

PAPER • OPEN ACCESS

Brief Analysis of Seasonal Correlations in the Power Generation Data of Photovoltaic Power Stations

To cite this article: Weihao Yu 2019 *IOP Conf. Ser.: Mater. Sci. Eng.* **484** 012061

View the [article online](#) for updates and enhancements.

Brief Analysis of Seasonal Correlations in the Power Generation Data of Photovoltaic Power Stations

Weihaio Yu

Department of Electrical Engineering, North China Electric Power University, School of Electrical and Electronic Engineering, room 8A111, No.2, Beinong Road, Changping District, Beijing, P.R.China..
Email: ywh0131@163.com

Abstract. Correlations in the power generation of photovoltaic (PV) power stations could improve the PV power utilization in electric power systems and reduce the occurrence of PV power dumping and throttling. In this study, the seasonal correlations in the PV power generation data of the Hanergy Xinhe PV power station in Xinjiang are investigated from a temporal perspective. Firstly, the PV output curves of the power station in typical weather conditions, i.e., the sunny day and non-sunny days, are analyzed. And then, after the presentation for the basic description of correlation theory, the correction of PV generation in the viewpoint of season is presented. The resulted with the field data shows that, for the sunny days, the correlation between autumn and spring data is much stronger than the correlation between winter and summer data, and for the non-sunny days, the correlation may be small.

1. Introduction

In the face of fossil fuel depletion, environmental pollution, and climate change, renewable energies have become an important part of global sustainable energy strategies. Photovoltaic (PV) power generation draws from an abundant source of energy, i.e., the sun, and is also relatively cheap and efficient. With the maturation of PV technologies, PV power generation is now expanding at a breakneck pace.

In China, PV power generation has reached 106.9 billion kWh, with a year-on-year growth of 72%. The annual PV power generation of China has already exceeded 100 billion kWh as of January 2, 2018. However, the randomness and intermittency of PV power generation induces large-scale fluctuations in power. These fluctuations significantly alter the distribution of power in electric power systems and thus have a significant impact on load regulation and frequency modulation. These effects will threaten the safety of the electric power system and subsequently lead to PV power dumping and throttling.

The PV power generation characteristics of different PV power stations are similar as they are spatiotemporally related to each other. These correlations could help electric power systems take pre-emptive measures to minimize PV power dumping and throttling. There are numerous studies about the correlations of PV power generation.

Studies on the correlations of PV power generation generally focus on correlations with sunshine intensity or sunshine duration. Although significant progress has been achieved by these studies, the requirements of electric power systems have yet to be satisfied. At present, studies about the correlations of PV power generation have yet to account for the seasonal characteristics of PV power generation, which have an immense impact on the characteristics of PV power generation. Furthermore, reports about the seasonal yield of PV power generation are almost non-existent. It is



therefore necessary to perform an in-depth study about the characteristics of large-scale PV power stations.

Based on our recognition of the above problems, the correlations of PV power generation were investigated from a temporal perspective in this work, using the on-site PV power generation data of the Hanergy Xinhe PV power station.

The rest of this paper is organized as follows. The characteristic curves of the PV power generation data are provided in the section 2. And then, the briefly introduction of concept of correlation and application in the PV generation is showed in section 3. Finally, the conclusion and discussions are given in section 4.

2. Analysis of the Typical Daily Characteristics

The Hanergy Xinhe PV Power Station lies in Xinhe County (Aksu Prefecture, Xinjiang), in the southwestern region of the Xinjiang Uyghur Autonomous Region. The power generation data indicate that the effective power output is sometimes a negative value. In the data for January and June (and the data of other months), it was observed that the effective power output is a constant value during unlit hours. The negative values were therefore balanced by simply adding an offset.

2.1. Sunny Days

To ensure that representative sunny days were selected for this analysis, we selected a period that contains several sunny days in succession, and thereafter selected a few days from this period. This ensures that the selected data are representative of sunny days in this region. Using this approach, we performed our preliminary analysis using the data of 7/13, 7/14, 7/15, 7/16, 7/21, 7/22, and 7/23 in July.

The weather, sunrise/sunset times, and PV output curves of these days are shown below.

Table 1. Weather of the Xinhe County in July (partial).

Date	Highest temperature	Lowest t temperature	Weather	Sunrise	Solar noon	Sunset	PV output curve
2015-07-13	34	17	Sunny	07:06:51	14:35:17	22:03:43	C1
2015-07-14	37	18	Sunny	07:07:38	14:35:24	22:03:11	C2
2015-07-15	36	18	Sunny	07:08:25	14:35:31	22:02:36	B1
2015-07-16	37	17	Sunny	07:09:14	14:35:36	22:01:59	A
2015-07-21	37	19	Sunny	07:13:30	14:35:58	21:58:26	B2
2015-07-22	38	21	Sunny	07:14:24	14:36:01	21:57:38	C3
2015-07-23	38	21	Sunny	07:15:18	14:36:03	21:56:48	B3

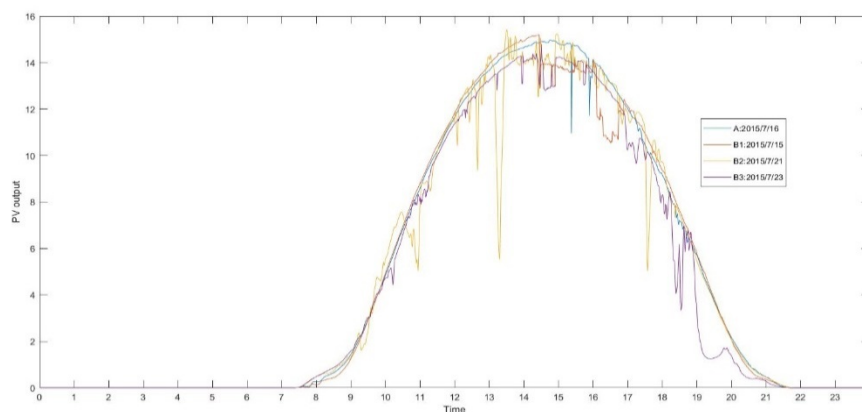


Figure 1. PV output curves of Group A and Group B.

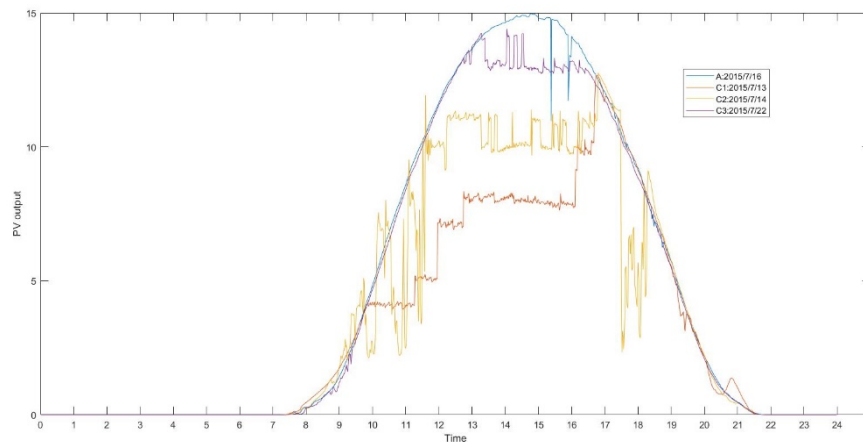


Figure 2. PV output curves of Group A and Group C.

The following conclusions were drawn from a preliminary analysis of figures 1 and 2 and tables 1.

(1) The PV output curves are generally parabolic in shape, and the peak occurs around 14:00–15:00. This is caused by the effects of sunshine on PV power generation, and the peak is related to the solar noon (the point when the solar elevation angle is at its highest). This result is consistent with our expectations.

(2) In the C1 (2015/7/13) data, it may be observed that the second half of the curve still shows signs of a parabolic shape. However, there are several periods of time in the first half of this data, where the numerical value stays at a constant value for a while before jumping to another value. Data of this type do not only appear in a single day. Owing to the abnormal stability of these numerical values, we suspect that this is the result of PV power dumping and throttling. Results of this type could produce a significant uncontrollable effect in the data of this work.

2.2. Non-sunny Days

As non-sunny days are atypical around the area of study, they were only briefly analysed in this work.

The data corresponding to 2015/7/28 (cloudy) and 2015/1/28 (light snow followed by cloudy weather) were used to plot the following figure.

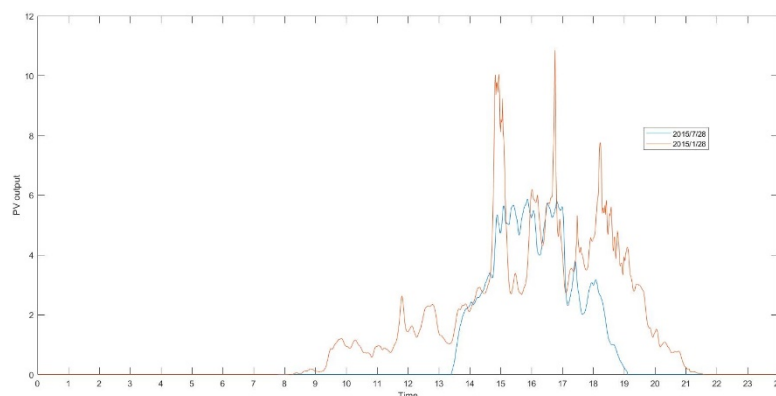


Figure 3. PV output curve on 2015/7/28.

The figure indicates that the PV output curves on non-sunny days do not exhibit any significant features. Therefore, the correlation analyses in the later sections of this paper simply focused on sunny days.

Thus, we analysed single-day data from summer and winter and obtained a simple understanding of PV power generation in these seasons.

3. Analysis of Seasonal Correlations

3.1. Correlation analysis on Summer and Winter Data

Please note that this analysis was performed using a sampling-based approach as there was an excessive amount of data. The data corresponding to the following dates (which had good weather) were included in this case analysis: 1/4, 1/11, 1/18, 1/25, 2/8, 2/15, 2/22, 6/7, 6/21, 6/28, 7/12, 7/19, 7/26, 8/9, 8/16, 8/23, 8/30, 12/13, 12/20, and 12/27. Here, December, January, and February are defined as winter months, whereas June, July, and August are defined as summer months.

The mathematical expression of the correlation coefficient is

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2} \sqrt{\sum(y_i - \bar{y})^2}} \quad (1)$$

The value of the correlation coefficient, r , varies between -1 and 1. If $r > 0$, the two variables are positively correlated; if $r < 0$, the two variables are negatively correlated. The value of $|r|$ reflects the closeness of the linear relationship between these variables. The linear correlation between the two variables becomes closer as $|r|$ approaches 1. Conversely; the linear correlation between the two variables becomes weaker as $|r|$ approaches 0.

Correlation analysis was performed after the selected data were preprocessed. Based on statistical tools, the correlation coefficient of these data is

$$r_1 = 0.5440.$$

This indicates that the selected data are correlated to an extent, but the correlation is not very strong. However, Section 4 indicates that there is significant similarity in the PV output trends of each day, despite the presence of large differences between summer and winter in terms of effective sunshine duration and sunshine intensity. Therefore, we believe that higher correlation coefficients can be obtained by processing the data.

The selected data were sorted into a continuous arrangement, as shown in figure 4..

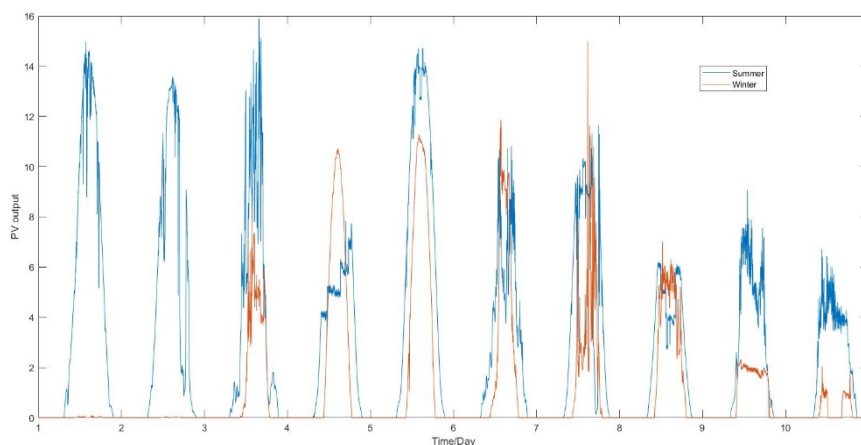


Figure 4. Continuous plot of (partial) summer and winter PV power generation data.

In this figure, it is shown that the days selected in December have very low PV output levels and are excessively different from the data for June, which exhibit very high PV output levels. Figures 5 and 6 are plotted to highlight the trends and numerical values of the selected December data.

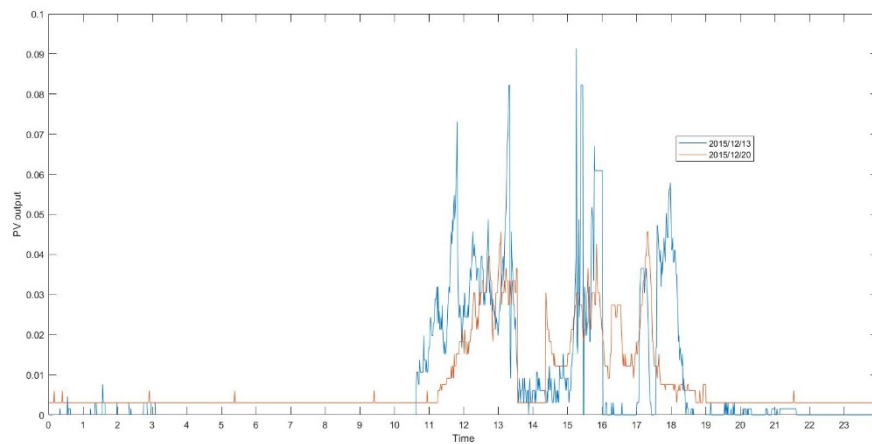


Figure 5. PV output curves on 12/13 and 12/20.

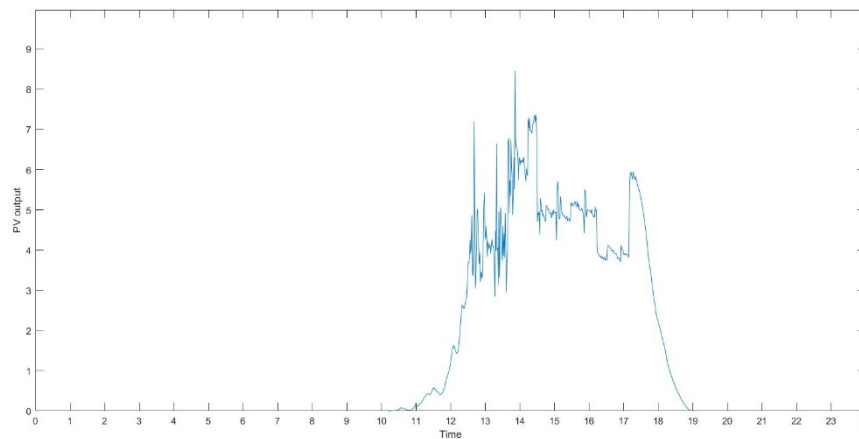


Figure 6. PV output curve on 12/27.

Please note that the vertical axes of figure 5 and 6 are very different from each other, and this is likely to have a detrimental and uncontrollable impact on our analysis. As this issue was previously discussed in the first subsection of Section 4, no further elaboration will be provided here.

The data for 12/13 and 12/20 were treated as bad data and thus excluded. The correlation coefficient thus becomes

$$r_2 = 0.7361.$$

The daily PV output trends obtained from the processed data are shown below:

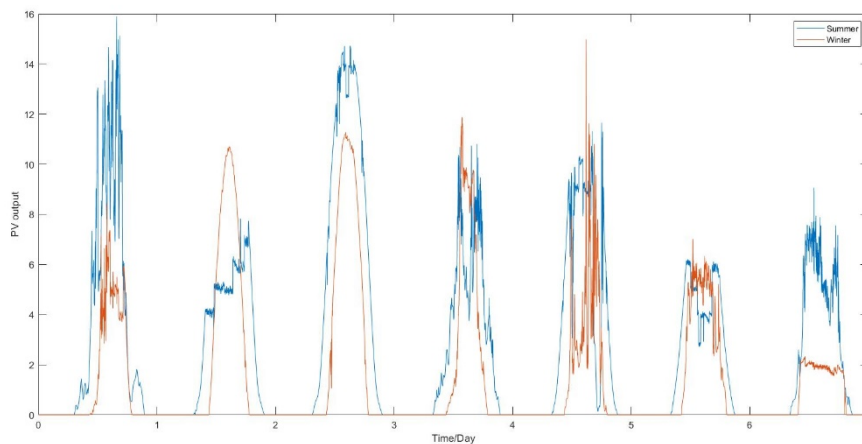


Figure 7. Continuous plot of processed summer and winter power generation data.

It may be observed that the data are now strongly correlated, and the shape of the plot in figure 7 confirms our hypothesis.

3.2. Analysis of Correlations between Spring and Autumn Data

As in the previous section, this analysis was performed using a sampling-based approach. The data corresponding to the following dates (which had good weather) were selected for this analysis: 3/1, 3/8, 3/15, 3/22, 3/29, 4/5, 4/12, 10/25, 11/1, 11/8, 11/15, 11/22, and 11/29. Here, March, April, and May are defined as spring months, whereas September, October, and November are defined as autumn months.

The following correlation coefficient was obtained after the data were processed (using the same procedures as in the previous section):

$$r_3 = 0.8674.$$

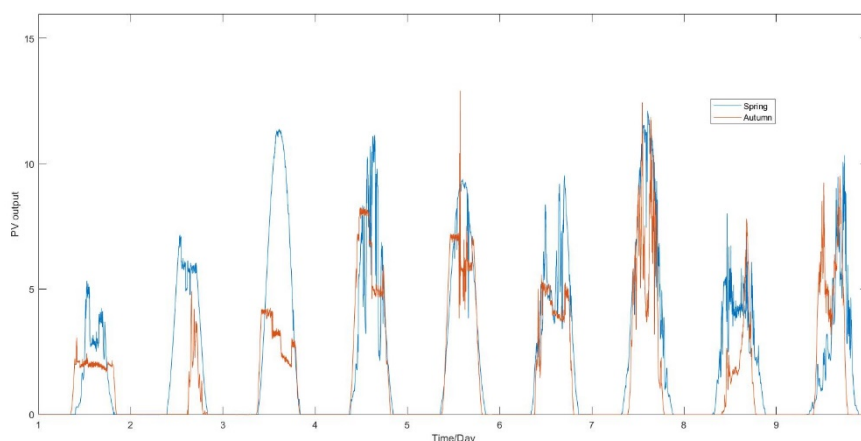


Figure 8. Continuous plot of processed spring and autumn power generation data.

The red line corresponds to the data for September, October, and November (autumn), whereas the blue line corresponds to the data for March, April, and May (spring).

4. Conclusion and Discussions

Summer and winter are extremely different in Xinjiang. Nonetheless, the PV outputs of the power station in summer and winter are still strongly correlated under normal circumstances, and this will be

very useful for PV output predictions. Although a considerable number of data points do not show any correlation, this appears to have been caused by human factors (PV power dumping and throttling). In correlation analyses and PV output predictions, it is necessary to account for this fact and take the corresponding data-processing measures

The correlation between autumn and spring data is much stronger than the correlation between winter and summer data. The possible causes for this difference are given below:

The correlation between spring and autumn weather is much stronger than the corresponding correlation between summer and winter. In this case study, the weather was sunny over long periods of time in summer, which is conducive for PV power generation. In winter, the weather was not sunny for several days in a row, and abnormal weather also occurred occasionally. These differences led to a lower level of correlation between the summer and winter PV power generation data. As the weather in spring and autumn is relatively similar, a much stronger level of correlation was observed between their PV power generation data.

Human-controlled PV power dumping and throttling did not occur very frequently in spring and autumn, which is beneficial for the data analyses of this work.

In addition to sunshine duration and sunshine intensity (which have been discussed previously in this work), we suspect that data deviations may be caused by other factors, such as temperature. Related work will be carried out in future research.

5. Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities (2018ZD01).

6. References

- [1] Baike entry: Photovoltaic power generation
<https://baike.baidu.com/item/%E5%85%89%E4%BC%8F%E5%8F%91%E7%94%B5/269917?fr=aladdin>
- [2] People's Daily Online: *PV power generation in China exceeds 100 billion kWh for the first time*,
<http://finance.people.com.cn/n1/2018/0102/c1004-29741194.html>
- [3] Zhu HL and Liu ZH *PV System Output Analysis of Environmental Factors Affect*
- [4] Zhang YP *An overview of research progress of short-term photovoltaic forecasts*
- [5] Wang D and Sun YJ *Progress in prediction of PV power generation systems*
- [6] Ni Q *Study on Some Photovoltaic Ultra Short-term Output Power Probability Forecasting*
- [7] Yang J *Numerical Study on Weather Types in the Prediction of Power Generation by Photovoltaic Power Systems*
- [8] Cui Y, Chen ZH and Sun PJ *Comparative Analysis of Short-Term Photovoltaic Power Generation Predictions in Regions with Different Latitudes in the Presence of Photovoltaic Power Dumping and Throttling*
- [9] Dai X, Xiao WB, Hu FY and Wu SY *Research of Photovoltaic Prediction Model*
- [10] Hu Y *the Output Forecast of Photovoltaic Power Based on Kernel Partial Least Squares*