Short-term PV/T module temperature prediction based on PCA-RBF neural network

Jiyong Li[1,2,3], Zhendong Zhao[1,2], Yisheng Li[1,2], Jing Xiao[1,2] and Yunfeng Tang[1,2]

[1] College of Electrical Engineering, Guangxi University, University Road 100, Xixiangtang District, Nanning, 530004, PR China;
[2] Guangxi Key Laboratory of Power System Optimization and Energy Technology, Guangxi University, University Road 100, Xixiangtang District, Nanning,530004, PR China.

[3] ji_yong_li@163.com

Abstract. Aiming at the non-linearity and large inertia of temperature control in PV/T system, short-term temperature prediction of PV/T module is proposed, to make the PV/T system controller run forward according to the short-term forecasting situation to optimize control effect. Based on the analysis of the correlation between PV/T module temperature and meteorological factors, and the temperature of adjacent time series, the principal component analysis (PCA) method is used to pre-process the original input sample data. Combined with the RBF neural network theory, the simulation results show that the PCA method makes the prediction accuracy of the network model higher and the generalization performance stronger than that of the RBF neural network without the main component extraction.

1. Introduction

Conversion efficiency of PV module is relatively low, usually 10% to 20%, and high PV module temperature will lead to further reduce the electrical efficiency [1]. The maximum temperature of the PV module is about 80 ℃ in the sunny weather in Nanning area of Guangxi (north latitude 22°50', east longitude 108°17'). In order to reduce the PV module temperature, improve the efficiency of power generation and recover the heat generated by the PV module, a photovoltaic-thermal(PV/T) system, which combines solar power generation and heat utilization, has become a research hotspot [2-3].

PV/T module temperature changes with the weather, showing significant volatility and intermittence. Through the temperature short-term prediction, control decisions can be made in advance, to optimize the control effect. There are two main ways to predict the temperature of PV/T module [4]. The first is to analyze the heat generation mechanism of the module and model the mechanism of heat production. As the module temperature is affected by complex environmental factors [5], making it difficult to model the heat generation mechanism, so the method is not practical. The second is the direct prediction by analyzing historical data and weather forecast data, which is widely used. In the literature [6], the correlation between PV module temperature and meteorological factors is established, and the regression analysis and prediction model is established. The method is only suitable for the rough calculation of PV module temperature in specific environment. In the literature [7], the back-propagation (BP) neural network is used to train the historical data to establish the forecasting model. Because the model does not consider the meteorological factors, the forecasting model is also limited. In this paper, the data preprocessing method and the prediction algorithm are

used to improve the prediction effect. The commonly used data preprocessing methods are data normalization, principal component analysis [8], wavelet decomposition [9] and so on.

Principal component analysis (PCA) is a statistical analysis method to transform high-dimensional complex correlation inputs into low dimension and weakly related inputs with the main information components [10]. With extracting the main data information components, ignoring the secondary components, the complex data can be made concise and clear and the data characteristics are more prominent [11], so as to improve the prediction accuracy.

2. PV/T system

Figure.1 shows the structure of the flat-type monocrystalline silicon PV/T module used in this study. It is mainly composed of monocrystalline silicon photovoltaic cells, copper pipe, insulator and so on. The heat absorbed by the PV module is stored by the insulator and taken away by cold water through the heat conduction copper tube. Figure. 2 is the PV/T system diagram, including the PV/T module, pumps, frequency converter, water tanks, and so on. According to environmental information and PV module temperature, the pump speed can be adjusted by control algorithm generated by the computer.

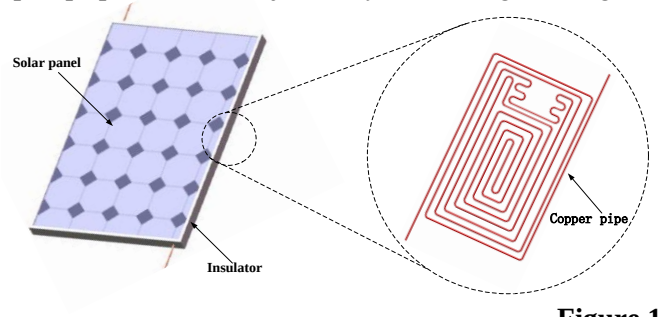

Figure 1. PV/T module structure.

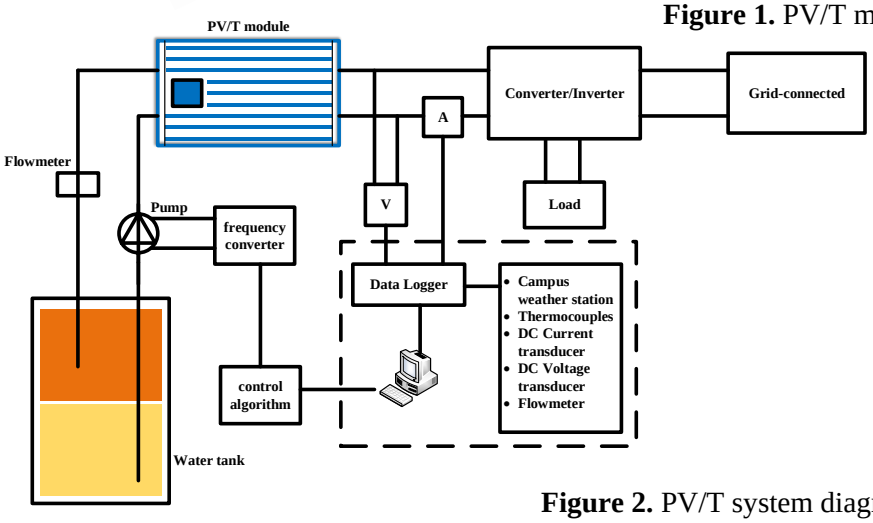

Figure 2. PV/T system diagram.

3. Basic principle of PCA

Assuming that there are m sets of research data, each of which contains n variables, constituting a $m \times n$ matrices.

Step1: Data normalization. Due to the difference between different types of variables will bring a lot of error, the original variable is required to be standardized, and then the standardized input matrix $\boldsymbol{X}_{m\times n}$ is obtained as follows.

$$\boldsymbol{X}_{m\times n} = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix} \tag{1}$$

Step2: Obtain covariance matrix $\boldsymbol{R}$ of $\boldsymbol{X}_{m\times n}$.

Step3: Obtain the eigenvalues $\lambda_1,\lambda_2\cdots\lambda_n$ of the covariance matrix $\boldsymbol{R}$ and the corresponding eigenvectors $[\boldsymbol{e}_1,\boldsymbol{e}_2,\cdots,\boldsymbol{e}_p]^T$.

Step4: Determine the number of principal components. Variance contribution rate and cumulative variance contribution rate can be obtained using

$$\eta_i = 100\%\lambda_i / \sum_i^n \lambda_i \tag{2}$$

$$\eta_{\Sigma(p)} = \sum_i^p \eta_i \tag{3}$$

Usually, when the cumulative variance contribution rate is greater than 85%, the principal component will contain the main information of the original variable.

Step5: The principal component expression is

$$\boldsymbol{Y} = \boldsymbol{EX} \tag{4}$$

$$\boldsymbol{E} = \begin{pmatrix} \boldsymbol{e}_{11} & \cdots & \boldsymbol{e}_{1p} \\ \vdots & \ddots & \vdots \\ \boldsymbol{e}_{n1} & \cdots & \boldsymbol{e}_{np} \end{pmatrix} \tag{5}$$

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{x}_1 \\ \boldsymbol{x}_2 \\ \vdots \\ \boldsymbol{x}_p \end{bmatrix} \tag{6}$$

Where: $\boldsymbol{E}$ is the eigenvector matrix of the principal component, $\boldsymbol{X}$ is the principal component extracted from the original input variable.

4. Model implementation

4.1. Influence factors of PV/T module temperature

The module is placed in the south direction and placed at an angle of 22°, in Nanning area of Guangxi (north latitude 22°50', east longitude 108°17'). The meteorological data are collected through the campus weather station, including solar irradiance ($\boldsymbol{I}_{\mathbf{r}}$), ambient temperature ($\boldsymbol{T}_{\mathbf{hj}}$), ambient humidity ($\boldsymbol{H}_{\mathbf{hj}}$), wind speed ($\boldsymbol{W}_{\mathbf{s}}$),wind direction ($\boldsymbol{W}_{\mathbf{d}}$).This paper firstly analyzes the correlation between PV/T module temperature and each meteorological factor, and then, analyzes the autocorrelation in time series. Two time series data of $\boldsymbol{T}_{(\mathbf{t-2})}$、$\boldsymbol{T}_{(\mathbf{t-1})}$ are selected to analyze the autocorrelation. The correlation coefficient $\boldsymbol{r}$ is calculated as follows:

$$\boldsymbol{r} = \frac{\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})(\boldsymbol{y}_i - \bar{\boldsymbol{y}})}{\sqrt{\sum_{i=1}^{n}(\boldsymbol{x}_i - \bar{\boldsymbol{x}})^2 \bullet \sum_{i=1}^{n}(\boldsymbol{y}_i - \bar{\boldsymbol{y}})^2}} \tag{7}$$

Where $\boldsymbol{x}_i, \boldsymbol{y}_i \in \boldsymbol{R}(\boldsymbol{i} \in (1,\boldsymbol{n}))$ $\bar{\boldsymbol{x}}$ is the mean value of $\boldsymbol{x}_i$, $\bar{\boldsymbol{y}}$ is the mean value of $\boldsymbol{y}_i$.

Table 1. Correlation between influence factors and PV/T temperature.

$\boldsymbol{r}$	$\boldsymbol{T}_{(\mathbf{t-2})}$	$\boldsymbol{T}_{(\mathbf{t-1})}$	$\boldsymbol{I}_{\mathbf{r}}$	$\boldsymbol{T}_{\mathbf{hj}}$	$\boldsymbol{H}_{\mathbf{hj}}$	$\boldsymbol{W}_{\mathbf{s}}$	$\boldsymbol{W}_{\mathbf{d}}$
$\boldsymbol{T}$	0.89	0.99	0.93	0.63	0.59	0.11	0.12

As can be seen from table 1, the self-correlation of PV/T module temperature is very strong. The correlation between the PV/T module temperature and the irradiation temperature, the ambient temperature, the ambient humidity, the wind speed, the wind direction is weakened in succession. From the above analysis, we can see that each meteorological factor has a certain degree of correlation with the PV/T module temperature. This paper selects the meteorological factors and history data as

the input. After the principal component extraction process, the principal component as the neural network model inputs to predict the PV/T module temperature. The process is shown in figure.3.

4.2. RBF neural network model

In the RBF network structure, there are three layers, namely, input layer, hidden layer, output layer, and each node connected with the input layer in hidden layer has its own center as shown in figure. 4. The role of the hidden layer node is to non-linear mapping of the input vector, and then to linear mapping of the output layer [12-13]. Each node of the hidden layer calculates the distance between input vector and its center, and then applies it to the excitation function to get the output of the hidden layer node. The output of the output node is the linear combination of the output of the hidden layer node. In this way, the hidden layer is obtained by a non-linear mapping of the input vector, and then through a linear mapping to the output. Excitation functions in the hidden layer are the same, except for their center. The following is a commonly used one-dimensional radial basis function:

$$\boldsymbol{f}(\boldsymbol{x}) = \exp(-\frac{\left\|\boldsymbol{x}-\boldsymbol{c}\right\|^2}{2\boldsymbol{\delta}^2}) \tag{8}$$

Determining this function requires two parameters: the center $\boldsymbol{c}$ and the variance $\boldsymbol{\delta}$. According to the network structure, the mathematical model of RBF neural network is:

$$\boldsymbol{y}_{kj}(\boldsymbol{X}_k) = \sum_{i=1}^{I} \boldsymbol{w}_{ij} \exp(-\frac{1}{2\boldsymbol{\delta}_i^2} \| \boldsymbol{X}_k - \boldsymbol{C}_i \|)^2 \tag{9}$$

Where: $\boldsymbol{w}_{ij}$ denotes the weights from hidden layer to output layer, $\boldsymbol{X}_k = \{\boldsymbol{x}_{k1}, \boldsymbol{x}_{k2}, ..., \boldsymbol{x}_{km}\}$ are the inputs of the neural network, $\boldsymbol{Y}_k = \{\boldsymbol{y}_{k1}, \boldsymbol{y}_{k2}, ..., \boldsymbol{y}_{km}\}$ are the outputs of neural network and $\boldsymbol{I}$ is the number of hidden layer nodes.

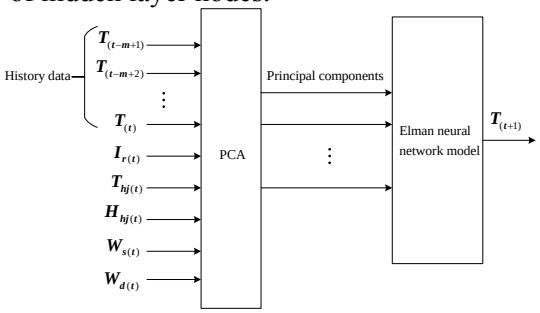

Figure 3. PV/T module temperature prediction flow chart

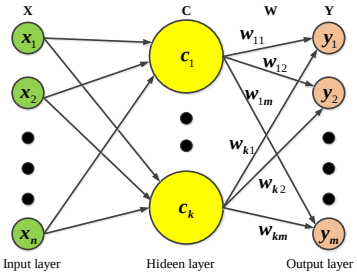

Figure 4.RBF neural network structure

5. Case analysis

5.1. Database description

According to the weather forecast information [14-15], 50 samples data are selected as network training data and 10 samples data are selected to test the prediction accuracy of the network, from July 1, 2016 to October 1, 2016 in sunny, cloudy weather.

5.2. Data normalization

According to the mathematical characteristics of the radial basis function, if the distance of the input data to the radial basis function center is smaller than a certain value, the RBF neural network can perform better. This article uses the "scale map method" for data normalization. Eq. (8) presents the normalization criterion,

$$\boldsymbol{y} = (\boldsymbol{x} - \boldsymbol{x}_{mid}) / (\boldsymbol{x}_{\max} - \boldsymbol{x}_{mid}) \tag{10}$$

$$\boldsymbol{x}_{mid} = 0.5(\boldsymbol{x}_{\max} + \boldsymbol{x}_{\min}) \tag{11}$$

Where: $x \in [x_{\min}, x_{\max}]$, x is the original data, y is the standardized value.

5.3. Key performance indices

The relative error (RE) and mean absolute error ($MAPE$) are used to evaluate the accuracy of the model. The calculation formula is shown as Eq. (10) and Eq. (11), respectively.

$$e_{RE} = \frac{|Y_i - Y_i^{'}|}{|Y_i|} \times 100\% \tag{12}$$

$$e_{MAPE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|Y_i - Y_i^{'}|}{|Y_i|} \times 100\% \tag{13}$$

Where n is the sample size, Y_i is the true value, $Y_i^{'}$ is the predicted value.

5.4. Principal component extraction

After the data is processed according to basic principle of PCA, the contribution rate of principal component and cumulative contribution rate is shown in figure 5.

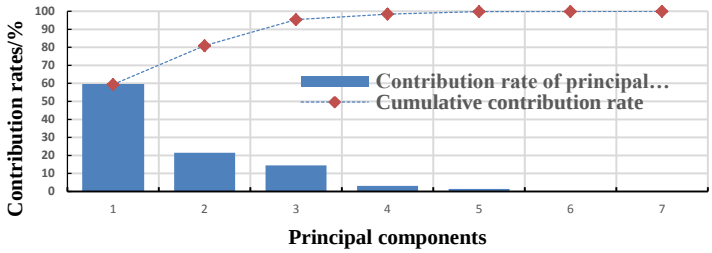

Figure 5. Contribution rates of principal components

It can be seen from figure 5 that the cumulative contribution rate of the first three principal components has reached 95% or so, including the main information of the input data. Therefore, the first three principal components are selected as the input of the neural network model. The result is shown in equation (14).

$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} 0.96 & 0.95 & 0.80 & 0.83 & -0.81 & 0.26 & 0.12 \\ 0.11 & 0.12 & 0.18 & -0.12 & 0.20 & -0.78 & 0.76 \\ 0.22 & 0.25 & 0.56 & -0.53 & 0.53 & 0.06 & -0.36 \end{bmatrix} \begin{bmatrix} T_{(t-2)} \\ T_{(t-1)} \\ I_r \\ T_{hj} \\ H_{hj} \\ W_s \\ W_d \end{bmatrix} \tag{14}$$

5.5. Experimental results analysis

Let model one: RBF model without principal component analysis, model two: PCA-RBF model with principal component analysis. The model comparison is carried out and the forecast results of different models are analyzed. The time step of prediction is 10 minutes.

5.5.1. Comparison of forecast results. The structure of the RBF model is 7-7-1, and the structure of the PCA-RBF model is 3-5-1. Figure 6, figure 7 is the prediction results of the RBF model and PCA-RBF model in sunny weather, figure 8, figure 9 is the prediction result of the RBF model and PCA-RBF model in cloudy weather, figure 10 is the relative error of two models in cloudy weather, figure 11 is the relative error of two models in sunny weather.

From the forecast results in sunny, cloudy weather, it can be seen that the PCA-RBF model is better than the RBF model, and its relative absolute error is small, especially when the temperature fluctuates seriously. The RBF model cannot follow the changes with the external environment timely and PCA-RBF model perform well in this situation, indicating generalization ability of the PCA-RBF

model is stronger. That is because the original data of the RBF model is not extracted by the PCA and the variables are highly correlated. The data feature is not obvious, which leads to the relatively complex network structure, reducing generalization performance and model sensitivity. In the PCA-RBF model, the PCA is used to extract the principal component variables instead of the original multivariate variables. The new principal component variables cover the vast majority information of the original variables and the correlation is weak, avoiding the cross-correlation between the original variables. In addition, the dimension of the input variable is reduced and the network structure is simplified. The PCA-RBF model is highly sensitive and the prediction performance is better. It can be seen from table 2 that the predictive evaluation indexes of the PCA-RBF model are lower than the RBF model.

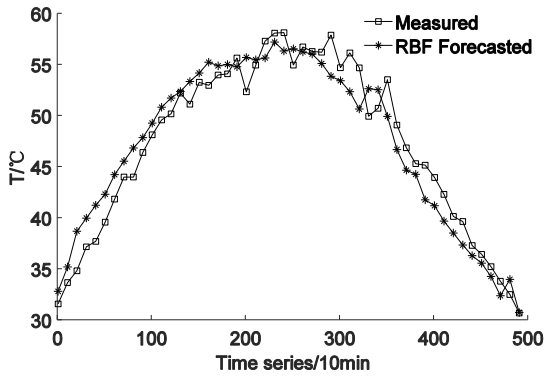

Figure.6 model RBF in sunny weather

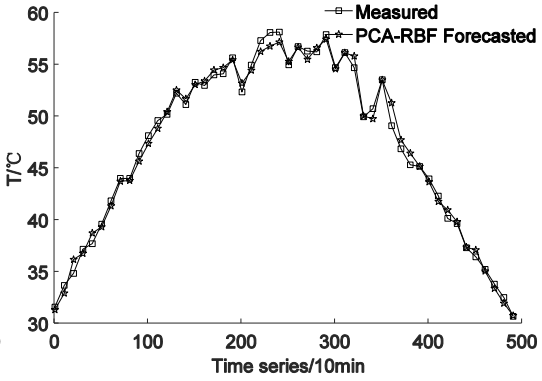

Figure.7 model PCA-RBF in sunny weather

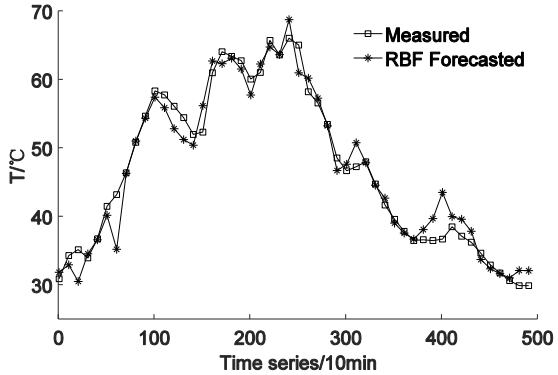

Figure.8 model RBF in cloudy weather

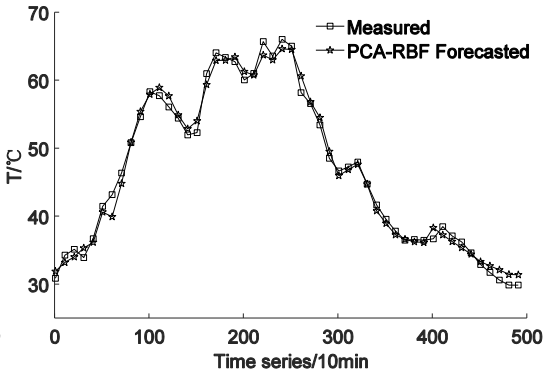

Figure.9 model PCA-RBF in cloudy weather

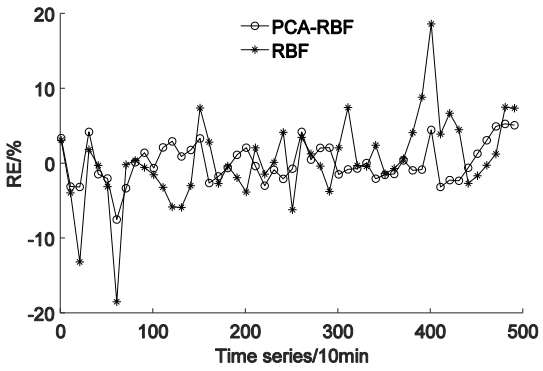

Figure 10. RE in cloudy weather

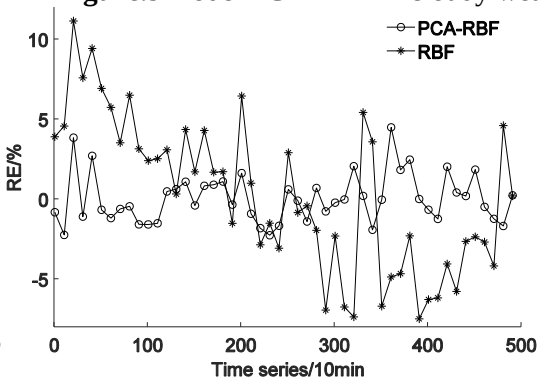

Figure 11. RE in sunny weather

6. Conclusions

In this paper, the meteorological factors and historical data correlated to the prediction temperature of PV/T module are taken into account. At the same time, the PCA technology is used to extract the

principal components of the input data. Compared with the model without PCA pretreatment, the forecasting effect of PCA-RBF is greatly improved, which can provide technical support for the further study of energy saving and emission reduction, and cascade utilization of solar energy.

Table 2. Evaluation index of predictive error.

Index /%	RBF		PCA-RBF	
	sunny	cloudy	sunny	cloudy
RE_{min}	11.12	18.58	4.47	7.53
RE_{max}	0.17	0.09	0.01	0.04
$MAPE$	3.78	4.08	1.18	2.16

Acknowledgement
This work was financially supported by the Guangxi Natural Science Foundation (2014GXNSFAA118372). Innovation Project of Guangxi Graduate Education (YCSW2017026)

References

[1] Xiao Wenbo, Wu Huaming, Fu Jianping et al. 2017 Effect of light intensity and temperature on the output characteristics of silicon photovoltaic cells *J. Huazhong Univ of Sci and Technol* **45**(1) 108-112 (in Chinese)
[2] Li Guangming, Liu Zuming, Li Jingtian, et al. 2013 Experimental study on new PV/T solar energy composite system *PROC CSEE* **33** (17) 83-89 (in Chinese)
[3] Khelifa A, Touafek K, Moussa H B, et al. 2016 Modeling and detailed study of hybrid photovoltaic thermal (PV/T) solar collector *Solar Energy* 135 169-176
[4] Gong Yingfei, Lu Zongxiang, Qiao Ying, et al.2016 Photovoltaic power prediction technology *J. MPCE* (04) 140-151 (in Chinese)
[5] Pan Jinjun, Shen Yanbo, Zeqiang, et al. 2014 In fluence of meteorological factors on the temperature of solar panels *J. Applied Meteorological Sci* **25** (2) 150-157 (in Chinese)
[6] Cheng Xu, Zhiping. 2013 Temperature prediction method for a photovoltaic cell module *Internet of Things Technol* **15**(11) 32-33 (in Chinese)
[7] Xu Ruidong, Dai You, Sun Xiaoyan. 2012 Prediction of PV array temperature based on BP neural network. *Industry and Mine Automation* **38** (7) 59-63 (in Chinese)
[8] Wang Lijie, Dong Lei, Gao Shuang. 2015 Short-term wind power forecast based on multi-location NWP and principal component analysis *Transactions of China Electrotechnical Society* **30** (5) 79-84 (in Chinese)
[9] Tian Zhong-da, Li Shu-jiang, Wand Yan-hong, et al. 2015 The short-term wind speed combination forecast of wind farm based on wavelet transform *Transactions of China Electrotechnical Society* **30** (9) 112-120 (in Chinese)
[10] Du Zifang. 2016 *Multivariate statistical analysis*. Beijing: Tsinghua University Press 239-269 (in Chinese)
[11] Qi Minfang, Fu Zhongguang, King Yuan, et al. 2013 Comprehensive evaluation method of thermal power units based on information entropy and principal component analysis *PROC CSEE* **33**(02) 58-64 (in Chinese)
[12] Li Qilong, Wang Lingmei, Shen Jianlin, et al. 2015 Prediction of vibration characteristics of wind turbine based on improved gray RBF neural network *Renewable Energy Resources* **33** (6) 876-882 (in Chinese)
[13] Liu Jingyan, Li Yudong, Guo Shunjing. 2016 Fault diagnosis of gearbox based on RBF neural network *Industry and Mine Automation* **42** (8) 47-51 (in Chinese)
[14] Li Bo, Men deyue, Yan Yaqin, et al. 2015 Bus load forecasting based on numerical weather forecasting *Automation of Electric Power Syst* (1) 137-140 (in Chinese)
[15] Wang Bo, Feng Shuanglei, Liu Chun. 2014 Wind power prediction method based on weather classification *Power Syst Technol* **38**(01) 93-98 (in Chinese)