Xiyun Yang[1], Feifei Jiang[1], Huan Liu[1]

School of Control and Computer Engineering,North China Electric Power University (1)

Short-Term Power Prediction of Photovoltaic Plant Based on SVM with Similar Data and Wavelet Analysis

***Abstract**. Photovoltaic(PV) power prediction is an important way to guarantee the stability for grid-connected PV power generation. In this paper, a power prediction method of PV plant was proposed. Firstly, the solar radiation was predicted based on support vector machine(SVM) combining the wavelet analysis with similar data. Similar data was extracted from large amounts of historical data. The original solar radiation signal was decomposed into trend signal of low frequency band and random signal of high frequency band by wavelet decomposition. Then different SVM radiation prediction models were trained respectively and the prediction result of every model was combined to obtain the final predicted radiation. Furthermore, the short-term power was predicted by the power curve to better fit the relationship between solar radiation and output power. Simulation results show that the improved SVM model can better fit the characteristics of solar radiation and improve the prediction accuracy; the power prediction model using the power curve has good generalization capability for engineering application.*

***Streszczenie.** W artykule przedstawiono metodę predykcji generowanej przez panele PV mocy. W algorytmie zastosowano maszynę wektorów pomocniczych oraz analizę falkową, które wykorzystano do predykcji promieniowania słonecznego, na podstawie danych historycznych. Dokonując porównania z krzywą mocy panelu określono zależność między promieniowaniem a mocą wyjściową. Badania symulacyjne wykazały, że proponowana nowa metoda pozwala na zwiększenie dokładności predykcji mocy. (**Krótkoterminowa predykcja mocy dla paneli fotowoltaicznych – zastosowanie maszyny wektorów pomocniczych z przykładowymi danymi i analizy Falkowej**).*

Keywords: solar radiation, short-term prediction, similar data, wavelet analysis, SVM, power curve
Słowa kluczowe: promieniowanie słoneczne, predykcja krótkoterminowa, similar data, analiza falkowa, SVM, krzywa mocy.

1 Introduction

Solar energy is rich in resources, pollution-free for environment and optimal choice for developing renewable energy sources. Grid-connected PV power generation is the development tendency for utilizing solar energy. PV power generation combined to the grid has caused severe challenges for the reliable and safe operation of power network [1,2]. Improving the accuracy of PV power prediction is an effective way to ensure the stability of power grid and enhance the operating efficiency of power station[3].

Solar panels use the photoelectric effect to transform the radiation into electrical energy. Solar radiation intensity is the major factor that affects PV power [4]. The solar radiation prediction mainly includes two kinds of methods: physical model-based methods and historical data-based methods [5]. The physical model is based on numerical weather prediction(NWP) to predict solar radiation. It does well in medium-term and long-term predictions. The historical data method does better in short-term prediction. It mainly includes autoregressive (AR) and autoregressive moving average (ARMA) [6], artificial neural networks(ANN) and support vector machine(SVM) [7] and so on. The artificial neural network method was used to predict solar radiation in [8], however, the neural network method needs long time for training and is easy to fall into local minimum. In [9], a mixed forecasting method combining wavelet analysis and neural network was adopted, but the degree of correlation between the training data and the predicted data will effect the prediction accuracy.

Recently, support vector machine (SVM) developed by Vapnik [10] is widely applied in computer science, bioinformatics, and environmental science [11]. Previous studies have proved that SVM shows better performance than neural networks and other statistical models [12]. SVM is a machine learning algorithm with better nonlinear fitting capability [13] and it is applied to predict solar radiation by some scholars. In this paper, a power prediction method of PV plant was proposed ,which included the solar radiation prediction based on support vector machine(SVM) combining the wavelet analysis and the solar output power prediction based on the power curve. Similar data was extracted from large amounts of historical data, mainly taking the factors such as ambient temperature, wind speed, cloud cover and the atmospheric transparency into account [14], which greatly increased the relativity between training sample and prediction data and enhanced the generalization capability of the model. Because solar radiation is nonstationary random sequence and the wavelet analysis has the performance of noise reduction, the original solar radiation data was decomposed into trend signal of low frequency band and random signal of high frequency band by wavelet decomposition. Different SVM radiation prediction models were trained respectively, which fit the low frequency and high frequency character of radiation data well and improved the prediction accuracy of the combined prediction data. Furthermore, an actual power curve，which can better map the relationship between solar radiation and output power, was used to predict the short-term power after obtaining the prediction of solar radiation. The power curve was exacted from historical data and it can reduce the influence of abnormal data in setting the power prediction model.

2 Methodology

2.1 SVM

SVM is a new-style computational learning theory, which has many advantages such as concise math expression form, solid theoretical foundation, intuitive geometric explanation, strong generalization ability and it can overcome the difficulty caused by nonlinearity and small sample [15].

SVM theory is proposed from the optimal hyperplane in linear classification cases. Given two sample sets (xi,yi), $I = 1,2,\ldots n$, $x \in R^d$, $y \in \{=1, -1\}$, the general type of linear discriminate function in the D dimensional space is:

$$\mathrm{g}(x) = w \cdot x + b \tag{1}$$

The hyperplane equation is:

$$w \cdot x + b = 0 \tag{2}$$

By normalizing the judging function, the two sets can meet the formula $|g(x)| \geq 1$, i.e., the samples which are nearest to the hyperplane can meet the formula $|g(x)| = 1$ and the class interval is $2/\|w\|$.Therefore, to make the interval maximum is equal to make $\|g(x)\|$ (or $\|g(x)\|^2$) minimum. All samples will be classified properly if equation (3) is met.

(3) $$y_i\left[\left(w\cdot x_i\right)+b\right]-1\geq 0$$

where: $i=1,2,\dots,n$

Thus, the problem of the optimal hyperplane can be expressed as equation (4) subjected to equation (3) to calculate the minimum.

(4) $$\Phi(w)=\frac{1}{2}\|w\|^2=\frac{1}{2}(w\cdot w)$$

The Lagrange function is:

(5) $$L\left(w,b,\alpha\right)=\frac{1}{2}\left(w\cdot w\right)-\sum_{i=1}^{n}\alpha_i\left\{y_i\left[\left(w\cdot x_i\right)+b\right]-1\right\}$$

where: α_i – the Lagrange coefficient.

Using inner product $k(x,x_i)$ to replace the inner product of optimal hyperplane, we can get the optimization function (6).

(6) $$Q\left(\alpha\right)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j k\left(x_i,y_j\right)$$

The discrimination function is :

(7) $$f\left(x\right)=\operatorname{sgn}\left(\sum_{i=1}^{n}\alpha_i^{*}y_i k(x_i,x)+b^{*}\right)$$

2.2 Wavelet Decomposition

Solar radiation possesses some randomness and it is mixed with noise and multiple signals of high volatility. Wavelet analysis with the ability of noise reduction and analyzing the local characteristics of the nonlinear and non-stationary signals is the localization analysis of the time (space) and frequency. Through the telescopic shifting operation, the signals will be gradually refined by multi-scale so as to achieve time subdivision at high frequency and frequency subdivision at low-frequency. It can automatically adapt to the analysis demand of time-frequency signal. Therefore, any details of the signal can be focused on, so that you can analyze the local features of the signal at various time and in various local range [16].

The expression of continuous wavelet transform for $x(t)$ signal is:

(8) $$W_{\Phi}x(b,a)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x(t)\overline{\Phi(\frac{t-a}{a})}dt$$

where: $a=1/2^s$, $b=k/2^s$, k and s belong to the set of integers Z.

At the time-scale plane, the value of continuous wavelet transform of $x(t)$ is a number which expresses the time-scale relationship of $x(t)$ and $\overline{\Phi}(t)$ and it is named as discrete wavelet transform (DWT) which generates a sparse set at the time-scale plane. Wavelet coefficients at the point $(b=k/2^s, a=1/2^s)$ can be expressed by equation (10).

(9) $$w_{k,s}=W_{\Phi}x\left(\frac{k}{2^s},\frac{1}{2^s}\right)=\frac{1}{\sqrt{a}}\int_{-\infty}^{\infty}x(t)\overline{\Phi(\frac{t-\frac{k}{2^s}}{\frac{1}{2^s}})}dt$$

Because redundant information would be generated by using continuous wavelet transform at the time-scale plane, the signal is more conveniently processed by discrete wavelet transform after selecting $(b=k/2^s, a=1/2^s)$. Discrete wavelet transform has been proved to maintain signal information adequately. Signal can be fully reconstructed according to wavelet coefficients.

The fast algorithms of discrete orthogonal wavelet transform can achieve orthogonal projection to the space V_j and W_j for signal $x(t)$. Then the discrete approximation signal $\alpha_j\left(t\right)$ and discrete detail signal $d_j(t)$ can be obtained respectively corresponding to resolution j of signal $x(t)$ Assuming that j gradually increased from 1, such as $j=3$ in this paper, the breakdown structure tree is shown in Fig.1.

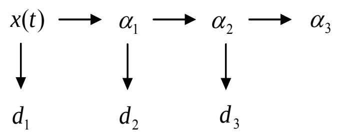

Fig.1. Multi-decomposition tree

Finally, wavelet decomposition of signal $x(t)$ can be obtained as equation (10).

(10) $$x(t)=\sum_{j=1}^{3}d_j(t)+a_3(t)$$

The effect diagram has been shown in Fig.2, where the original solar radiation signal is decomposed into the trend signal and random signal.

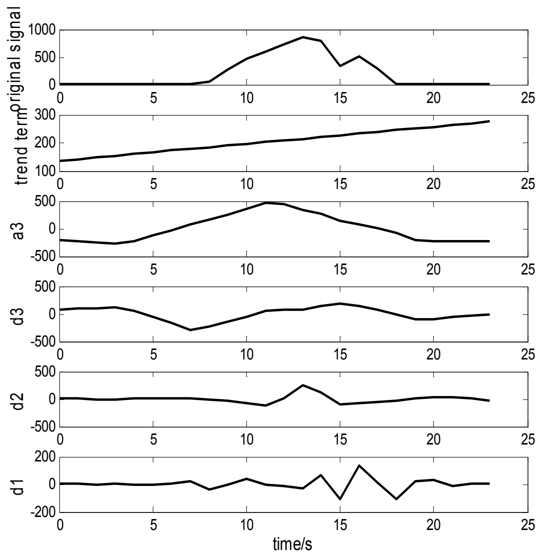

Fig. 2. Wavelet decomposition

2.3 Similar Data

It is obviously that solar radiation periodically changes with season. Even in one day, the change of radiation is also regular. SVM can map the input sample to high dimensional linear characteristic space through nonlinear mapping and make it linearly separable. What SVM records is the non-linear relationship between the input and output of training sample, so it can be seen that the prediction accuracy can be highly improved by choosing prediction sample which is similar with the training sample. Using the radiation curve in a certain time period as the sample and finding a set of curves which changed similarly with the sample curve, the wavelet SVM prediction model can be set using the curves obtained as the training sample.

2.4 Data Normalization

In order to improve the generalization capability of the model and reduce the training time of the program, data can be normalized when SVM modeling is established. That is to say the range of trend signal and detrended random signal is transformed to $[0,1]$.

The normalized formula is:

(11) $$\ddot{x}_i = \frac{x_i - x_{i\min}}{x_i - x_{i\max}}$$

where: $i=1,2,\ldots,n$

Among the formula $\ddot{x}_i$ is the value of normalized data x_i is the value of measured data $x_{imin}=min(x_i)$, $ximax=max(x_i)$ and n is the number of samples.

2.5 Performance Evaluation

The error formulas used in this paper are as follows:

a. Normalized root mean square error(nRMSE)

(12) $$e_{nRMSE} = \frac{\sqrt{\sum_{t=1}^{N}\left(v_t^{'} - v_t\right)^2}}{P_{cap}\sqrt{N}}$$

b. Mean relative error(MRE)

(13) $$e_{MRE} = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{v_t^{'} - v_t}{v_t}\right|$$

c. Mean absolute error(MAE)

(14) $$e_{MAE} = \frac{1}{N}\sum_{t=1}^{N}\left|v_t^{'} - v_t\right|$$

where: N – number of prediction data, v'_t – value of prediction data, v_t – value of measured data, P_{cap} – rated capacity power of the solar cell array.

3 Solar Radiation prediction model

3.1 Data Analysis

HOMER(Hybrid Optimization Model for Electric Renewable) is a model optimization software researched by National Renewable Energy Laboratory(NREL) for multi-renewable energy power generation system. The data used in this paper is all from a PV power station(Latitude: 24°49′ North，Longitude：103°20′ East), and the HOMER software simulates the radiation data in the whole year for analysis in the earlier stage.

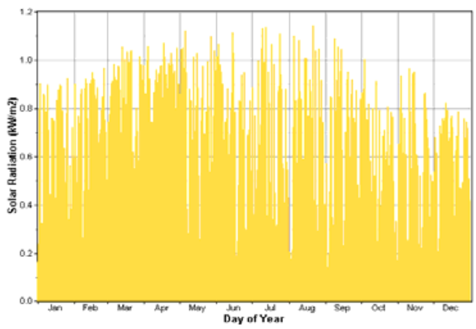

Fig. 3 Radiation data of everyday in one year

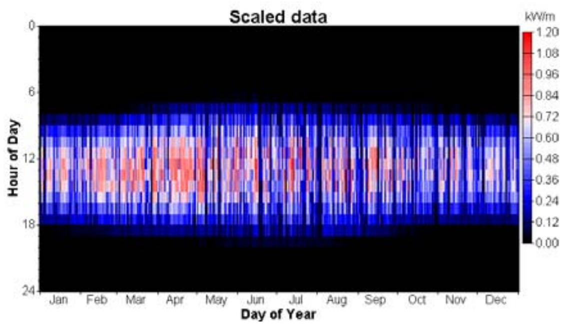

Fig. 4 Radiation intensity of each hour of a day in one year

Fig.3 shows the radiation data of everyday in one year, and Fig.4 shows the two-dimensional radiation intensity presentation of each hour of a day in one year. As shown in the two figures, taking no account of the severe weather of some days, the radiation of the other days almost changes similarly. There is no radiation during the night, so the solar radiation data from 6 am to 9 pm is used.

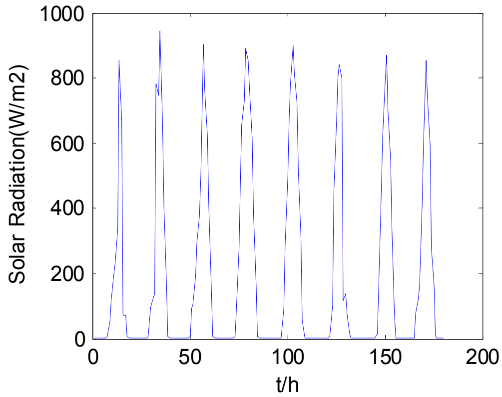

Fig. 5 Variation trend of radiation of similar days

Fig.5 shows the radiation data of 8 continuous days with similar weather situation, these days are similar with each other in ambient temperature, cloud cover and the atmospheric transparency. As shown in Fig.5, the radiation data has high relativity in the same hour of these days.

For this reason, similar data extracted from the historical data was used as training sample. The similar data extracted in this paper as the training sample was data of 8 days as shown in Fig.6, in which every 24 points in the abscissa represent one day. It can be seen from the figure that the variation trend of these days is very similar.

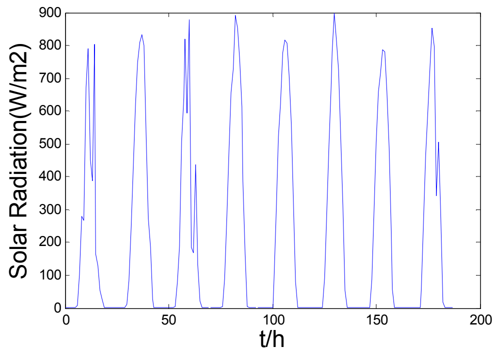

Fig.6 Training sample using simialr data

3.2 Solar Radiation Prediction Using SVM Model

The measured data of 2011 from a PV power station was used as the original data. There are 12 solar cell arrays in the PV station and the 3rd array was chosen for this research. 24 radiation points were recorded in one day, and every point was the mean value of six data in one hour. Then, similar data of 10 days is selected from the data before the day to predict. The radiation of the previous day was used as the prediction input.

The SVM kernel function used in the model is RBF, as shown in equation (15):

(15) $$k(x, x_i) = \exp\left(-\frac{\left\|x - x_i\right\|^2}{2\sigma^2}\right)$$

Take data of the first 8 days as the training sample as shown in figure 6 and data of the last two days as the prediction sample.

Firstly, the data is normalized to the range *[0,1]*, and then solar radiation is forecasted by using SVM model. Finally, the forecasting radiation data is obtained after anti-normalization. Figure 7 shows the prediction results based on SVM and the prediction error is shown in table 1.

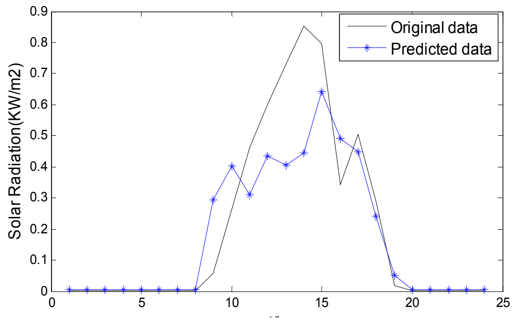

Fig.7 Predicted radiation based on SVM

3.3 Solar Radiation Prediction Using Wavelet SVM Model

Radiation signal can be divided into low-frequency trend signal and high-frequency random signal using the wavelet decomposition technique as shown in figure 2. Different SVM models were set respectively using these signals as the input sample. Figure 8 shows the five prediction results of every SVM model. The final predicted radiation combining these five results is shown in figure 9 and the prediction error is shown in table 1.

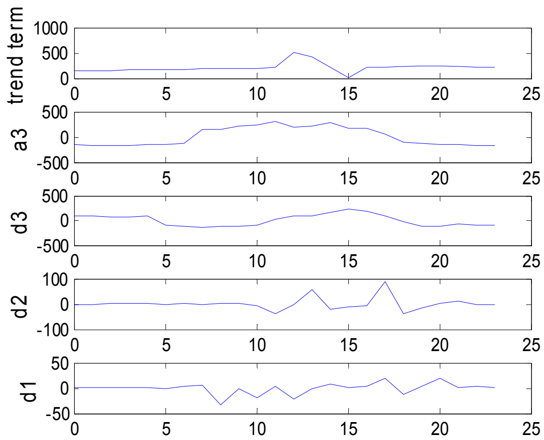

Fig.8 Different prediction results of five signals decomposed by the wavelet

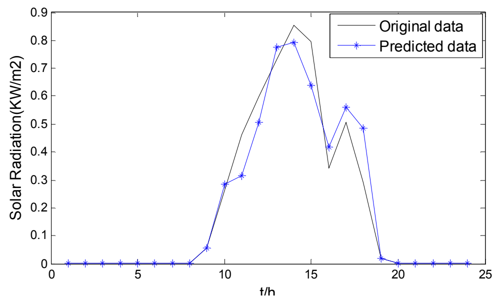

Fig.9 Predicted radiation based on wavelet decomposition

Table 1 Prediction error of solar radiation

Method	nRMSE(%)	MRE(%)	MAE(KW)
SVM	13.70	2.6439	0.0804
SVM and Wavelet	6.69	0.0780	0.0353

3.4 Analysis of Simulation Results

Figure 6 shows the prediction result of the SVM method, of which the nRMSE is 13.70 %, the MRE is 2.6439% and the MAE is 0.0804 KW/m^2. After adopting wavelet decomposition, the original solar radiation signal was decomposed into trend signal of low frequency band and random signal of high frequency band, so that the non-stationary of original data was reduced. What can be found from table 1 is that the nRMSE reduces from 13.70% to 6.69%, the MRE reduces from 2.6439% to 0.0780% and the MAE reduces from 0.0804 KW/m^2 to 0.0353 KW/m^2. It is apparent that the prediction accuracy of the model is obviously improved, which is of vital significance for applied engineering project.

4 Power prediction model

Solar radiation transforms to the output power through the solar cell panel. With the radiation being predicted, the output power will be obtained when the relationship between radiation and power is knowable. The power prediction model was set by using an actual power curve to establish the matching relation between the solar radiation and the output power.

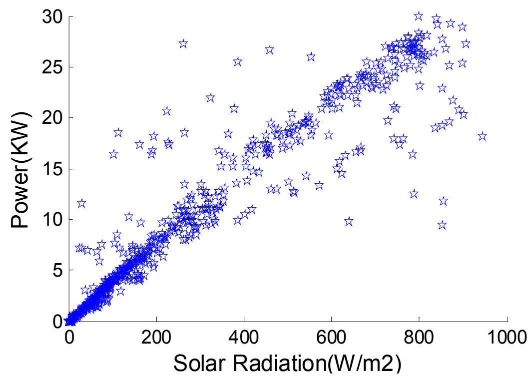

Fig.10 The relationship between solar radiation and power

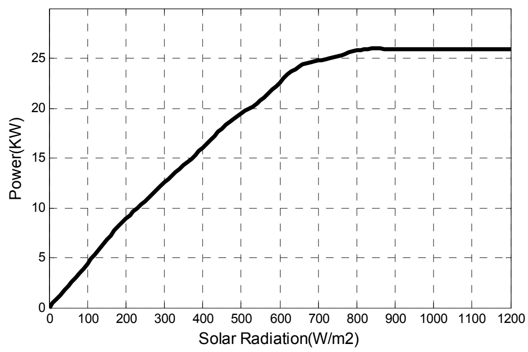

Fig.11 The extracted actual power curve

Figure 10 shows the relationship between solar radiation and power of the 3rd solar cell array. It can be seen from the figure that there exists a certain linear relationship between them, which has a high correlation coefficient of 0.9674. However, there are also many abnormal points deviating from the linear relationship, which may be caused by the unusual weather or the acquisition system fault. Directly using the historical radiation and power data for modeling will inevitably reduces the prediction accuracy. Therefore, it is important to determine the corresponding relation between the input and output samples. The historical data was handled by the L'Hospital's rule and the power curve is extracted from the handled historical data for modeling and predicting the power generation. Fig.11 shows the extracted actual power curve.

The extracted power curve was used to create the SVM model,which established the mapping function of radiation and power. Taking the predicted radiation as shown in Figure 7 as the input, the whole power prediction process will achieved. The final predicted output power is shown in Fig.12, the error shown in Table 2.

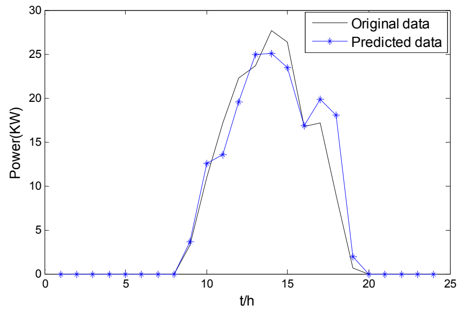

Fig.12 Predicted power based on the power curve

Table 2 Prediction error of output power

Method	nRMSE(%)	MRE(%)	MAE(KW)
Power curve	9.3	0.1608	1.1725

It can be seen from the results above that the power prediction effect using the power curve is satisfactory with the nRMSE of 9.3%, which can meet the demand of engineering application.

Conclusion

In this paper, a PV plant output power prediction method with high accuracy was proposed. Similar data was extracted from large amounts of historical data, which greatly increased the relativity between training sample and prediction data and enhanced the generalization capability of the model. Wavelet decomposition reduces the non-stationary of original data, fits the low frequency and high frequency character of radiation data well and improves the instantaneity of the algorithm. The prediction accuracy was highly improved by combining similar data and wavelet decomposition with SVM. An actual power curve was extracted from the handled historical data to set the power prediction model. The predicted power data shows that this method has high prediction accuracy and can meet the demand of engineering application. With the rapid development of PV power, this method will have high practical value in ensuring the stability of grid-connected PV power and increasing economic returns.

Acknowledgement

This paper is supported by the national natural science fund project (61074094, 51207049), Beijing Natural Science Foundation (4132061), the fundamental research funds for the central universities (12MG59, 12MS51).

REFERENCES

[1] Dr. Bernhard ERNST, Dr.Frank REYER, Wind power and photovoltaic prediction tools for balancing and grid operation, *Integration of Wide-Scale Renewable Resources Into the Power Delivery System, 2009 CIGRE/IEEE PES Joint Symposium*, 2009, 1-9

[2] G. S. Kinsey, K. Stone, J. Brown, Energy prediction of amonix cpv solar power plants, *Progress in Photovoltaics: Research and Applications,* 19(2011),No.7,794-796

[3] Elke Lorenz, Johannes Hurka and Detlev Heinemann, Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems, *IEEE Journal of Selected Topics In Applied Earth Observations and Remote Sensing,* 2(2009),NO.1,2-10

[4] M.G. Abraha, M.J. Savage, Comparison of estimates of daily solarradiation from air temperature range for application in crop simulations, *Agricultural and Forest Meteorology*, 148(2008),401–416

[5] Jianwu Zeng, Wei Qiao, Short-Term Solar Power Prediction Using an RBF Neural Network, *Power and Energy Society General Meeting,* 2011 IEEE.1-8

[6] C. Paolik, C. Voyant, M. Muselli and M. Nivet, Solar radiation forecast-ing using ad-hoc time series preprocessing and neural networks, *Proc. 5th International Conference on Emerging Intelligent Computing technology and applications*, 2009,898-907

[7] Zengxin Wang , Shi Su, Shaoquan Zhang, The Application of Photovoltaic Power Prediction Technology, *2011 International Conference on Communications and Control (ICECC),* 2011,2343-2346

[8] Ji-Long Chen, Hong-Bin Liu and Wei Wu, Estimation of monthly solar radiation from measured temperatures using support vector machines, *Renewable Energy*, 36(2011), No.1, 413–420

[9] A. Ahhi, M. Shamisi and M. Jama, Prediction of monthly average daily global radiation in Al Ain city - UAE using artificial neural networks, *Proc. 4th Wseas International Conference on Renewable Energy Sources (RES 10)*, 2010, 109-113

[10] Samsul Ariffin Abdul Karim, Balbir Singh Mahinder Singh, Radzuan Razali and Noorhana Yahya, Data Compression Technique for Modeling of Global Solar Radiation, *2011 IEEE International Conference on Control System, Computing and Engineering*, 2011,348-352

[11] H.B. Liu, D.T. Xie and W. Wu, Soil water content forecasting by ANN and SVM hybrid architecture Environ Monit Assess, *Environmental Monitoring and Assessment*, 143 (2008), 187–193

[12] B.Dong, C. Cao and S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, *Energy and Buildings*, 37 (2005), 545–553

[13] Rong Xiao, Jicheng Wang and Fuyan Zhang, An Approach to Incremental SVM Learning Algorithm, *12th IEEE International Conference on Tools with Artificial Intelligence*, 2000, 268 - 273.

[14] E.M. Crispim, Pedro M. Ferreira and A.E. Ruano, Solar radiation prediction using RBF neural networks and cloudiness indices, *2006 International Joint Conference on Neural Networks*, 2006, 2611-2618

[15] V. Vapnik, S.E. Golowich and A.J. Smola, Support vector method for function approximation, regression estimation and signal processing, *Adv Neural Inf Process Syst*, 9(1996), 281–287.

[16] V. Vapnik. Statistical learning theory Wiley, *New York* (1998).

***Authors**: prof. Xiyun Yang, School of Control and Computer Engineering, North China Electric Power University, E-mail:yangxiyun916@sohu.com; Feifei Jiang, School of Control and Computer Engineering, North China Electric Power University, E-mail:feiflyhi@yeah.net; Huan Liu, School of Control and Computer Engineering, North China Electric Power University, E-mail: liuhuanliuchang246@163.com.*

The correspondence address is:
Feifei Jiang, School of Control and Computer Engineering, North China Electric Power University, No.2 Beinong Road, Chang Ping District, Beijing, China, 102206. e-mail: feiflyhi@yeah.net