International Journal of Earth, Energy and Environmental Sciences
ISSN: 2517-942X
Vol:13, No:3, 2019

# Comparison of Power Generation Status of Photovoltaic Systems under Different Weather Conditions

Zhaojun Wang, Zongdi Sun, Qinqin Cui, Xingwan Ren

***Abstract*—**Based on multivariate statistical analysis theory, this paper uses the principal component analysis method, Mahalanobis distance analysis method and fitting method to establish the photovoltaic health model to evaluate the health of photovoltaic panels. First of all, according to weather conditions, the photovoltaic panel variable data are classified into five categories: sunny, cloudy, rainy, foggy, overcast. The health of photovoltaic panels in these five types of weather is studied. Secondly, a scatterplot of the relationship between the amount of electricity produced by each kind of weather and other variables was plotted. It was found that the amount of electricity generated by photovoltaic panels has a significant nonlinear relationship with time. The fitting method was used to fit the relationship between the amount of weather generated and the time, and the nonlinear equation was obtained. Then, using the principal component analysis method to analyze the independent variables under five kinds of weather conditions, according to the Kaiser-Meyer-Olkin test, it was found that three types of weather such as overcast, foggy, and sunny meet the conditions for factor analysis, while cloudy and rainy weather do not satisfy the conditions for factor analysis. Therefore, through the principal component analysis method, the main components of overcast weather are temperature, AQI, and pm2.5. The main component of foggy weather is temperature, and the main components of sunny weather are temperature, AQI, and pm2.5. Cloudy and rainy weather require analysis of all of their variables, namely temperature, AQI, pm2.5, solar radiation intensity and time. Finally, taking the variable values in sunny weather as observed values, taking the main components of cloudy, foggy, overcast and rainy weather as sample data, the Mahalanobis distances between observed value and these sample values are obtained. A comparative analysis was carried out to compare the degree of deviation of the Mahalanobis distance to determine the health of the photovoltaic panels under different weather conditions. It was found that the weather conditions in which the Mahalanobis distance fluctuations ranged from small to large were: foggy, cloudy, overcast and rainy.

***Keywords*—**Fitting, principal component analysis, Mahalanobis distance, SPSS, MATLAB.

## I. Introduction

THE reasonable selection of the characteristic parameters of the health state of the photovoltaic system is the prerequisite for the assessment of the health state of the photovoltaic system, and the selected characteristic parameters will have a direct or indirect effect on the performance state of the photovoltaic system.

Zhaojun Wang, Zongdi Sun, Qinqin Cui, and Xingwan Ren are with the Economy and Management College, Qingdao University of Science and Technology, Qingdao 266061 (e-mail: wzjun599@sina.com, sunzongdi1989@126.com, 2514037689@qq.com, 1124947154@qq.com).

In order to accurately reflect the health state of the photovoltaic system, this paper selects the solar radiation, ambient temperature, weather conditions, solar radiation intensity, air quality index and time as the characteristic parameters. According to different weather conditions, multivariate statistical analysis is used to classify these characteristics. The principal component analysis is used to preprocess the original multidimensional input variables, and the health degree of the system is characterized by mahalanobis distance for different weather conditions.

## II. Literature Review

Haiying and Xiaomin considered the randomness and discontinuity of photovoltaic system power generation. The Monte Carlo method was used to establish a solar irradiance model and an energy conversion model to evaluate the operational risks of photovoltaic grid-connected systems, and examples were verified [1].

Ning et al. analyzed the mechanism of harmonic generation, deduced the Thevenin equivalent circuit of the 1MVA power generation unit, and established an impedance network model that approximated the actual grid connection. The relationship between the generation of harmonics and irradiance and the temperature of the photovoltaic panels was studied. The effectiveness of the model was verified by the measured data of Qinghai 50 MVA grid-connected photovoltaic power stations [2].

Qiaona and Zhong proposed to use back propagation (BP) neural network model to improve the forecasting accuracy of photovoltaic power generation [3].

Dandan et al. comprehensively analyzed the factors that affect power generation, such as weather type, temperature, irradiance, etc. For the intermittent and easily impacted power grids in photovoltaic power plants, a SOFM-LM-BP neural network model was established to predict power generation, and the difference between the predicted value and the actual value was compared. It was found that the model had a better prediction effect [4].

Lin and Yingzi used the modified GM(1,1) residual correction model to establish a photovoltaic power generation forecasting model and predicted the power generation of a 5.6 kW photovoltaic system [5].

Jie and Yanxia proposed a feedback-based neural network short-term power generation forecasting model based on chaotic adaptive particle swarm optimization for the characteristics of photovoltaic power generation systems and related influencing factors [6].

Siqing et al. comprehensively consider factors such as photovoltaic panel installation factors, meteorological conditions, total solar radiation and other factors, establish a model of surface radiation acceptance of photovoltaic panels, and conduct statistical analysis of the meteorological conditions in North China [7].

## III. Data Preprocessing

Based on the existing data, the missing data are processed as follows: for the temperature, air quality, PM2.5, solar radiation intensity and power generation data, if there exists missing data, this article will take the average value of the two periods before and after the vacancy data to complement; for different weather data, the missing data can use the previous data to replace.

TABLE I
Symbol Description

| Symbols | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ | $Y$ |
|---|---|---|---|---|---|---|
| Meaning | temperature | AQI | pm2.5 | Solar radiation intensity | time | Photovoltaic Power generation |

## IV. Analysis of Factors Affecting Power Generation

Considering that the power generation may be affected by other variables, in order to reduce the complexity of the research, and to find out which factors in different weather conditions have the greatest impact on the health of the photovoltaic panels, in this paper, MATLAB software is used to plot the scatter diagram of the relationship between photovoltaic power generation and other variables under different weather conditions, take sunny weather condition as an example, the relationships are as follows:

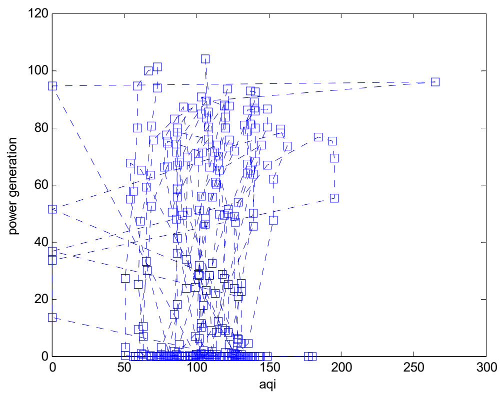


Fig. 1 Photovoltaic power generation and AQI

From Figs. 1-4, the relation between photovoltaic power generation and other variables is fuzzy.

From Fig. 5 it can be found that there is a more obvious nonlinear relationship between photovoltaic power generation and time. The relationship can be obtained by fitting method, as shown in Fig. 6.

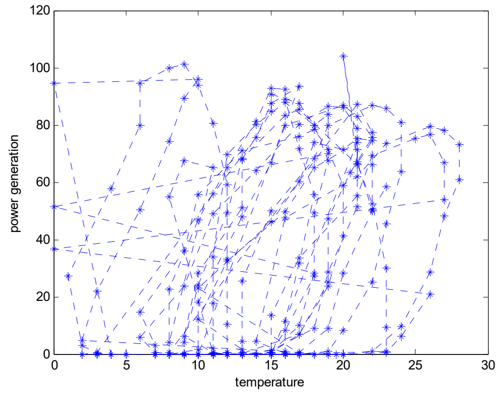


Fig. 2 Photovoltaic power generation and temperature

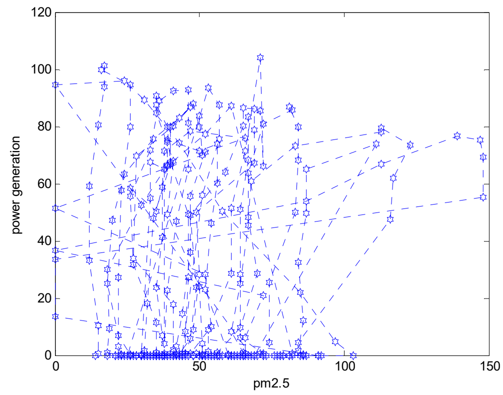


Fig. 3 Photovoltaic power generation and pm2.5

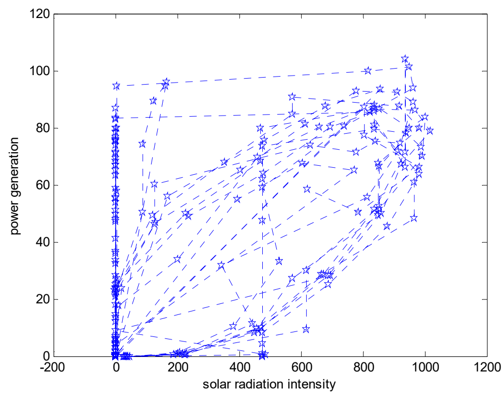


Fig. 4 Power generation and solar radiation intensity

It is ideal to find the effect of the fitting diagram in Fig. 6, the fitting value can cover as many observations as possible, and the relation between power generation and time under sunny weather condition is obtained.

$$Y = -0.0002X_5^6 + 0.0124X_5^5 - 0.3137X_5^4 + 3.4126X_5^3 - 14.6307X_5^2 + 20.187X_5 - 2.7288 \quad (1)$$

Similarly, the relationships of photovoltaic power generation with other variables under other conditions can be obtained, the results show that photovoltaic power generation has close relationship with time under different conditions, the fitting graphics and equations are as Figs. 7-11.

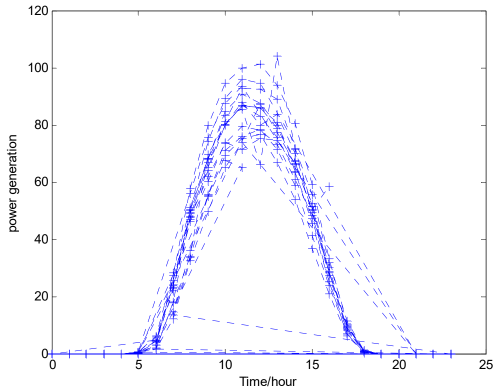


Fig. 5 Photovoltaic power generation and time

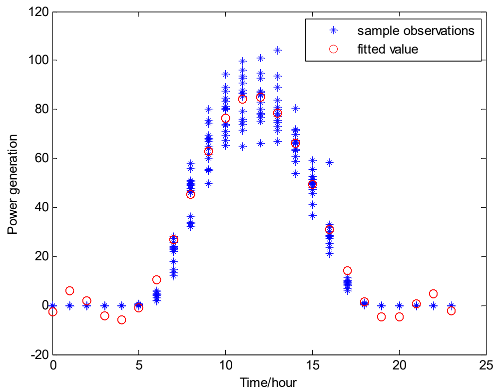


Fig. 6 Fitting effect of power generation and time

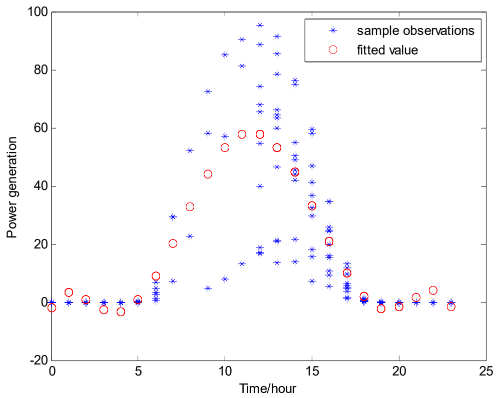


Fig. 7 Relation between power generation and time in overcast condition $Y=-0.0001X_5^6+0.0082X_5^5-0.2052X_5^4+2.202X_5^3-9.21X_5^2+12.430X_5-1.861$

From Fig. 11, it is found that the fitted values can not well reflect the relationship between the observed values and the time. In general, the relationship between power generation and time is not clear. In the same way, we get the relationship between the power generation and the time.

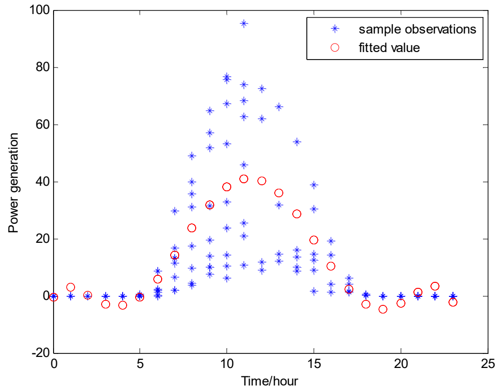


Fig. 8 Relation between power generation and time in cloudy weather
$Y=-0.0001X_5^6+0.0066X_5^5-0.1625X_5^4+1.714X_5^3-6.999X_5^2+8.621X_5-0.366$

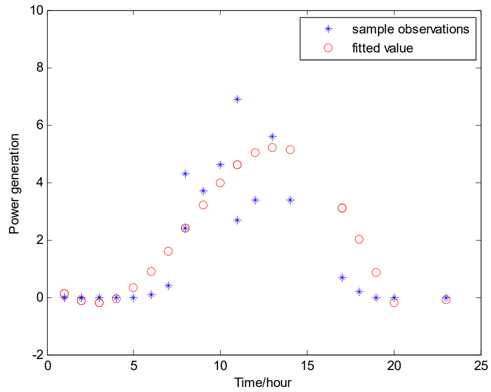


Fig. 9 Relation between power generation and time under foggy condition $Y=0.0027X_5^5-0.0791X_5^4+1.0717X_5^3-6.7231X_5^2+18.7095X_5-17.9459$

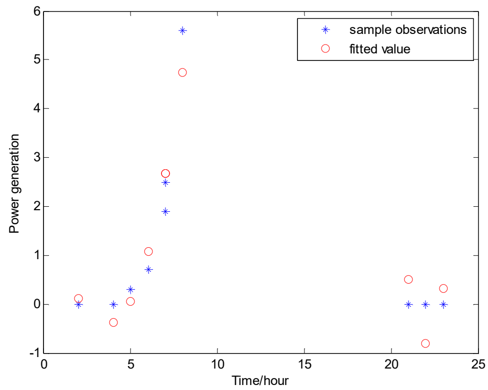


Fig. 10 Relation between power generation and time in rainy days
$Y=-0.0001X_5^5+0.0024X_5^4-0.0323X_5^3+0.3285X_5^2-1.2033X_5+1.0929$

$$Y=-0.0015X_5^4+0.1086X_5^3-3.7242X_5^2+56.316X_5+27.3791 \quad (2)$$

It is not difficult to find that the influence level of each factor on photovoltaic power generation is different, but there may be an overlapping correlation between the variables. In order to further reduce the complexity of the variables and find the factors which have great influences on the health of the photovoltaic board, this article carried out principal component analysis.

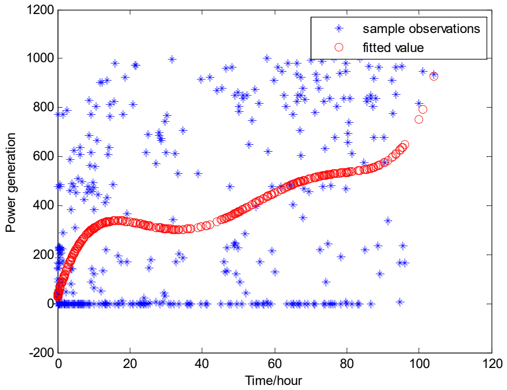


Fig. 11 The relationship between photovoltaic power generation and time for total sample data

## V. Principal Component Analysis Model

Factors affecting photovoltaic power generation include temperature, AQI, PM2.5, weather, and solar radiation intensity. In order to more clearly study the relationship between factors and the impact on power generation, this paper selects principal component analysis method to extract the main variables.

The principal component analysis calculates the data matrix formed by a plurality of sample input variables, obtains the correlation matrix of the data matrix and the contribution rate of the cumulative variance according to the eigenvalues of the correlation matrix. Finally, according to the feature vector of the correlation matrix, the main component is calculated. The specific steps are as follows:

### *A. Standardization of Original Data*

In order to eliminate the influence of the original variable dimension and the large numerical value difference, the raw data should be standardized. Assuming that the original variables have a total of M groups of samples, each group of samples contains a total of N feature parameters, the original variables can be expressed as:

$$X_{M\times N}=\begin{bmatrix} x_{11} & \cdots & x_{1N} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{MN} \end{bmatrix},$$

By converting the above matrix through the center normalization to the matrix $X^*_{M\times N}$, the data unit of the matrix $X^*_{M\times N}$ can be expressed as:

$$x^*_{ij}=\frac{x_{ij}-E(x_j)}{\sqrt{D(x_j)}}, \tag{3}$$

where $i=1,2,\cdots,M$; $j=1,2,\cdots,N$; $x_{ij}$ is the data unit of the matrix $X_{M\times N}$; $x_j$ is the $jth$ feature parameter; $E(x_j)$ and $D(x_j)$ are the expected and variance values of the variable $x_j$, respectively.

### *B. Creating a Correlation Matrix for Variables*

The correlation matrix of variables $R$ is established as:

$$R=\frac{X^{*T}_{M\times N}}{M-1}, \tag{4}$$

and calculate out the eigenvalues $\lambda_i(i=1:n)$ and eigenvectors $t_i(i=1:n)$ of the matrix.

### *C. Determination of the Number of Principal Components*

Variance contribution rate:

$$\eta_i=\frac{\lambda_i}{\sum_{i=1}^{n}\lambda_i}\times 100\% \tag{5}$$

Accumulated variance contribution rate:

$$\eta=\sum_{i=1}^{P}\eta_i \tag{6}$$

The number of principal components selected depends on the contribution of the cumulative variance. Usually, when the cumulative variance contribution rate is greater than 85%, the corresponding first $P$ principal components contain most of the information that the original variable can provide.

### *D. The Eigenvector Matrix*

The eigenvector matrix corresponding to the $P$ principal components is:

$$U_{N\times P}=(u_1,u_2,\cdots,u_p), \tag{7}$$

Then the matrix of $P$ principal components of $M$ samples of the photovoltaic system is:

$$U_{M\times N}=X^*{}_{M\times N}U_{N\times p} \tag{8}$$

## VI. Solution to Principal Component Analysis Model

Using Spss software for principal component analysis, the KMO test and total variance for five weather conditions can be obtained, as shown in Tables II-IX.

### *A. Cloudy Weather*

TABLE II
Examination of KMO and Bartlett in Cloudy Weather

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin metric for sampling enough | | 0.474 |
| Bartlett's sphericity test | Approximate Chi-square | 287.791 |
| | df | 10 |
| | Sig. | .000 |

For cloudy weather, it can be seen from Table II that the KMO test value is less than 0.5 and does not satisfy the conditions for the principal component analysis. Therefore, the principal component analysis method cannot be used to extract the main factor. Therefore, in the computational model for the photovoltaic panel health under cloudy weather, the variables have to be taken, such as $X_1$, $X_2$, $X_3$, $X_4$, $X_5$.

### B. Overcast Weather

From Table III, it can be found that the KMO test value is greater than 0.6, so the principal component analysis test can be performed under cloudy weather. The total variance explained is shown in Table IV.

TABLE III
EXAMINATION OF KMO AND BARTLETT IN OVERCAST WEATHER

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin metric for sampling enough | | 0.644 |
| Bartlett's sphericity test | Approximate Chi-square | 224.797 |
| | df | 10 |
| | Sig. | .000 |

TABLE IV
TOTAL VARIANCE EXPLAINED IN OVERCAST WEATHER

| Ingredients | Initial feature value | | | Extract and load | | | Rotation squared and loaded | |
|---|---|---|---|---|---|---|---|---|
| | Total | Variance% | Accumulation % | Total | Variance% | Accumulation % | Total | Variance% |
| 1 | 2.397 | 47.939 | 47.939 | 2.397 | 47.939 | 47.939 | 2.388 | 47.763 |
| 2 | 1.050 | 20.995 | 68.934 | 1.050 | 20.995 | 68.934 | 1.059 | 21.171 |
| 3 | 0.925 | 18.495 | 87.429 | | | | | |
| 4 | 0.433 | 8.650 | 96.079 | | | | | |
| 5 | 0.196 | 3.921 | 100.000 | | | | | |

As can be seen from Table IV, when the first three components are selected, the cumulative variance contribution rate reaches 87.429%, and the first three components have eigenvalues greater than 1 or close to 1, so the first 3 components are selected as the main components, i.e., temperature, AQI, pm2.5.

### C. Foggy Weather

Using the principal component analysis method, the total variance of the KMO test and the interpretation under foggy weather conditions can be obtained, as shown in Tables V and VI.

TABLE V
EXAMINATION OF KMO AND BARTLETT IN FOGGY WEATHER

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin metric for sampling enough | | 0.810 |
| Bartlett's sphericity test | Approximate Chi-square | 65.973 |
| | df | 6 |
| | Sig. | .000 |

TABLE VI
TOTAL VARIANCE EXPLAINED IN FOGGY WEATHER

| Ingredients | Initial feature value | | | Extract and load | | |
|---|---|---|---|---|---|---|
| | Total | Variance % | Accumulation % | Total | Variance % | Accumulation % |
| 1 | 3.887 | 97.181 | 97.181 | 3.887 | 97.181 | 97.181 |
| 2 | 0.085 | 2.122 | 99.303 | | | |
| 3 | 0.014 | 0.352 | 99.655 | | | |
| 4 | 0.014 | 0.345 | 100.000 | | | |

As can be seen from Table V, the KMO value is greater than 0.8 and is suitable for principal component analysis.

### D. Rain Weather

Using the principal component analysis method, the KMO test value under rainy weather conditions can be obtained, as shown in Table VII.

From Table VII, it can be seen that the KMO value is less than 0.5, which does not satisfy the principal component analysis conditions. Therefore, in the calculation model for the photovoltaic panel health degree on a rainy day, variables $X_1$, $X_2$, $X_3$, $X_4$, $X_5$ have to be taken into account.

### E. Sunny Weather

Under the conditions of sunny weather, the principal component analysis of the variables can obtain the KMO test value and the total variance value, as shown in Tables VIII and IX.

From Table VIII, it can be seen that the KMO value is greater than 0.55, so the principal component analysis can be done under sunny conditions.

From Table IX, we can find that the extraction of the first three components can significantly increase the cumulative variance contribution rate, and the first three components have eigenvalues greater than 1 or close to 1, so the first 3 components are selected as the main components, ie, temperature, AQI, pm2.5.

TABLE VII
EXAMINATION OF KMO AND BARTLETT IN RAIN WEATHER

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin metric for sampling enough | | 0.365 |
| Bartlett's sphericity test | Approximate Chi-square | 35.390 |
| | df | 10 |
| | Sig. | .000 |

After the principal component analysis was used to extract the main components, the power generation conditions under sunny conditions were used to as health standards to establish a Mahalanobis distance model to analyze the health of photovoltaic power generation under different weather conditions.

TABLE VIII
EXAMINATION OF KMO AND BARTLETT IN SUNNY WEATHER

| | | |
|---|---|---|
| Kaiser-Meyer-Olkin metric for sampling enough | | 0.551 |
| Bartlett's sphericity test | Approximate Chi-square | 496.804 |
| | Df | 10 |
| | Sig. | .000 |

TABLE IX
TOTAL VARIANCE EXPLAINED IN SUNNY WEATHER

| Ingredients | Initial feature value | | | Extract and load | | |
|---|---|---|---|---|---|---|
| | Total | Variance % | Accumulation % | Total | Variance % | Accumulation % |
| 1 | 2.253 | 45.065 | 45.065 | 2.253 | 45.065 | 45.065 |
| 2 | 1.296 | 25.928 | 70.993 | 1.296 | 25.928 | 70.993 |
| 3 | 0.827 | 16.547 | 87.540 | | | |
| 4 | 0.398 | 7.965 | 95.505 | | | |
| 5 | 0.225 | 4.495 | 100.000 | | | |

## VII. MAHALANOBIS DISTANCE MODEL

In order to reasonably assess the hourly health of photovoltaic panels, the degree of health needs to be measured by data under normal conditions and abnormal conditions, and the degree of health is determined by differences in data points.

Mahalanobis distance is a method for effectively calculating the similarity of two unknown sample sets. It can take into account the relationship between various characteristics and the scale is irrelevant, indicating the covariance distance of the data. The specific operation steps of the model are as follows:

### *A. Determine Observation Data and Sample Data Points*

Take the data under the sunny conditions as the observed data $Y$, and use the data under cloudy, foggy, overcast, and rainy conditions as the sample data point $X$. Both the sample observations and the sample data points have the same dimensions, but the number of samples does not have to be the same.

According to the principal component analysis method, the main components under the cloudy weather are $X_1$、$X_2$、$X_3$, and the main component under foggy weather is $X_1$. Since the cloudy and rainy conditions cannot satisfy the factor analysis conditions, the variables $X_1$、$X_2$、$X_3$、$X_4$、$X_5$ are sample variables under cloudy and rainy conditions. Considering that the dimensions between the observation data and the sample data points must be the same, the variable data under the sunny conditions in the same dimension as the sample data points are selected as observation data.

(1) Determine the Mean and Covariance Matrix of the Sample

The mean of the sample is:

$$\mu = (\mu_1, \mu_2, \mu_3, \cdots, \mu_p)^T,$$

that is:

$$\mu_j = \frac{1}{M}\sum_{i=1}^{M} P_{ij} \quad (j = 1, 2, \cdots, P), \tag{9}$$

where $P_{ij}$ is the data unit of the matrix.

The sample's covariance matrix is a square matrix with the same dimension as the sample dimensions. For two variables, the covariance of $x$ and $y$ is:

$$Cov(x, y) = E(x - E(x))(y - E(y)) \tag{10}$$

For multiple column vectors, the covariance matrix is:

$$\sum_{ij} = \mathrm{cov}(Dim_i, Dim_j) \tag{11}$$

(2) Calculate Mahalanobis Distance

$$MD = \sqrt{(x-\mu)^T \Sigma^{-1} (x-\mu)} \tag{12}$$

For the multivariate vector $x = (x_1, x_2, x_3, \cdots, x_p)$, the covariance $\sum$ and the mean $\mu$ and other related data are substituted into the formula to obtain the Mahalanobis distance between the sample observations and the sample data points.

## VIII. SOLUTION OF MAHALANOBIS DISTANCE MODEL

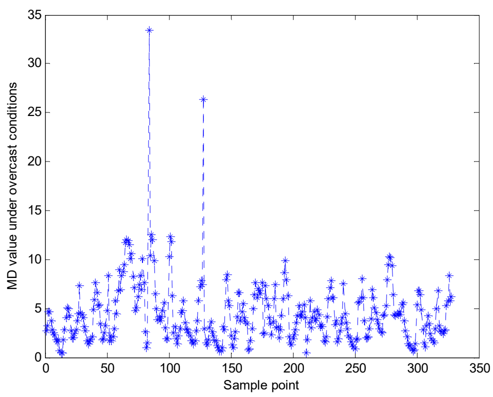


Fig. 12 Mahalanobis distance in overcast conditions

Substituting the relevant data into the model and using MATLAB software to solve it, we finally obtain the

Mahalanobis distance under conditions of overcast, cloudy, foggy, and rainy, as shown in Figs. 12-16.

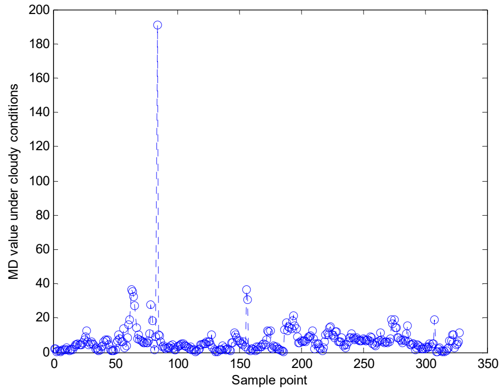


Fig. 13 Mahalanobis distance in cloudy conditions

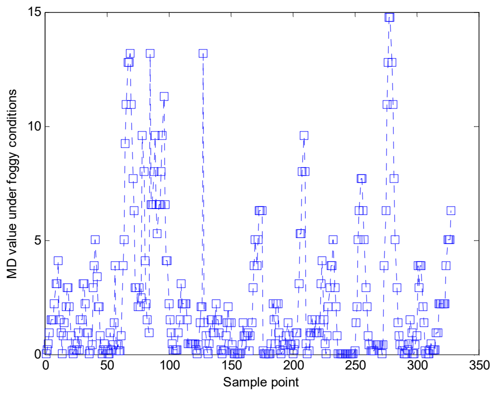


Fig. 14 Mahalanobis distance under foggy conditions

In general, the Mahalanobis distance fluctuates the most in rainy conditions, with the largest degree of deviation. Under cloudy conditions, the Mahalanobis distance between 50 and 100 and fluctuates fiercely, and fluctuations in other areas are relatively stable. Under overcast conditions, the Mahalanobis distance between 80 and 140 fluctuates relatively, and it is relatively stable in other areas. Under the foggy conditions, the Mahalanobis distance fluctuation is relatively small and stable.

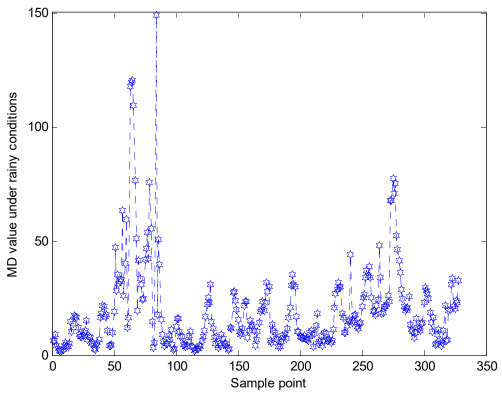


Fig. 15 Mahalanobis distance in rainy conditions

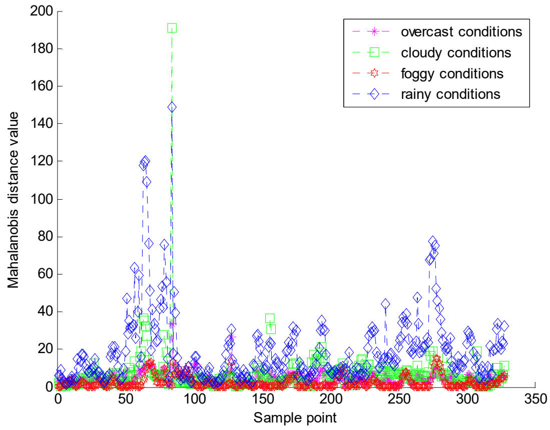


Fig. 16 Comparison of Mahalanobis distance under different weather conditions

From the results, it can be found that when the sample points are between 50-100 and 250-300, the Mahalanobis distance under rainy conditions and overcast conditions will fluctuate to a greater degree.

In summary, according to the established Mahalanobis distance model, it is known that the Mahalanobis distance fluctuation value is the smallest in the foggy weather, which indicates that the photovoltaic panel is in the most healthy condition in the foggy weather.

## IX. Conclusions

In this paper, we mainly used the principal component analysis method to extract principal components of variables affecting power generation in various weather conditions. The Mahalanobis distance model was established to evaluate the health of photovoltaic power generation under different kinds of weather. We can find the health status of photovoltaic power generation different under different kinds of weather. Therefore, for different cities, the photovoltaic power generation should consider many influencing factors. City managers should take different weather into consideration. Under different weather conditions, they should take different methods to obtain more photovoltaic power generation to promote sustainable urban development.

## References

[1] Wang Haiying, Bai Xiaomin. Short-term operational standby evaluation of grid-connected photovoltaics (J). Automation of Electric Power Systems, 2013, 37(05):55-60.
[2] Xie Ning, Luo An, Chen Yandong, Ma Fujun, Xu Xinwei, Lv Zhipeng, Shuai Zhikang. Dynamic modeling and harmonic characteristics analysis of large-scale photvolataic power plants (J). Proceedings of the CSEE, 2013, 33(36):10-17+4.
[3] Jiang Qiaona, Chen Zhong. Application of BP-Markov Combination Forecasting Method in PV Generation prediction (J). Electricity Demand Side Management, 2011, 13(06):21-24.
[4] Song Dandan, Ren Guochen, Zhang Mengsong, Su Renqi. Application of SOFM-LM-BP neural network in PV power generation forecasting (J). Journal of Liaoning University of Technology (Natural Science Edition), 2016, 36(03):163-167.
[5] He Lin, Li Yingzi. Application of Improved GM(1,1) Residual Correction Model in Short-term Forecast of Photovoltaic Generation (J). Journal of Beijing Institute of Civil Engineering and Architecture, 2008,

24(04):61-65.
[6] Zhao Jie, Zhang Yanxia. Prediction of short-term power generation of photovoltaic system based on CAPSO-RNN (J). Electric Power, 2012, 45(04):87-91.
[7] Sheng Siqing, Fan Lintao, Tan Xiaolin, Lu Li. Optimum installation angle calculation of photovoltaic panels considering weather conditions (J). Shaanxi Eletric Power, 2014, 42(01):42-45.