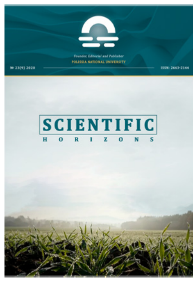
SCIENTIFIC HORIZONS

Journal homepage: https://sciencehorizon.com.ua
Scientific Horizons, 24(10), 9-16



UDC 621.316
DOI: 10.48077/scihor.24(10).2021.9-16

Determination of the Concepts of Building a Solar Power Forecasting Model

Alla Bosak[1], Dmytro Matushkin[1], Volodymyr Dubovyk[1], Sviatoslav Homon[2*], Leonid Kulakovskyi[1]

[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"
03056, 37 Peremohy Sq., Kyiv, Ukraine

[2]National University of Water and Environmental Engineering
33028, 11 Soborna Str., Rivne, Ukraine

Article's History:
Received: 15.10.2021
Revised: 13.11.2021
Accepted: 15.12.2021

Suggested Citation:
Bosak, A., Matushkin, D., Dubovyk, V., Homon, S., & Kulakovskyi, L. (2021). Determination of the concepts of building a solar power forecasting model. *Scientific Horizons*, 24(10), 9-16.

Abstract. Since in Ukraine there are fines for imbalances in solar power generation in the "day-ahead" energy market, the forecasting of electricity generation is an important component of the solar power plant operation. To forecast the active power generation of photovoltaic panels, a mathematical model should be developed, which considers the main factors affecting the volume of energy generation. In this article, the main factors affecting the performance of solar panels were analysed using correlation analysis. The data sets for the construction of the forecasting model were obtained from the solar power plant in the Kyiv region. Two types of data sets were used for the analysis of factors and model building: 10-minute time interval data and daily data. For each data set, the input parameters were selected using correlation analysis. Considering the determining factors, the models of finding the function of reflecting meteorological factors in the volume of electricity generation are built. It is established that through models with a lower discreteness of climatic parameters forecast it is possible to determine the potential volume of electricity production by the solar power plant for the day-ahead with a lower mean absolute error. The best accuracy of the model for predicting electric power generation over the 10-minute interval is obtained in the ensemble random of a forest model. It is determined that models without solar radiation intensity parameters on the input have an unsatisfactory coefficient of determination. Therefore, further research will focus on combining a model of forecasting the day-ahead solar radiation with 10-minutes discreteness with a model for determining the amount of electricity generation. The determined predicted values of solar radiation will be the input parameter of the forecasting model described in the article

Keywords: solar power plant, solar radiation, regression analysis, regularisation, model accuracy, coefficient of determination

*Corresponding author

INTRODUCTION

One of the largest sources of energy is the sun. Every year there is an increase in the exponential dependence of the electricity production of a solar power plant (SPP) in Ukraine [1]. According to [2], the main target parameters for the period up to 2035 include optimisation of the energy balance of the state, based on the requirements of security of energy supply and ensuring the share of renewable energy at 25%. A significant role is devoted to solar energy. However, with the development of renewable energy sources, there is a problem of ensuring the appropriate manoeuvrability of the power system. In [3], it is stated that the structure of the generating capacities of the Integrated Power System (IPS) of Ukraine in terms of ensuring effective frequency and power regulation in the power system is suboptimal. Among the reasons are the unregulated and variable operation of wind and solar power plants, aggravated by a lack of tools and approaches for forecasting electricity generation regimes.

According to the Law of Ukraine No. 1928-IX "Amendments to Certain Laws of Ukraine on Improving the Conditions for Supporting the Production of Electricity from Alternative Energy Sources" [4], in the day-ahead energy market, the fines for imbalances in the generation of SPP came into force since 2020. However, the Law does not provide a mechanism for short-term generation forecasting. There are no comments on databases, algorithms, techniques, and other grounds for effective day-ahead forecasting. In addition, Ukraine does not have an infrastructure base for a synoptic accurate short-term forecast for this field, which should be the basis for calculating the volume of generation, does not specify the degree of responsibility of third parties who provide data for forecasts, and there are no indications of the permissible accuracy of weather data for forecasting. At the same time, the Law imposes fines for actual hourly deviation from the projected schedules for the day-ahead and obliges the producer to be financially responsible for the imbalance of electricity to the Guaranteed Buyer.

That is why the issue of accurate forecasting of the possible electricity generation volume has become acute. However, solar energy forecasting is a rather difficult task, as it largely depends on climatic conditions that change over time. To overcome the above issues, it is important to use new intelligent methods to obtain reliable and accurate results.

Today, Machine Learning Methods have attracted considerable attention from many researchers and developers in solar radiance and power generation forecasting [5; 6]. Linear models based on the Autoregression method are mostly used to determine the radiation intensity. This method is simple but not flexible. An improved autoregressive integrated model with a moving average for determining the monthly solar radiation based on a set of radiation and temperature data for previous periods was proposed in [7]. A novel solar radiation prediction approach that combines two models, the Auto Regressive Moving Average (ARMA) and the Nonlinear Auto Regressive with eXogenous input (NARX) is presented in [8]. The effectiveness of combining a modification autoregression model and a convolutional neural network was studied in [9]. An empirical hybrid Autoregressive Integrated Moving Average (ARIMA) and Artificial Neural Network (ANN) approach shows a high correlation with experimental results and a relatively small error rate [10]. Deterministic and probabilistic forecasting of photovoltaic power based on a deep convolutional neural network is discussed in [11]. Also, recurrent neural networks are used for the hourly prediction of photovoltaic power output using meteorological information [12].

In addition, there are also some nonlinear methods based on time series. For example, a deep learning-based Photovoltaics (PV) power generation forecasting model based on Long Short-Term Memory (LSTM) model uses both outdated and forecast data by replacing the outdated weather data with the future weather forecast data during the testing phase for daily PV power generation forecasting [13]. In [14], the authors used a traditional recurrent artificial neural network and Support Vector Machine (SVM), based on a set of time series data, to increase forecasting accuracy for the next 24 hours. The special feature of time forecasting is that it considers the trend and seasonality of the predicted parameter. But the influence degree of the nature of changes in the values of climatic parameters in these models is mainly not considered. The question arises about the need in evaluating the application of other methods of Regression Analysis of Machine Learning, that will better account for the influence of disturbing factors on the further forecast through artificial neural networks. [15] presents Feature-Selective Ensemble Learning-Based Long-Term Regional PV Generation Forecasting. The Ensemble model that used simple multilayer perceptron and CNN with applied feature selection shows higher predictive power than the time series based single model.

The purpose of the study is to develop a concept and methodology for building a mathematical model for forecasting the amount of electricity generated by solar panels. It can be done by feature selection for active power generation parameters and selection of an adequate mathematical model for determining the target value of the energy generation function based on actual data.

MATERIALS AND METHODS

Data for the analysis of electricity generation by photovoltaic facilities and factors affecting the solar power plant (SPP) were obtained from Dymerska SPP in the village of Velyka Dymerka, Kyiv region. The data consisted of more than 26 thousand samples collected from July 1, 2020, to December 31, 2020, which characterize the operating conditions of solar panels with a capacity of 9 MW.

The data set consists of two types:

1. Weather condition data. The first part of the data consists of measured weather parameters such as temperature, humidity, solar radiation, atmospheric pressure, wind direction and wind speed.

2. Data from metering devices on the amount of electricity produced.

The analysed dataset consisted of the actual value of the output electric power for 10-minute intervals (the meter transfers the generation value to the monitoring point with the discreteness of 10 minutes) and the measured climatic parameters for the appropriate period. The data of 10-minute discreteness samples also need to be aggregated into daily samples for forecasting electric power generation on a day ahead.

To estimate the actual value of the influence of each parameter on the target function and separately on each of the input factors of the model, a correlation matrix is used. It is a structured approach to ranking the importance of predictors or input variables at the output. The correlation coefficient for the sample is determined from the equation [16]:

$$r_{xy} = \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{(n-1)s_x s_y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

where n is the sample size; $\bar{x}$ and $\bar{y}$ are the sample means of the individual sample points x and y accordingly; s_x and s_y are the sample standard deviations for x and y accordingly.

The first task of this study was to determine the nature of the dependence of the output target function, i.e., the volume of electricity generation on the set of input parameters (climate conditions) and to build mathematical models based on them. Thus, it is necessary to determine the amount of electricity generated (Y):

$$Y = \sum_{k=1}^{n} \beta_{jk}\psi_k(X_1, \dots, X_N) \quad (2)$$

where β_{jk} is an unknown constant, when $\psi_k(\cdot)$ is the set of basic functions, at $k \in \{1,\dots,N\}$, $X_1,\dots,X_N$ are the set of input parameters (temperature, humidity, solar radiation, wind speed etc.).

The determination of active power generation from solar panels is possible by the methods of linear regression, ridge regression, lasso regression and random forest regression. For linear regression, the relationship between the data was built using linear functions, and the unknown parameters of the model were estimated from inputs. Using the linear regression model, it is possible to obtain a process model:

$$y = \beta_0 + \beta_1 X_1 + \dots + \beta_i X_i + \varepsilon \quad (3)$$

where y is the target variable (the predicted value); $X_1,\dots,X_i$ are independent variables; β_0 is the bias coefficient; $\beta_1,\dots,\beta_i$ are coefficients of independent variables; ε is the error term (the residual).

The coefficient β_0 is the predicted value of y when X is 0. The coefficients β_i of the model were selected by the least-squares method (LSQ). This method minimizes the sums of the squares of the regression residuals.

In the case of increasing the number of model parameters, the linear regression does not differentiate between "important" and "less important" predictions in the model. So, it includes all variants. The model will be retrained, and it will be difficult to find unique solutions after. There will also be issues with the multicollinearity of data.

One of the solutions to the multicollinearity issue is to use L_2 regularisation. Ridge regression belongs to a class of regression tools that uses L_2 regularisation. L_2 regularisation adds an L_2 penalty, which is equal to the square of the value of the coefficients. All coefficients are reduced by a coefficient (so none are excluded) [17]:

$$Ridge_{loss} = \underbrace{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}_{Loss} + \lambda \underbrace{\sum_{i=1}^{k}\omega_i{}^2}_{Penalty} \quad (4)$$

where y_i is the actual value; $\hat{y}_i$ is the predicted value; λ is a non-negative tuning parameter; ω_j is the model weight.

Another modification of linear regression is lasso regression. In lasso regression, the loss function is modified to minimise the complexity of the model by limiting the sum of the absolute values of the model coefficients (the so-called L_1-norm):

$$Lasso_{loss} = \underbrace{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}_{Loss} + \lambda \underbrace{\sum_{j=1}^{k}|\omega_j|}_{Penalty} \quad (5)$$

L_1 regularisation will lead to zero weights of some features. So, the features selection is the result of the L_1 regularisation, which produces sparse coefficients.

The influence of random fluctuations in a single dimension is weakened by averaging the results of observations. This can provide a more stable and reliable assessment. Algorithms of combining models present a similar concept. The construction of their ensembles is one of the most powerful methods of Machine Learning. They are often the best models for the quality of forecasts compared to other methods. One of the most common ensemble methods is the random forest method. Random Forest Regression builds several decision trees of a regression model during training and obtains an average forecast as input. The basic concept of a random forest is that a set of random trees find a solution independently of each other and act together, surpassing any solution obtained by a single decision tree [18].

The results of testing models obtained using the considered methods must be checked for the accuracy of obtaining forecasts for the generation of electricity. For this purpose, authors applied such criteria as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percentage error (MAPE) and coefficient of determination (R^2). The MAE measures the average distance between $\hat{y}_i$ and y_i, i.e., directly describes the mean offsets. The R-squared (R^2) measures the level of correlation between $\hat{y}_i$ and y_i [19]. The MSE is a function that corresponds to the expected value of the error loss square. The MAPE is widely used as a loss function for regression problems and in model evaluation, because of its very intuitive interpretation in terms of relative error. In the case of SPP, it is normalised by power. Mathematical equations of indicators are formulated as follows:

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \tag{6}$$

$$RMSE = \sqrt{MSE} \tag{7}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{8}$$

$$MAPE = \frac{100}{N}\sum_{i=1}^{N}\left|\frac{y_i - \hat{y}_i}{P_0}\right| \tag{9}$$

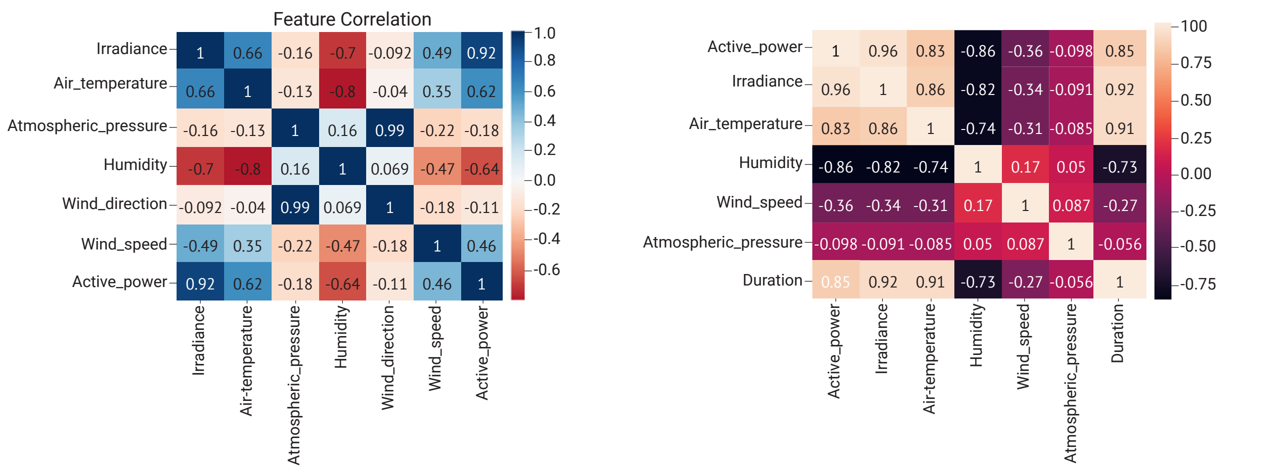
$$R^2 = 1 - \left(\frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{(y_i - \bar{y}_i)^2}\right) \tag{10}$$

where N is the sample size; y_i is the actual value; $\hat{y}_i$ is the predicted value; $\bar{y}$ is the sample mean; P_0 is the rated power of the SPP.

Sometimes, the equation (9) is normalised by the actual value, i.e., y_i, rather than P_0, which has the disadvantage as a zero power value has a meaning when the SPP does not generate energy.

RESULTS AND DISCUSSION

The thermal parameter correlation matrix for the power generation data for the 10-minute intervals is shown in Figure 1a and for the daily intervals in Figure 1b. For daily intervals, the values of wind speed, temperature, humidity, and atmospheric pressure were averaged per day, and the total value per day was calculated for solar radiation and generation. Weather condition data on 10-minutes discreteness consists with measured wind direction parameters for each period. Otherwise, values of this parameter cannot be aggregated on a daily period. But for a daily active power generation can be used another parameter – the duration of daylight. It is a time duration from sunrise to sunset.



a) b)

***Figure 1**. Correlation heatmaps for weather features and active power generation for a) 10-minutes intervals data; b) daily intervals data*

According to the results of the calculation, two parameters had positive correlations with the output power, namely solar radiation, air temperature on each data time discreteness. Also, such parameter as the duration of daylight on daily data has a huge correlation with active power generation (85%) and the wind speed parameter on 10-minutes intervals has almost 50% correlation. Relative humidity had a negative correlation. It should be noted that the wind direction is correlated with an atmospheric pressure of 99% on 10-minute discreteness. This dependence means that the value of one parameter changes almost completely as the value of another. Thus, these factors in the model will be duplicated. Such a case could increase the error of the model and the possibility of retraining. Therefore, the "wind direction" parameter was excluded. If ensemble models are built in combination with other climatic parameters, the atmospheric pressure factor can also affect them. However, there was no effect of atmospheric pressure on the model at a daily interval according to the correlation matrix and this factor was not included in the totality of output parameters. In this case, the degree of influence of most factors may vary depending on the season. Therefore, in further studies, both solar radiance and electricity generation values should consider seasonality and introduce the degree of importance of input parameters depending on the age of the data.

The data for the six months which were used in the experiment were divided into three segments: a training data set (60%), a validation data set (20%), and a test data set (20%). A training data set was used to train the models and a five-time test was performed as a resampling procedure. The experiments were implemented using the *scikit-learn library in Python 3.8* which allows implementing the Machine Learning Methods.

Construction models should be analysed on the accuracy of forecasting active power generation. MAE, MSE, RMSE could be used to characterise the difference in solar forecasting performance attributed to spatial aggregation [20]. A lower value of these indicators points to a higher quality of the forecast. For a day-ahead forecast, the value of these errors is important, as in absolute terms it shows how real data may differ from the forecast. However, it is necessary to consider the variability of data in the short term. There can be a significant difference between the possible minimum and maximum value of absolute errors for certain climatic conditions. In particular, at night in the winter months, the generation of active electricity varies from 0 to 10 W per hour. At the same time, during daylight hours in the summer months, the generation can vary from a few hundred W to 50 W per hour. For this period, the absolute errors are higher. In this case, it is also advisable to use relative errors to assess the effectiveness of the model. Percentage errors have the advantage of being unit-free, so they are frequently used to compare forecast performances between data sets [21]. MAPE metric can be used to compare the results from different spatial and temporal scales of forecast errors. Firstly, attention should be paid to absolute errors and the coefficient of determination. Then it is necessary to determine the MAPE value (Table 1). In case the model has a bad or insufficient forecast error of MAPE then it is better to choose another model with lower absolute errors [22]. Table 2 shows the results of the calculation of accuracy for all models.

***Table 1**. Interpretation of typical MAPE values*

MAPE, %	Forecast accuracy
Lower than 10%	High
10-20%	Good
21-40%	Satisfactorily
41-50%	Bad
More than 50%	Unsatisfactorily

***Table 2**. Comparison of errors criteria for the results obtained for each of the tests from the solar radiation forecast*

Model	MAE, W	MSE, W	RMSE, W	MAPE, %	R^2, abs.un.
10-minute interval					
Linear regression	243.25	259514.05	509.43	11.12	0.912
Ridge regression	242.71	260008.21	509.91	10.86	0.915
Lasso regression	231.51	266892.45	516.62	11.56	0.916
Random forest regression	146.91	172864.78	415.77	9.34	0.941
Daily interval					
Linear regression	28979.11	1.24e+09	35227.21	15.19	0.947
Ridge regression	25719.42	1.21e+09	34734.32	13.43	0.949
Lasso regression	35520.55	1.77e+09	42024.57	34.92	0.925
Random forest regression	25509.59	1.41e+09	37557.94	17.99	0.94

The simulation results showed that in the dataset, where a better correlation of model parameters was observed (more than 80%, and especially with the value of solar radiation of 0.96) with the target function, linear regression and ridge regression had better quality. This is mainly attributable to the fact that the model did not require a combination of features, and the sample was smaller. Such results were obtained for daily intervals datasets. The best results in a sample of the 10-minute interval were obtained by the method of random forest regression, where the overall correlation was not so close. In this case, sorting features and randomly searching for the optimal model by combining and crossing features allowed to get a more accurate model. The combined features were more correlated with the target function than when they were alone. The results of the forecast of active power generation on July 1st and July 2nd according to the initial parameters in the test sample and the real data are shown in Figure 2.

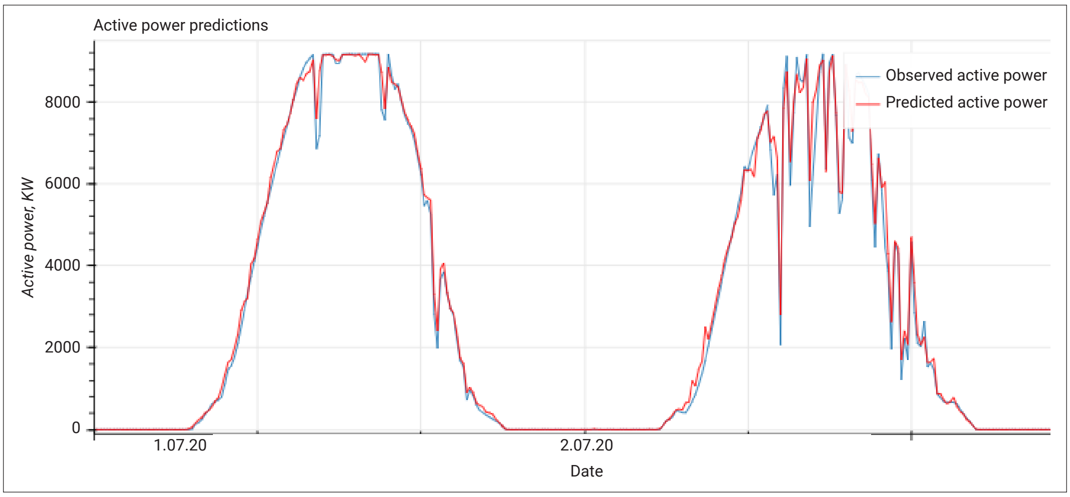


***Figure 2**. The forecast of active power generation by the method of random forest regression according to the initial parameters*

Figure 2 shows that in general, the model predicts the amount of electricity generation accurately. The results in Table 2 also show that the models obtained at the daily interval have a higher coefficient of determination – 0.95 (ridge regression), while on the 10-minute interval the model obtained by the random forest method was 0.94. However, to determine which time period was more effective, it is necessary to compare the value of the obtained deviation of values of the original function. If in the random forest regression model for a 10-minute interval the MAE value is 146.9 W, then it can have the following maximum value per day:

$$MAE_{day} = MAE_{10minute} \cdot 6 \cdot 24 = 146.9 \cdot 6 \cdot 24 = 21\ 155 [W], \quad (11)$$

because an hour has 60 minutes (six 10-minute intervals), and a day has 24 hours.

Obtaining accurate predictions of solar radiation can be quite a challenge for certain measurement sites. Therefore, models determining the amount of electricity generation according to the climatic data defined above, excluding the amount of solar radiation, were found. On the daily interval, instead of its value, it is possible to apply data on the duration of daylight (the correlation coefficient according to Figure 1b is 0.92). The 10-minute model excluded the parameter of radiation.

The values of the accuracy indicators of the models, excluding the solar radiation and including the duration of daylight at daily intervals, are given in Table 3.

***Table 3**. Comparison of criteria for results errors, excluding the solar radiation at 10-minute intervals and including the duration of daylight at a daily interval*

Model	MAE, W	MSE, W	RMSE, W	MAPE, %	R^2, abs.un.
10-minute interval					
Linear regression	1290.65	3.19e+06	1787.89	62.34	0.494
Ridge regression	1288.024	3.2e+06	1788.72	61.47	0.494
Lasso regression	1476.25	3.86e+06	1963.55	72.64	0.39
Random forest regression	735.5	1.86e+06	1362.17	83.12	0.61
Daily interval					
Linear regression	36549.11	1.24e+09	35227.21	27.75	0.793
Ridge regression	34719.42	1.21e+09	34734.32	16.88	0.823
Lasso regression	42521.52	1.77e+09	42024.57	21.14	0.82
Random forest regression	35512.19	1.41e+09	37557.94	15.95	0.827

The forecast result shows that the complete elimination of the radiation parameter from the model did not allow obtaining an adequate accuracy of the target function (R^2<0.75 and MAPE is 50% higher). The daily chart shows that the presence of the duration of daylight parameter made it possible to obtain a model with R^2>0.75. However, it was significantly worse than the model with a solar radiation parameter (R^2=0.827 compared to R^2=0.949 for a model with a solar radiation factor at the input). Accordingly, the MAE value is also lower by about 26% and MAPE has a higher evaluation error. The results showed the need for a solar radiation parameter in the model, and the daylight indicator did not replace it completely. Therefore, it is needed to predict solar radiation for a short period ahead. These results should then be used in a model for predicting active power generation from the data sets used at 10-minute intervals.

The next step in the research involves a forecasting model based on neural networks of long short-term memory. This network may allow providing time series in conditions when there are time delays of unknown duration between important events and when it is necessary to take into account the seasonality of parameters [23]. So, obtaining a model by random forest method for obtaining active power generation amount on 10-minutes intervals can be used for forecasting on a long short-term period by time series models.

CONCLUSIONS

It is established that the main factors for building a model of power generation forecasting are solar radiation, temperature, humidity, and wind speed. The obtained models by using both 10-minute and daily intervals were quite accurate, as the coefficient of determination was more than 0.94 for each of them. In addition, it was found that the models that used a 10-minute interval had a lower MAE value per day compared to the value from the daily interval. Therefore, the use of a model with a lower discreteness of the forecast of climatic parameters will determine the possible volume of electricity generation of SPP for the day-ahead with a lower forecast error. The best accuracy in models at a 10-minute interval was obtained in the ensemble model of a random forest, and among models, with daily interval, the best one was obtained based on linear regression and its regularisation. This results from the high correlation dependence of the main factors (solar radiation, temperature, and humidity) with the target function.

The models that did not use solar radiation as one of the input parameters had an unsatisfactory value of the coefficient of determination (R^2<0.75) and MAPE (>50%). By replacing the solar radiation parameter with the duration of daylight on the daily interval, it was possible to obtain an adequate model (R^2>0.75), although the MAE value increased by more than 25%. This indicated that the model should include solar radiation.

REFERENCES

[1] Butenko, V., Baidala, V., & Kozyrska, T. (2019). Factors of solar power development in Ukraine. *Investytsiyi: Praktyka ta Dosvid*, 31(477), 5-11. doi: 10.32702/2306-6814.2019.17.5.

[2] Cabinet of Ministers Ukraine, Energy Strategy of Ukraine for the Period Up to 2035 “Security, Energy Efficiency, Competitiveness”. (2017, August). Retrieved from https://zakon.rada.gov.ua/laws/show/605-2017-%D1%80#Text.

[3] “Ukrenergo” NPC, Draft document “Transmission System Development Plan for 2019-2028”. (2019, August). Retrieved from https://www.slideshare.net/Ukrenergo/2019-2028.

[4] Law of Ukraine No. 810-IX “On Amendments to Certain Laws of Ukraine Concerning Improving the Conditions for Supporting the Production of Electric Power from Alternative Energy Sources”. (2021, December). Retrieved from https://zakon.rada.gov.ua/laws/show/810-20#Text.

[5] Mellit, A., Massi Pavan, A., Ogliari, E., Leva, S., & Lughi, V. (2020). Advanced methods for photovoltaic output power forecasting: A review. *Applied Sciences*, 10(2), article number 487.

[6] Behera, M.K., Majumder, I., & Nayak, N. (2018). Solar photovoltaic power forecasting using optimized modified extreme learning machine technique. *Engineering Science and Technology, an International Journal*, 21(3), 428-438.

[7] Li, Y., Su, Y., & Shu, L. (2014). An ARMAX model for forecasting the power output of a grid connected photovoltaic system. *Renewable Energy*, 66, 78-89. doi: 10.1016/j.renene.2013.11.067.

[8] Sansa, I., Boussaada, Z., & Bellaaj, N.M. (2021). Solar radiation prediction using a novel hybrid model of ARMA and NARX. *Energies*, 14(21), article number 6920. doi: 10.3390/en14216920.

[9] Marikkar, U., Hassan, A.J., Maithripala, M.S., Godaliyadda, R.I., Ekanayake, P.B., & Ekanayake, J.B. (2020, November). Modified auto regressive technique for univariate time series prediction of solar irradiance. In *2020 IEEE 15th International conference on industrial and information systems (ICIIS)* (pp. 22-27). doi: 10.1109/ICIIS51140.2020.9342694.

[10] Belmahdi, B., Louzazni, M., & El Bouardi, A. (2020). A hybrid ARIMA–ANN method to forecast daily global solar radiation in three different cities in Morocco. *The European Physical Journal Plus*, 135, article number 925. doi: 10.1140/epjp/s13360-020-00920-9.

[11] Wang, H., Yi, H., Peng, J., Wang, G., Liu, Y., Jiang, H., & Liu, W. (2017). Deterministic and probabilistic forecasting of photovoltaic power based on deep convolutional neural network. *Energy Conversion and Management*, 153, 409-422.

[12] Lee, D., & Kim, K. (2019). Recurrent neural network-based hourly prediction of photovoltaic power output using meteorological information. *Energies*, 12(2), article number 215.

[13] Yu, D., Choi, W., Kim, M., & Liu, L. (2020). Forecasting day-ahead hourly photovoltaic power generation using convolutional self-attention based long short-term memory. *Energies*, 13(15), article number 4017.
[14] Das, U.K., Tey, K.S., Seyedmahmoudian, M., Idna Idris, M.Y., Mekhilef, S., Horan, B., & Stojcevski, A. (2017). SVR-based model to forecast PV power generation under different weather conditions. *Energies*, 10(7), article number 876. doi: 10.3390/en10070876.
[15] Eom, H., Son, Y., & Choi, S. (2020). Feature-selective ensemble learning-based long-term regional PV generation forecasting. *IEEE Access*, 8, 54620-54630. doi: 10.1109/ACCESS.2020.2981819.
[16] da Silva Filho, A.M., Zebende, G.F., de Castro, A.P.N., & Guedes, E.F. (2021). Statistical test for multiple detrended cross-correlation coefficient. *Physica A: Statistical Mechanics and Its Applications*, 562, article number 125285.
[17] Saunders, C., Gammerman, A., & Vovk, V. (1998). Ridge regression learning algorithm in dual variables. In *Proceedings of the fifteenth international conference on machine learning* (pp. 515-521). Madison: Morgan Kaufmann Publishers Inc.
[18] Izquierdo-Verdiguier, E., & Zurita-Milla, R. (2020). An evaluation of guided regularized random forest for classification and regression tasks in remote sensing. *International Journal of Applied Earth Observation and Geoinformation*, 88, article number 102051. doi: 10.1016/j.jag.2020.102051.
[19] Willmott, C.J., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), 79-82.
[20] Ahmed, R., Sreeram, V., Mishra, Y., & Arif, M. D. (2020). A review and evaluation of the state-of-the-art in PV solar power forecasting: Techniques and optimization. *Renewable and Sustainable Energy Reviews*, 124, article number 109792.
[21] Hyndman, R.J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice*. Melbourne: OTexts.
[22] Demchik, Y., & Rozen, V. (2019). Estimations of error of prognosis models and prognoses of the used electric energy are on objects of power market. *Power Engineering: Economics, Technique, Ecology*, 4, 69-78.
[23] Pan, C., Tan, J., Feng, D., & Li, Y. (2019). Very short-term solar generation forecasting based on LSTM with temporal attention mechanism. In *2019 IEEE 5th International conference on computer and communications (ICCC)* (pp. 267-271). doi: 10.1109/ICCC47050.2019.9064298.

Визначення концепції побудови моделі прогнозування сонячної енергії

Алла Василівна Босак[1], Дмитро Сергійович Матушкін[1], Володимир Григорович Дубовик[1], Святослав Святославович Гомон[2], Леонід Ярославович Кулаковський[1]

[1]Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського»
03056, просп. Перемоги, 37, м. Київ, Україна

[2]Національний університет водного господарства та природокористування
33028, вул. Соборна, 11, м. Рівне, Україна

Анотація. Оскільки в Україні діють штрафи за дисбаланс виробництва сонячної енергії на ринку електроенергії «на добу вперед», то виникає необхідність прогнозування виробництва електроенергії сонячною електростанцією. Для прогнозування виробництва активної потужності фотоелектричних панелей потрібно розробити математичну модель з урахуванням основних факторів, що впливають на величину генерації енергії. В статті виділення основних факторів, що впливають на роботу сонячних панелей, було здійснено за допомогою кореляційного аналізу. Набір даних для побудови моделі прогнозування було отримано на сонячній електростанції в Київській області. Для аналізу факторів і побудови моделі використовувалися два типи наборів даних: дані 10-хвилинних інтервалів часу та добові дані. Для кожного набору даних за допомогою кореляційного аналізу були обрані відповідні вхідні параметри. З урахуванням визначених факторів побудовано моделі знаходження функції відображення метеорологічних факторів від обсягу вироблення електроенергії. Встановлено, що моделі з меншою дискретністю прогнозу кліматичних параметрів дозволяють визначити можливий обсяг виробництва електроенергії сонячною електростанцією на добу вперед з меншою середньою абсолютною похибкою. Найкращу точність моделі прогнозу виробітку електричної енергії на 10-хвилинному інтервалі отримано в ансамблевій моделі випадкового лісу. Встановлено, що моделі, що не містять на вході параметру інтенсивності сонячного випромінення, мають незадовільний коефіцієнт детермінації. Тому подальші дослідження будуть зосереджені на поєднанні моделі прогнозування сонячного випромінення з 10-хвилинною дискретністю на добу наперед із моделлю визначення кількості виробленої електроенергії. Визначені прогнозовані значення сонячного випромінення будуть вхідним параметром описаної в статті моделі прогнозування

Ключові слова: сонячна електростанція, сонячне випромінювання, регресійний аналіз, регуляризація, точність моделі, коефіцієнт детермінації