УДК 621.311.26, 551.509.3, 551.509.316 **DOI 10.20535/1813-5420.3.2023.289651**

D. Matushkin[1], Ph. D. student, ORCID 0000-0003-4431-7862
A. Bosak[1], Cand. Sc. (Eng.), Assoc. Prof, ORCID 0000-0003-0545-9980
[1]National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute"

CONCEPT OF AN ENSEMBLE FORECASTING SYSTEM FOR OPTIMIZATION PROBLEMS OF CONTROL OF SOLAR MICROGRID

Accurate probabilistic forecasts of renewable generation are the driving force for optimizing the operation and management of MicroGrid systems. Combining forecasts of different individual models can improve forecast accuracy, but unlike combining point forecasts, for which simple weighted averaging is often a plausible solution, combining probabilistic forecasts is a much more complex task. Today, ensembles of forecasting models are one of the promising directions for problem solving, where forecasting accuracy is more important than the ability to interpret the model. The main idea of ensembles is the training of several basic models and the aggregation of the results of their work. Empirical studies show that combinations of forecasts, on average, are more likely to produce better forecasts than methods that are based on selecting only one forecasting model. When building ensembles, the issue of ensuring diversity of models and effective training of model members of the ensemble becomes especially relevant. The article is devoted to solving the issues of building an ensemble model for forecasting photovoltaic (PV) power, which combines the results of several basic probabilistic models. Using the ensemble method proposed by the authors can improve forecasting accuracy and reduce the time required for training and evaluation of ensemble member models. Directions and prospects of further research are formulated.

Keywords: *forecasting system, ensembles of forecasting models, solar power plant, MicroGrid, ensemble architecture*

Introduction. Forecasting plays a key role in the planning of any process, as it provides an insight into uncertainty. With the help of simulation, it is possible to assess whether the proposed strategies can achieve the desired goals within predetermined limits. However, determining the best prediction method is not an easy task and depends a lot on the user's goals and the constraints he/she is likely to face. Rather than trying to find a single best prediction method, an alternative approach might be to combine the results of independent predictors and take some average of the predictions.

This method of averaging a number of independent forecasts obtained from different forecasting methods is known as forecast combining, and the result is often called a consensus forecast. If a specific forecast model that yields smaller accuracy errors compared to other individual forecasts cannot be identified, adopting a consensus approach may be beneficial by achieving diversification. In recent decades, combined forecasts have attracted a lot of interest, supported by the publication of a large quantity of research on the accuracy of forecasts. Empirical researches have shown that combining forecasts has made it possible to increase the accuracy of forecasts [1-6].

One of the advantages of using such a forecast is that it can be useful if the situation has a significant degree of uncertainty or risk, and it is difficult to choose the most accurate forecast in advance. Even if one method is determined to be the best, the combination is still worthwhile if the other methods can make some positive contribution to forecast accuracy. Moreover, many factors can influence an independent prediction, and these, together with any additional useful information, can be obtained using a consensus approach.

A MicroGrid is a small-scale electrical network that integrates distributed energy sources and can operate connected to the United Energy Systems (UPS) or autonomously. MicroGrids are equipped with various types of energy generation and storage sources, including solar and wind power plants, diesel generators, batteries, etc. (Fig. 1).

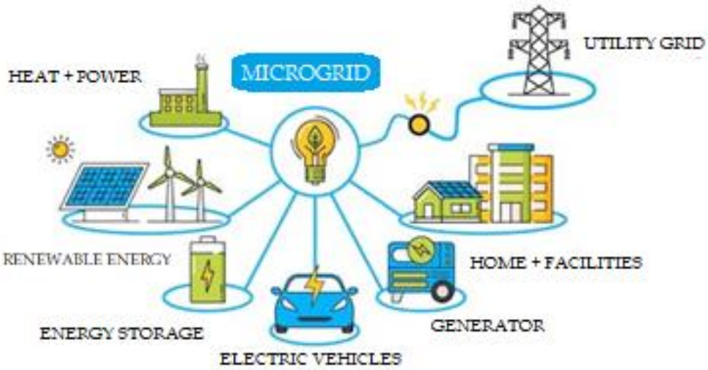

Figure 1 – *A typical MicroGrid architecture*

The generation of electrical energy by renewable energy sources, and a solar plant in particular, is very unstable, it depends on weather and geographical conditions, season, time, etc. The MicroGrid, in contrast to UPS, is sensitive to various disturbances, it does not have such a property as inertia, which is usually provided by powerful generators of large power plants. Therefore, the effective operation of the MicroGrid is related to accurate grid management. Due to the variety and complexity of the processes that take place in the local power system, an effective tool for managing such a network can be an ensemble (combined) system, the tasks of which can include:

–forecasting of generation and consumption, maximum load;

–management of energy source operating modes;

–problem detection and system management.

For the effective operation of such a control system, it is necessary to choose the right architecture, which depends on a number of parameters (type of input data and system goals).

Purpose of the work. Development of the architecture of the ensemble forecasting system for the tasks of optimizing the management of the Solar MicroGrid and the theoretical justification of the prospects for the implementation of such a forecasting procedure.

Material and research results. In electrical networks, data is most often obtained in the form of time series. They can be used as received or prepared by using data standardization and normalization. Forecasting is one of the main tasks in a MicroGrid.

As a rule, past data for a certain period are used for forecasting. The choice of the duration of this period is a compromise between the complexity of the system, because the longer this period is, the more complex the system is, and the quality of the forecast. In the Solar MicroGrid, the optimal forecast is made for 1-24 hours, and input parameters are taken for the last 3-72 hours, respectively [7].

An important issue is the detection of abnormal situations in MicroGrid. The problem is that when creating a training sample, there is much less data on extraordinary situations than on normal work, and it is impractical to get them specifically. The forecasting system does not necessarily need to know what each abnormal situation looks like, and the task of identifying problems is reduced to finding anomalies in time series. Therefore, the training sample should consist of examples without anomalies. To solve this problem, the models accepted in the forecasting system must process a certain period of time. The use of autoencoders is appropriate for this task. An autoencoder is a model that consists of two parts: an encoder and a decoder. Accordingly, it learns to encode data and reproduce it with the smallest possible error. Since the autoencoder cannot reliably reproduce anomalous data, a noticeable increase in error will immediately be obtained, indicating an anomaly in the data due to an abnormal situation.

The management problem corresponds to the classification problem. And in this case, the state and situations in the solar power system are classified according to which decisions are made. Therefore, the main types of models that can effectively solve this problem are K-Nearest Neighbor (KNN) or Neural Networks (NN) with architectures of Perceptron, Convolutional NN or Deep Machine Learning Network. Often the management problem is that vague categories are being operated on [8-10]. Therefore, the use of Neuro-Fuzzy Systems, such as ANFIS (Adaptive Network Based Fuzzy Inference System), is effective for this task. This method combines a clear interpretation of the raw data corresponding to Fuzzy Logic, and the membership functions are implemented by a Neural Network.

Overview of the Proposed Ensemble Method for Forecasting Solar MicroGrid Power. The predictions of the underlying probabilistic models are combined into a competitive ensemble method, illustrated in Fig. 2. Historical and current PV data and weather data together with calendar variables and the geographic location of the Solar MicroGrid are the inputs to the proposed forecasting system architecture. This data is used by the underlying probabilistic models to construct individual probabilistic power forecasts. Finally, the predicted quantiles returned from the underlying models are fed as inputs to the ensemble model to combine them appropriately. In the forecast pooling step, calendar variables and geographic location may or may not be used, which differentiates the parameter estimation. The output of the combination procedure is a probabilistic photovoltaic power forecast.

Database. The following types of datasets are used:

–meteorological data that can be obtained both from local sensors of weather parameters and from a centralized weather station [11];

–measurement data of electricity generation/consumption [11, 12];

–calendar information, according to which seasonality and time of day are determined;

–the geographical location of Solar MicroGrid.

Meteorological data is data that contains actual light and weather values with a resolution of 10 minutes. Characteristics of meteorological data can be solar radiation, air temperature, precipitation, relative humidity, wind speed, etc [11].

These data must be time-averaged to match the forecast interval. The measured real output power data of the PV system is also used as a data function.

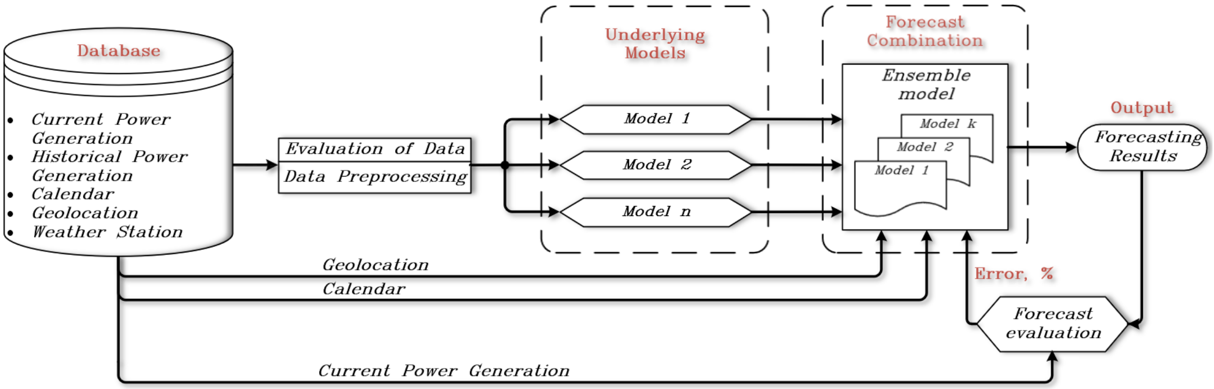

Figure 2 – *The concept of the general framework of an ensemble Solar MicroGrid power forecasting system*

Evaluation of Data and Data Preprocessing. A correlation matrix is used to estimate the actual value of the influence of each parameter on the output variable and separately on each of the input factors of the model. It is a structured approach to ranking the importance of predictors or input variables in the output. The Pearson Correlation Coefficient (PCC) and *t*-statistics are used to select the appropriate characteristics of the data. The PCC is used to calculate the correlation value. Values are in the range -1 to 1 [11]. Value $r = 1$ indicates a positive correlation, $r = 0$ indicates the absence of correlation, and $r = -1$ indicates a negative correlation. The PCC formula is as follows:

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

where $\bar{x}$, $\bar{y}$ – the average values for the sample x and y, s_x and s_y – unbiased (adjusted) estimate of the standard deviation for x and y.

Data preprocessing involves normalization, cleaning, recovery, and separation of data. At the stage of data preparation, data are normalized using minimum-maximum normalization. Minimum-maximum normalization is defined as [13]:

$$\bar{x}_n = \frac{x_n - x_{min}}{x_{max} - x_{min}} \tag{2}$$

де $\bar{x}_n$ – normalized data, x_n – output data, x_{max} і x_{min} – maximum and minimum value of x_n accordingly.

After data normalization, data cleaning removes outliers and data reconstruction replaces missing values using linear interpolation. The "good" data is further divided into training and testing sets.

The stages of data preprocessing are shown in Fig. 3.

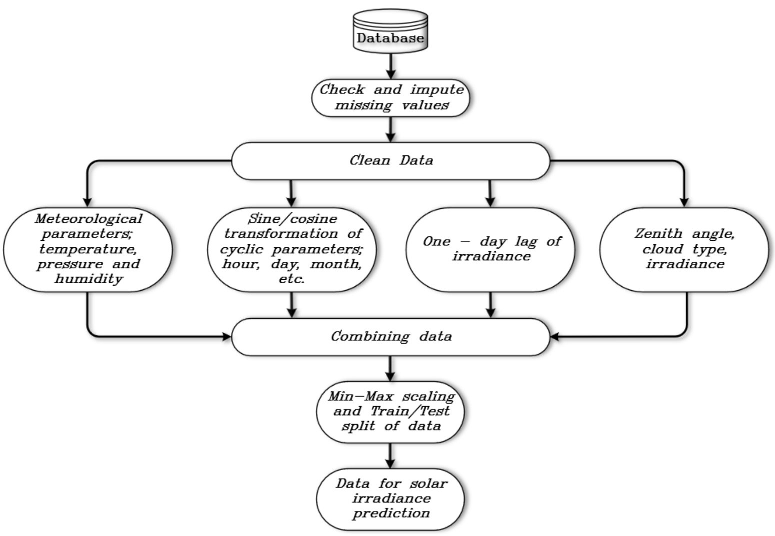

Figure 3 – *Preprocessing steps for cleaning and separating data into train test sets*

First, the missing values are identified and are replaced with the values from the same time the previous day. If the value for the same time of the previous day is also missing, then the missing data is imputed using the same time value of the last previous day with available data. Then further processing of the clean data is performed in the ways discussed above. The transformations of sines and cosines of cyclic parameters, such as hour of the day, day of the year, month of the year, season of the year, are determined. Binary encoding from the atmosphere model can be used to encode the Cloud Type feature. Meteorological parameters including pressure, temperature and wind speed are separated. One-day time lags of solar irradiance for the last day are organized as separate characteristics for solar irradiance models. These functions are concatenated and then normalized, i.e. the range of each input vector is bounded to (0, 1). The scaled data are then divided into training and test datasets for training and evaluating the proposed model, respectively.

Probabilistic Underlying Models. The selection of basic models is based on an individual assessment of each model, the number of which in the proposed architecture is unlimited. For each model, *Model 1..n*, it is assumed that the same subsequent training data are available at the starting point of the forecast *t*: *N* historical and current values $P_{t-(N-1)}, P_{t-(N-2)},\ldots, P_t$ of PV power; *N* vectors $p_{t-(N-1)}, p_{t-(N-2)},\ldots, p_t$ *M* predictors, corresponding to each of the *N* historical PV power values. In particular, the general *j*-th vector of predictors is $p_j = \left\{p_{1_j},\ldots,p_{M_j}\right\}$ for $j = t-(N-1), t-(N-2), \ldots, t$.

The Ensemble Model for Forecast Combination. The ensemble model is based on the quantile-weighted sum. In [14] proposes and compares eight different strategies:

–the Pure Chuantile Weighted Sum (PQWS);

–the Hourly QuantileWeighted Sum (HQWS);

–the Pure Constrained Quantile Weighted Sum (PCQWS);

–the Hourly Constrained QuantileWeighted Sum (HCQWS);

–the Pure Quantile Weighted Sum with Least Absolute Shrinkage and Selection Operator (LASSO) Regularization (PQWSLR);

–the Hourly QuantileWeighted Sum with LASSO Regularization (HQWSLR);

–the Pure Quantile Weighted Sum with Ridge Regularization (PQWSRR);

–the Hourly QuantileWeighted Sum with Ridge Regularization (HQWSRR).

In the "pure" approaches, the weights were estimated without any differentiation in terms of daily duration, while in the "hourly" approaches, weights were estimated using only one-hour observations. In this way, the weights are differentiated by hours of the day to account for the daily periodicity of the PV power scheme.

In [14], the last four approaches were extended, starting with the Least Absolute Shrinkage and Selection Operator (LASSO) quantile regression and quantile Ridge Regression, respectively, which allows adjusting the weights by assigning a penalty associated with the weights.

In the PQWS and HQWS strategies, the weights were estimated without any regularization restrictions or losses. Compared to the constrained or regularized strategies, the PQWS and HQWS strategies return the lowest net sum in the sample because the minimization problem is unconstrained. However, there is no certainty that these weights are the best choice for out-of-sample predictions. This is a common problem for regression applied to prediction, where the preponderance of training data has negative consequences when the model is used to predict unknown data.

Since the ensemble combination of probabilistic individual forecasts is evaluated in terms of relative improvement over the individual forecasts, the use of a specific combination block construction strategy depends on the selected underlying models.

Forecast Evaluation. When operating any predictive model, sooner or later the question of the possibility of correcting its parameters arises. Every year, the growth of electricity consumption and its generation increases, the operating conditions of the electricity market and capacity change, the energy systems themselves develop, as a result, at a certain stage, the forecasting system turns out to be inadequate to the current state. Taking into account the specifics of the electric power industry, namely the dependence of the amount of generation on many different factors (weather conditions; type of day – weekend, working, holiday; type of consumer – industrial and household sector), the problem of retraining the forecasting system is obvious, because sooner or later the model will cease to provide an adequate forecast .

To realize the possibility of retraining, correction of model parameters based on the error of the original forecast is introduced into the forecasting system. The output forecast of the system is compared with the actual value of the Solar Microgrid generation and when the error will exceed a certain set allowable level, the system is retrained on the new data.

As a block of forecast evaluation, standard criteria for determining a qualitative forecast such as mean square error (MSE), root mean square error (RMSE), mean absolute error (MAE), mean absolute percent error (MAPE), coefficient of determination (R2) can be used [12], as well as methods based on the genetic algorithm or fuzzy logic.

Conclusions. To improve the forecasting accuracy in the Solar Microgrid management strategy, this study proposes a concept of a short-term PV power forecasting system that uses an ensemble forecasting method. The ensemble model is built by combining individual forecasting models as basic models and then organizing them into an ensemble combination. The model also provides for adaptation based on a change in the output value, if the forecast becomes of inadequate quality.

Further development of methods of combining forecasts should be carried out within the framework of two approaches: by including new types of individual models in the basic set of the combined model and by improving and developing new methods of combining forecasts.

Prospective directions for improvement of methods of combining forecasts include combining based on technologies of intelligent data analysis, and combining forecasts based on stable statistical estimates.

References

1.P. Mandal, S. T. S. Madhira, A. Ul Haque, J. Meng and R. L. Pineda, "Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques", *Proc. Computer Science*, vol. 12, pp. 332 – 337, 2012.

2.Y. Zhengque, C. Yapei, and X. Jiapeng, "Power generation forecasting model for photovoltaic array based on generic algorithm and BP neural network", *IEEE 3rd Int. Conf. on Cloud Computing and Intelligence Systems (CCIS)*, Shenzhen, pp.380-383, Nov, 2014.

3.M. Monfared, M. Fazeli, R. P. Lewis and J. R. Searle, "Fuzzy Predictor With Additive Learning for Very Short-Term PV Power Generation", *IEEE Access*, vol. 7, pp. 91183-91192, 2019.

4.A. K. Pani and N. Nayak, "Photovoltaic Power Forecasting by Evolutionary Algorithm-Based Improved Extreme Learning Machine", Adv. in El. Contr. and Sign. Syst., pp.109-129, 2020.

5.K. S. Sivhugwana and E. Ranganai, "Intelligent techniques, harmonically coupled and SARIMA models in forecasting solar radiation data: A hybridization approach", *Journal of Energy in Southern Africa*, vol. 31, pp. 14-37, 2020.

6.U. B. Tayab, F. Yang, A. S. M. Metwally and J. Lu, "Solar photovoltaic power forecasting for microgrid energy management system using an ensemble forecasting strategy", *Energy Sources*, *Part A: Recovery, Utilization, and Environmental Effects*, vol. 44, no. 4, pp. 10045-10070, Nov. 2022. doi.org/10.1080/15567036.2022.2143945.

7.V. Sharyfullyn and A. Sharyfullyna, "Neirosetevoe prohnozyrovanye potreblenyia elektroenerhyy promyshlennim predpryiatyem", *Vestnyk KHEU*, № 2, 2012. (Rus)

8.A. Kroshylyn, A. Babkyn and S. Kroshylyna, "Osobennosty postroenyia system podderzhky pryniatyia reshenyi na osnove nechetkoi lohyky", Nauchno-tekhnycheskye vedomosty SPbHTU, №2, pp. 58-63, 2010. (Rus)

9.V. Kyryk, "Matematychnyi aparat shtuchnoho intelektu v elektroenerhetychnykh systemakh: pidruchnyk", Igor Sikorsky KPI, Vyd-vo «Politekhnika», 224 p., 2019. (Ukr)

10.I. Kaya, M. Çolak and F. Terzi, "A comprehensive review of fuzzy multi criteria decision making methodologies for energy policy making", *Energy Strategy Reviews*, vol. 24, pp. 207-228, 2019. doi:10.1016/j.esr.2019.03.003.

11.D. Matushkin, A. Bosak and L. Kulakovskyi, "Analysis of factors for forecasting electric power generation by solar power plants", *POWER ENGINEERING: economics, technique, ecology*, vol. 4, pp. 64-69, 2020. doi:10.20535/1813-5420.4.2020.233597.

12.A. Bosak, D. Matushkin, V. Dubovyk, S. Homon and L. Kulakovskyi, "Determination of the concepts of building a solar power forecasting model", *Scientific Horizons*, vol. 24, no. 10, pp. 9-16, 2021. doi:10.48077/scihor.24(10).2021.9-16.

13.A. A. H. Lateko, H.-T. Yang, C.-M. Huang, "Short-Term PV Power Forecasting Using a Regression- Based Ensemble Method", *Energies*, vol. 15, no. 11, pp. 1-21, 2022. https://doi.org/10.3390/en15114171.

14.A. Bracale, G. Carpinelli and P. De Falco, "Developing and Comparing Different Strategies for Combining Probabilistic Photovoltaic Power Forecasts in an Ensemble Method", *Energies,* vol. 12, pp. 1-16, 2019. doi:10.3390/en12061011.

Д. С. Матушкін[1], аспірант, ORCID 0000-0003-4431-7862
А. В. Босак[1], канд. техн. наук, доцент, ORCID 0000-0003-0545-9980
[1]**Національний технічний університет України**
«Київський політехнічний інститут імені Ігоря Сікорського»

КОНЦЕПЦІЯ АНСАМБЛЕВОЇ СИСТЕМИ ПРОГНОЗУВАННЯ ДЛЯ ЗАДАЧ ОПТИМІЗАЦІЇ УПРАВЛІННЯ MICROGRID СОНЯЧНОЇ ЕНЕРГІЇ

Точні ймовірнісні прогнози відновлюваної генерації є рушійною силою для оптимізації експлуатації та управління MicroGrid системами. Комбінація прогнозів, наданих різними індивідуальними моделями, може дозволити підвищити точність прогнозів, однак, на відміну від комбінування точкових прогнозів, для яких просте зважене усереднення часто є вірогідним рішенням, комбінування ймовірнісних прогнозів *є набагато складнішим завданням. Сьогодні ансамблі моделей прогнозування є одним із перспективних напрямів розв'язання задач, де прогностична точність важливіша, ніж можливість інтерпретувати модель. Головна ідея ансамблів полягає в навчанні кількох базових моделей та агрегації результатів їх роботи. Емпіричні дослідження показують, що комбінації прогнозів у середньому частіше дають кращі прогнози, ніж методи, які базуються на виборі лише однієї моделі прогнозування. При побудові ансамблів особливої актуальності набуває питання забезпечення різноманітності моделей та ефективного навчання моделей-членів ансамблю. Стаття присвячена розв'язанню питань побудови ансамблевої моделі для прогнозування фотоелектричної (PV) потужності, яка поєднує результати декількох базових імовірнісних моделей. Використання запропонованого авторами ансамблевого методу може підвищити точність прогнозування та зменшити час, необхідний для тренування й оцінки моделей-членів ансамблю. Сформульовані напрями та перспективи подальших досліджень.*

Ключові слова: система прогнозування, ансамблі моделей прогнозування, генерація електричної енергії, сонячна електростанція, MicroGrid, ансамблева архітектура

Список використаної літератури

1.P. Mandal, S. T. S. Madhira, A. Ul Haque, J. Meng and R. L. Pineda "Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques", Proc. Computer Science, vol. 12, pp. 332 – 337, 2012.

2.Y. Zhengque, C. Yapei, and X. Jiapeng "Power generation forecasting model for photovoltaic array based on generic algorithm and BP neural network", IEEE 3rd Int. Conf. on Cloud Computing and Intelligence Systems (CCIS), Shenzhen, pp.380-383, Nov, 2014.

3.M. Monfared, M. Fazeli, R. P. Lewis and J. R. Searle "Fuzzy Predictor With Additive Learning for Very Short-Term PV Power Generation." IEEE Access, vol. 7, pp. 91183-91192, 2019.

4.A. K. Pani and N. Nayak, "Photovoltaic Power Forecasting by Evolutionary Algorithm-Based Improved Extreme Learning Machine", Adv. in El. Contr. and Sign. Syst., pp.109-129, 2020.

5.K. S. Sivhugwana and E. Ranganai, "Intelligent techniques, harmonically coupled and SARIMA models in forecasting solar radiation data: A hybridization approach", *Journal of Energy in Southern Africa*, vol. 31, pp. 14-37, 2020.

6.U. B. Tayab, F. Yang, A. S. M. Metwally and J. Lu, "Solar photovoltaic power forecasting for microgrid energy management system using an ensemble forecasting strategy", Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, vol. 44, no. 4, pp. 10045-10070, Nov. 2022. doi.org/10.1080/15567036.2022.2143945.

7.Шарифуллин В. Ш., Шарифуллина А. В. Нейросетевое прогнозирование потребления электроэнергии промышленным предприятием // Вестник КГЭУ, 2012. № 2 (13). С. 6-11.

8.Крошилин А. В., Бабкин А. В., Крошилина С. В. Особенности построения систем поддержки принятия решений на основе нечёткой логики // Научно-технические ведом. СПбГТУ, 2010. №2. С. 58-63.

9.Кирик В. В. Математичний апарат штучного інтелекту в електроенергетичних системах: підручник. Київ : КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2019. 224 с.

10.I. Kaya, M. Çolak and F. Terzi, "A comprehensive review of fuzzy multi criteria decision making methodologies for energy policy making", Energy Strategy Reviews, vol. 24, pp. 207-228, 2019. doi:10.1016/j.esr.2019.03.003.

11.D. Matushkin, A. Bosak and L. Kulakovskyi, "Analysis of factors for forecasting electric power generation by solar power plants", *POWER ENGINEERING: economics, technique, ecology*, vol. 4, pp. 64-69, 2020. doi:10.20535/1813-5420.4.2020.233597.

12.A. Bosak, D. Matushkin, V. Dubovyk, S. Homon and L. Kulakovskyi, "Determination of the concepts of building a solar power forecasting model", *Scientific Horizons*, vol. 24, no. 10, pp. 9-16, 2021. doi:10.48077/scihor.24(10).2021.9-16.

13.A. A. H. Lateko, H.-T. Yang, C.-M. Huang, "Short-Term PV Power Forecasting Using a Regression-Based Ensemble Method", *Energies*, vol. 15, no. 11, pp. 1-21, 2022. https://doi.org/10.3390/en15114171.

14.A. Bracale, G. Carpinelli and P. De Falco, "Developing and Comparing Different Strategies for Combining Probabilistic Photovoltaic Power Forecasts in an Ensemble Method", *Energies,* vol. 12, pp. 1-16, 2019. doi:10.3390/en12061011.

Надійшла: 21.03.2023
Recieved: 21.03.2023