A solar energy resources estimation method based on EOF and ARMA model

Qiang Zhou[1,a], Libin Yang [2], Hai Zhou[1] and Jie Ding[1]

[1]Renewable Energy Department, China Electric Power Research Institute, Nanjing, China
[2]Electric Power Research Institute of State Grid Qinghai Electric Power Company, Xining, China

Abstract. A method to estimate solar energy resources was introduced in this paper. Based on the solar data observed by meteorological stations, the temporal-spatial characteristics of solar energy resources in the west of China were analysed by EOF. The main modal was determined and its time series was forecasted by ARMA model. Using the time series forecasted by ARMA model and the space factors of observed data computed by EOF, the prediction of solar resources field was estimated. Experiment results show that the predicted field tends to reflect the characteristics of the measured field and the area of solar energy resources can be estimated by this method properly.

Keywords: EOF; ARMA model; solar energy; resources calculation.

1 Introduction

At the end of 2015, China's solar photovoltaic power generation cumulative capacity reached 41.58GW, an increase of 67.3% than in 2014.There are 11 provinces and autonomous regions cumulative PV installed capacity reaching 1 GW. Centralized photovoltaic power generations are mainly distributed in the west of China, including Gansu, Xinjiang, Qinghai installed capacity more than 5 GW. Distributed photovoltaic power generations are mainly constructed in the middle and east of China, including Jiangsu, Zhejiang, Shandong, Anhui installed capacity more than 1 million kilowatts.

An important task for the development of solar energy resources is the estimation of radiation resources. This paper introduces a method based on solar observed data by using the EOF to analyze the temporal and the spatial distribution characteristics of solar resources and determine the principal modal. The future solar field can be predicted by the spatial coefficients of principal modal and corresponding time series which calculated by ARMA model.

2 Empirical orthogonal function

2.1 Method summary[1-4]

Empirical Orthogonal Function(EOF)is a method to decompose original variable field into linear

---

[a] Corresponding author : zhouqiang5@epri.sgcc.com.cn

combination of orthogonal functions constituting a few irrelevant typical modals instead of the original variable fields and each typical modal contains the original field information as much as possible.

Matrix $X_{(m \times n)}$ is composed of m interrelated variables and each variable has n samples. After linear transformation of the X, a new variable is linear combined by p variables:

$$Z_{p \times n} = A_{p \times m} X_{m \times n}$$

Z is the principal component of the original variable and the A is the linear transformation matrix. In this process, most of the information of the original variables is centralized to the principal component of the minority independent variables.

The matrix $X_{(m \times n)}$ which is formed by m spatial points and n times observed record is regarded as a linear combination of p spatial feature vectors and corresponding time coefficients:

$$X_{m \times n} = V_{m \times p} T_{p \times n}$$

T is the time coefficient and V is the spatial feature vector. In this process, the main information of the variable field is represented by several typical feature vectors.

2.2 Computational procedure

The observation data of a climate variable field is given in the form of matrix:

$$X = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... \\ x_{m1} & x_{m2} & ... & x_{mn} \end{pmatrix} \tag{1}$$

In this matrix, m is the number of spatial observation points and n is the number of observations and x_{ij} represents the first j observation value at the first i observation point..

Matrix X minus the mean matrix equals to the anomaly matrix $\overline{X}$:

$$\overline{X} = X - M = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1n} \\ x_{21} & x_{22} & ... & x_{2n} \\ ... & ... & ... & ... \\ x_{m1} & x_{m2} & ... & x_{mn} \end{pmatrix} - \begin{pmatrix} m_1 & m_1 & ... & m_1 \\ m_2 & m_2 & ... & m_2 \\ ... & ... & ... & ... \\ m_m & m_m & ... & m_m \end{pmatrix} \tag{2}$$

$$m_i = \frac{\sum_{j=1}^{n} x_{ij}}{n} .$$

Computing covariance matrix: $C = \frac{1}{n} \overline{X}\overline{X}^T$ (X^T is the rank of X).According to matrix theory, C is symmetric matrix.

Computing eigen values and eigenvectors (K) of C: $CK = V \Lambda$,

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & ... & 0 \\ 0 & \lambda_2 & ... & 0 \\ ... & ... & ... & ... \\ 0 & 0 & ... & \lambda_m \end{pmatrix} (\lambda_1 > \lambda_2 > ... > \lambda_m) \tag{3}$$

The number j column of the matrix K is the eigenvectors of the eigen value λ_j.

Computing principal component: The variance contribution and calmative variance contribution can be obtained according to eigen values. The eigenvector corresponding to the largest variance contribution of eigen value is regarded as the principal modal. The variance contribution and calmative variance could be calculated as following:

$$\frac{\lambda_i}{\sum\limits_{k=1}^{m}\lambda_k}, \quad \frac{\lambda_i}{\sum\limits_{k=1}^{m}\lambda_k} \quad (i=1,2,\cdots,m) \tag{4}$$

The time series can be obtained by spatial matrix V multiplies X after principal modal has been determined.

$$T = S^T \overline{X}. \tag{5}$$

The T is a matrix which has p rows and n columns and the first i row elements of the T is the time series of the first i mode.

3 Radiation resource estimation

3.1 Estimation principle

By the equation (5) we can infer $X = VT$. and by the equation (2) we can know $X = \overline{X} + M = VT + M$.

$$X_f = VT_f + M. \tag{6}$$

The X_f is the prediction matrix of solar resources and the T_f is the prediction matrix of time series.

According to (6), the future solar field could be calculated after the time series of principal modal has been forecasted [5-6].

3.2 Time series model [7-8]

The ARMA model namely Auto Regressive Moving Average model is a finite linear model with a finite parameter linear model to describe the autocorrelation structure of time series and it is easy to carry out statistical analysis and mathematical processing. Most of the random phenomenon could be described by finite parameter linear model and its accuracy can meet the practical engineering requirements. The linear prediction theory can be derived from the finite parametric linear model. This model is suitable for stationary time series.

$$X_t = \sum_{i=1}^{p}\varphi_i X_{t-i} - \sum_{j=0}^{q}\theta_j \varepsilon_{t-j} \tag{7}$$

ε_t: Zero means white noise

$\{X_t\}$: The autoregressive-moving average sequence whose orders are p and q is called ARMA

(*p*,*q*)

Equation (7) shows that this model is the memory of the system to the past and the noise entry into the system. The value of a time series at a certain time can be represented by a linear combination of p observed records and a *q* term moving average of a white noise sequence.

The modeling and forecasting of ARMA model can be roughly divided into: model identification, model order, parameter estimation, model checking, and model prediction.

4 Case application

EOF method was used to analyze the total radiation data of 23 meteorological stations in Northwest China from 1974 to 2003.
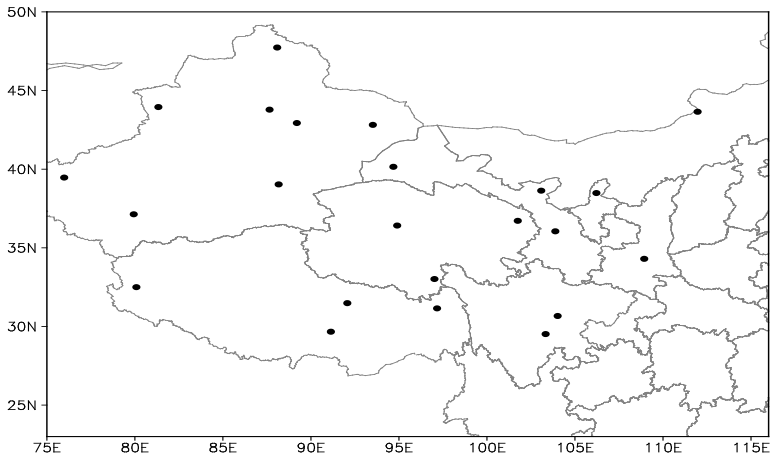

Figure 1. Meteorological observation stations

Through the data preprocessing, the covariance matrix was obtained and the first five modal variance contribution rate calculated as table 1:

Table 1. Top five modal variance contribution

Modal	Variance contribution	Cumulative variance contribution
1	0.5911	0.5911
2	0.098	0.6891
3	0.0758	0.7649
4	0.0521	0.817
5	0.0356	0.8526

The main modal was selected which contribution rate of variance was 59.11% and time series corresponding to the main modal was also obtained. Using the data of the first 27 years as a learning sample, the data of the last 3 years was used as a test sample, and the time series of the main mode was predicted (Figure 2).
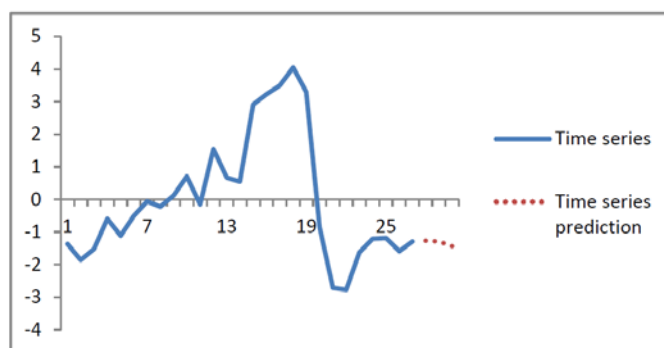
Figure 2. Main modal time series and predictions

The time coefficient of the main modal in the first 15years tends to rise with fluctuation, reaching a peak and then declined rapidly.
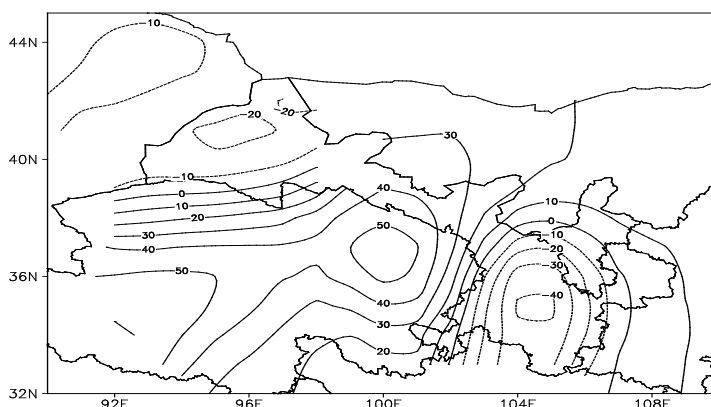

Figure 3. Average distribution of prediction error from 2001 to 2003 (MJ/m^2·a)

The average distribution of the radiation prediction field minus the average distribution of the measured field shows that the calculated results were smaller in the southeastern and northwestern parts of Gansu but larger in most areas of Qinghai (Figure 3). The predicted field basically reflects the characteristics of the measured field and has a strong representation.

5 Summary

After determining the principal modal by EOF, not only the temporal and spatial distribution characteristics of solar resources can be analyzed, but also the time series of the main modal was predicted which can forecast the future distribution of resources in the region. In this paper, the ARMA model was used in the prediction of time series, and the support vector machine, neural network and other methods could also be used in the application.

References

1. Wu Hongbao,Wu Lei. *Climate variability diagnosis and prediction method*[M].(2005)
2. Wei Fengying. *Modern climate statistical diagnosis and prediction technology*[M].(2007)
3. Huang Jiayou. *Meteorological statistical analysis and forecast method* [M]. (2004)
4. Shi Neng. *Multivariate analysis methods in meteorology research and forecast*[M].(2002)

5. GAO Jian-yun, DENG Zi-wang, ZHOU Xiao-lan, et al. Spatial/Temporal Features of Drought/Flood in Fujian for the Past 4 Decades[J]. Journal of Tropical Meteorology, **22**(4)**: 491-497**(2006)
6. He Lin. Study of the Spatio-temporal Variation Characteristics of the Solar Radiation in Growing Season over Northeast China [D]. Changchun: Northeast Normal University(2013)
7. Yi Danhui. Time series analysis method [M]. Beijing: China Renmin University Press(2011)
8. ZHENG Zonghan, ZHENG Xiaoming. *Algorithm design and analysis* [M]. Beijing: Tsinghua University Press(2011)