Artificial Neural Network Approach to Photovoltaic System Power Output Forecasting

Tatjana Konjic
Institute for Systems and Computer Engineering of Porto (INESC Porto)

Amra Jahic
Helmut Schmidt University / University of the Federal Armed Forces Hamburg

Artificial Neural Network Approach to Photovoltaic System Power Output Forecasting

Tatjana Konjić
Faculty of Electrical Engineering
University of Tuzla
Franjevacka 2, 75000 Tuzla, Bosnia and Herzegovina
tatjana.konjic@untz.ba

Amra Jahić
Graduate School
University of Applied Sciences Offenburg
Badstr. 24, 77652 Offenburg, Germany
ajahic@stud.hs-offenburg.de

Jože Pihler
Faculty of Electrical Engineering and Computer Science
University of Maribor
Smetanova 17, 2000, Maribor, Slovenia
joze.pihler@um.si

Abstract **- Installed global capacity of grid connected photovoltaic (PV) power plants is increasing each year. In order to successfully integrate these power plants into the system, power output forecasts need to be more accurate. Several approaches to forecasting have been developed throughout previous years, from physical methods using weather forecasts and satellite images to purely statistical, numerical methods and lately even artificial intelligence. In this paper, models based on different artificial neural network (ANN) architectures, different training algorithms and different time horizon of training data set are used to forecast the hourly power output of real PV system. This paper explores the accuracy of those models. A confidence interval around forecasted curve has been determined.**

Key words - Artificial Neural Network, PV system, power output forecasting, confidence interval

I. Introduction

2013 was yet another historic year for PV power plants, with over 38.4 GW of newly-installed capacity worldwide. The total global capacity is now 138.9 GW and it puts PVs right behind hydro and wind as the most commonly used renewables [1]. Taken into consideration that all global policies clearly promote solar energy and that it is on the verge of becoming fully competitive in so many markets, it is clear that the global installed PV capacity can only go higher. Penetration of PVs into national power systems is now inevitable and it raises a question of successful integration. Only in 2013, 11 GW of PV capacity was connected to the grid in Europe. PV now covers 3% of the electricity demand in Europe and 6% of the peak electricity demand in Europe [1].

PV production is variable and unpredictable and therefore it can be a threat to system stability, especially if that system has a large share of renewables [2]. One of the solutions for better integration of PV power plants in the power system is forecasting their power output.

II. PV Forecasting

The field of PV forecasting is evolving very fast and today we can talk about a variety of different types of models and approaches, depending on the final use of predictions. Forecasts can refer to a single PV power plant, a group of them in a wider geographical area or to a different time horizons. Intra-day forecasts, specially 0-6 hours ahead, and day ahead forecasts are crucial for successful grid integration [3]. According to them system operator can organize the work of the complete production portfolio and maintain the needed balance in the system. Longer forecasts, such as week, season or a year ahead, are no less important and allow utilities, distributors and system operator to make long-term plans.

Forecasting methods can generally be categorized as physical or statistical, depending on the type of used input information [3]. The physical approach uses weather forecasts, PV models and characteristics, whereas the statistical approach relies primarily on past data and numerical methods, with little or no reliance on PV models [4]. Although physical methods, such as Total sky imagery and Satellite imagery show great success in short-term forecasts (up to one day ahead) longer forecasts cannot be done without the use of statistical methods [3].

Most of the published papers in the early stage of PV forecasting were dealing with solar radiation forecasting [5]-[7]. As the number of connected PV power plants in the power system increased, research interests focused more in development of various models for the PV power output forecasting. Models based on advanced techniques, such as artificial neural networks [8]-[10], support vector machine [11], [12] or different hybrid models [13]-[15] have shown success in this field. A comparative study on performance of different forecasting techniques for forecasting the power output of grid-connected PV systems was presented in [16]-[19].

III. Artificial Neural Networks

Artificial Neural Network (ANN) is a form of artificial intelligence mimicking biological neural system, human brain in particular. This self-adaptive method has shown great pattern recognition capabilities. When given the data, ANN manages to learn from examples and form functional relations among data, which is just like the process of human learning. Therefore ANN is suitable for all kind of forecasting [20]-[21]. Just like the network of biological neurons in a human brain, ANN consists of a number of interconnected processing elements called artificial neurons. Each of these neurons receives an input signal, processes it through activation function and then passes the output signal to another neuron.

Many models of ANNs have been proposed throughout history but the most commonly used one today is Multi-Layered Perceptrons (MLP) [22]. MLP consists of several layers of neurons which could generally be classified as input layer, hidden layer(s) and output layer (Fig. 1). Number of layers, number of neurons in each layer, activation functions, as well as the learning algorithms, all depends on a particular problem and needs to be properly chosen and adjusted. The input information (data) passes from nodes in input layer through nodes in hidden layer(s) to node(s) in output layer. The nodes from different layers are interconnected among each other. These connections are characterized by *weight*, parameter which indicates the strength of the connections. Network "learns" by changing the weights. Procedure, through which the weights are adjusted, is called learning or training algorithm [23].

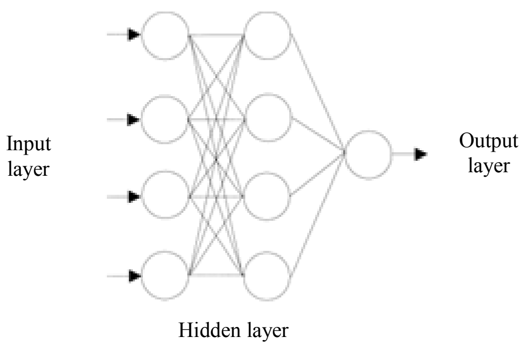

Fig. 1. Multi-Layered Perceptrons

Available database could be divided in three sets (training data set, validation data set and test data set). In case the available database does not contain data of a long historical period, the database is usually divided only in the training data set and the test data set. Once the ANN is trained it can predict the output with a set of completely new input data.

Accuracy of developed ANN model is usually assessed by some of the following measures: root mean square error, normalised root mean square error, mean absolute percentage error, mean absolute deviation percentage.

IV. DEVELOPED MODELS

Analysed PV system is installed at the roof of a parking at the Faculty of Electrical Engineering and Computer Science in Maribor, Slovenia (Fig. 2). PV system consists of three fields, each one with installed capacity of 2.5 kW.

All developed ANN models have three inputs (solar radiation, ambient temperature and module temperature) and one output (power output of PV system). Block diagram of the models is given in Fig. 3.

Fig. 2. Analysed PV system

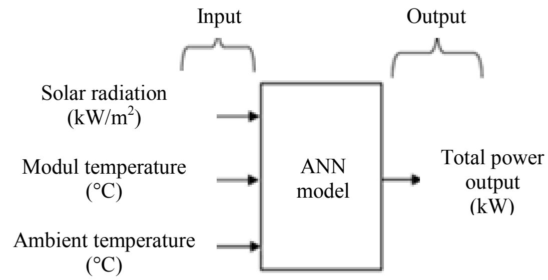

Fig. 3. ANN model used for the forecast

Available database contains hourly data of four previously mentioned parameters from 6 am to 8 pm for years 2012 and 2013. Model developments have been done with two different training data sets. The first training data set contains data from period of May-July in 2012 and 2013. It is 6 months period covering summer season in two consecutive years. Total number of data per observed parameter is 2760. The second one contains data from period of one year, July 2012 - July 2013. Total number of data per observed parameter is 5475. Predictions were done for August 2013. MATLAB was used for simulations.

In order to get as possible as better forecast many different models (more than 50) were tested. Different learning algorithms, activation functions, number of layers and number of neurons have been used in developed models. The following learning algorithms were tested: Levenberg-Marquardt back propagation (*trainlm*), Scaled conjugate gradient backpropagation (*trainscg*), Gradient descent backpropagation (*traingd*), Polak-Ribier conjugate gradient backpropagation (*traincgp*), Fletcher-Powell conjugate gradient backpropagation (*traincgf*), BFGS Newton back propagation algorithm (*trainbfg*) [24]. Levenberg-Marquardt (LM) back propagation learning algorithm gave the far best results and it was further used. Tested activation functions were *logsig*, *tansig* and *purelin*. Number of used layers was up to 3 and number of neurons in the layers was up to 20.

According to the obtained results during many simulations, only 5 models developed on the base of 6 months training data set and 5 models developed on the base of 12 months training data set have been chosen for the following analyses. The main characteristics of the chosen models are shown in Table I.

Table I. Developed models for the study

Training data set	Model	Learning algorithm	Activation Function	Hidden Layers	Neurons
6 months	1	trainlm	tansig	1	4
	2	trainlm	tansig	1	5
	3	trainlm	logsig-tansig	2	4-4
	4	trainlm	logsig-logsig-tansig	3	4-4-4
	5	trainlm	logsig-tansig-tansig	3	4-4-4
12 months	1	trainlm	logsig	1	3
	2	trainlm	logsig	1	4
	3	trainlm	logsig	1	5
	4	trainlm	tansig	1	3
	5	trainlm	tansig	1	10

V. Obtained Results

Hourly forecast of power output was done for each day in August 2013. The assessment of the ANN prediction performance was done by quantifying the prediction obtained on independent data sets. The root mean square error (RMSE), normalised root mean square error (NRMSE) and mean absolute deviation percent (MADP) were used to represent accuracy of the models:

$$\mathrm{RMSE}=\sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(P_{\mathrm{i,m}}-\mathrm{P}_{\mathrm{i,f}}\right)^{2}}\,, \tag{1}$$

$$\mathrm{NRMSE}=100\sqrt{\frac{1}{\mathrm{N}}\sum_{\mathrm{i}=1}^{\mathrm{N}}\left(\frac{P_{\mathrm{i,m}}-\mathrm{P}_{\mathrm{i,f}}}{P_{\mathrm{inst}}}\right)^{2}}\,, \tag{2}$$

and

$$\mathrm{MADP}=\frac{\sum_{\mathrm{i}=1}^{\mathrm{N}}\left|P_{\mathrm{i,f}}-\mathrm{P}_{\mathrm{i,m}}\right|}{\sum_{\mathrm{i}=1}^{\mathrm{N}}P_{\mathrm{i,m}}}\,, \tag{3}$$

where

$\mathrm{P}_{\mathrm{i,f}}$ is an hourly forecasted power output of the PV system, $\mathrm{P}_{\mathrm{i,m}}$ is an hourly measured power output of the PV system, $\mathrm{P}_{\mathrm{inst}}$ is the installed capacity of the PV system, N is the total number of observed hourly power output data in August.

The errors of the models presented in Table I are given in Table II.

According to the results, RMSE, NRMSE and MADP for models based on 6 months database show pretty similar values. Analysing the models architectures it can be noticed that increasing the number of layers and number of neurons in presented developed models, as well as in the other models that are not presented in the paper, does not influence a lot the model's accuracy. However, a little bit better errors were obtained in the case of 3 or 2 layers, while number of neurons should not exceed 4. Logsig and tansig activations functions were adopted for the models while models with purelin activation function did not lead to acceptable results.

Errors obtained for the models based on 12 months database are generally higher than the errors of the models based on 6 months database. Increasing number of layers and numbers of neurons more than 5 per layer does not improve the error. Therefore only models with one hidden layer were chosen for presentation. Again purelin activation function is not acceptable for use.

Using presented models, forecasting was done for each day in August 2013 but due to clearer presentation, only predicted hourly power output of PV system during four days of August was shown in the following figures. Two days (06.08.2013. and 09.08.2013.) were chosen as days with significant weather variations during the day while other two days (01.08.2013. and 12.08.2013.) presented typical summer days in August. Prediction of the power output obtained by the best two models (Model 3 and Model 5) developed on the 6 months database was shown in Fig. 4. Fig. 5 shows predicted curves obtained by Model 1 and Model 3 developed on 12 months data base. Predicted curves deviate more from the real measured curve during peak periods and in the case of sudden change of PV system production due to weather variations.

Table II. Forecasting errors

Training data set	Model	RMSE (W)	NRMSE (%)	MADP (%)
6 months	1	281.39	3.75	8.44
	2	268.42	3.58	8.18
	3	267.08	3.56	7.85
	4	274.96	3.67	8.40
	5	255.95	3.41	7.86
12 months	**1**	273.83	3.65	9.33
	2	291.81	3.75	9.88
	3	281.56	3.83	9.39
	4	287.56	3.89	9.78
	5	314.95	4.20	10.41

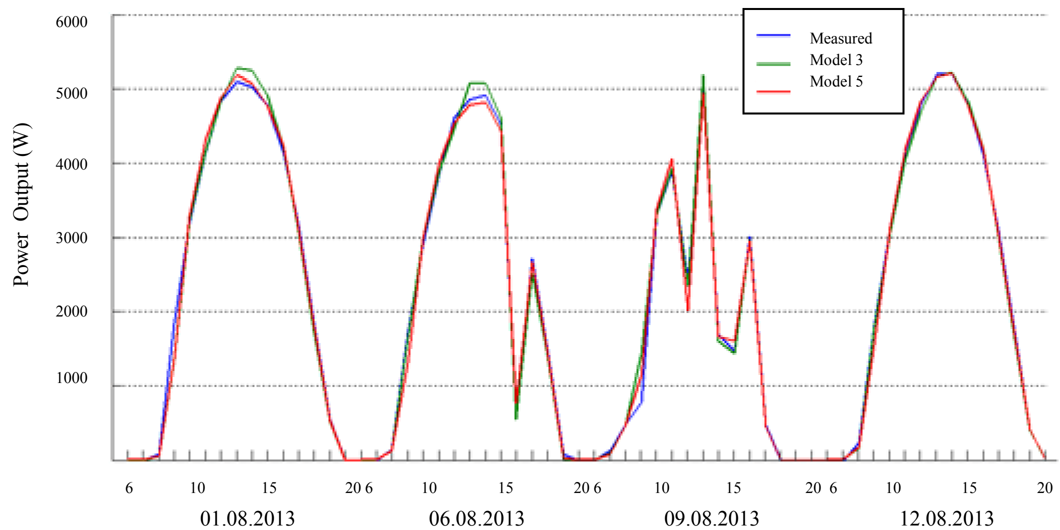

Fig. 4. Forecast with Model 3 and Model 5 (6 months database)

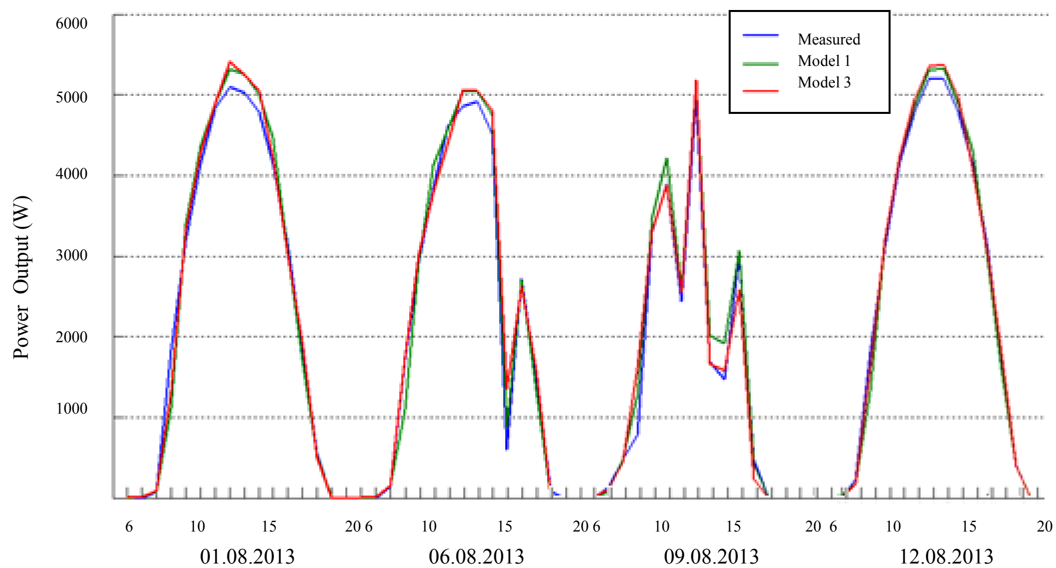

Fig. 5. Forecast for Model 1 and Model 3 (12 months database)

Making comparison of results obtained from numerous reported experiments is not always easy and valid. The forecast accuracy depends on accuracy and availability of data for the model development, chosen test period and chosen measures for model assessment that varies from author to author. Considering the NRMSE of some presented models in [14], the accuracy of the simulations in this paper, taking into account smaller input data set, is acceptable.

VI. Confidence Intervals

One possible approach which we adopted is to define a confidence band around the forecasted curve. This confidence band is not the same for each hour during the day because of the varying forecasting accuracy [25].

Standard deviation of each hour k from 6 am to 8 pm, σ^k, during observed month August, was calculated using the following formula:

$$\sigma^k = \sqrt{\frac{1}{M}\sum_{i=1}^{M}(P_{i,f}^k - \overline{P^k})^2} \,, \tag{4}$$

where,

$P_{i,f}^k$ is the forecasted power output for a certain hour k during each day i in August,

$\overline{P^k}$ is the mean value of forecasted power output of hour k for all M days in August,

M is number of observed days in August (i= 1, 2, …, 31 days).

These standard deviations are applied around the hourly forecasted power output obtained from test data set, and then the confidence intervals of different breadths (e.g. ±0.5σ, ±σ, ±2σ) can be made. Percentage of actual power output points from test data set that fall inside the confidence interval for the best models (Model 3, Model 5) in the case of 6 months database and for the best model (Model 1) in the case of 12 months database is presented in Table III. According to the results presented in Table III it is possible to expect that, with a band factor 1, 93% of actual hourly power outputs of PV system will be inside the confidence interval in the case of using Model 3 with 6 months database.

Confidence interval of different breadths (e.g. ±0.5σ, ±σ, ±2σ) was also calculated for each model from Table I and results were found to be from 88% to 96% for models based on 6 months database and from 82% to 94% for models based on 12 months database.

Fig. 6 shows confidence band P_f±σ, actual and forecasted curve of PV power output for Model 3 (6 months database).

Table III. Percentage of actual power output points that fall inside confidence band of different breadths and the best models for the forecasting period

Data base	Model	P_f±0.5σ (%)	P_f±σ (%)	P_f±2σ (%)
6 months	3	84.67	93.11	97.11
	5	84.22	88.89	92.00
12 months	1	80.89	86.22	91.78

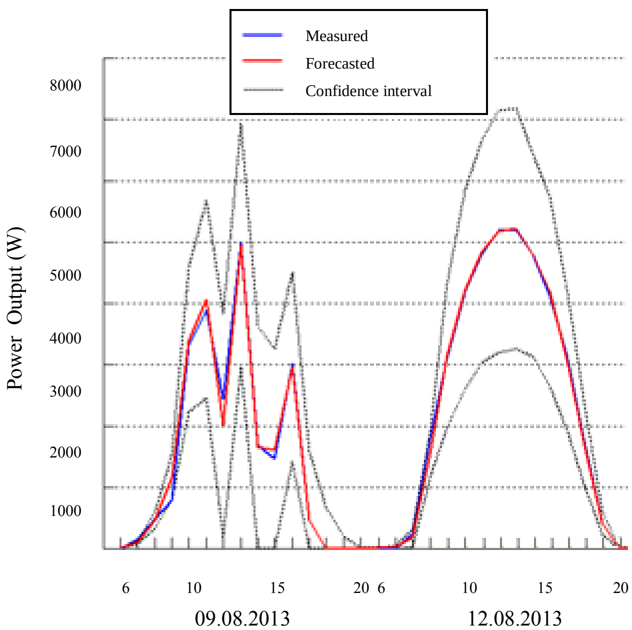

Fig. 6. Confidence band P_f±σ with actual and forecasted curve of PV power output for Model 3 (6 months database)

VII. Conclusion

Photovoltaic power output forecasting became an area of great interest due to fast increase of grid-connected PV systems.

In the paper, different architectures of multilayer perceptrons and training algorithms have been tested in order to forecast the hourly electric power output of PV system in real conditions. Influence of differently formed databases in accordance to available measured historical data of PV power output, solar radiation, ambient temperature and module temperature for model development have been considered, too. Models based on ANN successfully reduced the effect of uncertainty and randomness in input database.

Results demonstrate that simple artificial neural network, such as MLP trained by LM learning algorithm, is able to predict the PV power output with acceptable accuracy. In terms of needed database for model development, it was shown that seasonal database is more suitable than yearly database. In the paper, data of summer seasons was used. However, the same procedure could be repeated for other seasons. Presented models with up to three hidden layers, four neurons per layers and tansig and logsig activation functions gave approximately similar errors in the case of seasonal database.

Confidence interval was adopted around forecasting curve. The result offers a high confidence level of prediction that is important and useful for system operation, planning and power management.

References

[1] European Photovoltaic Industry Association (EPIA), "Global market outlook for photovoltaics 2014-2018," June 2014.

[2] European Photovoltaic Industry Association (EPIA), "Connecting the sun," September 2012.

[3] International Energy Agency (IEA), "Photovoltaic and solar forecasting, state of the art," October 2013.

[4] J. Kleissl, "Current state of the art in solar forecasting," University of California, 2013.

[5] A. Mellit and A. Massi Pavan, "A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste," Italy. Sol. Energy 2010, 84, pp. 807–821.

[6] S. Kaplanis and E. Kaplani, "Stochastic prediction of hourly global solar radiation for Patra, Greece," Appl. Energy 2010, 87, pp. 3748–3758.

[7] F. Wang, Z. Mi, S. Su and H Zhao, "Short-term solar irradiance forecasting model based on artificial neural network using statistical feature parameters," Energies 2012, 5, pp. 1355–1370.

[8] E. Izgi, A. Öztopal, B. Yerli, M. K. Kaymak, and A. D. Şahin, "Short-mid-term solar power prediction by using artificial neural networks," Solar Energy, Vol. 86, No. 2, 2012, pp. 725–733.

[9] V. Lo Brano, G. Ciulla and M. Di Falco, "Artificial neural networks to predict the power output of a PV panel," International Journal of Photoenergy, Vol. 2014, Article ID 193083, 12 pages. Available: http://dx.doi.org/10.1155/2014/193083

[10] J. Shi, W. J. Lee, Y. Liu, Y. Yang, and P Wang, "Forecasting power output of photovoltaic systems based on weather classification and support vector machines," IEEE Transactions on Industry Applications, Vol. 48, No. 3, 2012, pp. 1064–1069.

[11] J.G. Da Silva Fonseca, T. Oozeki, T. Takashima, G. Koshimizu, Y. Uchida and K. Ogimoto, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan," Progress in Photovoltaics, Vol. 20, No. 7, 2012, pp. 874-882.

[12] P. Mandal, S.T.S. Madhira, A.U.I Haque, J. Meng and R.L. Pineda, "Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques," Procedia Comput. Sci. 2012, 12, pp. 332–337.

[13] P. Krömer, L. Prokop, V. Snášel, S. Mišák, J. Platoš and A. Abraham, "Evolutionary prediction of photovoltaic power plant energy

production," in Proceedings of International Conference on Genetic and Evolutionary Computation, Philadelphia, USA, July 2012, pp. 35–42.

[14] E. Ogliari, F. Grimaccia, S. Leva, and M. Mussetta, "Hybrid predictive models for accurate forecasting in PV systems," Energies, Vol. 6, No. 4, 2013, pp. 1918–1929.

[15] W. Yuan-Kang, C. Chao-Rong and H. Abdul Rahman, "A novel hybrid model for short-term forecasting in PV power generation," International Journal of Photoenergy, Vol. 2014, Article ID 569249, 9 pages. Available: http://dx.doi.org/10.1155/2014/569249

[16] H.T.C. Pedro and C.F.M. Coibra, "Assessment of forecasting techniques for solar power production with no exogenous inputs." Solar Energy 2012, 86, pp. 2017–2028.

[17] M. Paulescu, E. Paulescu, P. Gravila and V. Badescu, "Forecasting the power output of PV systems, weather modeling and forecasting of PV systems operation," Green Energy and Technology 2013, pp. 325-345.

[18] C. Monteiro, J.I. Ramirez-Rosado, A. Muñoz-Jimenez and P.M. Lara-Santillan, "Short-term forecasting models for photovoltaic plants: Analytical versus soft-computing techniques," Mathematical Problems in Engineering Vol. 2013, Article ID 767284, 9 pages. Available: http://dx.doi.org/10.1155/2013/767284

[19] S. A. Kalogirou and A. Sencan, "Artificial intelligence techniques in solar energy applications," ISBN: 978-953-307-142-8, InTech, 2010.

[20] A. Saberian, H. Hizam, M. A. M. Radzi, M. Z. A. Ab Kadir and M. Mirzaei, "Modelling and prediction of photovoltaic power output using artificial neural networks," International Journal of Photoenergy, Vol. 2014, Article ID 469701, 10 pages. Available: http://dx.doi.org/10.1155/2014/469701

[21] P. van der Smagt and B. Kröse, "An introduction to neural networks," University of Amsterdam, 1996.

[22] S. Haykin, "Neural networks - A comprehensive foundation," IEEE Press, 1994.

[23] N. Hodžić, T. Konjić and V. Miranda, "Artificial neural networks applied to short term load diagram prediction," Neurel – 8th Seminar on Neural Network Applications in Electrical Engineering, Belgrade 2006.

[24] M. T. Hagan, H. B. Demuth, M. H. Beale and O. Jesus, "Neural network Design, 2nd Edition," Oklahoma University, eBook, 2013, ISBN-13: 978-0-9717321-1-7.

[25] T. Konjic, V. Miranda and I. Kapetanovic, "Fuzzy inference systems applied to LV substation load estimation," IEEE Transaction, May 2005, Vol. 20, No. 2, pp. 742-749.