EREM 81/2
Journal of Environmental Research, Engineering and Management
Vol. 81 / No. 2 / 2025
pp. 64–77
10.5755/j01.erem.81.2.38542

Use of Regression Models for Estimation of Electric Power Generation by Photovoltaic Power Plants

Received 2024/08 | Accepted after revisions 2025/05

https://doi.org/10.5755/j01.erem.81.2.38542

Use of Regression Models for Estimation of Electric Power Generation by Photovoltaic Power Plants

Andriy Bandura[1]*, Yaroslav Batsala[2], Petro Kurliak[2]

[1] Department of Advanced Mathematics, Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine

[2] Department of Electrical Power Engineering, Electrical Engineering, and Electromechanics, Ivano-Frankivsk National Technical University of Oil and Gas, Ukraine

***Corresponding author:** andriykopanytsia@gmail.com

The choice of an approach for accurate forecasting of photovoltaic power plants and modeling of power systems with renewable energy sources depends on the availability of input data, time horizon, installation location, and weather variables. Our goal is to improve mathematical models and find new solutions to improve the performance of predicting the operation of photovoltaic power plants in energy systems using regression models. There is a problem of predicting the amount of electricity generated by photovoltaic plants in Ukraine. The data for 3 years of daily electricity production is used. The problem is solved by the application of the least squares' method to estimate unknown parameters of the suggested dependence between the length of daylight and daily solar power generation. The main assumption is the following: daily solar power generation can be given as a linear combination of some exponential polynomials with the independent variable as the duration of the sunny day. The regression model is reduced to a system of significantly non-linear equations, which is solved numerically by the iteration method. Regression models were built using R software for big data analysis. Another novelty moment concerns grouping of the data according to the same length of daylight, and then three values were found for each such group: maximum, minimum, and average value. The proposed moving average regression models with the usage of exponential polynomials as approximating functions admit a small standard residual error between the exact values and the predicted values of solar power generation (0.3022 kWh). The forecasting horizon is one year. The significance of the created mathematical forecasting models demonstrates a possibility of using the daylight duration as a parameter in forecasting tasks, as well as evaluating the prospects to consider the parameters that affect the performance of photovoltaic power plants.

Keywords: photoelectric power plant, regression forecasting model, length of daylight, cloudiness factor, exponential polynomial.

Introduction

Most of the world's energy is changing as the share of solar and wind power in the energy system increases. These generation sources are characterized by the variation and variability of the amount of electricity that is supplied to the grid depending on the generation time. The main factors affecting the production and efficiency of photovoltaic power plants are variables of the sunlight during the daytime, temperature and humidity (Erten and Aydilek, 2022; Meghea, 2023; Meghea and Mihai, 2018). In the case of using time variable, it is possible to get mathematical dependencies for evaluation of their effect on main parameters. The obtained time dependencies also allow determining the exact limits of changes in the generation capacity of photovoltaic stations (PES) for improved predictions used by operators in energy markets.

In Ukraine, the importance of balancing loads due to missile attacks on the power system has increased. In February 2022, Russia attacked Ukraine and started a full-scale war. From September 2022 to the present, the Russian army has carried out more than a thousand strikes on thermal and hydroelectric power plants, substations and gas storage facilities, and occupied the Zaporizhzhia nuclear power plant (its capacity is 6 GW). Most of the power units of thermal power plants and the Kakhovka hydroelectric power plant were destroyed, and the total loss of installed capacity of the Ukrainian power system is more than 20 GW (for comparison, peak consumption in Ukraine per day is 14–20 GW), although energy workers are restoring power equipment every day. Most often, the generation of photovoltaic power plants is limited (about 500 MW). Depending on the balance of electricity consumption and generation, operators give commands to balancing group and individual power plants to reduce generation during certain hours. This leads to the loss of their profits (Kim et al., 2021; Batsala et al., 2021a; Batsala et al., 2023). Operators turn off some grid loads manually at local power plants or automatically turn off inverters, and limit inverter capacity where it is possible. The issue of power loss can be partially resolved by using storage systems, but their cost does not allow for quick results. Therefore, the role of accurate forecasts in balancing groups is growing.

The laws «About the EU Battery Regulation» and «On the electricity market» have become a big incentive for scientific research. They encourage the use of local energy storage systems for energy market participants and introduce the need to predict hourly electricity consumption for a day ahead.

Artificial intelligence (Bracale et al., 2013; Chen et al., 2011) and neural networks (Leva et al., 2017; Dolara et al., 2015) are popular and known forecasting methods. NASA uses a huge amount of satellite data for weather forecasting as well as forecasting electricity generation. Through data science and machine learning, users gain access to modern predictive models that use different inputs. When input parameters change, the system applies "overfitting" to refine output parameters. A separate group of forecasting methods is generated by support vector machines (De Giorgi et al., 2016; Lin and Pai, 2016), splines (Massidda and Marrocu, 2017; Wang et al., 2016), wavelets (Zhu et al., 2016; Luo et al., 2021), and other optimization methods (Das et al., 2018, El Hendouzi and Bourouhou, 2020). Since weather conditions themselves are subject to the chaos theory, insolation changes chaotically from day to day. Therefore, ordinary differential equations (Liang et al., 2020) and partial differential equations (Said et al., 2021) are rarely used for a prediction of electric power generation of photovoltaic stations.

Regression forecasting models (Zhang et al., 2015, Pedro and Coimbra, 2012) have been often used for various tasks in energy, such as predicting energy consumption. For example, Erten and Aydilek (2022) have demonstrated the results of using four different regression models to predict solar energy: linear regression, logistic regression, Lasso regression and elastic regression. The use of different input data was justified, including wind speed, sun position, temperature, direct irradiation, diffuse irradiation, reflected irradiation, relative humidity, ambient temperature, accumulated dust, dew point temp, total cloud cover, date, time, radiation, sun rise time and sun set time, percent cloud, solar azimuth and elevation, and dataset clear sky index. Many input parameters increase the calculation time, and some functions can reduce the accuracy of the forecast. To decrease the time of calculation, it is possible to use genetic algorithms (Ratshilengo et al., 2021).

Meghea (2023) and Meghea and Mihai (2018) have used the division of experimental data into stationary time series with seasonality using the CUSUM method (cumulative sum control chart). The considered time series were formed by hourly data, and this was used to forecast the average specific power. The data were smoothed out using a moving average regression model and forecast specific power for the next day, next week and next month.

In a study by Kim et al. (2021), 12 regression models were developed for monthly data, and the calculation of electricity generation by a solar power plant when weather conditions change was estimated using R language software. The dependent variable was the daily generation of electricity by the solar power plant in kWh, and the independent variables were the intensity of insolation during daylight hours (MJ/m^2), light time (h), average relative humidity (%), minimum relative humidity (%), and amount of evaporation (mm). Prediction for photovoltaic stations can be performed using harmonic functions, using autoregression-integral moving average (ARIMA) models, as well as considering weather correlation coefficients (Batsala et al., 2021a). Combining forecast models of photovoltaic station performance with accurate weather forecasts and analysis of electricity quality allows increasing the efficiency of photovoltaic stations by controlling the main parameters of local networks (Batsala and Hlad, 2023). Comparison of approximation of electric power generation curves using the trigonometric function and the unction of normal distribution law for a day ahead forecast is shown in other scholars' works (Batsala et al., 2021b).

In a study by Mei et al. (2018), the authors developed a hybrid online model for ultra-short-term prediction of the production of photovoltaic stations for network dispatching every 5 minutes. The online model uses weather monitoring data, which is divided into four typical weather conditions (sunny, cloudy, rainy weather and cloudy day). In online prediction, autoregression (ARIMA) was used to predict solar irradiation. The accuracy of forecast weather models is highly dependent on the sensitivity of the weather forecast input. Leva et al. (2017) have shown a method of predicting photovoltaic energy based on CNM using the cloud index. The index affects the accuracy of the forecast depending on the chosen training data set. More extended literature reviews on modern investigations in the topic have been published (Iheanetu, 2022; Tuohy et al., 2015; Raza et al., 2016).

The splines (Massidda and Marrocu, 2017; Wang et al., 2016) can also be referred to as regression methods. Specifically, the method of multivariate adaptive regression splines involves building a functional relationship as a set of coefficients and basic functions solely from available data using a "divide and conquer" strategy in which the input space is divided into regions and a regression equation is evaluated for each of these regions. The method allows the selection of appropriate levels of the complexity of the model. A simple model was presented with only two input parameters, namely the solar radiation on the plane of array and the total cloud cover albeit without any limitations on the maximum degree of the interpolating function. A more complex model was obtained using global horizontal radiation, total cloud cover, and values of pressure, temperature, and wind speed, in conjunction with the cosine of the angle of incidence between the sun rays and the surface.

Another study (Lin and Pai, 2016) has developed an evolutionary seasonal decomposition least-square support vector regression to forecast monthly solar power output. The construction of the regression used a seasonal decomposition and the least-square support vector regression, and its parameters were chosen by genetic algorithms.

Monitoring and management of a photovoltaic system is very important for sending information that allows owners to maintain, operate and control these systems in order to reduce maintenance costs and avoid unwanted electric power disruptions. Different monitoring and management systems are discussed (Bekirov et al., 2019; Beranek et al., 2018; Rahman et al., 2018; Reatti et al., 2019). The published works consider the following requirements: circuit complexity, availability of friendly graphical user interface, easy-to-understand system architecture, maintenance facility and customization possibility for an end-user. The most detailed review of various monitoring and management technologies with system attributes and working structures has been presented by Beranek et al. (2018) to get a clear view of merits and demerits of existing photovoltaic monitoring and management systems.

Recently, many researchers have implemented various predictive modeling techniques such as artificial neural networks, fuzzy predictions, and support vector regressions for photovoltaic generation. However, most of these models are not applicable in making accurate predictions because they do not have sufficient primary source data. This suggests that the predictability of the models could be improved if more raw data are accumulated. This study aimed to propose a non-linear regression model to easily predict solar power according to the changes in the duration of the sunny days and weather conditions. To achieve this objective, a non-linear regression analysis technique was applied to the big data on the solar power generation around the area where the solar power plant was installed. Non-linear regression has some benefits: regression is relatively simple to understand and realize, and it can be applied to model the relationship between a continuous outcome variable and one or more predictor variables. Moreover, it is widely used and well-understood, so there is a wealth of resources available for learning about it and using it effectively. Together with some benefits, the regression also has drawbacks. It can be sensitive to outliers, which can affect the estimated coefficients and the predictions made by the model. Given this, the robustness of the approach to noise can be low. But it can be higher for several types of regression analysis. For example, lasso regression (Erten and Aydilek, 2022) is a regression analysis method that uses a regularization term in the optimization process. The regularization term is a penalty applied to the coefficients of predictor variables in the model, which helps to prevent overfitting by reducing the complexity of the model. It is given by the l_1-norm. Lasso regression is particularly useful for selecting important features in a dataset since it tends to drive the coefficients of unimportant features to zero. In the context of solar power prediction, Lasso regression can be used to select the most important features in the dataset, which can improve the accuracy of predictions. Lasso regression uses the tuning positive parameter $t \geq 0$, which controls the amount of shrinkage. Elastic regression is a type of regression that combines the strengths of both lasso and ridge regression. Like lasso regression, it uses a regularization term in the optimization process to prevent overfitting. However, unlike lasso regression, which uses the l1 norm as the regularization term, elastic regression uses a linear combination of the l1 and l2-norms. This allows elastic regression to balance the trade-off between model complexity and goodness of fit, which can be beneficial in some situations.

Since weather patterns and locational atmospheric conditions vary considerably both spatially and temporally, solar forecasting accuracy is dependent on the geographic location and timescale of the data. Two key factors that impact the accuracy of solar forecasting are geographic locations and forecast timescales. There is a very detailed analysis (Zhang et al., 2015) of solar power plant generation at multiple geographic regions at multiple timescales to quantify the effects of geographic location and forecast horizon on the forecasting accuracy of regression methods. Different time and geographic scales influence the severity of up- or down-ramps in solar power output. Forecasting solar power can help reduce the uncertainty involved with the power supply. In addition, the distribution of errors at a larger geographic area has a more pronounced peak, slimmer shoulders, and longer tails (Zhang et al., 2015). This observation indicates that relative forecast errors are smaller for a large geographic area, which shows the smoothing effect from geographic diversity for solar. The normalized root mean square error values become smaller with an increasing geographic area, which shows that the solar forecast performs relatively better for larger regions. Another key gap in developing solar forecasting models is the unavailability of a consistent and robust set of metrics to measure and assess the improvement in forecasting accuracy, because different researchers use improvements described by different metrics as their own evaluation criteria. In addition, it is not clear whether the traditional statistical metrics used to evaluate forecasts best represent the needs of power system operators. One should be aware of a few important considerations when selecting the appropriate metrics for evaluating the performance of solar power forecasting of regression methods. First, because skewness and kurtosis are not standalone metrics, it is recommended that the metrics of mean bias error, standard deviation, skewness, kurtosis, and the distribution of forecast errors should be used as a group. In addition, it is important to select at least one metric from each class, determined through the nonparametric statistical testing. Thus, for a comprehensive, consistent, and robust assessment of the performance of solar power forecasts, a suite of metrics consisting of mean bias error, standard deviation, skewness, kurtosis, distribution of forecast errors, and Rényi entropy is recommended. One of the biggest concerns associated with integrating a large amount of solar power into the grid is the ability to handle large ramps in solar power output, which are often caused by cloud events and extreme weather events (Mills and

Wiser, 2010). Different time and geographic scales influence the severity of up- or down-ramps in solar power output. Forecasting solar power can help reduce the uncertainty and noise involved with the power supply. The swinging door algorithm was suggested to identify ramps over varying time frames because of its flexibility and simplicity (Florita et al., 2013). The swinging door algorithm extracts ramp periods in a series of power signals by identifying the start and end points of each ramp. The user sets a threshold parameter ε that influences the algorithm's sensitivity to ramp variations. The tunable parameter ε directly characterizes the threshold sensitivity to noise and/or insignificant fluctuations to be specified. With a smaller ε value, many small ramps will be identified; with a larger ε value, only a few large ramps will be identified.

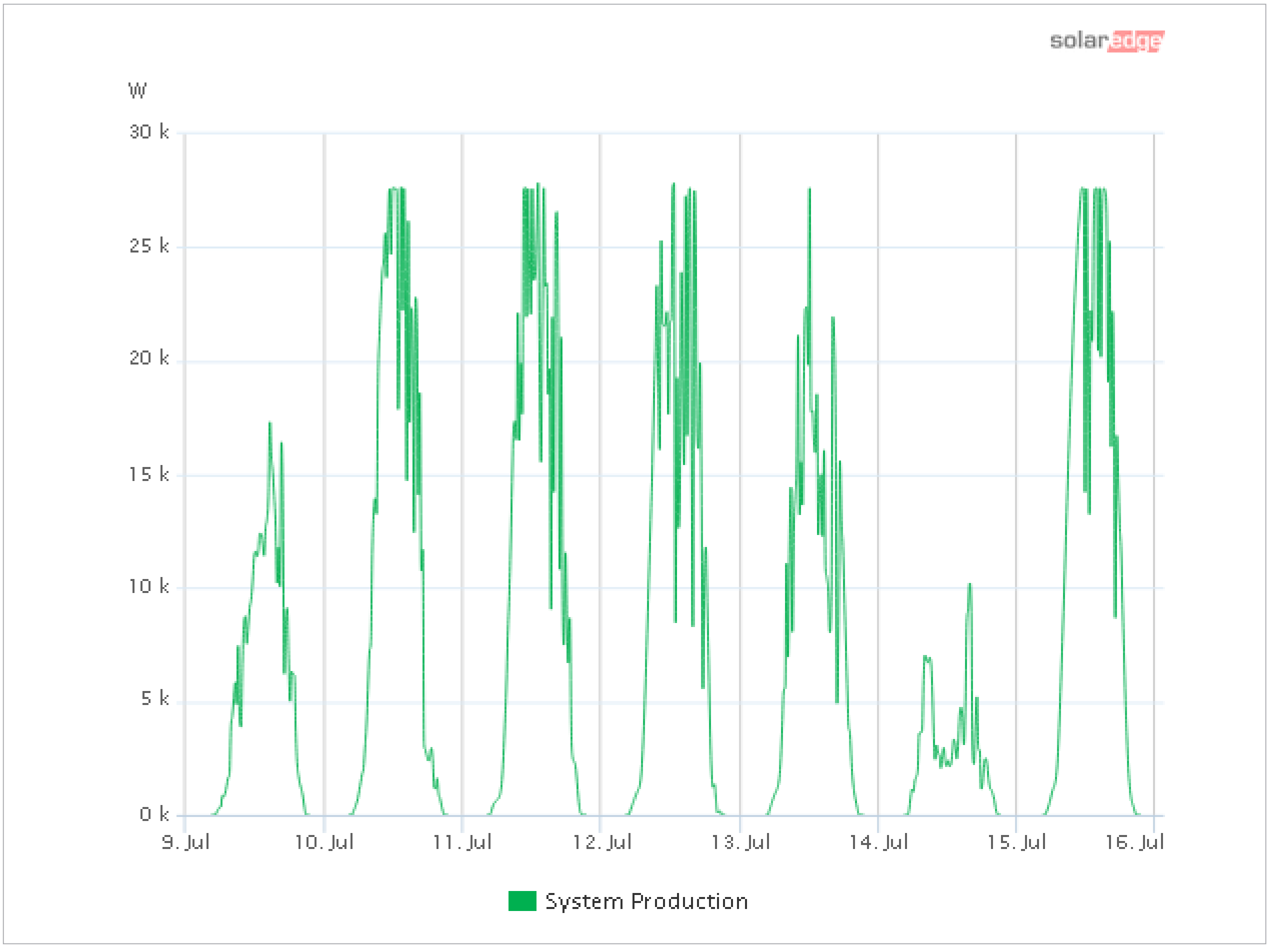
In view of this, the approach proposed in this article below requires further research on robustness to the noise with use of a suite of statistical metrics and the swinging door algorithm. Determining the necessary main input parameters for predicting the electricity production of photovoltaic stations can help to simplify forecast models. The selection of the solar day duration input parameter has not been used before. Determining the time interval of electricity production by photovoltaic plants helps to improve daily forecast models. Perhaps, the correctness of building the regression model that uses only the duration of a sunny day raises questions, since it does not use meteorological factors. Therefore, the accuracy of the model can be significantly lower than that of similar short-term solutions. However, this study does not aim to make traditional short-term forecasts. Its goal is to provide a long-term average forecast of electricity production by solar power plants throughout the year. Temperature, insolation, cloud cover, etc. are all weather factors that have a short duration. They change chaotically every day. Therefore, it is difficult to take them into account separately in long-term forecasting. For a particular area, weather factors have a certain average impact on solar electricity production over several years. At the same time, the length of daylight hours is a constant factor that has a periodic effect. In view of this, we tried to build a two-component model, based on the maximum values of electricity production, which refers to the best weather conditions for a given area, and on the average values of electricity production, which is due to the average weather conditions inherent in a particular area. Other research studies (Kim et al., 2021) show that such meteorological variables as the insolation intensity at the peak time (MJ/m^2), insolation intensity during daylight hours (MJ/m^2), daylight time (h), average relative humidity (%), minimum relative humidity (%), and amount of evaporation (mm) demonstrate a sufficiently strong correlation with solar power generation. In the study by Kim et al. (2021) the biggest correlation – 0.833 – was obtained for daylight time. Other mentioned factors showed correlations between 0.442 and 0.770. But their nature is completely random and/or chaotic. They are very sensitive to the initial data and a small error can cause significant changes, as is usually the case with meteorological factors. Given this, only a model using the duration of sunny days is considered.

Methods

Data sampling is a very important part of research. In this work, statistical data were obtained at photovoltaic stations located near the cities Ivano-Frankivsk and Kryvyi Rih for different periods of time: from 2014 to 2016 and from 2019 to August 2023. The first period was the training dataset, and the second period was the test dataset. In the paper, we calculated standard residual errors for training and test datasets separately to compare models and take into account overfitting. Grid-connected photovoltaic stations were selected for analysis. A typical photovoltaic plant includes solar modules (such as Chaori Solar, JinkoSolar, SUNOWE and others) and inverters (Fronius, Schneider, Huawei, KASO POWADOR). For photovoltaic plants with a capacity of more than 1 MW, inverters are connected to the low-voltage sections of transformer substations with a capacity of 1000 kVA. Photoelectric stations with lower power are connected directly to the 0.38 kV network. The productivity of power plants depends on the total amount of solar insolation for the given period.

The obtained statistics contain specific power values for each hour in each period, and all these data were reduced to a power of 1 kW. The data were collected using open platforms for monitoring the performance of photovoltaic stations (Solar Edge), as well as an electricity metering system from a local energy company. Half of the data correspond to zero at night. *Fig. 1* shows changes for a week in July 2023 in the performance of a 37-kW solar power plant in Zarichchia village, Ivano-Frankivsk region, Ukraine. This photovoltaic station

Fig. 1. *Daily generation of solar energy in July 2023 (37 kW photovoltaic power plant in Zarichchia village, Ivano-Frankivsk region, Ukraine. Figure source:) SolarEdge, (n.d.)*



uses Risen RSM120-330M solar panels and a Huawei SUN2000-33KTL-A inverter. On summer days with a reduced amount of solar insolation due to cloudiness, the performance of a photovoltaic station decreases.

For comparison, we used another 36 kW photovoltaic plant in the city of Kryvyi Rih, Ukraine. This photovoltaic plant on Tabyrna street in the city of Kryvyi Rih is designed for 66 Trina Solar panels with a capacity of 540 W, which are connected through the Huawei SUN2000-30KTL-M3 inverter. Below, approximating formulas (Eq. 1 and Eq. 2) are presented to estimate solar power generation for this station. Due to Russian attacks on the Ukrainian energy sector, all information on power plants is in closed access and not publicly available. Therefore, it is impossible to give a more comprehensive and deep analysis for diverse geographical locations in Ukraine.

Fig. 2 demonstrates a weekly schedule of daily electricity production at this power plant in December 2022. The maximum power of the power plant is reduced by almost 10 times, and due to a decrease in the duration of a sunny day, the productivity of the power plant is even lower.

To build the model, data for 3 years of daily electricity production by a solar power plant depending on the duration of daylight hours were used. As a result of mathematical processing of statistical data, these indicators were grouped by the same duration of daylight hours with an accuracy of up to a minute, and then an average value was found for each group. Accordingly, 278 different durations were obtained.

Figs. 3, 4 and *5* show 3 diagrams obtained after the initial grouping and estimation of the specified values (maximum, average and minimum values).

Fig. 2. *Daily solar power generation in December 2022 (37 kW photovoltaic power plant in Zarichchia village, Ivano-Frankivsk region, Ukraine. Data source: SolarEdge, (n.d.)*

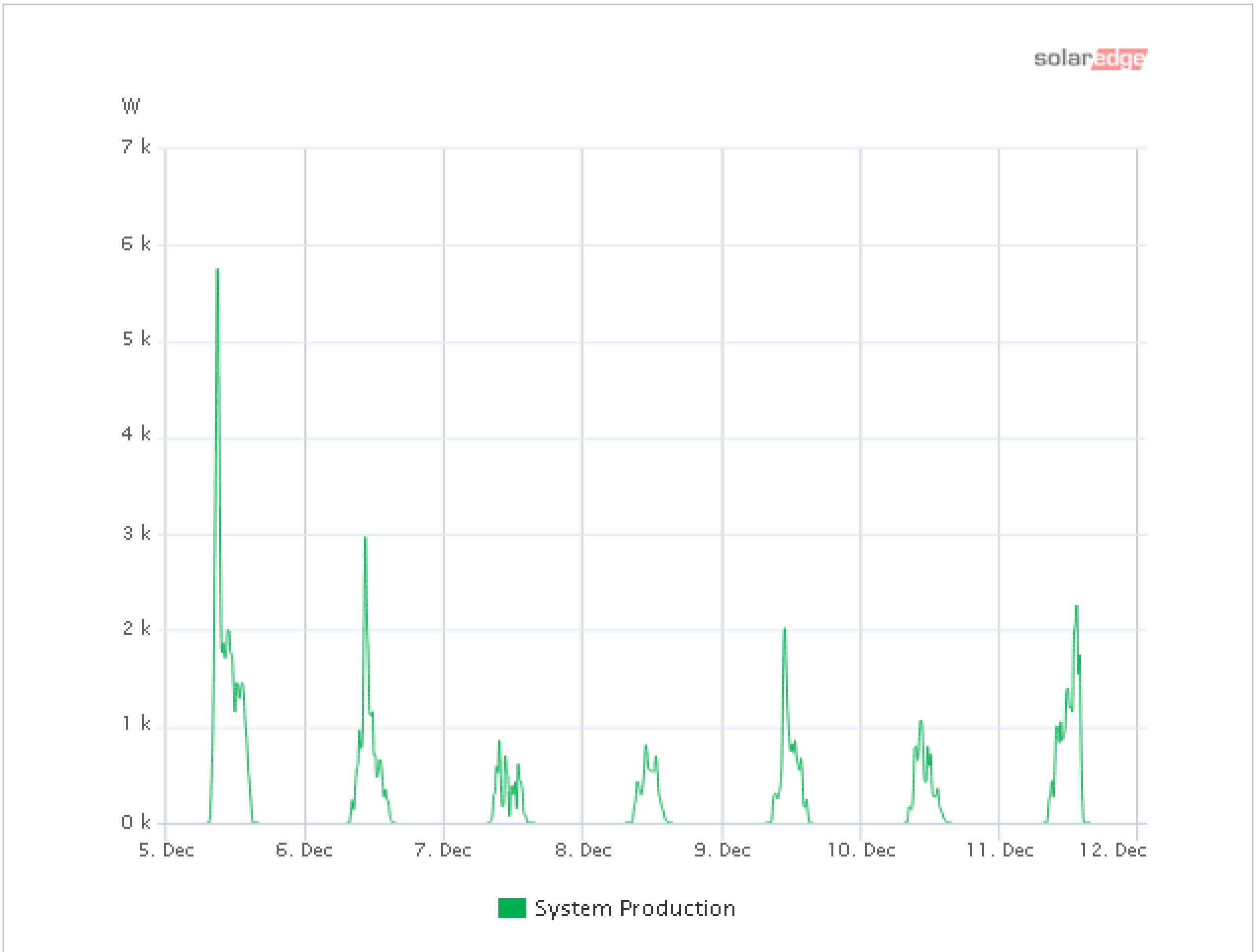


Fig. 3. *Maximum electricity production by the photovoltaic plant depending on the length of daylight (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*

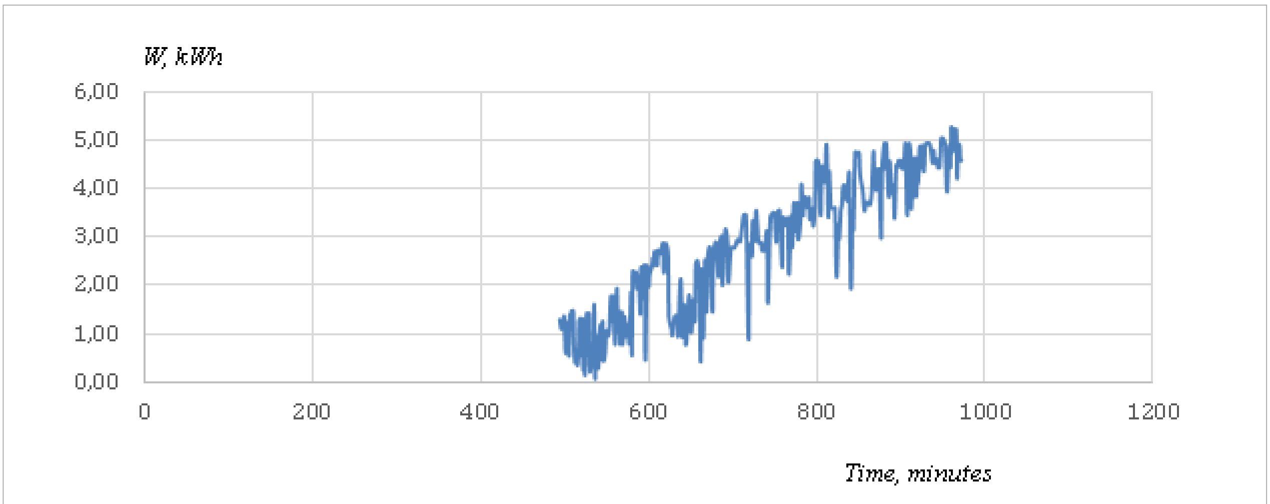


Fig. 4. *Average electricity production by the photovoltaic plant depending on the length of daylight (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*

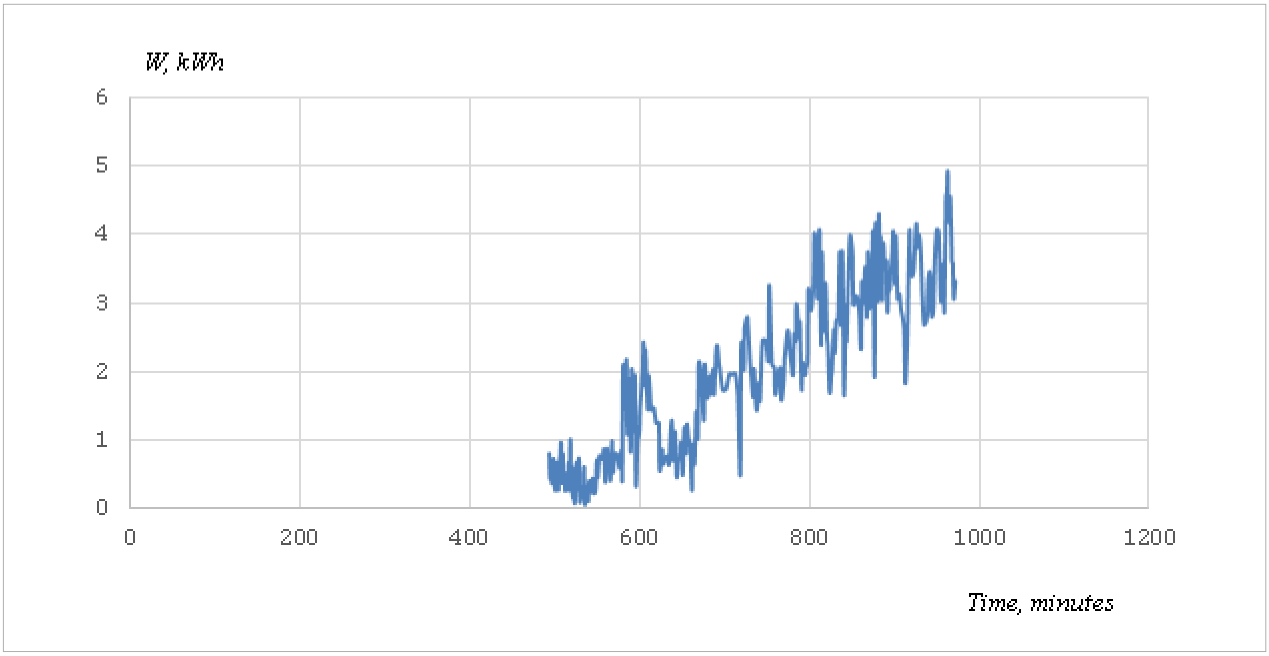
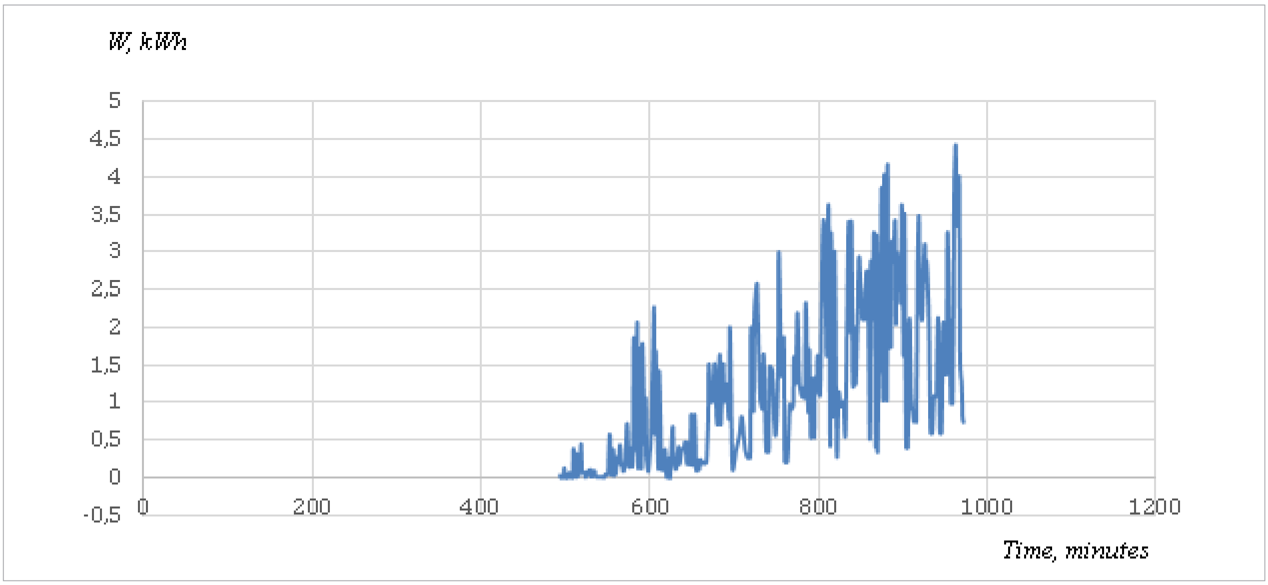


Fig. 5. *Minimum electricity production by the photovoltaic plant depending on the length of daylight (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*



From these diagrams, even the highs, lows and averages of the electricity generated are subject to various perturbations and it is difficult to find a successful curve that would best describe them. However, the least squares method was applied to determine the parameters of such a function, which would provide a minimum sum of squares of the differences between the predicted and real fluctuations in production. The initial assumptions about the shape of such a function were:

1 due to fluctuations in the diagram, the search function should contain periodic functions, such as cosines and sines;

2 due to the gradual growth of graphs, an unknown function must also contain polynomials, as factors of periodic functions, and a power function, as terms.

Below, regression models (1), (4) which overlap conditions 1)-2) are presented. Mathematically, these curves are entire functions of the exponential type. Moreover, they have a bounded index (for more details on fine properties of this function class see Bandura, 2017; Bandura and Skaskiv, 2017).

The data of power generation will be fitted with such a curve of power generation:

$$y = (a_1x + a_2)sin(a_3x + a_4) + a_5x + a_6x^{a_7} \quad (1)$$

where y is the electricity production, x is the duration of daylight hours, and a_j are the unknown real parameters. They were estimated by the least squares method, j ∈ {1,2,...,7}.

Results and Discussion

One should take into consideration that the main criteria to select models are a minimized standard residual error and a maximized coefficient of determination. For example, these values were calculated for simpler regressions: a linear function $y = ax + b$, a power function $y = a_1x^{a_2}$, an exponential function $y = a_1e^{a_2x}$, a logarithmic function $y = a_1 + a_5\ln x$, the first degree trigonometric polynomial $y = (a_1x + a_2)\sin(a_3x + a_4)$, a sum of the trigonometric polynomial and linear function $y = (a_1x + a_2)\sin(a_3x + a_4) + a_5x$, and, at last, the function from Eq. (1). Consistently applying the procedure described below, the parameters and statistical characteristics of the above functions were estimated until we arrived at formula (1).Since the corresponding calculations are rather cumbersome due to the complex form of the function, and the corresponding system of 7 equations compiled for the least squares method l are also nonlinear (primarily the parameters for sine and the exponent for x), the authors decided to use the software implementation of this method for nonlinear expressions. One such option is the nls procedure in the R language and the Rstudio environment for statistical analysis and data processing. Running this procedure for maxima, using these initial approximations – $a_1 = 100$, $a_2 = 200$, $a_3 = 0.1$, $a_4 = 200$, $a_5 = 1$, $a_6 = 1000$, $a_7 = 0.4$ – and formula (1) resulted in the parameter estimates presented in *Table 1*.

Substituting this data in formula (1), we get a function describing the predicted electricity production

$$y = (0.0006479\,\mathrm{x} - 0.6055)\sin(0.1058x + 197) + \\ + 0.007581x - 32.02x^{-0.3768} \quad (2)$$

with a standard residual error of 0.5597 kWh for the training dataset and 0.5732 kWh for the test dataset, which for a scale of 6 parameters is not too large. Moreover, the R^2 equals 0.7612 in this case. It is calculated for Zarichchia, Ivano-Frankivsk region, Ukraine. Similar results for Kryvyi Rih are presented as:

$$y = (0.0006951\,\mathrm{x} - 0.5011)\sin(0.1113x + 207) + \\ + 0.008011x - 31.12x^{-0.3812} \quad (3)$$

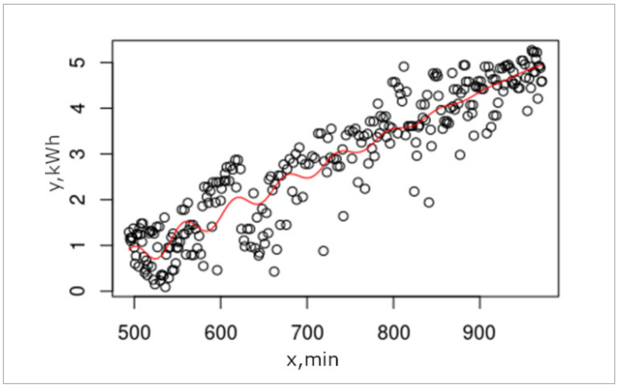
with a standard residual error of 0.4597 kWh for the training dataset and 0.5106 kWh for the test dataset, which for a scale of 6 parameters is not too large, and the R^2 equals 0.7912.

If we plot the function from formula (2) and plot the production data and the duration of daylight hours in the form of points, we obtain the diagram shown in *Fig. 6*.

Table 1. *Estimations of parameters for maxima and formula (1)*

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
a_1	$6.479\cdot10^{-4}$	$3.084\cdot10^{-4}$	2.101	0.0366
a_2	$-6.055\cdot10^{-1}$	$2.288\cdot10^{-1}$	−2.646	0.0086
a_3	$1.054\cdot10^{-1}$	$3.534\cdot10^{-3}$	29.946	$2\cdot10^{-16}$
a_4	$1.970\cdot10^{2}$	2.107e+00	93.514	$2\cdot10^{-16}$
a_5	$7.581\cdot10^{-3}$	$1.417\cdot10^{-3}$	5.351	$1.83\cdot10^{-7}$
a_6	$-3.202\cdot10^{1}$	$9.510\cdot10^{1}$	−0.337	0.551339
a_7	$-3.768\cdot10^{-1}$	$5.112\cdot10^{-1}$	−0.737	0.165890

Fig. 6. *Maximum electric power generation by the photovoltaic station vs the duration of daylight hours and the forecast function (red line) with one harmonic (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*



The current goal is to reduce the standard error of residuals. To do this, some parameters were searched for such a curve for electricity production:

$$y = (a_1x + a_2)\sin(a_3x + a_4) + a_5x + a_6x^{a_7} + \\ + (a_8x + a_9)\sin(a_{10}x + a_{11}) \quad (4)$$

where y is electricity production, x is the duration of daylight hours, a_j are unknown real parameters, to be estimated by the least squares meth od, j∈{1,2,...,7}. By running nls procedures for maxima with initial approximations of a_1 = 100, a_2 = 200, a_3 = 0.1, a_4 = 200, a_5 = 1, a_6 = 1000, a_7 = 0.4, a_8 = 300, a_9 = 100, a_{10} = 0.17, and a_{11} = 300, the parameter estimates were obtained.

The function describing maximum electric power generation takes the form for Zarichchia:

$$y = (0.0007572\,\mathrm{x} - 0.6795)\sin(0.1064x + 196.7) + \\ + 0.007449x - 41.02x^{-0.4199} + \\ +(-0.0001653x + 0.05109)\sin(0.1571x + 41.12) \quad (5)$$

The standard residual error is 0.5613 kWh for the training dataset and 0.5135 kWh for the test dataset. In addition, the R squared equals 0.8215.

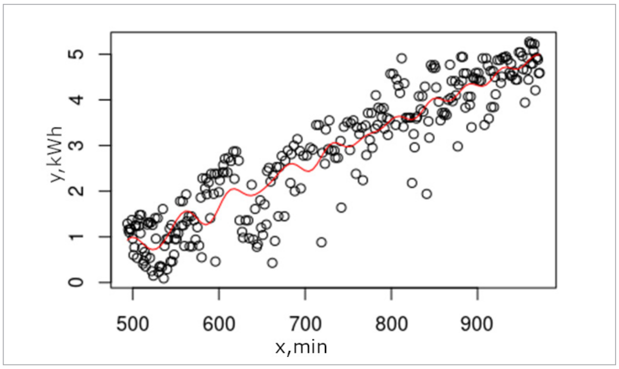
The corresponding function takes the form for Kryvyi Rih:

$$y = (0.0007812\,\mathrm{x} - 0.6805)\sin(0.1142x + 5.9) + \\ + 0.007812x - 40.22x^{-0.4013} + \\ +(-0.0001653x + 0.05109)\sin(0.1571x + 1.34) \quad (6)$$

The standard residual error is 0.4812 kWh. In addition, the R squared equals 0.8753.

Its plot has the form shown in *Fig. 7*.

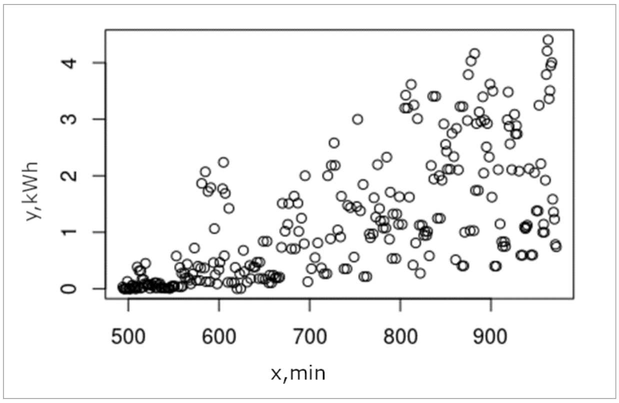
Fig. 7. *Maximum electric power generation by the photovoltaic station by duration of daylight hours and forecast function (red line) with two harmonics (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*



Due to the increase in the error of the residuals, it is obvious that this option of clarifying the formula does not yield any improvement.

If a point diagram for minimal electricity production depending on the duration of a bright day is constructed, then a distribution shown in *Fig. 8* is obtained.

Fig. 8. *Minimum electricity production by the photovoltaic station by duration of daylight hours (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*



Due to the large loosening of the data, including many days with zero production, the authors did not select the parameters of formula (1) by the least squares method in this case.

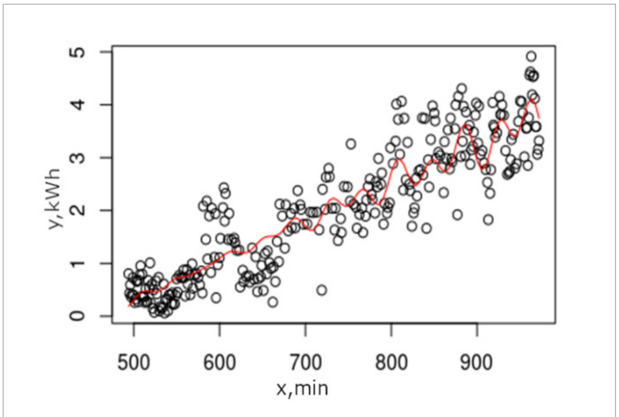
In view of this, the original formula (1) was further investigated for the average values of electricity production. According to the scheme described above, the least squares method (nls procedure) was used to obtain parameter estimates for average electricity production by the duration of days with two harmonics.

At initial approximation of $a_1 = 100$, $a_2 = 200$, $a_3 = 0.1$, $a_4 = 200$, $a_5 = 1$, $a_6 = 1000$, $a_7 = 0.4$, $a_8 = 300$, $a_9 = 100$, $a_{10} = 0.17$, $a_{11} = 30$, the corresponding prediction function from formula (4) takes the form for Zarichchia:

$$y = (0.0005861\,\mathrm{x} - 0.3143)\sin(0.092x + 206.6) + \\ + 0.006502x - 10.73x^{-0.2078} + (-0.001073x + \\ + 0.6211)\sin(0.1593x + 39.51) \quad (7)$$

The standard residual error of 0.5173 kWh for the training dataset and 0.4921 kWh for the test dataset. Also, the coefficient of determination R^2 is equal to 0.8956. A point diagram with a plot of the function is shown in *Fig. 9*.

Fig. 9. *Average electricity production by the photovoltaic station by duration of daylight hours and forecast function (red line) with two harmonics (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*



Finally, the advantages of the smoothing of fluctuations in electricity production due to the moving average were used. To do this, a seven-day moving average for average data on daylight durations was obtained as the usual arithmetic mean without additional smoothing coefficients.

$$y_t = \frac{1}{7}\sum_{i=0}^{6} x_{t+i} \quad (8)$$

where x_{t+i} is the average electricity production for the corresponding daylight hours; y_t is the calculated moving average.

Therefore, nls was run again for function (1) with these initial values: $a_1 = 100$, $a_2 = 200$, $a_3 = 0.1$, $a_4 = 200$, $a_5 = 1$, $a_6 = 1000$, $a_7 = 0.4$.

Estimations obtained are presented in *Table 2*.

Substituting these data in (1), a function describing the predicted electricity production was obtained for Zarichchia:

$$y = (0.0004747\,\mathrm{x} - 0.4806)\sin(0.1054x + 197.2) + \\ + 0.007559x - 37.04x^{-0.39999} \quad (9)$$

at a standard residual error of 0.3023 kWh for the training dataset and 0.2741 kWh for the test dataset. In the last case, the coefficient of determination R^2 is 0.9752.

The results for Kryvyi Rih are the following:

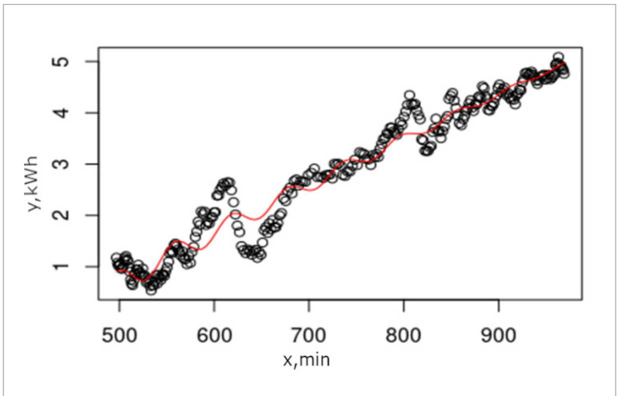
$$y = (0.0004981\,\mathrm{x} - 0.4901)\sin(0.1023x + 2.3) + \\ + 0.007819x - 36.12x^{-0.381} \quad (10)$$

The standard residual error is 0.2504 kWh for the training dataset and 0.2714 kWh for the test dataset. In the last case, the coefficient of determination R^2 is 0.9832.

Table 2. *Estimations of parameters for 7 day moving average and formula (1)*

Parameter	Estimate	Std. Error	t value	Pr(>|t|)
a_1	$4.747 \cdot 10^{-4}$	$1.756 \cdot 10^{-4}$	2.702	0.007318
a_2	$4.806 \cdot 10^{-1}$	$1.296 \cdot 10^{-1}$	−3.710	0.000252
a_3	$1.054 \cdot 10^{-1}$	$1.737 \cdot 10^{-3}$	60.673	$2 \cdot 10^{-16}$
a_4	$1.972 \cdot 10^{2}$	1.078	182.904	$2 \cdot 10^{-16}$
a_5	$7.559 \cdot 10^{-3}$	$7.789 \cdot 10^{-4}$	9.705	$2 \cdot 10^{-16}$
a_6	$-3.704 \cdot 10^{1}$	$6.210 \cdot 10^{1}$	−0.596	0.551339
a_7	$-3.999 \cdot 10^{-1}$	$2.879 \cdot 10^{-1}$	−1.389	0.165890

Fig. 10. *7-day moving average of electricity production by the photovoltaic plant by daylight durations and prediction function (red line) with one harmonic (Source: built by the authors based on data from SolarEdge Monitoring Platform for Zarichchia, Ivano-Frankivsk region, Ukraine)*



Conclusions

The use of regression models allows obtaining high accuracy of forecasts. The best results were obtained for a 7-day moving average of electricity production by the photovoltaic plants with the coefficient of determination R^2 0.9752 and 0.9832.

Comparison of the results revealed that the model is less dependent on weather factors. Other models did not group the data by daylight hours and did not take any averaging for the respective duration. Therefore, their estimates were more dependent on weather conditions compared with ours. The accuracy of weather forecasts, especially cloud cover and solar insolation, can be quite low, which can affect the results.

Except for the case of maximum production by the duration of daylight hours, in the cases of average production and average moving for 7 days, the formula in the form of trigonometric polynomials with two harmonics was chosen for the estimation, if the least squares calculations yielded a smaller standard error of the residuals, and in the remaining cases, the formula in the form of trigonometric polynomials with one harmonic was used.

The results obtained prove the importance of monitoring energy and meteorological parameters at photovoltaic power plants. The measured parameters allow us to consider trend or seasonal variation, to obtain new regression dependencies to improve the accuracy of models or to simplify calculations in the absence of paid accurate forecasts. The prospect of improving the model is to create daily forecasts considering seasonality, forecasts for shorter periods of time (3–6 hours), and the possibility of retraining the model. This allows the owners of local photovoltaic power plants to become advanced active consumers in Smart Grid with the ability to choose forecast models depending on their financial and technical capabilities.

The research confirms the usefulness of ARIMA models for predicting the operation of photovoltaic power plants, as well as changes in power and insolation of photovoltaic power plants in different geographical locations characterized by different climatic conditions. It is important to identify an adequate ARIMA model for locations with different climatic conditions each time. The accuracy of the model depends on the stability of meteorological conditions (fully sunny or cloudy day). The differences in the observed results indicate that ARIMA models are best suited for forecasting in stable conditions (sunny or cloudy), i.e., when the cloud cover changes frequently during the day, such models do not provide satisfactory forecasts. The modeling results presented in this article using ARIMA models show high accuracy for medium- and long-term forecasts but can be effective when combined with additional models.

It is important to note that when changing locations, it is necessary to select an appropriate ARIMA model, estimate its parameters, and check the model fit and forecasting ability. In future, further validation of the model for diverse geographical locations both in Ukraine and worldwide under varying weather conditions is needed to enhance its generalizability.

The limitations of the suggested model are the following:

- It is not applicable to estimate minimum power generation by the duration of daylight hours because cloud and weather conditions can lead to near-zero power output;
- It does not allow building-short term predictions for maximum and average power output because they also depend on the weather conditions.

Its potential application in real-world scenarios is the long-term estimation of 7-day moving average power generation throughout the whole year.

Acknowledgements

The team of authors express their gratitude to the anonymous reviewers for valuable recommendations that were implemented to significantly improve the quality of this paper.

References

Bandura A. I. (2017) Some improvements of criteria of L-index boundedness in direction. Mat. Stud. 47(1): 27–32. Available at: https://doi.org/10.15330/ms.47.1.27-32

Bandura A. and Skaskiv O. (2017) Entire functions of bounded L-index: Its zeros and behavior of partial logarithmic derivatives. J. Complex Analysis 2017: Article ID 3253095. Available at: https://doi.org/10.1155/2017/3253095

Batsala Ya. V., Hlad I. V., Yaremak I. I. and Kiianiuk O. I. (2021a) Mathematical model for forecasting the process of electric power generation by photoelectric stations. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu 2021(1): 111–116. Available at: https://doi.org/10.33271/nvngu/2021-1/111

Batsala Ya. and Hlad I. (2023) Power quality issues of photovoltaic stations in electric grids and control of main parameters electromagnetic compatibility. In: Pawelczyk M., Bismor D., Ogonowski S., Kacprzyk J. (eds) Advanced, Contemporary Control. PCC. Lecture Notes in Networks and Systems, vol 708. Springer, Cham. Available at: https://doi.org/10.1007/978-3-031-35170-9_25

Batsala Ya., Hlad I. and Yaremak I. (2021b) Forecasting day-ahead of power generation from photovoltaic stations and use weather apps. Journal of New Technologies in Environmental Science 2021(4): 143–149. Available at: https://doi.org/10.53412/jntes-2021-4-3

Bracale A., Caramia P., Carpinelli G., Rita A., Fazio D. and Ferruzzi G. (2013) A Bayesian Method for Short-Term Probabilistic Forecasting of Photovoltaic Generation in Smart Grid Operation and Control. Energies 6: 733–747. Available at: https://doi.org/10.3390/en6020733

Bekirov E., Asanov M. and Zaliskyi M. (2019) System for Real-Time Monitoring the Quality of Electrical Energy Generated by Wind, Solar and Thermal Power Plants. 2019 Signal Processing Symposium, SPSympo 2019, Article ID 8882048: 177–180. Available at: https://doi.org/10.1109/SPS.2019.8882048

Beranek V., Olšan T., Libra M., Poulek V., Sedláček J., Dang M.-Q. and Tyukhov I. (2018) New Monitoring System for Photovoltaic Power Plants' Management. Energies 11(10): Article ID 2495. Available at: https://doi.org/10.3390/en11102495

Chen C., Duan S., Cai T. and Liu B. (2011) Online 24-h solar power forecasting based on weather type classification using artificial neural network. Sol. Energy 85: 2856–2870. Available at: https://doi.org/10.1016/j.solener.2011.08.027

Das S., Tey K., Seyedmahmoudian M., Mekhilef S., Idris M., Deventer W., Horan B. and Stojcevski A. (2018) Forecasting of photovoltaic power generation and model optimization: A review. Renew. Sustain. Energy Rev. 81: 912–928. Available at: https://doi.org/10.1016/j.rser.2017.08.017

Dolara A., Grimaccia F., Leva S., Mussetta M. and Ogliari E. (2015) A Physical Hybrid Artificial Neural Network for Short Term Forecasting of PV Plant Power Output. Energies 8: 1138–1153. Available at: https://doi.org/10.3390/en8021138

De Giorgi M., Malvoni M. and Congedo P. (2016) Comparison of strategies for multi-step ahead photovoltaic power forecasting models based on hybrid group method of data handling networks and least square support vector machine. Energy 107: 360–373. Available at: https://doi.org/10.1016/j.energy.2016.04.020

El Hendouzi A. and Bourouhou A. (2020) Solar Photovoltaic Power Forecasting. J. Electr. Comput. Eng. 2020: 8819925. Available at: https://doi.org/10.1155/2020/8819925

Erten M. and Aydilek H. (2022) Solar power prediction using regression models. International Journal of Engineering Research and Development, Special Issue 2022: 333–342. Available at: https://doi.org/10.29137/umagd.1100957

Florita A., Hodge B.-M. and Orwig K. (2013) Identifying wind and solar ramping events. IEEE 5th Green Technologies Conf., Denver, Colorado. Available at: https://doi.org/10.1109/GreenTech.2013.30

Iheanetu K. J. (2022) Solar Photovoltaic Power Forecasting: A Review. Sustainability 14(24): 17005. Available at: https://doi.org/10.3390/su142417005

Kim Y. S., Joo H. Y., Kim J. W., Jeong S. Y. and Moon J. H. (2021) Use of a big data analysis in regression of solar power generation on meteorological variables for a Korean solar power plant. Applied Sciences 11(4): article ID 1776. Available at: https://doi.org/10.3390/app11041776

Leva S., Dolara A., Grimaccia F., Mussetta M. and Ogliari E. (2017) Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power. Mathematics and Computers in Simulation 131: 88–100. Available at: https://doi.org/10.1016/j.matcom.2015.05.010

Liang J., Qiao K., Yu K., Ge S., Qu B., Xu R. and Li K. (2020) Parameters estimation of solar photovoltaic models via a self-adaptive ensemble-based differential evolution. Sol. Energy 207: 336–346. Available at: https://doi.org/10.1016/j.solener.2020.06.100

Lin K. and Pai P. (2016) Solar power output forecasting using evolutionary seasonal decomposition least-square support vector regression. J. Clean. Prod. 134: 456–462. Available at: https://doi.org/10.1016/j.jclepro.2015.08.099

Luo X., Zhang D. and Zhu X. (2021) Deep learning based forecasting of photovoltaic power generation by incorporating domain knowledge. Energy 225: 120240. Available at: https://doi.org/10.1016/j.energy.2021.120240

Massidda L. and Marrocu M. (2017) Use of Multilinear Adaptive Regression Splines and numerical weather prediction to forecast the power output of a PV plant in Borkum, Germany. Sol. Energy 146: 141–149. Available at: https://doi.org/10.1016/j.solener.2017.02.007

Meghea I. (2023) Comparison of statistical production models for a solar and a wind power plant. Mathematics 11(5): 1–16. Available at: https://doi.org/10.3390/math11051115

Meghea I. and Mihai M. (2018) Time series model applied to predict the solar power production. In: 18th International Multidisciplinary Scientific GeoConference SGEM 2018, 02-08 July, 2018, pp. 759–766. Available at: https://doi.org/10.5593/sgem2018/4.1/S17.098

Mei F., Pan Y., Zhu K. and Zheng J. (2018) A Hybrid Online forecasting model for ultrashort-term photovoltaic power generation. Sustainability 10(3): article ID 820. Available at: https://doi.org/10.3390/su10030820

Mills A. and Wiser R. (2010) Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. LBNL-3884E. Lawrence Berkeley National Laboratory, Environmental Energy Technologies Division, Berkeley, CA. Available at: https://escholarship.org/uc/item/9mz3w055

Pedro H. and Coimbra C. (2012) Assessment of forecasting techniques for solar power production with no exogenous inputs. Sol. Energy 86: 2017–2028. Available at: https://doi.org/10.1016/j.solener.2012.04.004

Rahman M. M., Selvaraja J., Rahima N. A. and Hasanuzzamana M. (2018) Global modern monitoring systems for PV based power generation: A review. Renew. Sustain. Energy Rev. 82(3): 4142–4158. Available at: https://doi.org/10.1016/j.rser.2017.10.111

Ratshilengo M., Sigauke C. and Bere A. (2021) Short-Term Solar Power Forecasting Using Genetic Algorithms: An Application Using South African Data. Appl. Sci. 11: 4214. Available at: https://doi.org/10.3390/app11094214

Raza M., Nadarajah M. and Ekanayake C. (2016) On recent advances in PV output power forecast. Sol. Energy 136: 124–144. Available at: https://doi.org/10.1016/j.solener.2016.06.073

Reatti A., Kazimierczuk M. K., Catelani M. and Ciani L. (2017) Monitoring and field data acquisition system for hybrid static concentrator plant. Measurement 98: 384–392. Available at: https://doi.org/10.1016/j.measurement.2015.06.022

Said M., Shaheen Abdullah M., Ginidi A. R., El-Sehiemy R. A., Mahmoud K., Lehtonen M. and Darwish M. M. F. (2021) Estimating Parameters of Photovoltaic Models Using Accurate Turbulent Flow of Water Optimizer. Processes 9: 627. Available at: https://doi.org/10.3390/pr9040627

SolarEdge (n.d.) SolarEdge Monitoring Platform. Available at: https://monitoringpublic.solaredge.com

Tuohy A., Zack J., Haupt S., Sharp J., Ahlstrom M., Dise S., Grimit E., Mohrlen C., Lange M., Casado M. et al. (2015) Solar Forecasting: Methods, Challenges, and Performance. IEEE Power Energy Mag. 13: 50–59. Available at: https://doi.org/10.1109/MPE.2015.2461351

Wang G., Su Y. and Shu L. (2016) One-day-ahead daily power forecasting of photovoltaic systems based on partial functional linear regression models. Renew. Energy 96: 469–478. Available at: https://doi.org/10.1016/j.renene.2016.04.089

Zhang J., Florita A., Hodge B., Lu S., Hamann H., Banunarayanan V. and Brockway A. (2015) A suite of metrics for assessing the performance of solar power forecasting. Sol. Energy 111: 157–175. Available at: https://doi.org/10.1016/j.solener.2014.10.016

Zhu H., Li X., Sun Q., Nie L., Yao J. and Zhao G. (2016) A Power Prediction Method for Photovoltaic Power Plant Based on Wavelet Decomposition and Artificial Neural Networks. Energies 9(10): 11. Available at: https://doi.org/10.3390/en9010011