

One day ahead forecasting of energy generating in photovoltaic systems

Grzegorz Draluś^{1,*}, *Grzegorz Dec*², and *Damian Mazur*¹

¹Rzeszow University of Technology, Department of Electrical and Computer Engineering Fundamentals, 35-959 Rzeszów, Poland

²Rzeszow University of Technology, Department of Computer and Control Engineering, 35-959 Rzeszów, Poland

Abstract. The article presents selected methods for forecasting energy generated by a solar system. Short-term forecasts are necessary in planning the work of renewable energy sources and their share in the energy market. Forecasting from the one-day horizon is one of the short-term forecasts. Rear-round prognostic models have been designed using various forecasting methods such as regression, neural networks or time series. On the basis of one day ahead forecasts the accuracy of designed models was assessed. The influence of selected weather factors on forecasts accuracy is also presented, only for models implemented by MLP neural networks. As well as the results of research on the impact of the model structure (as MLP neural network) on the accuracy of forecasts are presented.

1 Introduction

Currently, renewable energy is an important element of the energy market in the world. Its share in the total energy balance is constantly growing. One of the significant sources of renewable energy is solar energy, from which electric energy is generated in solar systems. Solar energy is considered to be inexhaustible, free and clean. For these reasons, photovoltaic systems have gained great acclaim and importance as an alternative to conventional sources. However, the production of energy from photovoltaic cells is characterized by great variability and is strongly dependent on weather conditions. Changes in the volume of this production must be regularly monitored and supplemented with conventional energy sources. Short-term changes in the amount of energy generated from renewable sources as a result of sudden weather changes may be particularly dangerous for the power system. On the other hand, an important problem is improving the control performance of energy transmission from PV sources. Therefore, an important issue in the energy balance is accurate forecasting of energy production in the short term.

Short-term forecasting in the energy sector means forecasting in the time horizon up to several hours. Therefore, forecasting one day ahead, is classified as short-term forecasting.

Except the time horizons, the parameters subject to forecasting are also important. In the case of short-term forecasts, average power values are usually applied in periods from 15

* Corresponding author: gregor@prz.edu.pl

minutes to 1 hour. In the case of predictions with a day-ahead, the average daily power value or quantity of energy produced within 24 hours is used. Forecasts of energy production in PV systems are usually based on the value of solar irradiation during the period considered, weather conditions and others.

For forecasting energy generated in photovoltaic systems, different methods are used, for example, regression, time series, mathematical models, artificial intelligence, hybrid methods and many others. Among the artificial intelligence methods, neural networks achieved great success in forecasting. Therefore, different types of neural networks are used to build forecasting models, as well as combinations of neural networks with fuzzy logic or wavelet transform. The choice of neural network type usually does not depend on the horizon of forecasts, access and scope of data and the availability of numerical forecasts (NWP).

To forecast the amount of energy generated in solar PV installations, feed-forward neural network is usually used [1-4]. Other types of neural networks, such as radial basis function neural network (RBF) [5-6], recurrent neural network (RNN), e.g. the Elman network [6], dynamic neural networks of the type FTDNN and DTDNN [7 -9] are also successfully used. Forecasting can also be used with Support Vector Machine (SVM) [10], physical models [4], time series analysis [11 - 13] with recurrent neural network [14 - 15]. The connection of neural network with fuzzy-logic can also be used in the power forecasting of photovoltaic panels [16].

2 PV systems and methods

This chapter describes the solar system on which our research is carried out. In addition, this chapter presents a brief description of the forecasting methods, as well as a description of the prognostic models designed for research.

2.1 Description of a solar PV system

The photovoltaic system from which the data needed to build forecasting models and for which forecasts will be made is presented in Fig.1. Fig. 1 shows a block diagram of a real low power system, which is located in Rzeszów (50°02'N 22°17'E).

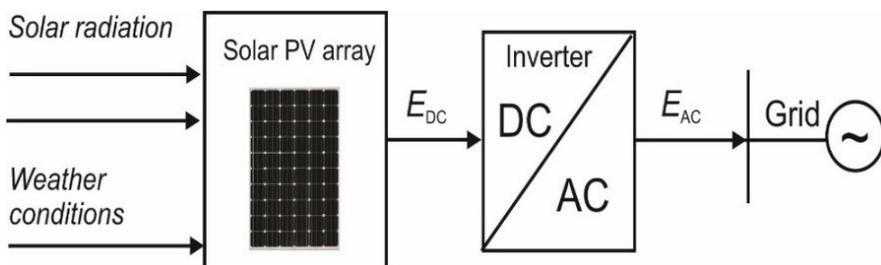


Fig. 1. The solar photovoltaic system.

The presented system consists of three monocrystalline modules with a total peak power of 330Wp, connected to a voltage inverter. Photovoltaic modules mounted at a fixed angle of 30 degree to the horizon. In the PV system the electric power of the solar system before and behind the inverter and the efficiency of the system were measured. A measuring system installed in the solar system collected DC voltage and DC current measurements before the inverter, AC voltage, AC current, and AC power behind the inverter. At the same time, the solar irradiation was also measured. All parameters were measured every 10 seconds and averaged to 1 minute values.

Data from the measurement system were prepared so that they could be used to build forecasting models with the horizon of one day ahead. On the basis of measurements of AC voltage, AC current and AC power, the total amount of AC energy (behind the inverter) generated during one day by the system was determined. Additionally, for the objectivity of the results obtained, the generated energy was scaled to one square meter of the active surface of a solar PV panel. These are measurement data from the PV system and meteorological conditions collected in one year (from November 2013 to October 2014 year). The data set collected for the study covers 340 days (25 days are missing, due to incorrect measurement readings or lack of reading). In addition, the data from the whole year was divided into training and testing data sets.

In order to determine the parameters of the forecasting models, in addition to the electrical data of the PV system, the data on weather conditions at the system's place of work are needed. Data on weather conditions on a given day are input data for models. In this work, seven weather factors taken from the local meteorological station (see Table 1) are used as input data for predicting models. As the eighth explanatory factor, the daily value of irradiation in the system's working place is used.

Table 1 contains a list of explanatory factors and their correlation coefficients with the amount of energy generated in the PV system. Correlation calculations were performed for data for the year-round model. Correlation coefficients were calculated separately for the training data set and test data set. The correlation coefficient determines which selected explanatory factor affects the amount of energy produced in the solar PV system.

Table 2. Formatting sections, subsections and subsubsections.

No	Spacing	Correlation coefficient
1	Average daily air temperature	0.570
2	The highest air temperature	0.689
3	Visibility	0.584
4	Daily wind speed	-0.399
5	Daily rainfall	0.164
6	Cloudiness	-0.720
7	Wind direction	-0.070
8	Daily solar irradiation	0.993

2.2 One day ahead forecasting year-round models

Acquisition of renewable energy in solar systems strongly depends on the weather conditions that prevail throughout the year, as well as the seasons and other factors. In different seasons the length of the day changes, that is, the action of the sun, the height of the sun on the

horizon, cloud cover and temperature also change. In literature, you can find different strategies for building prognostic models.

Most studies on short-term prediction of power generation in solar systems in the 24-hour horizon. There are works describing the models that generate 24 hour ahead forecastings with a 10-minute step, [7], with a 15-minute step [4, 14] or with a one-hour step [3, 5, 6]. These studies concerned forecasting models designed for selected months [4] or for all months of the year [5, 6, 14].

Forecasts in the 24-hour horizon are performed for the needs of current control and planning of renewable energy source (RES) work. However, such accurate forecasts are not needed for electricity trading in local or regional energy markets. For energy sales operators, forecasts of the daily energy production volume from a given RES source are sufficient to be able to receive this energy to the grid and sell it to another customer. Therefore, the amount of RES energy purchased and sold can be based on forecasts in the horizon one day ahead. There are not many studies in the subject in the literature so far. Few works present methods and forecasting models in the horizon of one day ahead [2, 5]. So, taking this issue at work seems to be justified. Therefore, our goal is to develop several methods to forecast one day ahead of the amount of energy generated in PV systems.

In order to examine the suitability and accuracy of selected predicting methods, it was assumed that the studies will be conducted on a year-round mode.

The following methods of one day ahead forecasting were selected for the research:

- Multiple regression
- Neural network MLP
- Neural network RBF
- Support Vector Machine SVM
- Nonlinear autoregressive model with exogenous inputs and MLP neural networks

2.3 Forecasting models

The following methods and tools are used to build predictive models for the PV system.

2.3.1 Linear regression model

Regression is a statistical method that permit to predict unknown values of a dependent variable based on known values of an independent (explanatory) variable. Regression, in which there is more than one independent variable, is also called multiple regression.

In the case of multiple regression models, where there is a greater number of explanatory variables, the following formula is applied to the regression line [17]:

$$Y = b_1X_1 + b_2X_2 + \dots + b_nX_n + a \quad (1)$$

where: Y is a dependent variable (output), X_1, X_2, X_n , - independent (explanatory) variables, b_1, b_2, b_n - regression coefficients, a - constant, random factor.

In the year-round model using linear regression, the forecasts of energy production in the PV system are adjusted according to the equation (1) for eight explanatory factors listed in Table 1. Optimization of regression coefficients b_1, b_2, \dots, b_n of the model was made using the least squares method.

2.3.2 MLP neural network model

Neural networks due to their many features are successfully used in many, very different fields such as finance, medicine, technology or physics. Neural networks can be used

wherever tasks related to prediction, classification or control appear. The basic feature of neural networks is the ability to generalize or generalize knowledge for new data previously unknown, i.e. not published during the course of learning. This is also referred to as the ability of neural networks to approximate the value of a function of several variables. Thus, neural networks are also very well-suited for building prognostic models.

Among many types, feedforward neural networks are often used for forecasting tasks, in particular neural network called multilayer perceptron (MLP) [1,2,4].

A feedforward neural two-layers MLP neural network is shown in Fig. 2.

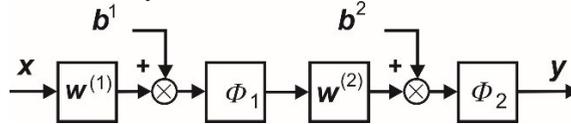


Fig. 2. A feedforward MLP neural network.

The output of j -th neuron in of neural network is given by following formula (2):

$$y_j(t) = f_2 \left[\sum_k w_{jk}^{(2)} \cdot f_1 \left[\sum_{i=0}^I (w_{ki}^{(1)} x_i(t) + b_k^{(1)}) \right] + b_j^{(2)} \right] \quad (2)$$

where: $w^{(1)} \in R^{K \times I}$, $w^{(2)} \in R^{J \times K}$, $b^{(1)} \in R^K$, $b^{(2)} \in R^J$, $x \in R^I$, $y \in R^J$, $x_i(t)$ is the i -th input of the neural network, y_j is the j -th output of the neural network, $w_{ki}^{(1)}$ and $w_{jk}^{(2)}$ - the weights between neuros, $b_k^{(1)}$ and $b_j^{(2)}$ are the biases of neuros, t is a sample of data.

In the operation of MLP neural networks, an important issue is the selection of the transfer function. In equation (2) there are two transfer functions denoted as $f_1() \in \Phi_1$ in first layer and denoted as $f_2() \in \Phi_2$ in second layer. Typically, the first layer called the hidden layer uses non-linear transfer functions such as log-sigmoid or tan-sigmoid. In the second layer called the output layer, a linear function is usually used, but other functions can be used successfully.

The MLP neural networks was used to build a model for predicting the energy of the PV system. Thus, the neuronal model uses eight explanatory factors (see Table 1), which are inputs the model. The model output is a forecast of the daily amount of energy generated in the PV system.

For the base forecasting model (e.g MLP neural network) was adopted the following structure: 8-7-1 (8 inputs, 7 neurons in the hidden layer, 1 neuron in the output layer). All neurons in hidden layer use the log-sigmoidal transfer function. A linear transfer function is used in the output layer.

The parameters of the forecasting model were determined by minimizing quality criterion such as a Root Mean Square Error (RMSE).

2.3.3 RBF neural network model

FBF neural network is a variation of the MLP neural network, in which neurons in the signal processing uses radial basis functions. RBF neural networks have found application in solving classification problems, tasks of approximation of functions of many variables, as well as in prediction problems.

The architecture of RBF neural networks has a structure analogous to the structure of a multilayer neural network with one hidden layer, which represents a nonlinear mapping carried out by neurons with a radial base function. In practice, several types of radial basis

functions are used, but the Gaussian function is most commonly used, which is described by following formula:

$$\phi(r) = \exp(-\beta r^2) \quad (3)$$

where: β is the parameter of smoothness of the Gaussian function.

Other examples of radial basis functions are the polynomial function, the multiquadratic or the inverse multiquadratic function.

RBF neural networks were also used to build a forecasting model, in which eight explanatory factors are used (see Table 1). As the radial basis function, the Gaussian function was accepted.

2.3.4 SVM model

Support Vector Machine (SVM) was also used to build a model for predicting production of electricity by the solar PV system. This method can be successfully used to estimate multidimensional functions, for approximation and prediction.

Based on the knowledge in the field of SVM, a forecasting model was designed, which uses eight explanatory factors as input data (see Table 1). The SVM model uses a polynomial 3rd order kernel function and data normalization.

2.3.5 Nonlinear autoregressive model

The forecasting model of the amount of energy production by a PV system can also be built by using nonlinear autoregressive model with exogenous inputs (NARX). The NARX model is based on the linear autoregressive model, which is commonly used in time-series analysis.

The NARX model is defined by the following equation [18]:

$$y(t) = f(y(t-1), \dots, y(t-n), u(t-1), \dots, u(t-m)) \quad (4)$$

where: $f()$ is some nonlinear function, $y(t)$ is the output of the model at the t time, $y(t-1)$, $y(t-n)$ are the past values of the model output, $u(t-1)$, $u(t-m)$ are the inputs of the NARX model, n and m are the maximum time lag for the output and input. Therefore, in the NARX model, the next value of the dependent output signal $y(t)$ is regressed on previous values of the output signal and previous values of an independent (exogenous) input signal.

Nonlinear function $f()$ in formula (4) is performed by MLP neural networks with a log-sigmoid transfer function in hidden layers. Two forecasting models with the first and second order recursion were designed.

The first model is the first order NARX model with the structure 5-8-3-1. The model has 5 inputs for 5 explanatory factors (factor from Table 1 e.g.: average air temperature, the highest air temperature, visibility, daily irradiation, and energy generated on the previous day). In addition, the model has two hidden layers, 8 neurons in the first and 3 neurons in the second hidden layer, nonlinear transfer functions of the log-sigmoid type. The model has one output, on which we obtain the forecast of energy production on a given day.

3 Results

In this chapter, errors of energy production forecasts made by the developed models are presented.

To estimate the accuracy of forecasting models a Mean Absolute Percentage Error (MAPE) of the forecasts was determined. The MAPE error was calculated by following formula:

$$MAPE = \frac{1}{T} \sum_{i=1}^T \frac{|E_i^p - E_i|}{E_i} \cdot 100\% \tag{5}$$

where: E_i is measured energy in PV system in the i -th day, E_i^p is prediction of energy generated in the i -th day, T - number of days.

Table 2 presents forecasting errors of energy for all developed models. Forecast errors were estimated by the measure of MAPE error (5). MAPE errors were determined for the training data of the learning phase of the models, as well as for the test data of the verification phase.

Table 2. The errors forecasts of energy generation in the year-round models.

Model	Estimation MAPE [%]	Verification MAPE [%]
Multiple regression	16.2	11.7
MLP neural network(8-7-1)	3.61	3.30
RBF neural network(8-136-1)	4.32	3.67
SVM	5.03	4.65
1 st order NARX (5-8-3-1)	4.19	3.16

The complete set of data was divided into a training set containing data from 280 days and a test set containing data from 60 days. The training and testing set contained eight explanatory factors (except NARX’ models) and the daily amount of energy produced in the solar PV system. Parameters of the MLP neural model were determined using Levenberg Marquard (LM) learning algorithms [19]. The weights and number of neurons in the hidden layer in the RBF neural network were determined by the least squares method and the back propagation learning algorithm. Thus, RBF neuron model has 8 inputs, 136 neurons in the hidden layer and 1 neuron in the output layer. The parameters of the 1st order NARX model were determined using the LM algorithm. Initial weights were generated randomly in the models, which use neural network.

For the test data, the most accurate is the 1st order NARX model, in which the MAPE error of the forecast is equal to 3.16%. The MPAE error around 3 percent means that year-round model generates accurate forecasts. The MLP neural network model is the second most accurate (MAPE error is 3.30%). The RBF model generates predictions with slightly larger errors than the top two models, but their accuracy is quite good. In the SVM model, the MAPE error is below five percent, i.e. the model can be considered to be sufficiently accurate. A relatively large error in the forecast is generated in the linear regression model (around 11.7%), but this model is the least computationally demanding one.

Fig. 3 presents an example of the forecast of and the real amount of energy production by the solar photovoltaic panels for test data (verification phase) for a selected model, i.e. for RBF neural networks model. The forecast covers 60 selected days of the whole year, includes 15 days from each season.

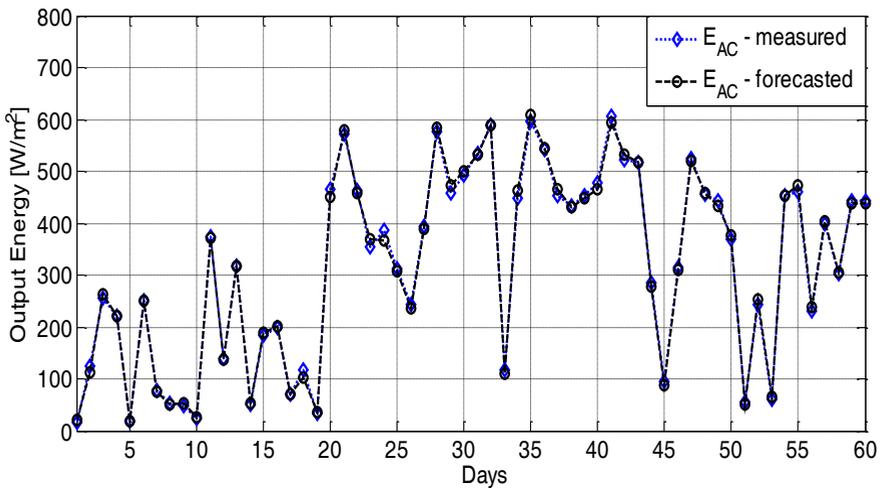


Fig. 3. The course of the AC energy (E_{AC}) forecast generated by RBF neural network model for the test data.

4 Research on the impact of model structure and explanatory factors on the accuracy of forecasts

This chapter describes the results of studies on the impact of the MLP neural network structure on the accuracy of the generated forecasts, as well as on the influence of the number of explanatory factors on the accuracy of forecasts.

Studies on the impact of the neural network structure on the accuracy of the generated forecasts were carried out on six models built of MLP neural networks. Three models were built from a network with one hidden layer and contained 7, 9 and 13 neurons in hidden layer, respectively. The other three models consisted of two hidden layers that contained the following number of neurons in the first and second hidden layer, 5-3, 7-3 and 9-3 neurons, respectively. The basic research on the impact of the model structure on the accuracy of forecasts was carried out for the maximum number of explanatory factors, i.e. up to eight factors (see Table 1). In addition, an impact study was conducted for four and two explanatory factors (see below). The results of these studies are presented in Table 3, which contains MAPE errors of predictions for the test data.

On the six MLP neural models, studied on the influence of the amount of the explanatory factors on the accuracy of the prognostic model was conducted. The point of reference are forecast errors for the 8 input explanatory factors (the first variant). In the second variant, only for four explanatory models generate the forecasts. This four selected explanatory factors have a large correlation coefficient (greater than 0.2) and these are following factors: average air temperature, the highest air temperature, visibility, daily irradiation. In the third variant, the number of explanatory factors was reduced only to two factors, with have the highest non-negative correlation (e.g. the highest air temperature and daily irradiation).

The influence of the number of explanatory factors on the accuracy of forecast in MLP neural network models is shown in Table 3 and Fig. 4.

Fig. 4 shows the influence of the number of explanatory factors on the accuracy of forecasts (i.e. MAPE error values) for six MLP neural networks as predictive models. At the same time, in Fig. 4, it can be observe the influence of the structure of neural models on the accuracy of forecasts.

Table 3. *MAPE* errors of energy forecast in MLP neural network models (x-number of inputs).

Structure of the model	2 factors	4 factors	8 factors
x-7-1	3.35	3.45	3.30
x-9-1	3.64	3.51	3.35
x-13-1	3.34	3.46	3.29
x-5-3-1	3.36	3.57	3.05
x-7-3-1	3.25	3.31	3.25
x-9-3-1	3.38	3.12	3.13

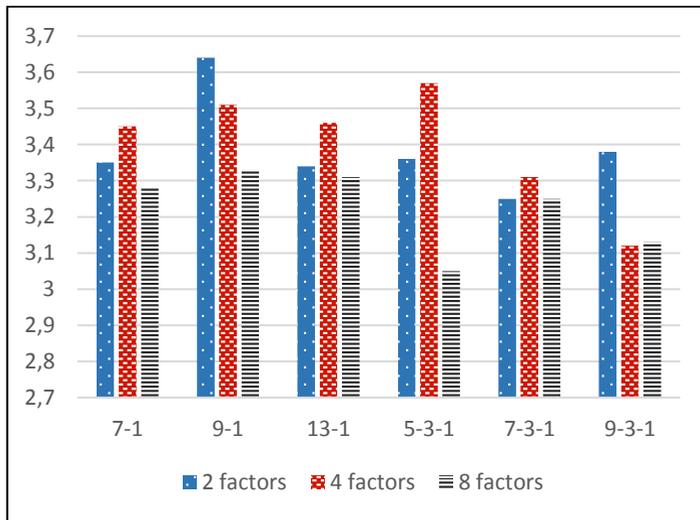


Fig. 4. Values of *MAPE* forecast errors depending on the structure of the MLP neural network model.

The analysis of results obtained shows that the MLP neural network models with two hidden layers generate more accurate forecasts than models with one hidden layer, but only in the case of forecasts for eight explanatory factors. In other cases, models with two hidden layers are usually more accurate.

The analysis obtained results from the fact that a decrease in the number of explanatory factors causes an increase in the errors of forecasts. Models using eight explanatory factors to generate forecasts are the most accurate, but with one exception. The only exception is the model with structure (4-9-3-1) using four explanatory factors.

Conclusions

The work presents the results of research on one day ahead forecasting of electricity production by the solar PV system. For this purpose, various forecasting methods have been

developed. The forecasting of AC energy production was generally based on eight explanatory factors (seven weather factors and daily irradiation level). In the NARX models, smaller number of explanatory factors were used. The developed methods and training sets were used to build year-round models. The models were verified using a test set that was 60 days long.

Due to the fact that for MLP neural networks it is difficult to determine the optimal structure in advance, therefore the influence of the model structure (i.e. MLP neural networks) on the accuracy of forecasts was examined in the work. The work also includes the results of research on the influence of the number of explanatory factors (weather factors) on the accuracy of energy forecasting.

The models provide forecasts of energy production by the solar system with satisfactory accuracy. Forecast errors in models using artificial intelligence range from 3.16% to 4.65%, but errors in the regression model are at the level of 11.7%.

The MLP neural networks and the first order NARX model based on MLP neural network are the most effective tool for one day ahead forecasting. However, most of the developed methods can successfully apply the tasks of forecasting of energy generation in photovoltaic systems for practical application.

References

1. F. Almonacid, C. Rus, P. J. Perez, L. Hontoria, *Renewable Energy*, vol. **34**, pp. 2743-2750 (2009)
2. G. Drałus, Z. Gomółka, D. Mazur, A. Smoleń, *Proceedings of International Interdisciplinary PhD Workshop* (2017),
3. P. Piotrowski, *Electrical Review (in Polish)*, **5**, pp. 162-165 (2015)
4. Y. Huang J. Lu, Ch. Liu, X. Xu, W. Wang, X. Zhou, *International Conference on Power System Technology (POWERCON)* (2010)
5. Ch. Chen, S. Duan, T. Cai, B. Liu, *Solar Energy* vol. **85**, Issue 11, pp. 2856-287 (2011)
6. A. Yona, T. Senju, A. Saber, T. Funabashi, H. Sekine, Ch-H. Kim, *Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century*, IEEE (2008)
7. N. Al-Messabi, Y. Li, I. El-Amin, C. Goh, *Comput. Intell.*, pp.10-15 (2012)
8. Z. Gomolka, E. Dudek-Dyduch, Y.P. Kondratenko, *From Homogeneous Network to Neural Nets with Fractional Derivative Mechanism. Artificial Intelligence and Soft Computing*, vol **10245**, pp 52-63, (2017)
9. Z. Gomolka, B. Twarog, E. Zeslowska, *Cognitive Investigation on Pilot Attention During Take-Offs and Landings Using Flight Simulator. Artificial Intelligence and Soft Computing. Lecture Notes in Computer Science*, vol **10246**, pp 432-443, (2017)
10. J. Shi, W-J. Lee, Y. Liu, Y. Yang, P. Wang, *2011 IEEE Industry Appl. Society Annual Meeting (IAS)*, IEEE, pp.1-6 (2011).
11. L. Martin, L.F. Zarzalejo, J. Polo, A. Navarro, R. Marchante, M. Cony, *Solar Energy*, **84**, pp. 1772-1781 (2010).
12. Z.Gomolka, B.Twarog, E. Zeslowska, A. Lewicki, T. Kwater, *Using Artificial Neural Networks to Solve the Problem Represented by BOD and DO Indicators*, *Water* 2018, **10(1)**, 4
13. Z. Gomolka, *Neurons' Transfer Function Modeling with the Use of Fractional Derivative. Contemporary Complex Systems and Their Dependability. DepCoS-RELCOMEX 2018. Advances in Intelligent Systems and Computing*, vol **761**. Pp. 218-227, (2019)

14. M. Cococcioni, E. D'Andrea, B. Lazzerini, *11th International Conference on Intelligent Systems Design and Applications* (2011)
15. T. Cai, S. Duan, Ch. Chen, Power Electronics for Distributed Generation Systems (PEDG), 2nd IEEE International Symposium, pp.773-777 (2010)
16. F. Grimaccia, M. Mussetta, R. Zich, Fuzzy Systems (FUZZ), IEEE International Conference, pp.2454-2457 (2011)
17. D.A. Freedman, *Statistical Models: Theory and Practice* (2005)
18. L. Ljung, *System Identification: Theory for the user* (1987)
19. M.T. Hagan, M. Menhaj, IEEE Trans. on Neural Networks, vol. **5**, No. **6**, pp. 989-993 (1994)