

## MODELING OF PHOTOVOLTAIC MODULES USING A GRAY-BOX NEURAL NETWORK APPROACH

by

**Aleksandar M. RANKOVIĆ\*** and **Dragan N. ĆETENOVIĆ**

Faculty of Technical Sciences, University of Kragujevac, Cacak, Serbia

Original scientific paper  
<https://doi.org/10.2298/TSC160322023R>

*This paper proposes a gray-box approach to modeling and simulation of photovoltaic modules. The process of building a gray-box model is split into two components (known, and unknown or partially unknown). The former is based on physical principles while the latter relies on functional approximator and data-based modeling. In this paper, artificial neural networks were used to construct the functional approximator. Compared to the standard mathematical model of photovoltaic module which involves the three input variables – solar irradiance, ambient temperature, and wind speed- a gray-box model allows the use of additional input environmental variables, such as wind direction, atmospheric pressure, and humidity. In order to improve the accuracy of the gray-box model, we have proposed two criteria for the classification of the daily input/output data whereby the former determines the season while the latter classifies days into sunny and cloudy. The accuracy of this model is verified on the real-life photovoltaic generator, by comparing with single-diode mathematical model and artificial neural networks model towards measured output power data.*

**Keywords:** photovoltaic module, artificial neural networks, gray-box model, output power, data clustering, functional approximator

### Introduction

Due to the fast growing amount of photovoltaic (PV) systems in global electricity production, accurate prediction of these systems is becoming very significant. Reliable forecasting of PV systems in electricity production can help the grid operators maintain a better balance between power demand and supply. The amount of power generated by a PV system dominantly depends on the intensity of the incident solar irradiance. Unpredictable variations in PV power output may increase the operating costs of the power system and reduce the operational reliability of electricity supply.

Conventional models of PV module performance are based on physical principles. These models use irradiance and temperature of the module (or ambient temperature and wind speed) as input variables. Circuit-based PV models generally include both linear and non-linear components. Series and shunt resistances describe the linear part of the equivalent circuit. Single-diode, double-diode or three-diode current-voltage (I-V) characteristics describe the non-linear part of the equivalent circuit [1-4].

Most PV module manufacturers provide only a list of basic technical data measured at standard test conditions (STC), while only a few manufacturers provide data for nominal

\*Corresponding author, e-mail: [aleksandar.rankovic@ftn.kg.ac.rs](mailto:aleksandar.rankovic@ftn.kg.ac.rs)

operating cell temperature (NOCT). The performance of PV modules at a particular site under real operating conditions can differ dramatically from the one measured at STC and NOCT [5, 6]. The different factors affecting PV module performance are as follows: location (longitude and latitude), environmental factors (solar irradiance, temperature, wind, humidity, atmospheric pressure, pollution, dust, rain, *etc.*), inclination angle, ways of cooling, and the type of PV technology used [7, 8]. Thus, the accuracy of a conventional model of PV module is limited. Obviously, the output power of PV module is not only a function of irradiance and temperature but also of a number of environmental factors, therefore, determining physical principles based on a model involving these interdependent factors is very difficult [7-9].

An alternative approach is using a functional approximator artificial neural networks (ANN) [2, 10-15], genetic algorithm [16], neuro-fuzzy inference system [13], and particle swarm optimization [17, 18]) to simulate PV module. The input and output layer of functional approximator consists of weather data (historically recorded) and measured output power of PV module, respectively. Various input parameters, important to the operation of PV module, are used in these studies.

In [10] the authors proposed ANN based model to estimate the produced power of a Si-polycrystalline PV module, using the solar irradiance and temperature as an input. The ANN approach is used to predict solar cell short-circuit current, open-circuit voltage and thus set current-voltage characteristics of the PV generator (for an example, see [2, 11]). Also, ANN are used for PV output power and solar irradiance forecasting [12-14], with the esteemed influence of input parameters, such as humidity, atmospheric pressure, wind speed, *etc.* The results obtained with the functional approximator are generally better than those obtained by a conventional model based on physical principles.

In this paper, we propose a gray-box approach to improve the prediction accuracy of a conventional model of PV module. In gray-box modeling, the system model is partitioned into two components: known, and unknown or partially unknown. The former relies on physical principles, whereas the latter is modeled by using a functional approximator. One of the pioneer works in the application of grey-box modeling concept to engineering problems was reported in [19]. The use of grey-box approach for modeling PV power system via particle swarm optimization has been only recently proposed by [17, 18]. Particle swarm optimization was chosen to optimize the parameters in the clear-box model and to find the best coefficients in the grey-box model.

Our approach for PV module modeling uses a conventional single-diode model as a clear-box model. The three main input variables (solar irradiance, ambient temperature, and wind speed) are used as the inputs of the clear-box model. The output of the clear-box model and the three additional environmental variables (wind direction, atmospheric pressure, and humidity) are used as inputs of ANN based black-box model in order to improve the accuracy of the conventional model to predict the output power of the PV module. The input/output data used for training and validation of gray-box model were collected from January 1, 2014 to December 31, 2014. To examine the influence of the type of day on the accuracy of the gray-box model we proposed two criteria for the classification of the daily input/output data: season-based criterion (winter, summer and spring/autumn) and the additional criterion for classifying days into sunny and cloudy. Therefore, three gray-box cases were investigated. In the first case, we used all training input/output data for training one ANN. We used input/output data classified by season for training three ANN in the second case and input/output data classified by the both proposed criteria for training six ANN in the third one.

## Mathematical modeling

### Single-diode five parameter PV generator model

The PV generators are typically formed by combinations of  $N_{pg}$  parallel and  $N_{sg}$  series connections of PV modules, where PV generator series and parallel resistance are, respectively:  $R_{sg} = (N_{sg}/N_{pg})R_{sm}$ ,  $R_{pg} = (N_{pg}/N_{sg})R_{pm}$ . A PV module typically consists of  $N_s$  cells in series [1, 3, 20, 21]. Please note that in order to reduce negative impacts of shading and eliminate the hot-spot phenomena accordingly, it is common to include one bypass diode per up-to 20 series-connected cells.

The basic current-voltage (I-V) characteristics of the PV generator can be described [21]:

$$I_g = N_{pg} I_{scm} - \frac{N_{pg} I_{scm}}{\exp\left(\frac{V_{ocm}}{nN_s V_T}\right)} \left[ \exp\left(\frac{V_g + I_g R_{sg}}{nN_s N_{sg} V_T}\right) - 1 \right] - \frac{V_g + I_g R_{sg}}{R_{pg}} \quad (1)$$

where  $I_g$  [A] and  $V_g$  [V] are current and voltage of PV generator, respectively,  $I_{scm}$  [A] and  $V_{ocm}$  [V] – the short-circuit current and open-circuit voltage of the module, respectively,  $n$  – the diode coefficient (for ideal diode is  $n = 1$ ),  $V_T = k_B T_m / q$  [V] – the thermal voltage,  $k_B$  – the Boltzmann's constant ( $k_B = 1.380622 \cdot 10^{-23}$  J/K),  $T_m$  [K] – the temperature of the module,  $q$  – the quantity of electric charge ( $q = 1.60 \cdot 10^{-19}$  C), and  $R_{sm}$  [ $\Omega$ ] and  $R_{pm}$  [ $\Omega$ ] – the series and parallel resistance of the module, respectively.

The PV module temperature can be calculated by solar irradiance incident on a tilted PV module surface, ambient temperature and wind speed [22]:

$$T_m = 3.12 + 0.25 \frac{G_{G,tilt}}{G^{stc}} + 0.899 T_a - 1.3 v_w + 273 \quad (2)$$

where  $G_{G,tilt}$  [ $W/m^2$ ] is the solar irradiance incident on a tilted PV module surface,  $G^{stc}$  [ $W/m^2$ ] – the solar irradiance at STC,  $T_a$  [ $^{\circ}C$ ] – the ambient temperature, and  $v_w$  [ $ms^{-1}$ ] – the wind speed.

Manufacturers of PV modules usually provide basic information, measured at STC and NOCT, *i. e.*, the  $I_{scm}^{stc}$ ,  $V_{ocm}^{stc}$ , the temperature coefficient of the short-circuit current  $\alpha_I$  the temperature coefficient of the open-circuit voltage  $\alpha_V$ , the voltage  $V_{MPP}^{NOCT}$  and current  $I_{MPP}^{NOCT}$  at maximum power point (MPP), and the maximum power output  $P_{max}^{NOCT}$ . The values of  $I_{scm}$  and  $V_{ocm}$  are provided with reference to STC temperature of the module  $T_m^{stc} = 25$   $^{\circ}C$  and STC solar irradiance  $G^{stc} = 1000$   $W/m^2$ . For details see [3, 20, 23]. For constant solar irradiance and module temperature, MPP represents an operating point on the V-I characteristics, see eq. (1), where maximum output power is achieved by:

$$P_{max} = V_{MPP} I_{MPP} \quad (3)$$

The MPP varies among different combinations of atmospheric conditions. A maximum power point tracking (MPPT) algorithm adjusts the voltage at which a PV module generates its maximum output power at all times and under varying atmospheric conditions.

### Solar irradiance

The amount of solar irradiance incident on a tilted PV module depends on the angle between the PV module and the Sun position, as shown in fig. 1.

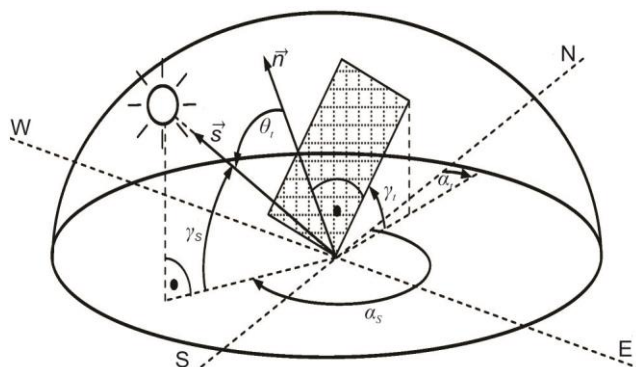


Figure 1. Defining angles of PV module

The total (also called global) irradiance on a horizontal surface  $G_{G,hor}$  represents the sum of the direct irradiance  $G_{dir,hor}$  and the diffuse irradiance  $G_{diff,hor}$  on the horizontal surface:

$$G_{G,hor} = G_{dir,hor} + G_{diff,hor} \quad (4)$$

Diffuse irradiance in a given location can be calculated through

the diffuse transposition factor by several models (*i. e.* Perez or Hay-Davies model). These models

depend on complex atmospheric properties. Given that the data for such models were not available for the given location of the PV system, the diffuse irradiance  $G_{diff,hor}$  on the horizontal surface was determined based on known irradiance on a horizontal surface  $G_{G,hor}$  by empirical formulae [24]:

$$G_{diff,hor} = \begin{cases} G_{G,hor} (1.02 - 0.254k_T + 0.0123 \sin \gamma_s), & k_T \leq 0.3 \\ G_{G,hor} (1.4 - 1.749k_T + 0.177 \sin \gamma_s) & , 0.3 < k_T < 0.78 \\ G_{G,hor} (0.486k_T - 0.182 \sin \gamma_s) & , k_T \geq 0.78 \end{cases} \quad (5)$$

where  $k_T = (G_{G,hor}/G_0 \sin \gamma_s)$  is the clearness index,  $G_0$  – the extraterrestrial irradiance ( $G_0 = 1361 \text{ W/m}^2$ ) [25], and  $\gamma_s$  – the angle of solar altitude (Sun height).

The total irradiance on a tilted surface of PV module  $G_{G,tilt}$  is:

$$G_{G,tilt} = G_{dir,tilt} + G_{diff,tilt} + G_{refl,tilt} \quad (6)$$

The direct irradiance,  $G_{dir,tilt}$ , the diffuse irradiance,  $G_{diff,tilt}$ , and the ground reflected irradiance,  $G_{refl,tilt}$ , on the tilted surface can be calculated from the irradiance on the horizontal surface as [26, 27]:

$$G_{dir,tilt} = G_{dir,hor} \frac{\cos \theta_t}{\sin \gamma_s} \quad (7)$$

$$G_{diff,tilt} = 0.5G_{diff,hor} \left( 1 + F \cos^2 \theta_t \cos^3 \gamma_s \right) \left[ 1 + F \sin^3 (0.5\gamma_t) \right] (1 - \cos \gamma_t) \quad (8)$$

$$G_{refl,tilt} = 0.5G_{G,hor} A (1 - \cos \gamma_t) \quad (9)$$

where the angle between the vector  $\vec{s}$  in the direction of the Sun and the normal vector  $\vec{n}$  perpendicular to the surface is:

$$\theta_t = \arccos \left[ \sin \gamma_s \cos \gamma_t - \cos \gamma_s \sin \gamma_t \cos (\alpha_s - \alpha_t) \right] \quad (10)$$

where  $F = 1 - (G_{dir,hor}/G_{G,hor})^2$ ,  $\gamma$  – the tilt angle of PV module,  $A$  – the albedo (reflection coefficient),  $\alpha_s$  – the solar azimuth angle (relative to north, in the eastward direction), and  $\alpha_t$  – the PV module orientation angle (relative to north, in the eastward direction).

The angles which define Sun position ( $\gamma_s, \alpha_s$ ) and the angle between the Sun and normal on tilted surface of PV module,  $\theta_t$ , depend on time and geographic data for the location of the analyzed PV module. Calculating total irradiance on a tilted surface of PV module at any given time necessitates recalculating respective angles.

### Gray-box modeling

#### Overall structure of the proposed model

Besides the material performance (cell semiconductor material and glazing material) and solar irradiance, the proposed physical principles-based model of PV generator uses module temperature as input data. The module temperature is calculated as a function of solar irradiance incident on a tilted PV module surface, ambient temperature and wind speed, see eq. (2). The analysis of the simultaneous effect of other weather variables, such as wind direction, atmospheric pressure, humidity, etc., is very complex [7], therefore, physical principles-based model which obtains the actual module temperature and output power from the previously mentioned ambient data is not available.

In order to solve this problem, this paper proposes a gray-box approach that combines physical principles (denoted as *Clear-box model*) and simplicity of functional approximator (denoted as *Black-box model*), as shown in fig. 2.

The *Clear-box model* is a single-diode mathematical model of PV generator (see section *Mathematical modeling*), while the *Black-box model* is a 3-layer feed forward artificial neural network (FF ANN) with backpropagation training algorithm. *Clear-box model* of PV module requires three input variables: irradiance,  $G_{G,hor}$ , ambient temperature,  $T_a$ , and wind speed,  $v_w$ . Gray-box model allows the use of additional input weather variables: wind direction,  $dir_w$ , humidity,  $RH$ , and atmospheric pressure,  $P_a$ , whereby all the inputs can be easily measured at a real weather station. The indices 'c' and 'g' (fig. 2) denote the clear-box model and gray-box model maximum output power of PV generator, respectively.

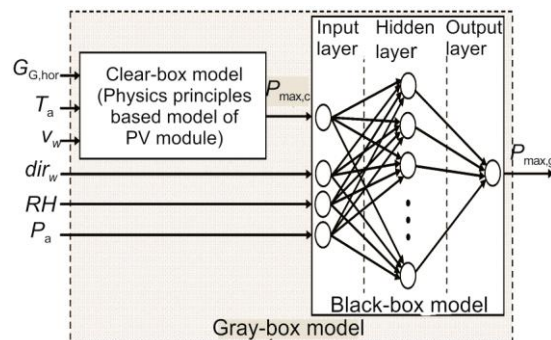


Figure 2. Gray-box model of PV generator

The population of training sets used to train *Gray-box model* is prepared as follows. Matrices of measured time-dependent input (weather variables) and output (maximum output power of PV generator) data are, respectively:

$$\mathbf{I}^{\text{meas}}(t) = \begin{bmatrix} G_{G,hor}(t_1) & T_a(t_1) & v_w(t_1) & dir_w(t_1) & RH(t_1) & P_a(t_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{G,hor}(t_j) & T_a(t_j) & v_w(t_j) & dir_w(t_j) & RH(t_j) & P_a(t_j) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ G_{G,hor}(t_J) & T_a(t_J) & v_w(t_J) & dir_w(t_J) & RH(t_J) & P_a(t_J) \end{bmatrix}, \quad \mathbf{O}^{\text{meas}}(t) = \begin{bmatrix} P_{\max,g}(t_1) \\ \vdots \\ P_{\max,g}(t_j) \\ \vdots \\ P_{\max,g}(t_J) \end{bmatrix} \quad (11)$$

where  $J$  is the number of time samples for ANN training (number of rows in input/output training matrices).

In the proposed PV generator model a set of pre-processed input data (fig. 2) is used to train ANN based *Black-box model*. Input patterns for ANN are populated by measured weather data and active power calculated by physical principles-based PV generator model (output of *Clear-box model*):

$$\mathbf{I}^{\text{black-box}}(\mathbf{t}) = \begin{bmatrix} P_{\max,c}(t_1) & dir_w(t_1) & RH(t_1) & P_a(t_1) \\ \vdots & \vdots & \vdots & \vdots \\ P_{\max,c}(t_j) & dir_w(t_j) & RH(t_j) & P_a(t_j) \\ \vdots & \vdots & \vdots & \vdots \\ P_{\max,c}(t_J) & dir_w(t_J) & RH(t_J) & P_a(t_J) \end{bmatrix} \quad \mathbf{O}^{\text{black-box}}(\mathbf{t}) = \mathbf{O}^{\text{meas}}(\mathbf{t}) \quad (12)$$

The ANN is trained using the Levenberg-Marquardt backpropagation algorithm [28]. A standard approach is to normalize all time-dependent input and output vectors to the range [0, 1]. Following the application of each input, the ANN computes its output which is subsequently compared with the target output to produce an error.

#### Clustering of weather data

In order to improve the accuracy of the proposed *Gray-box model* of PV generator, the measured time-dependent input/output data should be classified into clusters. The goal is to group days of the year in such a manner that they contain the most similar weather conditions. Different methods for classification of typical meteorological days have been proposed in a number of recent works [29]. In [10] the authors used annual average solar irradiance on a horizontal surface to classify days into sunny and cloudy. Days with irradiance higher than annual average solar irradiance are considered as sunny days.

In this paper we propose grouping the meteorological data based on two criteria. The first one groups data into three calendar-based seasons – winter (denoted as index, w), summer (denoted as index, s), and spring/autumn (denoted as index, s/a). This criterion grouped days of the year with a similar tilt of the Earth's rotational axis and the length of days. For each season, maximum value of average daily total irradiance was calculated on a horizontal surface:

$$G_{G,\text{hor},w(s,s/a)}^{\text{ad,max}} = \max \left\{ G_{G,\text{hor}}^{\text{ad}}(i_{w(s,s/a)}) \right\}, \quad i_{w(s,s/a)} = 1, 2, \dots, n_{w(s,s/a)} \quad (13)$$

$$G_{G,\text{hor}}^{\text{ad}}[i_{w(s,s/a)}] = \frac{1}{J[i_{w(s,s/a)}]} \sum_{j=1}^{J[i_{w(s,s/a)}]} G_{G,\text{hor}}(t_j) \quad (14)$$

where

$$G_{G,\text{hor}}^{\text{ad}}[i_{w(s,s/a)}]$$

is the average daily total irradiance on a horizontal surface,  $n_w$ ,  $n_s$ ,  $n_{s/a}$  – the number of days in season, respectively, and  $J[i_{w(s,s/a)}]$  – the number of time samples in one day.

Previously calculated values of irradiance ( $G_{G,\text{hor},w}^{\text{ad,max}}$ ,  $G_{G,\text{hor},s}^{\text{ad,max}}$ , and  $G_{G,\text{hor},s/a}^{\text{ad,max}}$ ) define the sunniest day of the each season.

The second criterion classifies the days of the season into sunny and cloudy. Days of the season (sunny and mostly sunny) fulfilling the condition  $G_{G,\text{hor}}^{\text{ad}}(i_{w(s,s/a)}) > 0.5G_{G,\text{hor},w(s,s/a)}^{\text{ad,max}}$

are classified as sunny. All other days are classified as cloudy. In addition, all time-dependent input/output data are clustered using the proposed criteria. Flow-chart of clustering input and output time series data is shown in fig. 3.

Note that independent *Gray-box model* corresponds to each cluster of measured time-dependent input and output data. In *Operation phase*, the additional vector of input data  $I^{\text{meas}}$  is classified into one of the clusters.

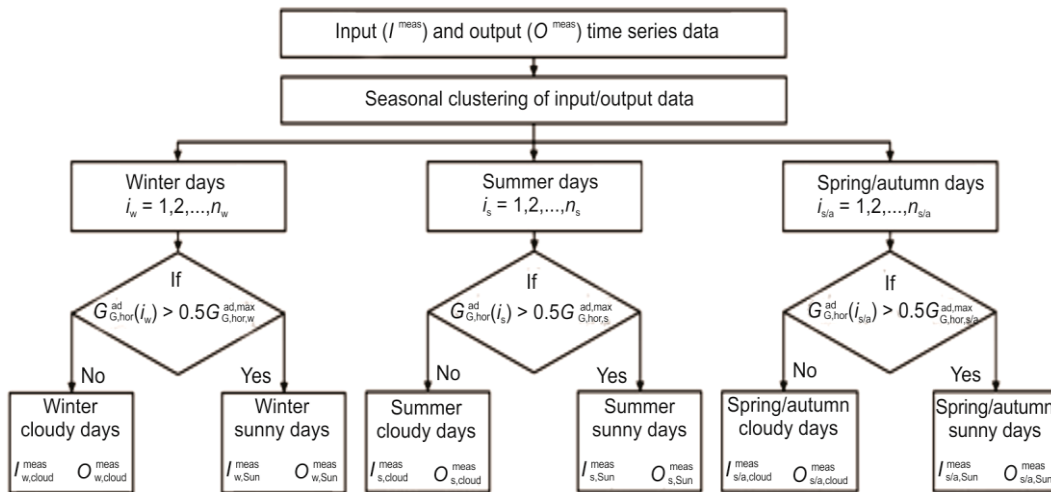


Figure 3. Flowchart of clustering input and output time series data

### Evaluation criteria

To evaluate the proposed *Gray-box model*, three different standard statistical measures were used: the relative root means square error (RMSE), relative mean bias error (MBE), and correlation coefficient,  $\rho$ . The RMSE is a measure of the average spread of the errors. The *RMSE* [%] were calculated using the following equation:

$$RMSE = \sqrt{\frac{1}{K} \sum_{k=1}^K \left[ \frac{P_p^{PV}(t_k) - P_{\text{meas}}^{PV}(t_k)}{P_{\text{meas}}^{PV}(t_k)} \right]^2} \cdot 100 \quad (15)$$

where  $P_p^{PV}$ ,  $P_{\text{meas}}^{PV}$  are predicted and measured value of maximum output power, respectively, and  $K$  – the number of test time samples.

The MBE represents the average deviation of the predicted values from the measured values. Also, *MBE* [%] indicates the average amount of over- or under-estimation, given:

$$MBE = \frac{1}{K} \sum_{k=1}^K \frac{P_p^{PV}(t_k) - P_{\text{meas}}^{PV}(t_k)}{P_{\text{meas}}^{PV}(t_k)} \cdot 100 \quad (16)$$

The correlation coefficient,  $\rho$  [%], is a way to measure the strength and the direction of the relationship between the predicted and measured values, expressed:

$$\rho = \frac{[\text{Cov}(P_p^{\text{PV}}, P_{\text{meas}}^{\text{PV}})]^2}{\text{Var}(P_p^{\text{PV}})\text{Var}(P_{\text{meas}}^{\text{PV}})} = \frac{\sum_{k=1}^K [P_p^{\text{PV}}(t_k) - \bar{P}_p^{\text{PV}}][P_{\text{meas}}^{\text{PV}}(t_k) - \bar{P}_{\text{meas}}^{\text{PV}}]}{\sqrt{\sum_{k=1}^K [P_p^{\text{PV}}(t_k) - \bar{P}_p^{\text{PV}}]^2} \sqrt{\sum_{k=1}^K [P_{\text{meas}}^{\text{PV}}(t_k) - \bar{P}_{\text{meas}}^{\text{PV}}]^2}} 100 \quad (17)$$

where  $\bar{P}_p^{\text{PV}}$ ,  $\bar{P}_{\text{meas}}^{\text{PV}}$  are the mean value of  $P_p^{\text{PV}}$  and  $P_{\text{meas}}^{\text{PV}}$ , respectively.

The range of correlation coefficient is such that  $-1 \leq \rho \leq +1$ . A positive correlation coefficient close to +1 indicates a strong positive linear relationship between the predicted and the measured values.

## Results and discussion

The proposed *Gray-box model* was tested on a 5 kWp PV generator with 20 modules installed on the roof of Technical School in city of Cacak, Serbia. All PV modules are set in the southeast direction ( $\alpha_t = -80^\circ$ ) with tilt angles of  $\gamma_t = 35^\circ$ . The grid-connected PV generator system is equipped with an MPPT unit to obtain the maximum power. The parameters of PV modules under STC are given by manufacturers (Luxor Solar GmbH, Germany):  $I_{\text{scm}}^{\text{stc}} = 8.59 \text{ A}$ ,  $V_{\text{ocm}}^{\text{stc}} = 37.31 \text{ V}$ ,  $\alpha_1 = 0.05\%/^\circ\text{C}$ ,  $\alpha_V = -0.32\%/^\circ\text{C}$ ,  $R_{\text{sm}} = 0.02 \Omega$ , and  $R_{\text{pm}} = 19.50 \Omega$ .

Data-acquisition system is used to collect output power of the installed PV generator on the AC side. The weather data are obtained from an automatic weather station near the location of PV generator. Both electrical and meteorological data are updated every 10 minutes. Data were collected for the period from January 1, 2014 to December 31, 2014. The solar position is calculated from time and geographic data for the location of the analyzed PV generator (latitude N43°53'40", longitude E20°20'32", and height of 300 m above sea level).

About seventy percent of the input/output data in each season is used for training: winter (January 1-February 28), summer (June 1-July 31), and spring/autumn (March 1-April 30, September 1 - October 30). The remaining data are used to validate the *Gray-box model*: winter (December 1-31), summer (August 1-31), and spring/autumn (May 1-31, November 1-30).

The accuracy of the *Gray-box model* for predicting the maximum output power of PV generator is compared with that from the *Clear-box model* only and neural network only (denoted as *FF ANN model*). The *Clear-box model* has three input variables ( $G_{\text{G,hor}}$ ,  $T_a$ , and  $v_w$ ) as explained in the section *Mathematical modeling*. In *FF ANN model* six input variables ( $G_{\text{G,tilt}}$ ,  $T_a$ ,  $v_w$ ,  $dir_w$ ,  $RH$ , and  $P_a$ ) were used for ANN training. Note  $G_{\text{G,tilt}}$  is calculated from  $G_{\text{G,hor}}$  using eqs. (6)-(10). Figure 2 shows ANN which is designed for *Gray-box model* and has four input vectors eq. (12). The chosen number of neurons in the hidden layer of ANN was 20 and the chosen number of training cycles was 1000. This is a tradeoff between the complexity of a model and its predictive accuracy.

In order to examine the effect of classifying time-dependent input/output data into clusters, we proposed three cases:

– In the *Training phase* all training input/output data are used for training one ANN (denoted as Case 1).



- In the *Training phase* training, input/output data are classified by one criterion, based on the season (winter, summer and spring/autumn) and by methodology defined in the section *Clustering of weather data*. A particular ANN (denoted as Case 2) was created for each season.
- In the *Training phase*, each day is classified based on the season (winter, spring/autumn, and summer). In addition, the second criterion classifies the days as sunny and cloudy in six clusters, by the methodology defined in the section *Clustering of weather data*. The six separate ANN are trained for each cluster (denoted as Case 3).

The proposed method of seasonal clustering, Case 2 allows calculation of PV maximum output power in real-time. However, additional criterion which classifies the days of the season into sunny and cloudy, Case 3, requires  $G_{G,hor}^{ad}$ , eq. (14), which can be calculated after the end of the day. The alternative way is to forecast  $G_{G,hor}^{ad}$  using numerical weather prediction models [10], which is out of the scope of our paper.

After the ANN were trained, we analyzed the prediction errors by comparing the results obtained by *Gray-box model* with those obtained by other models. This was done on validation data set (denoted as *Validation phase*). In Case 1, in *Validation phase* all test samples (in all seasons) were used for the calculation of *RMSE*, *MBE*, and  $\rho$ . The error statistics of the *Gray-box model*, *FF ANN model*, and *Clear-box model*, regarding the Case 1, are shown in tab. 1.

**Table 1. Statistics error of prediction results (*Validation phase*) for Case 1**

	RMSE [%]	MBE [%]	$\rho$ [%]
Clear-box model	6.19	-4.15	99.74
FF ANN model	5.96	-1.87	99.56
Gray-box model	4.83	-1.04	99.85

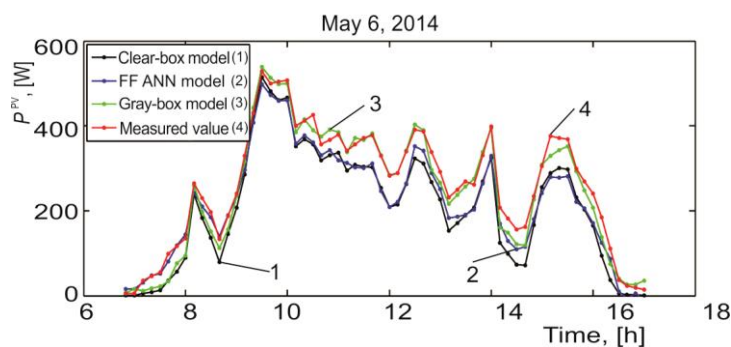
The *RMSE* results show that *Clear-box model* performed better than *Clear-box* and *FF ANN* models with respect to the prediction of the maximum output power of PV generator. In *Gray-box model*, *MBE* values show slight underestimation in the maximum output power of PV generator. The *MBE* rate is high in *Clear-box model*, especially in the first and the last hour of the solar day (from sunrise to sunset). If we neglect first and last hour of the solar day, *MBE* rate of *Clear-box model* is reduced from -4.15% to -1.73%, which indicates that the *Gray-box model* and *FF ANN model* are more reliable than the *Clear-box model* at predicting PV generator maximum output power under poor sunlight conditions.

As shown in tab. 1, the results for correlation coefficient in different models show similar values, indicating a strong positive correlation between the predicted and the measured values of PV generator maximum output power.

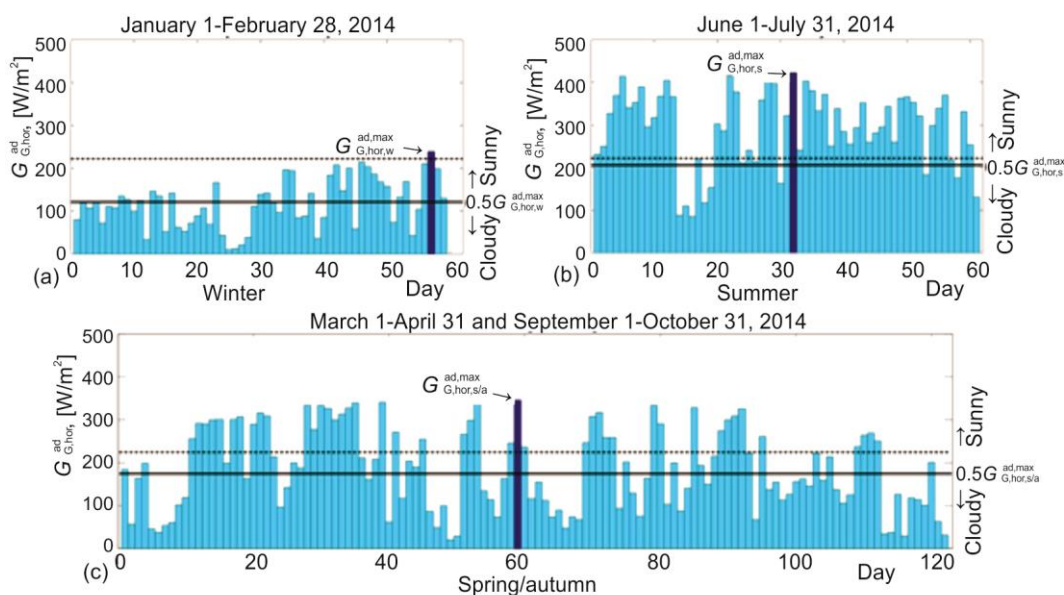
Figure 4 shows a sample of PV generator maximum output power prediction results (*Validation phase*) for Case 1. The results are for cloudy spring day on May 6, 2014.

For Case 3, in the *Training phase*, each day was classified as sunny or cloudy within a season by methodology defined in the section *Clustering of weather data*, as shown in fig. 5. Note, the criterion for classifying days into sunny or cloudy,  $0.5G_{G,hor,w(s,s/a)}^{ad,max}$ , is denoted by a horizontal solid line in fig. 5.

For the location of the analyzed PV generator, the annual average solar irradiance on a horizontal surface in 2014 was 225 W/m<sup>2</sup> per day. The obtained value is denoted as a horizontal dashed line in fig. 5. If we apply the criterion defined according to [10] almost all winter days can be classified as cloudy.



**Figure 4. Sample of prediction results (Validation phase) and comparison with measurements for Case 1**  
(for colour image see journal web site)



**Figure 5. Sunny/cloudy classification of days within season (in Training phase)**

Figure 5 shows a significant impact of days classification. In this paper, the proposed approach classifies days under more realistic assumptions. For example, the sunniest winter day has a much lower value of maximum average daily total irradiance compared to the sunniest summer day.

In [30], the impact of wind direction on the convective heat transfer from a roof-mounted flat plate solar collector was analyzed. It was pointed out that the change in wind direction led to a slight but distinct variation in convective heat transfer process, particularly at higher wind speed when the plate surface was on the leeward side of the building. To investigate the influence of wind direction on the maximum output power of PV generator, we compared the results obtained for the two scenarios:

The ANN (for both Case 2 and Case 3) were trained and validated with all the input data (denoted as *Complete input data* scenario);

In the *Training and Validation phase* of the ANN (for both Case 2 and Case 3) wind direction data is omitted (denoted as *Omitted wind direction* scenario).

The influence of clustering input/output data (Case 2 and Case 3) on PV generator maximum output power for both proposed scenarios is shown in the tab. 2.

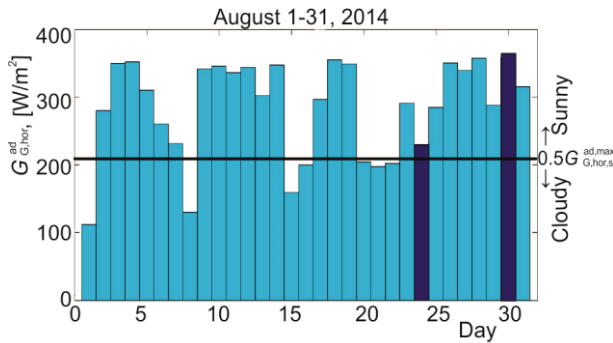
It can be observed that the clustering of input/output data by season (Case 2) significantly affects the results. In Case 2, *RMSE* and *MBE* of the *Gray-box model* were significantly less, compared to the *Clear-box model* and *FF ANN model*, tab. 2. Additional criteria which classify days into sunny or cloudy (Case 3) lead to further decrease in *RMSE* and *MBE*.

The results shown in tab. 2 show that wind direction has a negligible effect on the performance of the PV generator maximum output power, although it also may add noise to the model. A possible reason for such error behavior can be the minor impact of wind direction on convective heat transfer process at lower wind speed. The average measured wind speed for winter, summer and spring/autumn season was 0.6 m/s, 0.9 m/s, and 1.1 m/s, respectively.

**Table 2. Seasonal statistics error of prediction results (*Validation phase*) for different models and differently classified input/output data**

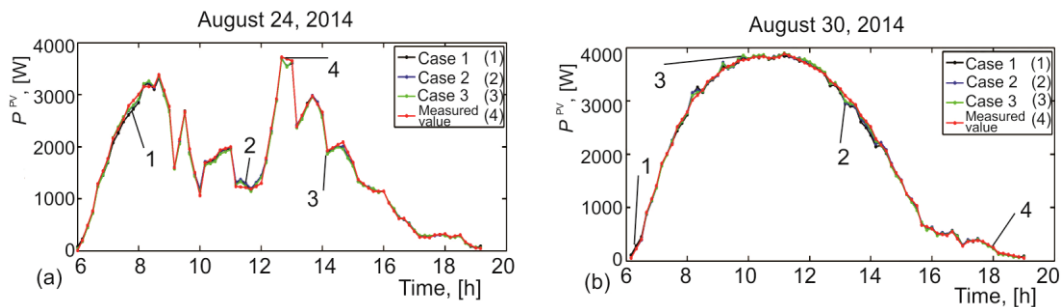
Scenario	Classification			Clear-box model			FF ANN model			Gray-box model		
	Season	Case	Day	RMSE [%]	MBE [%]	$\rho$ [%]	RMSE [%]	MBE [%]	$\rho$ [%]	RMSE [%]	MBE [%]	$\rho$ [%]
Complete input data	Winter	Case 2	All	7.02	-5.54	99.71	5.75	-2.81	99.67	4.21	-1.59	99.87
		Case 3	Sunny	4.99	-4.74	99.78	4.91	-2.47	99.72	3.59	0.21	99.86
			Cloudy	7.97	-6.23	99.61	5.37	-2.17	99.55	3.71	-1.22	99.81
	Summer	Case 2	All	4.34	-2.57	99.85	3.60	2.08	99.93	2.64	0.81	99.90
		Case 3	Sunny	4.06	-2.36	99.85	3.51	1.15	99.17	2.27	0.73	99.91
			Cloudy	4.87	-3.30	99.81	3.19	-1.68	99.73	2.59	0.23	99.82
	Spring/autumn	Case 2	All	6.17	-4.04	99.82	3.36	2.88	99.86	3.18	-0.45	99.88
		Case 3	Sunny	4.83	-4.01	99.88	3.26	1.11	99.85	2.81	-0.11	99.89
			Cloudy	7.10	-4.08	99.64	3.27	1.22	99.48	3.06	-0.37	99.87
Omitted wind direction	Winter	Case 2	All	/	/	/	5.96	-2.83	99.41	4.69	-2.16	99.65
		Case 3	Sunny	/	/	/	5.60	-1.69	98.75	3.40	-1.05	99.46
			Cloudy	/	/	/	5.34	-2.37	98.84	4.09	-2.15	99.63
	Summer	Case 2	All	/	/	/	3.22	1.51	99.77	2.51	1.46	99.84
		Case 3	Sunny	/	/	/	3.19	1.97	99.75	2.23	0.33	99.83
			Cloudy	/	/	/	3.10	-1.80	99.77	2.36	-1.02	99.78
	Spring/autumn	Case 2	All	/	/	/	3.32	2.42	99.83	3.02	-0.28	99.87
		Case 3	Sunny	/	/	/	3.08	-1.67	99.61	2.65	-0.15	99.89
			Cloudy	/	/	/	3.15	1.21	99.35	2.48	-0.24	99.85

Furthermore, the obtained correlation coefficient values,  $\rho$ , are almost the same as the ones previously obtained in Case 2.



**Figure 6.** Sunny/cloudy classification of summer days in *Validation phase*

shown in fig. 7. The given results include two summer days, classified as sunny, denoted as dark blue bar in fig. 6: August 24, 2014, mostly sunny and August 30, 2014, the sunniest day of the month.



**Figure 7.** Samples of prediction results (*Validation phase*) and comparison with measurements for *Gray-box model* trained with differently classified input/output data (for colour image see journal web site)

The daily statistics error for *Gray-box model* which was trained with differently classified input/output data on PV generator maximum output power is shown in tab. 3.

**Table 3.** Daily statistics error of prediction results (*Validation phase*) for *Gray-box model* trained with differently classified input/output data

Day	Case 1			Case 2			Case 3'		
	RMSE [%]	MBE [%]	$\rho$ [%]	RMSE [%]	MBE [%]	$\rho$ [%]	RMSE [%]	MBE [%]	$\rho$ [%]
August 24, 2014.	3.94	0.71	99.89	3.26	-0.44	99.87	2.93	0.08	99.87
August 30, 2014.	3.43	-0.55	99.94	2.33	0.15	99.93	2.06	-0.05	99.93

The results presented in the tab. 3 suggest that the predicted and measured values for Case 2 and Case 3 are in better agreement, compared with non-classified input/output data (Case1). Also, *RMSE* and *MBE* for the partially sunny day are higher than for the sunniest day

Figure 6 shows sunny/cloudy classification of summer days in *Validation phase*. The criterion for classifying days into sunny or cloudy is obtained from the *Training phase*, ( $0.5G_{G,hor,s}^{ad,max} = 205 \text{ W/m}^2$  per day, see fig. 5b) and denoted as a horizontal solid line in fig. 6.

The characteristically samples of PV generator maximum output power prediction results (*Validation phase*) obtained using *Gray-box model* which was trained with differently classified input/output data are

of the month. For each case analyzed separately, the *RMSE* results are in the range of less than 1%. These results, as well as their comparison with results for seasonal *RMSE* shown in tab. 2 (marked in tabs. 2 and 3), indicate that the errors of the predicted PV generator maximum output power are within an acceptable range for both analyzed days.

## Conclusions

A new approach for predicting output power of PV module has been presented. The proposed model consists of the physical principles-based model and ANN-based functional approximator. The convenience of this model is that the functional approximator part would allow using additional input environmental variables. The test results demonstrated that the gray-box model exceeds the accuracy of both mathematical and ANN model.

In addition, the two criteria for the classification of the daily input/output data are used to improve the accuracy of gray-box model: season-based criterion and the one classifying days into sunny and cloudy. The validation of the classification methods is carried out by seasonal analysis using the standard statistical measures: *RMSE*, *MBE*, and correlation coefficient. The obtained results show that gray-box model combined with both classification methods helps to determine the output power of PV module more accurately.

The application of the proposed gray-box model can be significant for increasing profit by appropriate construction of solar energy systems suitable to the climate of a particular site.

## Acknowledgment

The authors hereby express their sincere gratitude to the Ministry of Education and Science of the Republic of Serbia for their support to this work provided within projects III-42009.

## References

- [1] Chenni, R., *et al.*, A Detailed Modeling Method for Photovoltaic Cells, *Energy*, 32 (2007), 9, pp. 1724-1730
- [2] Karatepe, E., *et al.*, Neural Network Based Solar Cell Model, *Energy Conversion and Management*, 47 (2006), 9-10, pp. 1159-2871
- [3] Castaner, L., Silvestre, S., *Modelling Photovoltaic Systems Using PSpice*, John Wiley & Sons, 1<sup>st</sup> ed., Chichester, UK, 2002
- [4] Nishioka, K., *et al.*, Analysis of Multicrystalline Silicon Solar Cells by Modified 3-Diode Equivalent Circuit Model Taking Leakage Current through Periphery into Consideration, *Solar Energy Materials and Solar Cells*, 91 (2007), 13, pp. 1222-1227
- [5] Celik, A. N., Acikgoz, N., Modelling and Experimental Verification of the Operating Current of Monocrystalline Photovoltaic Modules Using Four- and Five-Parameter Models, *Applied Energy*, 84 (2007), 1, pp. 1-15
- [6] Dolara, A., *et al.*, Comparison of Different Physical Models for PV Power Output Prediction, *Solar Energy*, 119 (2015), 1, pp. 83-99
- [7] Mekhilefa, S., *et al.*, Effect of Dust, Humidity and Air Velocity on Efficiency of Photovoltaic Cells, *Renewable and Sustainable Energy Reviews*, 16 (2012), 5, pp. 2920-2925
- [8] Ali, H. M., *et al.*, Effect of Dust Deposition on the Performance of Photovoltaic Modules in Taxila, Pakistan, *Thermal Science*, 21 (2017), 2, pp. 915-923
- [9] Omubo-Pepple, V. B., *et al.*, Effects of Temperature, Solar Flux and Relative Humidity on the Efficient Conversion of Solar Energy to Electricity, *European Journal of Scientific Research*, 35 (2009), 2, pp. 173-180
- [10] Mellit, A., *et al.*, Artificial Neural Network-Based Model for Estimating the Produced Power of a Photovoltaic Module, *Renewable Energy*, 60 (2013), 1, pp. 71-78
- [11] Almonacid, F., *et al.*, Characterisation of PV CIS Module by Artificial Neural Networks, *Renewable Energy*, 35 (2010), 5, pp. 973-980

- [12] Chen, C., et al., Online 24-h Solar Power Forecasting Based on Weather Type Classification Using Artificial Neural Network, *Solar Energy*, 85 (2011), 11, pp. 2856-2871
- [13] Chen, S. X., et al., Solar Radiation Forecast Based on Fuzzy Logic and Neural Networks, *Renewable Energy*, 60 (2013), 1, pp. 195-201
- [14] Ji, W., Chee, K. C., Prediction of Hourly Solar Radiation Using a Novel Hybrid Model of ARMA and TDNN, *Solar Energy*, 85 (2011), 5, pp. 808-817
- [15] Huang, C., et al., Improvement in Artificial Neural Network-Based Estimation of Grid Connected Photovoltaic Power Output, *Renewable Energy*, 97 (2016), 1, pp. 838-848
- [16] Aybar-Ruiz, A., et al., A Novel Grouping Genetic Algorithm-Extreme Learning Machine Approach for Global Solar Radiation Prediction from Numerical Weather Models Inputs, *Solar Energy*, 132 (2016), 1, pp. 129-142
- [17] Al-Messabi, N., et al., Grey-Box Identification for Photovoltaic Power Systems via Particle Swarm Algorithm, *Proceedings*, 21<sup>st</sup> International Conference on Automation and Computing (ICAC), Glasgow, UK, 2015, Vol. 1, pp. 1-7
- [18] Al-Messabi, N., et al., Heuristic Grey-Box Modelling for Photovoltaic Power Systems, *Systems Science & Control Engineering*, 4 (2016), 1, pp. 235-246
- [19] Tan, K. C., Li, Y., Grey-Box Model Identification via Evolutionary Computing, *Control Engineering Practice*, 10 (2002), 7, pp. 673-684
- [20] Villalva, M. G., et al., Comprehensive Approach to Modeling and Simulation of Photovoltaic Arrays, *IEEE Transactions on Power Electronics*, 24 (2009), 5, pp. 1198-1208
- [21] Ranković, A., Sarić, A. T., Modeling of Photovoltaic and Wind Turbine Based Distributed Generation in State Estimation, *Proceedings*, 15<sup>th</sup> International Power Electronics and Motion Control Conference and Exposition – EPE-PEMC 2012, Novi Sad, Serbia, 2012, Vol. 1, pp. 1-6
- [22] Qi, C., Ming, Z., Photovoltaic Module Simulink Model for a Stand-alone PV System, *Proceedings*, International Conference on Applied Physics and Industrial Engineering (Physics Procedia), Amsterdam, The Netherlands, 2012, Vol. 24, pp. 94-100
- [23] Di Piazza, M. C., et al., Identification of Photovoltaic Array Model Parameters by Robust Linear Regression Methods, *Proceedings*, International Conference on Renewable Energies and Power Quality (ICREPO'09), Valencia, Spain, 2009, Vol. 1, pp. 1-7
- [24] Reindl, D. T., et al., Diffuse Fraction Correlations, *Solar Energy*, 45 (1990), 1, pp. 1-7
- [25] Myers, D., *Solar Radiation: Practical Modeling for Renewable Energy Applications*, 1<sup>st</sup> ed., CRC Press/Taylor & Francis Group, Boca Ration, Fla., USA, 2013
- [26] Quasching, V., *Understanding Renewable Energy Systems*, 2<sup>nd</sup> ed., Earthscan Publications Ltd., London, 2016
- [27] Klucher, T. M., Evaluation of Models to Predict Insolation on Tilted Surfaces, *Solar Energy*, 23 (1979), 2, pp. 111-114
- [28] Hagan, M. T., et al., *Neural Network Design*, 2<sup>nd</sup> ed., Martin Hagan Publisher, Oklahoma State University, Stillwater, Okla., USA, 2014
- [29] Badescu, V. *Modeling Solar Radiation at the Earth's Surface*, 1<sup>st</sup> ed., Springer-Verlag, Berlin, 2008
- [30] Sharples, S., Charlesworth, P. S., Full-Scale Measurements of Wind-Induced Convective Heat Transfer from a Roof-Mounted Flat Plate Solar Collectors, *Solar Energy*, 62 (1998), 2, pp. 69-77