time-shifted.

Impact of geographical location: Different geographical locations cause differences in longitude and latitude, weather, and sunrise time so that photovoltaic power stations at two different places have different photovoltaic output at the same time, and the correlation between them is also reduced.

Due to the geographical proximity, the light intensity and output of different photovoltaic power stations are correlated. The mathematical relationships can be described by correlations between variables. It is an important in mathematical statistics to study the correlation between variables. The simplest way to describe correlations between variables is to rely on a measure index, which uses one or more numerical statistics to reflect the direction or degree [13] of correlation. The linear correlation coefficient is the most commonly used numerical correlation statistic.

III. INTRODUCTION OF PHOTOVOLTAIC POWER STATION DATA

A. General Introduction

The data used in this paper come from two photovoltaic substations in Korla, Xinjiang, namely the Hanergy Korla

photovoltaic power station in Korla city and Guangrun Yanqi photovoltaic power station in Yanqi county.

The geographical location of the two photovoltaic substations is shown in TABLE I

TABLE I. THE GEOGRAPHICAL POSITION

Area	longitude	latitude
Korla	E86° 10' 15.94"	N41 ° 43 ' 29.87"
Yanqi Country	E86° 34' 19.15"	N42 ° 03 ' 28.45"

From TABLE I, it can be seen that the two geographic locations have a longitude difference of 24 '3.21' and latitude difference of 19 '58.58, the straight-line distance between the two photovoltaic power stations is 78.85 km, the east-west distance is 60.78 km, and the north-south distance is 50.23 km.

The data of photovoltaic power station is the active power generation value of each photovoltaic power station calculated every minute. The sampling time is from January 23 to 30, 2015.

B. Data of A Power Station

The active power of every minute of power generation between January 23 and 31, 2015, of Hanergy Korla photovoltaic power station is shown in FIGURE I.

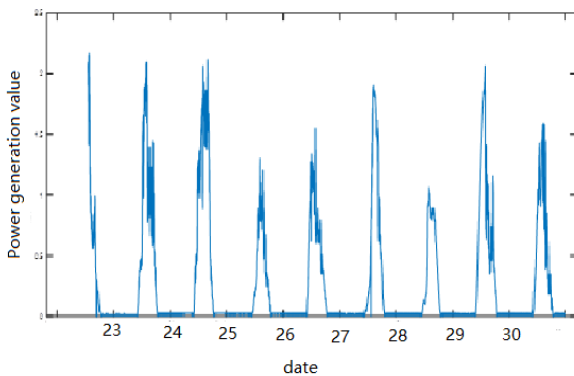


FIGURE I. ACTIVE VALUE OF PHOTOVOLTAIC POWER GENERATION

In particular, the weather in Korla during this period is shown in TABLE II

TABLE II. THE WEATHER IN KORLA

Data	Maximum temperature [°C]	Minimum temperature [°C]	Weather
2015/01/23	2	-7	Sunny to cloudy
2015/01/24	2	-7	Sunny to cloudy
2015/01/25	0	-5	Cloudy
2015/01/26	0	-6	Yin

2015/01/27	0	-7	Yin
2015/01/28	-1	-8	Cloudy and cloudy
2015/01/29	-1	-10	Yin
2015/01/30	-2	-9	Yin
2015/01/31	-2	-10	Cloudy and cloudy

C. Data of B pOwer Station

The active power value of the photovoltaic power station between January 23 and 30, 2015, is shown in FIGURE II.

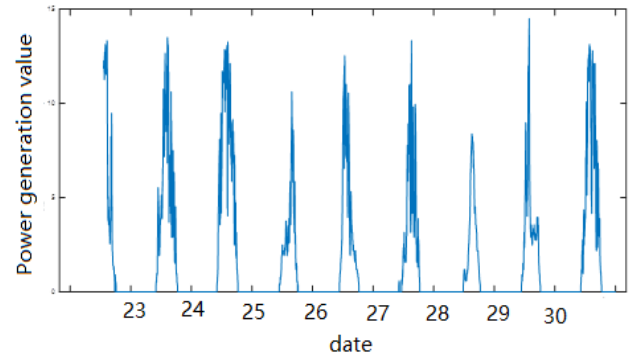


FIGURE II. ACTIVE VALUE OF PHOTOVOLTAIC POWER GENERATION

IV. ANALYSIS OF AUTOCORRELATION OF PHOTOVOLTAIC POWER STATION

This section takes the Hanergy Korla photovoltaic power station as an example to analyze the correlation characteristics of output data of photovoltaic power station.

A. The Effect of Weather Characteristics

The active power generation value of the Hanergy Korla photovoltaic power station on January 28 and 29, 2015, is shown in FIGURE III.

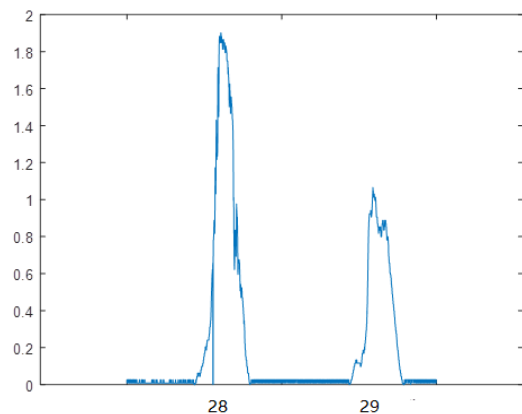


FIGURE III. TWO DAYS' ACTIVE VALUE OF PHOTOVOLTAIC POWER GENERATION

FIGURE III shows that the photovoltaic output generated on January 19 is lower than that on January 28. The reason is that the output of photovoltaic power generation is determined by radiation intensity, and the active power value of power generation is affected by various factors such as weather change and cloud cover. According to the weather forecast for Korla, the temperature is 2 degrees lower on the 29th than on the 28th, and 29, cloud cover is serious, the autocorrelation coefficient values of 28 day with other days is calculated, as shown in FIGURE IV, the autocorrelation coefficient was 0.6338. The low correlation between two days may be due to changes in the weather and increased cloud cover.

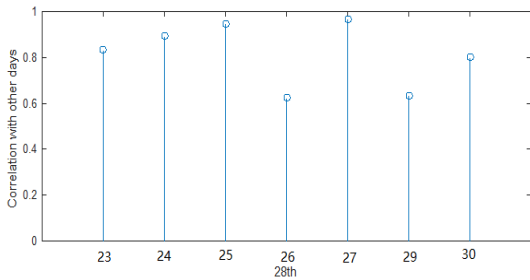


FIGURE IV. AUTOCORRELATION COEFFICIENT VALUES OF 28 DAY WITH OTHER DAYS

B. Different Time Effects

Furthermore, although the output of the substation varies across different dates, it still has a certain correlation. For the data of Hanergy korla photovoltaic power station from January 23 to 30, the autocorrelation coefficient values of 24 day with other days is calculated, as shown in FIGURE V.

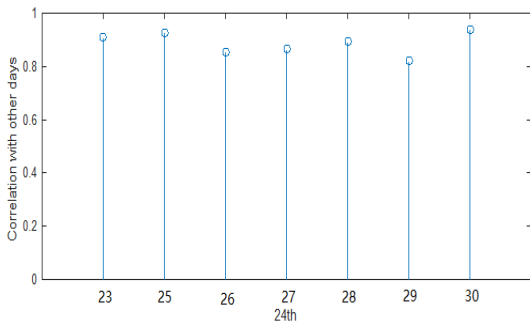


FIGURE V. AUTOCORRELATION COEFFICIENT VALUES OF 24 DAY WITH OTHER DAYS

As can be seen in FIGURE V, the correlation coefficient with January 30 is higher. Due to a similar temperature for the two days and same weather conditions, the photovoltaic output is similar.

Considering that the time of sunrise for January 24 and 30 is different, the correlation coefficient with delay is used to analyze the correlation. The correlation coefficient for different delay values for January 24 and 30 is shown in FIGURE VI.

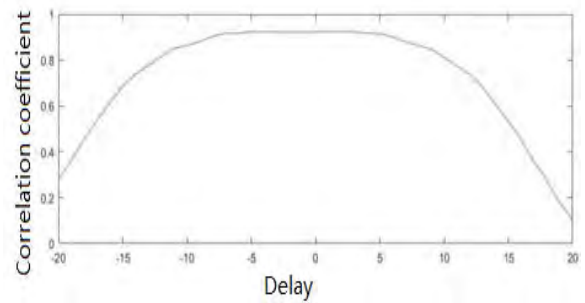


FIGURE VI. THE CORRELATION COEFFICIENT FOR DIFFERENT DELAY VALUES CONSIDERING THE TIME OF SUNRISE

It can be seen from FIGURE VI that the maximum correlation between the delay for January 24 and 30 occurs at the time of -5 min, that is, the time of sunrise on January 24 occurs later than on January 30 by 5 min. The real-time delay was 4 min 46 s; thus, it can be judged that the delay in the correlation coefficient between these two days is caused by the different sunrise time.

V. CASE ANALYSIS OF MUTUAL CORRELATION BETWEEN PHOTOVOLTAIC POWER STATIONS

This section takes the Hanergy Korla photovoltaic power station as an example to analyze the correlation characteristics of output data of a photovoltaic power station. Case analysis of mutual correlation between photovoltaic power stations.

A. Same Day Cross-correlation Analysis

When only considering the influence of geographical location on the number of mutual relations, i.e., with no delay effect, the number of mutual relations between the two photovoltaic power stations at the same time is as shown in FIGURE VII.

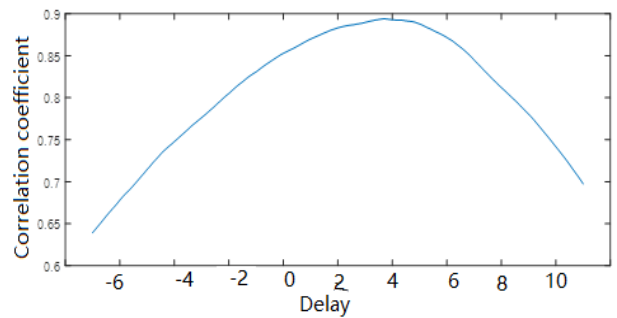


FIGURE VII. THE CORRELATION COEFFICIENT FOR DIFFERENT DELAY VALUES DUE TO DIFFERENT GEOGRAPHICAL LOCATIONS

It can be seen from FIGURE VII that the mutual relations between the Hanergy Korla and Guangrun Yanqi power stations have the highest value of 0.8939 when the correlation is the strongest, the two places are different by 4 min, with a geographical distance of approximately 80 km.

B. Correlation Analysis of Different Days

Considering the difference in sunrise time, the correlation coefficient with delay is used to analyze mutual correlation. The sun rose 6 min later on the 23rd than on the 30th, and the correlation coefficient was 4 min longer. Therefore, the correlation coefficient of the 23rd for Hanergy should be 2 min later than that for Yanqi for the 30th due to the difference in the sunrise time. The correlation coefficient of different delays between January 23 for Hanergy and January 30 for Yanqi is shown in FIGURE VIII. It can be seen that the maximum correlation (0.9155) between these two dates is -2 min.

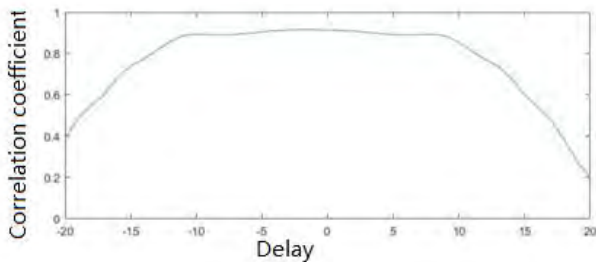


FIGURE VIII. THE CORRELATION COEFFICIENT FOR DIFFERENT DELAY VALUES DUE TO DIFFERENT SUNRISE TIME

After considering the effect of sunrise time, the relationship between the two power stations increases. It can be seen that the difference in sunrise time between the two photovoltaic power stations affects the active power value of the two photovoltaic power stations to a large extent.

VI. CONCLUSION

Based on the field data of the photovoltaic power station in Korla, Xinjiang, the following conclusions can be reached:

- a) without considering the delay, the output of pv power generation is determined by the radiation intensity. The active power value of pv power generation is affected by the geographical location, weather changes, cloud cover, and other factors.
- b) in the case of delay, the difference in sunrise time caused by different weather conditions is relatively high after the overall delay.
- c) the correlation between the two power stations is determined by the time of sunrise and the geographical location.

REFERENCES

[1] China Economic Times. New and accumulated electrical capacity in China over the years 2001-2002 [EB/OL][2013-09-3].<http://jjsb.cet.com.cn/articlecontent.aspx?articleID=152079>

[2] Renewable Energy "Twelfth Five-Year Plan" [EB/OL].[2013-09.03].<http://newsbjx.com.cn/html/20120810/37961.shtml>.

[3] K. G. T Hollands, R. G. Huget, "A probability density function for the clearness index with applications," Solar Energy, 1983.

[4] C. Zhijun, "Discussion on the model of monthly clear sky index in China," J. Nanjing Inst. Meteorol., vol. 5, 2005.

[5] S. A. Karim, B. S. Singh, Razali, R, et al, "Data compression technique for modeling of global solar radiation," IEEE Int. Conf. Control Syst. Comput. Eng., 2011.

[6] V. A. Graham, and K. G. T. Holland, "Method to generate synthetic hourly solar radiation globally," Solar Energy, 1990.

[7] G. Tina, S. Gagliano, and S. Raiti, "Hybrid solar/wind power system probabilistic modelling for long-term performance assessment," Solar Energy, vol. 5, 2005.

[8] J. Peng, A. Xin, and X. Jiajia, "Isolated microgrid economic operation model based on sequential operation theory," J. China Electrical Engineering, vol. 25, 2012.

[9] H. Wang, and X. Bai, "Short-term operational standby assessment of grid-connected photovoltaics," Power Syst. Auto., vol. 5, 2013.

[10] X. Fang, Q. Guo, D. Zhang, and S. Liang, "Evaluation of confidence capacity of photovoltaic power plants considering weather uncertainty," Power Syst. Auto., vol. 10, 2012.

[11] Z. Zhao, Y. Lei, F. He, Z. Lu, and Q. Tian, "Overview of large-capacity grid-connected photovoltaic power station technology," Automat. Electric Power Syst., vol. 12, 2011.

[12] H. Wang, J. Su, M. Ding, and G. Zhang, "Photovoltaic grid-connected power regulation system," J. China Electrical Engineering, vol. 2, 2007.

[13] Y. Li, "Wind speed models considering dependence and their applications in reliability evaluation of generating systems, Chongqing University, 2012.