SICE Annual Conference 2012
August 20-23, 2012, Akita University, Akita, Japan

Short-term Photovoltaic Prediction by using $\mathcal{H}_\infty$ Filtering and Clustering

Yasuhiko Hosoda[1] and Toru Namerikawa[1]

[1]Department of System Design Engineering, Keio University, Kanagawa, Japan
(Tel: +81-45-563-1151; E-mail: yasuhiko@nl.sd.keio.ac.jp, namerikawa@sd.keio.ac.jp)

Abstract: This paper deals with prediction algorithm applying for photovoltaic (PV) systems in smart grid. This prediction is aim to predict the amount of the next day of generation using the previous data and the weather forecast which get from Japan Meteorological Agency. The procedure of prediction consists of two steps, the data processing and the unknown parameters estimation. In the data processing, our proposed method considers the characteristics of PV generation using cluster ensemble. We propose the cluster ensemble based on k-means to choose the groups with a correlation with previous data. In the unknown parameters estimation, we provide the regression model for PV generation and the unknown parameters are estimated via $\mathcal{H}_\infty$ filtering. The effectiveness of the proposed prediction method is demonstrated through numerical simulations.

Keywords: PV, Short-term, Prediction, Smart Grid, Clustering, k-means, Estimation, $\mathcal{H}_\infty$ Filtering

1. INTRODUCTION

In recent years, energy and environmental problem have become the hottest worldwide social problems. Therefore, the energy technologies such as smart grid, micro grid and sensor networks are interested and growing all over the world. For example, it is required in the smart grid (see Fig. 1) that different power generators (photo-voltaic generator, wind farm, fuel cell, micro gas turbine, battery and so on) connected with each other and cooperative in the energetically and environmentally. At the same time, it is well known that PV system is one of the best solution for such an environmental and energy problems. The problem in operational management of PV system is reverse power flow to the power system. The reverse power flow adversely affects the hole of power system. Moreover, from the point view of improving the control performance of power system, generation power from PV system should be estimated as accuracy as possible. Thus, the predict amount of power in next day is essential to stably operation for PV system [1-3].

A great deal of efforts to developed application of prediction, regression prediction techniques for time series data like PV generation data has been studied for a long time. In machine learning, the regression using Kernel method has been developed. Support Vector Machine (SVM) is a kind of the Karnel method and they are used in PV prediction [4]. The neural networks have been investigated for the PV prediction and it is easy to be applied for PV forecasting [3, 5]. On the other hand, a linear regression is one of the simplest method and they are used in various situations. It is expansible to treat a nonlinear regression, and they have been successful in past many problems. Our proposed method employs the linear regression model and in their estimation via $\mathcal{H}_\infty$ filtering which make it possible to estimate more accurately against initial errors, disturbances and noises.

The main objective of this paper is to present the algorithm for PV system prediction which can be useful for the amount of generation in the next day. It is well known that the PV generation has a specific characteristics and difficult to predict accuracy. Thus, the prediction needs to consider this characteristics of PV generation. Hence, the prediction algorithm has two steps, the data processing and the unknown parameters estimation. In the data processing, our proposed method considers the characteristics of PV generation using cluster ensemble. At the same time, the PV system has been influenced by cell temperature. Hence, there have 2-dimensional data temperature and generation data. On the other hand, representative cluster ensemble k-means method is not be able to separate the data stably because their algorithm select the initial value randomly. Our proposed method also has 2-dimension, hence we modeled PV system, and treat the 2-dimensional data as 1-dimension. After that, we propose the cluster method based on k-means considering initial value. In estimation, we provide the linear regression model for PV generation. The unknown parameters are estimated via $\mathcal{H}_\infty$ filtering and simulations were held and got good results in prediction.

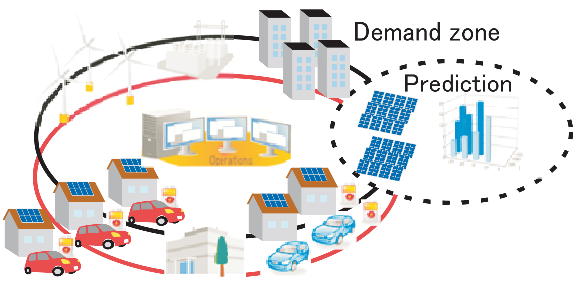


Fig. 1 Smart Grid

2. PROBLEM FORMULATION

2.1 Modeling of PV System

In this subsection, we present a general plant model for PV system as a space case. The efficiency of PV generation is effected by the cell temperature [6-8]. Let us define the efficiency of power generation $\eta_k^d(t_k^d)$ in day d at time k.

$$\eta_k^d(t_k^d) = \eta_0(1-\zeta(T_r - t_k^d)) \tag{1}$$

where ζ is constant and which is given in the coefficient of a cell temperature and battery, [1/K]. t_k^d is cell temperature, [K]. T_r is the reference temperature which is 298[K], η_0 is the efficiency of power generation at T_r [4, 9]. The equation (1) means that the power generation efficiency gets lower if the temperature rises.

The following assumption is set to design the prediction model and the processes of previous data.

Assumption 1: The classified data is correlated with each other.

Assumption 2: The weather forecast will be announced in three time intervals up to 24 hours ahead by the Japan Meteorological Agency, sunny, cloudy, rainy pattern classification by three patterns are known.

The generated power y_k^d is defined in the regression model, equation (2).

$$y_k^d = a_0 + \sum_{i=1}^{p} a_k^{d-i} y_k^{d-i} \frac{\eta_k^d}{\eta_k^{d-i}} + v_k^d \tag{2}$$

where $y_k^d \in R^1$ is composed of the previous measured data y_k^{d-i}, a_0 is the error of the model , η_k^{d-i} is efficiency of power generation and measurement noise is assumed to be zero mean white Gaussian noise. This model (2) is configured on the assumption 1. The process dynamics of plant is given in equation (3).

$$x_{k+1} = A_k x_k + w_k \tag{3}$$

where $x_k \in \mathbb{R}^{\,p}$ is the unknown parameters as seen in equation (5), the process noise $w_k \in \mathbb{R}^{\,p}$ is assumed to be zero mean white Gaussian noise. The matrix $A_k \in \mathbb{R}^{\,pp}$ represents efficiency of the power generation, seen in equation (6). The measurement equation on day d at time k is given in equation (4).

$$z_k = C_k x_k + v_k \tag{4}$$

where $z_k \in \mathbb{R}^{\,1}, v_k \in \mathbb{R}^{\,1}$ are the measurement output of PV. Additionally, $C_k \in \mathbb{R}^{\,1p}$ represents the previous PV data as seen in equation (7). The detail of above matrix is indicated in following equations (5)-(7).

$$x_k = \left[a_0, a_k^{d-1}, \cdots, a_k^{d-p}\right]^T \tag{5}$$

$$A_k = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \frac{\eta_k^d}{\eta_k^{d-1}} & & 0 \\ 0 & & \ddots & \\ 0 & 0 & & \frac{\eta_k^d}{\eta_k^{d-p}} \end{bmatrix} \tag{6}$$

$$C_k = \left[1, y_k^{d-1}, \cdots, y_k^{d-p}\right] \tag{7}$$

The purpose of the prediction is to schedule the next day of power generation efficiency and to know the amount of power generation. Let us define the evaluation of error as in equation (8).

$$J = \sum \mid y_k^d - \hat{y}_k^d \mid / \sum y_k^d \tag{8}$$

Now, we deal with the prediction problem of this system as estimation of unknown parameters problem.

2.2 Prediction Algorithm

To solve this estimation problem, process of estimation algorithm is described as in Fig. 2. There have two processes, cluster ensemble for measured previous data and unknown parameters estimation. In estimation, we assume that the model has a correlation with previous data. However, PV generation depends on temperature and weather as in Figs. 3-4. Fig. 3 shows the power generation data for 7 days and Fig. 4 shows the characteristics of PV generation. The shape of generation data looks identical, but difference between the amounts of generation is too big. Therefore our prediction algorithm consists of two steps, $\mathcal{H}_\infty$ filtering and improved cluster analysis based on k-means method. In prediction algorithm, we utilized the generation data and the next day's weather forecast which can be obtained from Japan Meteorological Agency. The detail of prediction algorithm is described in the following section.

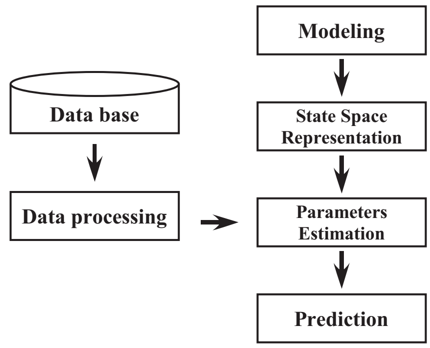


Fig. 2 Prediction Process

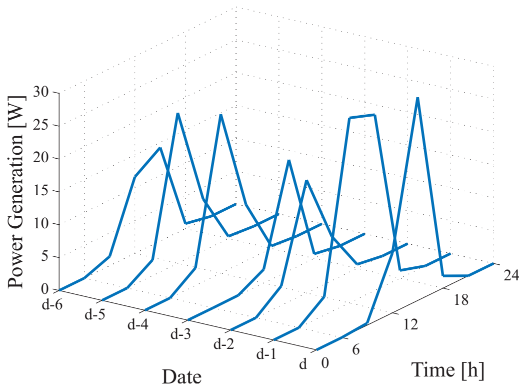


Fig. 3 Power Generation data for 7 days

3. PARAMETER ESTIMATION

This section, We account for how to estimate unknown parameters. Then, the component of parameters estimation is described. In the process of prediction, The credibility of prediction result is strongly required. However, it is impossible to ensure the full reliability because the prediction is treated in the future. Accordingly, the parameters estimation is employed as a kind of an index

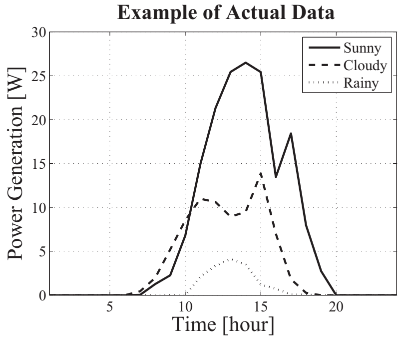


Fig. 4 Characteristics of PV Generation

of prediction. It is able to know how the unknown parameters changes. If they converge to the constants, the prediction should be expected to be succeed. Parameters estimation leads to reliability of estimation.

The step of estimation in the model describes as the following. An example of the model can be described as in equation (9).

$$y_k^d = a_k^{d-1} y_k^{d-1} \frac{\eta_k^d}{\eta_k^{d-2}} \tag{9}$$

This model which indicates that the generation y_k^d is composed the day before prediction day and the coefficient normalized by the temperature. The first step, the initial value of unknown parameters set, in this case the initialize number is set as $a_k^0 = 1$.

$$y_k^2 = a_k^1 y_k^1 \frac{\eta_k^2}{\eta_k^1} \tag{10}$$

$$y_k^3 = a_k^2 y_k^2 \frac{\eta_k^3}{\eta_k^2} \tag{11}$$

$$\vdots$$

$$y_k^n = a_k^{n-1} y_k^{n-1} \frac{\eta_k^n}{\eta_k^{n-1}} \tag{12}$$

The step from equations (11) to (12), filtering theory can be used, get the most likely value of the unknown parameters and then the next step y_k^4 will be calculate. This processes calculate repeatedly and get the unknown parameters.

3.1 Estimation algorithm based on $\mathcal{H}_\infty$ filtering

To estimate unknown parameters, we propose to use estimation algorithm based on $\mathcal{H}_\infty$ filtering which is constrained to solve the estimation of unknown parameters problem. This filter make it possible to estimate it accurately against modeling error margin. Like PV prediction, a lot of indefinite factors make precision of prediction bad. Thus we use $\mathcal{H}_\infty$ filter for estimating parameters. The unknown parameters matrix x_k can be treated as in the following minimax problem. Let design an estimator to satisfy in the equation (13).

$$\sup_{x_0, v} \frac{\sum_{k=0}^{N} \|x_k - \hat{x}_k\|^2}{\|x_0 - \hat{x}_0\|^2_{P_0^{-1}} + \sum_{k=0}^{N} \|w_k\|^2_{W_k^{-1}} + \sum_{k=0}^{N} \|v_k\|^2_{V_k^{-1}}} < \gamma^2 \tag{13}$$

where P_0 is a weight matrix for the uncertainty of initial state x_0, W_k is weight matrix for the state noise w_k, V_k is a weight matrix for the observation noise v_k and $\hat{x}_k$ means the estimated value of x_k. According to equation (13), the maximum value of the ratio of the difference between true value x_k and estimated value $\hat{x}$ and the difference between true value x_k and estimated value $\hat{x}$ plus noise energy can be refrained under γ^2. Namely, this equation make the ratio of the energy and noise energy estimation error which are all the noise energy is bounded refrain under certain value. In generally, it is impossible to get optimal solution in Left-hand side of equation (13) and get minimize about $\hat{x}$. Hence, we will find sub-optimal solution under equation (13). The γ need to set under appropriate value so as not to choose too small and can not be found the sub-optimal solution.

The $\mathcal{H}_\infty$ filter algorithm is represented as in equations (15)-(20).

$$\begin{aligned}
\hat{x}_{k+1|k} &= A_k \hat{x}_{k|k} && (14)\\
\hat{z}_{k+1} &= C_k \hat{z}_k && (15)\\
S_k &= z_k - \hat{z}_k && \\
&= C_k P_{k|k} C_k^T + V && (16)\\
K_k &= P_{k|k} C_k^T S_k^{-1} && (17)\\
x_{k+1|k} &= \hat{x}_{k+1|k} + K_k \left(z_k - C_k \hat{x}_k\right) && (18)\\
P_{k+1|k} &= A_k P_{k|k} \psi_k^{-1} A_k^{-1} + W && (19)\\
\psi_k &= I + \left(C_k^T V^{-1} C_k - \gamma^{-2} I\right) P_{k|k} && (20)
\end{aligned}$$

Above estimation algorithm for solving the sub-optimal solution and unknown parameters which satisfied equation (13) can be calculated according to equations (15)-(20) repeatedly. Furthermore, our proposed algorithm considers the effect of state dependent noises and responds to predicted estimate value sharing algorithm. First, we predict the next step based on the model and show as in equations (15), (16). Second, the filter gain can be calculated to get the optimal weight for the observation error and the estimation error as in equations (18), (19). Third, covariance matrix will be updated as in equation (20).

3.2 Cluster ensemble based on k-means method

In this subsection, we propose cluster ensemble which can be used to consider the characteristics of PV generation. The purpose of clustering is to make the previous data satisfy the Assumption 1. Clustering is an effort to classify similar objects in the same groups. In original k-means, it's enable to treat n-dimension data and they can be used in any other problems. But there have the initial value problem, they choose the initial value randomly [13, 14]. Thus sometimes they can not separate relevant. Therefore, our proposed algorithm improve this problem. Our proposed cluster analysis is based on k-means considering initial value to extract the previous generation data that are correlated with each other. The feature of this clustering constrained for 1-dimension and overcome the initial value problem.

Let the group $\mathbf{C_1}, \ldots, \mathbf{C_k}$ be a set of partitions $\mathbf{k}$ of a data set $\mathbb{Z} = \{Z_1, \ldots, Z_n\}$ which is of 1-dimension. The details of the clustering algorithm can be described as follows.

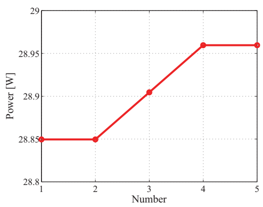
1. Given is a data set $\mathbb{Z} = \{\mathbf{Z}_i \mid i = 1, \ldots, n\}$, where n is the size of $\mathbb{Z}$.
2. Arranged the data in order of size. Then divided into $\mathbf{k}$ arranged data groups. This initial groups $\mathbb{C} = \{\mathbf{C}_j \mid j = 1, \ldots, k\} \in \mathbb{Z}$ is generated.
3. Calculate $\mathbf{c_j}$ which is defined as the center of $\mathbf{C_j}$.
4. Calculate the distance of $d_j(\mathbf{C_j})$.
$$d_j(\mathbf{C_j}) = \| \mathbf{E}[\mathbf{C_j}] \| \tag{21}$$
Provided that E means expectation.
5. Replace to minimize d_i between groups and assign to a group.
6. Go back to Step.3, or stop.

First, the data is given and set the number of groups and the Assumption 2 provides the number of partition $k = 3$. Second, to overcome initial value problem in original k-means, it is necessary to choose the proper initial value. This algorithm selects the initial value in an appropriate manner. Third, the center of gravity and the distance from the center of gravity are calculated, after that, the data are belonged to the group in accordance with the rules. Finally, we calculate the distance the center of gravity and replace the groups themselves repeatedly. This clustering ensemble makes the group which has minimize variance of the data.

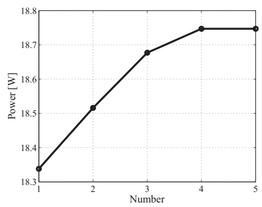
4. PREDICTION

4.1 Result of the Data Processing

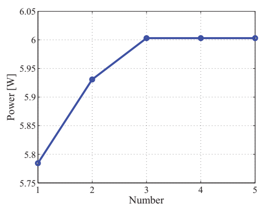
Fig. 5 shows the changes in the position of gravity. In the data processing, the position of the center of gravity is calculated repeatedly and this simulation indicates that each clustering repeat about five times to get classified groups. Fig. 6 shows the result of clustering. We pro-



[1] Sunny



[2] Cloudy



[3] Rainy

Fig. 5 Changes in position of the center of gravity

cess the data which sampled 3 hours because the weather forecast can be obtained every 3 hours. From this result, the data is divided and the group is selected uniquely.

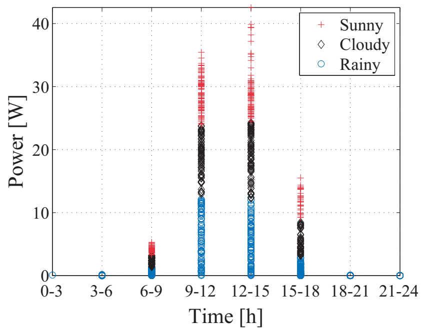


Fig. 6 Result of Clustering

4.2 Results of The Unknown Parameters Estimation

After the data processing and modeling. The predictive calculation was carried out for 20 days. The length of data which are estimated the unknown parameters is depend on their groups, because the data parcel out the weather group after the clustering. To consider calculation simple and the number of repeat data, the model is set as the following equation (22).

$$\begin{aligned} y_k^d &= a_k^{d-1} y_k^{d-1} \frac{\eta_k^d}{\eta_k^{d-2}} + a_k^{d-2} y_k^{d-2} \frac{\eta_k^d}{\eta_k^{d-2}} \\ &\quad + a_k^{d-3} y_k^{d-3} \frac{\eta_k^d}{\eta_k^{d-3}} \end{aligned} \tag{22}$$

The unknown parameters $a_k^{d-1}, a_k^{d-2}, a_k^{d-3}$ are estimated by $\mathcal{H}_\infty$ filtering. The matrix of plant in the model (22) are set in the following equations (23)-(25).

$$x_k = \left[a_k^{d-1}, a_k^{d-2}, a_k^{d-3}\right]^T \tag{23}$$

$$A_k = \begin{bmatrix} \frac{\eta_k^d}{\eta_k^{d-1}} & 0 & 0 \\ 0 & \frac{\eta_k^d}{\eta_k^{d-2}} & 0 \\ 0 & 0 & \frac{\eta_k^d}{\eta_k^{d-3}} \end{bmatrix} \tag{24}$$

$$C_k = \left[y_k^{d-1}, y_k^{d-2}, y_k^{d-3}\right] \tag{25}$$

The initial condition need to set the state noise and the measurement noise. There noises are set as in the following equations (26), (27).

$$W = 10^{-7} I_3 \tag{26}$$

$$V = 15 \tag{27}$$

Generally, the design parameter γ should be set as in smaller. But actually, too smaller designed parameter γ makes bad prediction. We design this parameter as in simulation repeatedly and decide $\gamma = 8.5$.

Fig. 7 shows the result of the unknown parameters $a_{12-15}^{d-1}, a_{12-15}^{d-2}, a_{12-15}^{d-3}$ estimation. At the same time, Fig. 8 shows determinant of covariance P. From these results show that each parameter estimation well converges to constant value. Final value of estimated unknown parameters will be assigned to the model and calculated. The magnitude of unknown parameters indicate the a correlation and when they approach magnitude 1, they have a strong coefficient. Furthermore, trace P is converged when they repeat 40 times. Therefore, it becomes possible to estimate by repeating the calculation of 40 times.

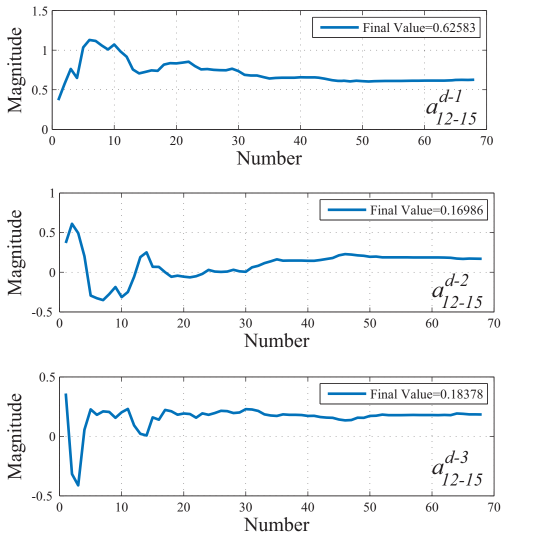


Fig. 7 The result of the Unknown Parameters $a^{d-1}_{12-15}, a^{d-2}_{12-15}, a^{d-3}_{12-15}$ Estimation

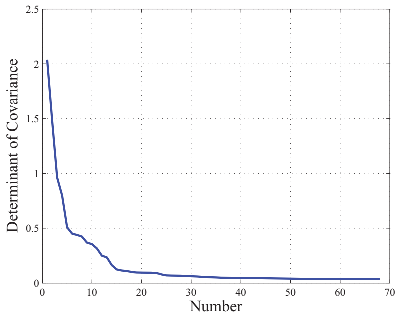


Fig. 8 Determinant of Covariance

4.3 Prediction Result and Discussion

This subsection shows prediction result and discussion. The predicted results are shown in Figs. 9-10. Fig. 9 shows an example of result for one day. The sampling time of prediction held 3-hours and it expresses that our proposed prediction enable to predict with high accuracy. Fig. 10 shows the result of prediction for 20 days and the average accuracy of prediction was about 13.9[%]. In the same condition, the errors of short term PV power output forecasting are in the range of 15-20 [%] [15]. In this result, our proposed method is also effective in PV prediction.

5. CONCLUSION

The short term prediction of PV system generation is considered. We deal with the prediction algorithm which treat not only the processes but also the modeling of the PV systems. The proposed prediction algorithm consists of the two steps, the past data processes and the unknown parameters estimation. In the past data processes, the past

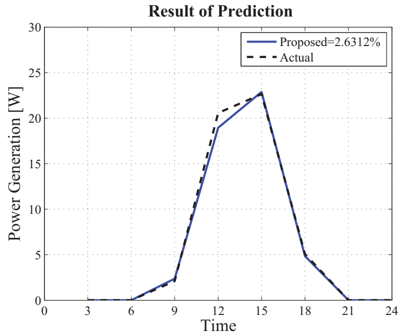


Fig. 9 Result of prediction

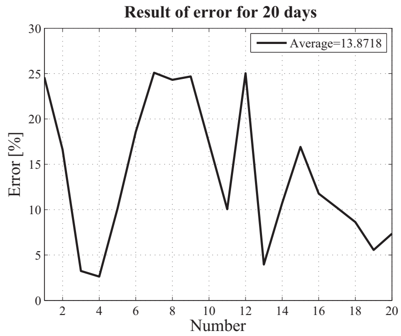


Fig. 10 Result of prediction for 20 days

data is divided according to its characteristics automatically using the cluster ensemble based on k-means. In the parameters estimation, we design the PV generation model which is considered the efficiency of generation effect from temperature, and estimates unknown parameters via $\mathcal{H}_\infty$ filtering. The simulation results show the effectiveness of proposed algorithm.

REFERENCES

[1] G. S. Kinsey, K. Stone, J. Brown and V. Garboushian, "Energy prediction of amonix cpv solar power plants", *Press in Photovoltaics: Research and Application*, 2010.

[2] T. C. Yu and H. T. Chang, "The forecast of the electrical energy generated by photovoltaic systems using neural network method", *Electric Information and Control Engineering (ICEICE), International conference*, pp.2758-2761, 2011.

[3] M. Kawaguchi, S. Ichikawa, M. Okuno, and T. Jimbo, "Predict for electric power generation of solar cell using the neural network", KES, pp.387-392, 2006.

[4] J. Shi, W. J. Lee, Y. Liu, Y. Yang and P. Wang, "Forecasting Power Output for Photovoltaic System Based on Weather Classification and Support Vector Machine", *IEEE Trans. Industrial Application*, Issue: 99, 2011.

[5] A. Mellit and A. M. Pavan, "A 24-h of solar irradiance using artificial neural network: Application for Performance prediction of a grid-connected PV plant at Trieste, Italy", *ELSEVIER. Solar Energy*, pp. 807-821, 2010.

[6] F. M. G. Longatt, ”Model of Photovoltaic Module in Matlab”, *CIBELEC*, 2006.
[7] G. N. Tiwari and S. Dubey, ”Fundamentals of Photovoltaic Modules and Their Applications”, *Centre for Energy Studies, Indian Institute of Technology (IIT)*, RSC publishing, pp. 99-100, 2010.
[8] H. Tsai, C. Tu, and Y. Su, ”Development of Generalized Photovoltaic Model Using MATLAB / SIMULINK”, *Proceedings of the World Congress on Engineering and Computer Science*, pp. 22-24, 2008.
[9] G.Notton, C. Cristofari, M. Mattei and P. Poggi, ”Modelling of a double-glass photovoltaic module using finite differences”, *Appl. Therm Eng*, 25, 2854-2877, 2005.
[10] W. Xiao, N. Ozog, W. Dunford and G. Topology, ”Study of Photovoltaic Interface for Maximum Power Point Tracking”, *IEEE Trans. Industrial Electronics*, vol. 54, no. 3, pp. 1696-1704, 2007.
[11] H. Patel and V. Agarwal, ”Maximum Power Point Tracking Scheme for PV Systems Operating Under Partially Shaded Conditions”, *IEEE Transactions on Industrial Electronics*, Vol. 55, No. 4, pp. 1689-1698, 2008.
[12] M. Azab, ”Optimal Power Point Tracking For Stand-Alone PV System Using Particle Swarm Optimization”, *IEEE Int Symposium on, in Industrial Electronics (ISIE)*, pp. 969-973, 2010.
[13] D. Steinley and M. J. Brusco, ”Initializing k-means Batch Clustering: A Critical Evaluation of Several Techniques”, *Journal of Classification*, Vol.24, No.1,pp.99-121, 2007.
[14] J.B.MacQueen, ”Some methods of classification and analysis of in multivariate observations”, *Proc. of 5th Berkley Symposium on Math. Stat. and Prob*, pp.281-297, 1967.
[15] A. Yona, T. Senjyu, A. Yousuf Saber and T. Funabashi, ”Application of Neural Network to 24-hour-Ahead Generating Power Forecasting for PV System”, *Power and Energy Society General Meeting Conversion and Delivery of Electrical Energy in the 21st Century, IEEE*, pp.281-297, 2008.