Short-Term photovoltaic system power forecasting based on ECSVM optimized by GA

Yun-jun Yu*, Yun-tao Xue, Sui Peng, Chao Tong, Zi-heng Xu

College of Information Engineering, Nanchang University Xuefu Rd. 999, Nanchang, Jiangxi, 330031, China

Received 1 May 2014, www.cmnt.lv

Abstract

It is of great significance to research PV forecasting techniques for mitigating the effects of the randomness of the Photovoltaic output. This paper analyses many factors from PV which impact photovoltaic output and extracts the main factors, forming sample data combined with the historical database generation data from PV monitoring system. And an error correction SVM method (ECSVM) is used to calculate the open integration of photovoltaic power storage system in advance or after the time in order to try to eliminate the system error between the predicted and actual values. At the same time, using genetic algorithm to optimize kernel function parameter and the error penalty factor and other parameters in this model, the establishment of the GA-ECSVM model improves portfolio optimization model parameter prediction accuracy and efficiency of the selected type. Finally examples verified and compared with standard SVM methods and ECSVM method, predicting effects show that: The GA-ECSVM optimization model presented in this paper has better learning ability and generalization ability in the short term prediction of photovoltaic power generation, with the prediction accuracy of 95.2016%.

Keywords: PV, SVM, error correction, GA, forecast

1 Introduction

Along with the intensity of environmental pollution caused by global energy shortage and the use of fossil resources, solar power as one of the important renewable energy, has received wide attention [1, 2]. Photovoltaic system has more practice value in this area because of its technical improvement. Because of the stochastic weather and fluctuation of illumination intensity, it will inevitably influence generation scheduling greatly [3-5]. The accurate PV power generation forecast is one of the effective means to improve the PV power capacity, the stability and economy of the power grids operation.

The theory of PV power generation forecasting has been researched in recent years, and a lot of forecasting methods are proposed. In [6], using historical weather data sources to make predictions about illumination intensity in future. But the accuracy of PV power generation prediction needs to be improved to satisfy the actual production. In [7], a back propagation (BP) neural network forecasting model was proposed whose input parameters were ambient temperature, humility and cloud cover, which improve accuracy of PV power generation prediction. But photovoltaic (PV) generation forecasting models need to take cloud cover as their input parameters. However, they were difficult to implement in China due to insufficient weather stations available.

Extraction temperature, Solar Radiation Intensity and sunshine time were the main factors, which affect PV power generation. This paper provides error correction SVM method (ECSVM) according to historical data. But error correction SVM method (ECSVM) is still deficits in the technical parameter selection. In allusion to these problems, this paper use genetic algorithm to optimize kernel function parameters and the error penalty factor and other parameters in this model, the establishment of the GA-ECSVM model improves portfolio optimization model parameter prediction accuracy and efficiency of the selected type.

2 Important factors affecting PV generation

For different districts and locations, meteorological factors, including solar irradiance, temperature, and so forth, are always changing [8, 9]. In order to efficiently utilize renewable energy using solar energy, an analysis of the characteristics of meteorological factors at a potential site should be considered. Data from the solar energy grid-connected PV plants, including the amount of solar electricity being produced and the meteorological factors, will be made available to analyse.

The output power of photovoltaic power station on sunny days is presented in Figure 1. It has strong randomicity, but similar on the whole.

* ***Corresponding author*** e-mail: yuyunjun@ncu.edu.cn

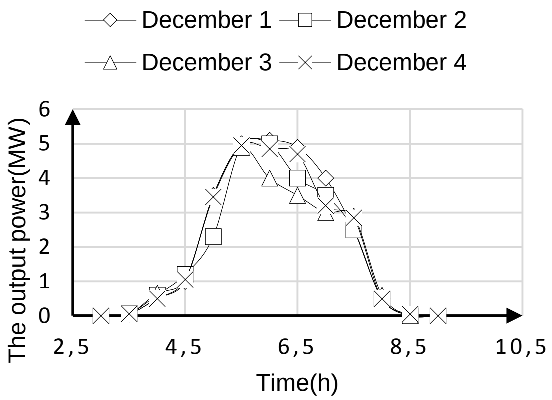

FIGURE 1 Daily power output in sunny days

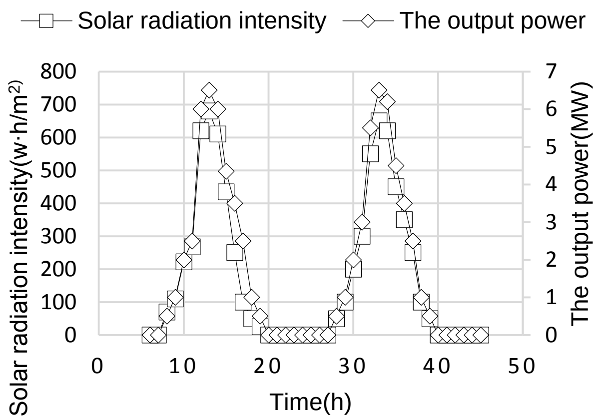

FIGURE 2 Solar radiation and power output of PV system

2.1 SOLAR RADIATION INTENSITY

As shown in Figure 2, The intensity of solar radiation has a significant impact on photovoltaic power generation. In photovoltaic system, the output of photovoltaic battery is affected greatly by the Solar Radiation Intensity, it has obvious non-linear features [10]. Thus, the intensity of solar radiation will be used as one significant impact of the predictive model.

2.2 CLOUD

The effects of the cloud on solar radiation is considered. The response time of a PV (photovoltaic) plant is very short and its output power follows the abrupt change in solar irradiance level due to alternate shadow by clouds [11, 12]. As shown in Figure 3, When the cloud covers a large area, the output power will be reduced relatively, also the shadow by clouds will be used as one significant impact of the predictive model.

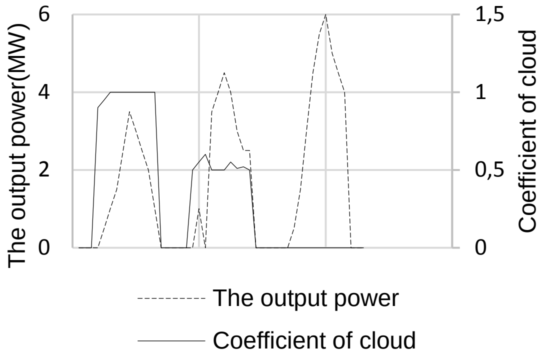

FIGURE 3 Coefficient of cloud and power output of PV system

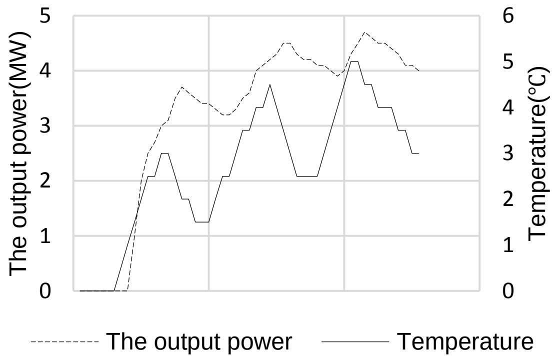

FIGURE 4 Temperature and power output of PV system

2.3 TEMPERATURE

The change rules of the output power were similar with the change tendency of temperature [13, 14]. As shown in Figure 4, there is a positive correlation between temperature and the output power. In addition, the temperature will be used as one significant impact of the predictive model.

3 The theory of SVM

The advantage of support vector machine predictive model can overcome the shortcoming of traditional methods, which are over-fitting, non-linear, disaster of dimensionality, local minimum.

3.1 THE THEORY OF SVM REGRESSION

Regression problems was based on a new input stylebook data x to inference corresponding output y.

Data sample set is $\{(x_i, y_i),...,(x_l, y_l)\}$, where $x_i \in R^n, y_i \in R, i = 1,2,3,...,l$.

The basic idea of SVM estimate regression means, performing the data of input space into a high-dimensional feature space through non-linear mapping relationship.

The value of sample data $\{x_i, y_i\}$, $i = 1,2,3,..., s(x_i \in R^n, y_i \in R)$, where y_i is the exception, s is the total number of data points.

Regression problem is solved by Loss function in SVM. Use function: $y = f(x) = [w\varphi(x)] + b$.

Take the extreme value of optimization goal:

$$\min Q = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i + \zeta_i), \tag{1}$$

$$s.t.\begin{cases} y_i - [wk\varphi(x_i)] - b \le \varepsilon + \xi_i \\ [wk\varphi(x_i)] + b - y = \varepsilon_i + \zeta_i \\ \xi_i, \zeta_i \ge 0, i = 1, l, s \end{cases},$$

where C is error penalty factors; ξ_i and ζ_i are relaxation factor; ε is loss function. Loss function can show decision function through sparse data points. The loss Equation (2) introduced has a good effect.

$$L_\varepsilon(y) = \begin{cases} 0 & |f(x) - y| < \varepsilon \\ |f(x) - y| - \varepsilon & |f(x) - y| \ge \varepsilon \end{cases}. \tag{2}$$

Using Lagrangian multiplication a_i and b_i, which facilitate convex optimization problems into quadratic maximum.

TABLE 1 Common kernel function

	Name	Expression
1	Liner kernel function	$k(u,v) = (u \cdot v)$
2	Polynomial kernel function	$k(u,v) = (r(u \cdot v) + coef0)^d$
3	RBF kernel function	$k(u,v) = \exp(-r\|u - v\|^2)$
4	Sigmoid kernel function	$k(u,v) = \tanh(r(u - v) + coef0)$

Note: the r is kernel function parameter.

Support vector machines(SVM) method is developed for solving highly nonlinear classification,whose program are mapped into a high-dimensional feature space through a certain transformation function to nonlinear data variables and then from line regression of high dimension using kernel function $k(x, x_i) = \phi(x)\phi(x_i)$. Introduce kernel function to replace inner product computation. optimization goal integrated as shown below :

$$\max W(a_i, b_i) = \sum_{i=1}^{n} y_i(a_i - b_i) - \varepsilon\sum_{i=1}^{n}(a_i + a_i') - \frac{1}{2}\sum_{i,j=1}^{n}(a_i - a_i')(a_j - a_i')(x_i - x_j'), \tag{3}$$

$$s.t.\begin{cases} \sum_{i=1}^{n} a_i = \sum_{i=1}^{n} a_i' & i = 1, 2, \dots, n \\ 0 \le a_i, a_i' \le C \end{cases},$$

where C is used to control the complexity of the model and compromises approximation errors. When the C larger, the fitting degree higher; ε was used to control regression approximate error and generalization ability models.

3.2 THE METHOD OF PREDICTING

The time series are $\{x_1, x_2, x_3, \dots, x_n\}, i = 1, 2, 3, \dots n$ and predictive goal value is $\{x_n\}$. This paper set up the functional projective relationship which is $R^m \to R$ from the input to the output .m is embedded dimension. We can get samples for learning SVM:

$$X = \begin{bmatrix} x_1 & x_2 & L & x_m \\ x_2 & x_3 & L & x_{m+1} \\ L & L & L & L \\ x_{n+m} & x_{n+m+1} & L & x_{n-1} \end{bmatrix}, Y = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ M \\ x_n \end{bmatrix}.$$

Regression function:

$$y_i = \sum_{i=1}^{n-m}(a_i - a_i')k(\overline{a_i} - \overline{a_j}) + b, j = m+1, \dots, n. \tag{4}$$

The first step:

$$x_{n+1} = \sum_{i=1}^{n-m}(a_i - a_i')k(\overline{a_i} - \overline{a_{n-m+1}}) + b, \tag{5}$$

where $\overline{a_{n-m+1}} = \{x_{n-m+1}, x_{n-m+2}, \dots, x_n\}$.

Then:

$$\overline{a_{n-m+2}} = \{x_{n-m+2}, x_{n-m+3}, \dots, x_n, x_{n+1}\}.$$

Therefore:

$$x_{n+2} = \sum_{i=1}^{n-m}(a_i - a_i')k(x_i, x_{n-m+2}) + b.$$

Result:

$$x_{n+l} = \sum_{i=1}^{n-m}(a_i - a_i')k(\overline{a_i} - \overline{a_{n-m+l}}) + b, \tag{6}$$

where $x_{n-m+l} = \{x_{n-m+l}, x_{n-m+l+1}, \dots, x_{n+l-1}\}$.

3.3 PV GENERATION PREDICTION MODEL BASED ON ECSVM

PV system affected by the environment, there are many uncertain disturbance sources. The errors between predictive value and the actual value were unavoidable. In order to minimize error, error integral method is proposed. Its characteristics are as follows:

$$Y_y = [x(1)x(2)x(3)\dots x(120)], \tag{7}$$

$$Y_s = [y(1)y(2)y(3)\dots y(120)], (n \le 120). \tag{8}$$

Y_y – predictive power, Y_s – real generation power,

$$E_r = \sum_{i=t}^{n}[Y_y(i) - Y_s(i)], (t \le n \le 120), \tag{9}$$

where $Y_y(i)$ – real generation power at time i equals; – Predictive power at time i equals; E_r – the errors between predictive value and the actual value.

$$t = \frac{E_r - E_g \times \eta_{inv}}{P_d \times \eta_r}, \tag{10}$$

where t – energy storage opening time error; E_g – the errors between predictive value and the actual value; P_d – energy storage equipment output power; η_{inv} – efficiency of PV power generation system; η_r – efficiency of energy storage output energy.

4 PV Generation prediction model based on GA-ECSVM

SVM can have different performances of classification through choosing different Kernel Functions and parameters [15, 16]. The performance of the SVM is influenced by kernel function parameter r [17, 18]. Meanwhile, kernel function is the key technology of SVM. Using different kernel function will affect the learning ability and generalization ability of SVM. In view of the present insufficiency in the selection technology aspect of SVM, several proposals were put forward, which will improve SVM learning ability, generalization performance and the ability to choosing the kernel function and function's parameters.

Genetic Algorithm is based on the nature selection and genetic transmission mechanisms, whose advantages is high collateral, stochastic, self-reliance [19, 20]. To improve the SVM, ECSVM optimized by GA are given in Figure 5.

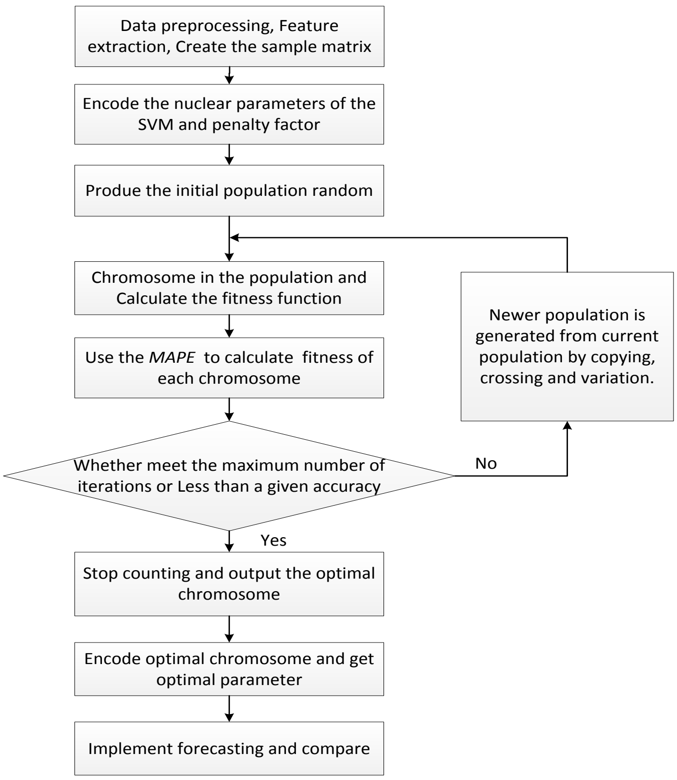

FIGURE 5 Flow chart

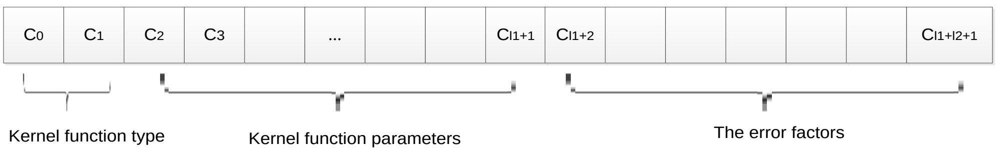

FIGURE 6 Decoding process

Step 1: In PV power generation data processing, the programme checks and modifies the bad-data.

Step 2: The factors which affect PV power generation multiple. The main factors, such as solar radiation intensity, sunshine time, and temperature, are considered in the paper.

Step 3: Consider the couple role of above factor. We can get an input matrix.

Step 4: The training set data and test data consist of kernel function and penalty factor, which is produced by a population of randomly generated each individual genes in the string of decoding, were input into model for training and simulation. Decoding process as show in fig. (6).

Step 5: The mean absolute percentage error of test sample *MAPE* is considered mean function to evaluate the population:

$$MAPE = \frac{1}{N}\sum_{k=1}^{N}\frac{\left|p_{\mathrm{simu}.k} - p_{trag.k}\right|}{p_{trag.k}} \times 100\% , \quad (12)$$

Step 6: The next generation of individuals were choose based on the individual fitness and selection principle.

Step 7: Selection, crossover and mutation operations were used to produce the next generation.

Step 8: judging whether the terminal condition is meet, if meet, turn into the next step, otherwise go to step 5.

Step 9: The best individual were exported and considered as its proximate optimum solution in this problems.

Step 10: Then, the forecast is realized by the use of Genetic algorithm that used to derive the corresponding kernel function , kernel function parameter and error warning factor.

Step 11: MAPE is used to predict and evaluate the performances, and last, they were compared with the traditional methods.

5 Case study

PV monitoring system historical electricity generation data and environmental parameters are included in the sample data of prediction system. Experimental samples are selected from Nanchang University rooftop photovoltaic power generation data, which recorded for 5 min interval. Foe the training sample the period from August 1 2011 to August 6 2011 was selected. The data from August 7 2011 are tested as forecast samples.

A quick look at the Table 2 above indicates that forecast accuracy and training accuracy of GA-ECSVM is better than ECSVM. Forecast effect of GA-ECSVM is given in Figure 7, which has high accuracy of prediction. In order to verify the superiority of the GA-ECSVM, Comparative forecast effect of SVM, ECSVM and GA_ECSVM can help rank forecast accuracy. The error of measurement using different methods are compared in Figure 8.

From Figure 8 the accuracy of prediction of GA-ECSVM model is higher than SVM and ECSVM. In order to calculate the errors between predictive value and the actual value accurately, mean absolute error method are proposed.

From Table 3 we know that ECSVM model, which has optimized the GA is more exactitude. The relative prediction error precisions are all below 5%.

TABLE 2 Contrast GA-ECSVM with ECSVM predictions

Experiment	SVM parameters			Training accuracy (%)	Forecast accuracy (%)
	Kernel function type	**Kernel function parameter *r***	**Penalty factor**		
Experience choice	1		1	62.5	86.67
	1		10	72.5	83.33
	2	1	10	95	90
	2	10	1	86	56.67
	3	1	10	95	86.67
	3	10	1	94	73
	4	1	10	57.5	73.33
	4	10	1	52.5	33.33
GA optimization	2	7.375	11.91	98	96

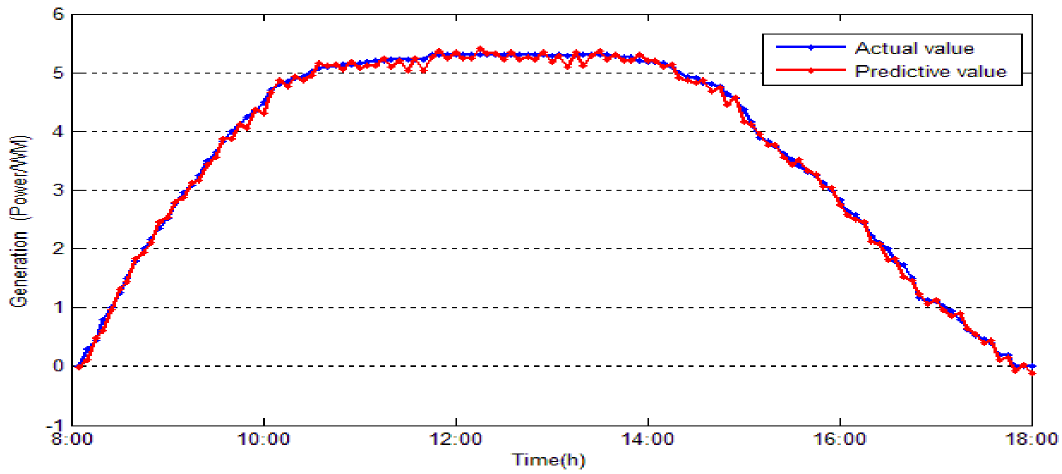

FIGURE 7 The power generation forecasting

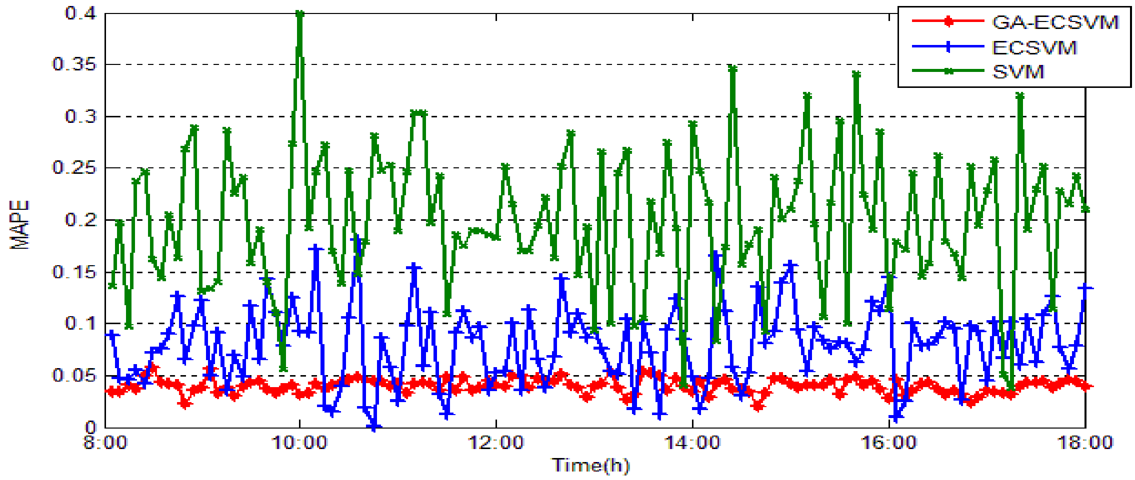

FIGURE 8 Forecasting error curve

TABLE 3 Forecasting error

	GA-ECSVM	ECSVM	SVM
MAPE (%)	4.7984	9.6255	15.3654

6 Conclusions

Power forecasting has received a great deal of attention due to its importance for planning the operations. In this paper, the GA-ECSVM model is proposed to forecast the output of PV power generation. Firstly, an error correction SVM method (ECSVM) is used to calculate the open integration of photovoltaic power storage system in advance or after the time in order to try to eliminate the system error between the predicted and actual values. Meanwhile, using GA to optimize kernel function parameters and the error penalty factor and other parameters in this model, GA-ECSVM model improves portfolio optimization model parameter prediction accuracy. Finally, an application example shows that : the GA-ECSVM model has better learning ability and generalization ability in the short term prediction of photovoltaic power generation than SVM and ECSVM, with the prediction accuracy of 95.2016%.

Acknowledgements

This work is supported by International Science and Technology Cooperation Project (2014DFG72240), and Jiangxi Science Support Project (2013BBE50102).

References

[1] Liu Q 2000 Strategic Choice for National Adjustment of Energy Structure *Advance In Earth Sciences* **15**(2) 154-64

[2] Xue F, Chen K, Wang N 2011 Coordinated Control Frame of Large-scale Intermittent Power Plant Cluster *Automation of Electric Power System* **35**(22) 45-33 *(in Chinese)*

[3] Zhang L, Hurley W G, Wöfley W H 2011 A new approach to achieve maximum point tracking for PV system with a variable inductor *IEEE Transactions on Power Electronics (Special Issue on Microgrids)* **26**(4) 1031-7

[4] Li J, Xu H, Zhao H 2008 Dynamic modeling and simulation of the grid-connected PV power station *Automation of Electric Power Systems* **32**(24) 83-7 *(in Chinese)*

[5] Chakraborty S, Weiss M D, Simoes M G 2007 *IEEE Transactions on Industrial Electronics* **54**(1) 97-109

[6] Yona A, Senjyu T, Funabashi T 2007 *IEEE Power Engineering Society General Meeting* 2007 1-6

[7] Dai Q, Duan S, Cai T, Chen C, Chen Z, Qiu C 2011 Short-term PV generation system forecasting model without irradiation based on weather type clustering Proceedings of the CSEE **31**(34) 28-35 *(in Chinese)*

[8] Lu Jing, Zhai H, Liu C, Wang X 2010 Study on Statistical Method for Predicting Photovoltaic Generation Power *East China Electric Power* **38**(4) 563-7 *(in Chinese)*

[9] Yang D, Wang Y, Jiao Y 2013 Generation forecasting for photovoltaic system based on wavelet neural networks *Renewable Energy Resources* **31**(7) 1-5 *(in Chinese)*

[10] Sharan A M 2009 Efficiency enhancement of stationary solar energy based power conversion systems in Canada *Applied Energy* **86**(9) 1405-9

[11] Dai Q, Duan S, Cai T, Chen C, Chen Z, Qiu C 2011 Short-term PV generation system forecasting model without irradiation based on weather type clustering Proceedings of the CSEE **31**(34) 28-35 *(in Chinese)*

[12] Li Y, Wang R, Wang T, Wu J, Xu Y 2007 Experimental Performance Analysis and Optimization of A Direct Expansion Solar Assisted Heat Pump Water Heater *Acta Energiae Solaris Sinica* 2007 **28**(5) 464-71 *(in Chinese)*

[13] Chen C, Duan S, Cai T 2011 Short-Term Photovoltaic Generation Forecasting System Based on Fuzzy Recognition Transactions of China Electrotechnical Society **26**(7) 83-9 *(in Chinese)*

[14] Armstrong A, Hurley W G 2010 A thermal model for photovoltaic panels under varying atmospheric conditions *Applied Thermal Engineering* **30**(12) 1488-95

[15] Mao M, Gong W, Chang L 2013 Short-term photovoltaic forecasting based on EEMD-SVM combined method. Proceeding of the Chinese Society of Electrical Engineering 2013 **33**(34) 17-24

[16] Fukushi D, Shichiri M, Sugiyama S, Yoshino T, Hagiwara S, Ohtani T 2013 Scanning near-field optical/ atomic force microscopy detection of fluorescence in situ hybridization signals beyond the optical limit. *Exp. Cell Res* **289** 237-44

[17] Chapelle O, Vapnik V, Bousquet O, Mukherjee S 2002 Choosing Multiple Parameters for Support Vector Machines *Machine Learning* **46**(1) 131-59

[18] Huang C, Lee Y, Lin D, Huang S 2007 Model selection for support vector machine via uniform design *Computational Statistics & Data Analysis* **52**(1) 335-46

[19] Saini L M, Aggarwal S K, Kumar A 2010 Parameter optimisation using genetic algorithm for support vector machine-based price-forecasting model in National electricity market *IET Generation, Transmission & Distribution* **4**(1) 39-49

[20] Cheng G, Wang H, Sun X 2012 Model selection for SVM classification based on kernel prototype and adaptive genetic algorithm *Journal of Graduate University of Chinese Academy of Sciences* **29**(1) 62-9

Authors

Yunjun Yu, born in June, 1978, Jiangxi, China

Current position grades: Doctor, lecturer at Nanchang University, Nanchang, Jiangxi, China.
University studies: Automation in Nanchang University. Ph.D degree at Chinese Academic of Science, 2013.
Scientific interest: photovoltaic inverter, fault diagnosis, micro-grid.
Publications: 30 Papers.

Yuntao Xue, born in December, 1991, Henan, China

Current position, grades: Junior student at Nanchang university, Jiangxi, China.
University studies: Nanchang University, Jiangxi, China
Scientific interest: power system and its automation.

Sui Peng, born in November, 1992, Jiangxi, China

Current position, grades: Junior student in Nanchang university, Jiangxi, China.
University studies: Nanchang University, Jiangxi, China.
Scientific interest: power system and its automation

Tong Chao, born in July, 1992, Jiangxi, China

Current position, grades: Junior student in Nanchang university, Jiangxi, China.
University studies: Nanchang University, Jiangxi, China
Scientific interest: power system and its automation

Ziheng Xu, born in September, 1992, Nanchang, China

Current position, grades: Junior student in Nanchang University, Jiangxi, China.
University studies: Computer science and technology in School of Electronics and Information Engineering.
Scientific interest: optimization and application of the algorithm.