Comparative analysis of models for short-term forecasting of electricity consumption

Mykola Korablyov[1], Igor Kobzev[2], Oleksandr Chubukin[1], Danylo Antonov[1], Vladyslav Polous[1] and Oleksandr Tkachuk[1]

[1] *Kharkiv National University of Radio Electronics, Kharkiv 61166, Ukraine*
[2] *Simon Kuznets Kharkiv National University of Economics, Kharkiv 61166, Ukraine*

Abstract

Forecasting electricity consumption is an urgent task, and the solution significantly affects the efficiency of the use of energy resources. The paper considers short-term forecasting of electricity consumption, which predicts the amount of energy that will be used in a short period, from several hours to several days in advance. There are various short-term forecasting models, so it is important to reasonably choose a model that provides analysis and effective forecasting of electricity consumption to optimize the use of energy resources. The purpose of the work is to analyze the main forecasting models, such as statistical models (autoregressive model, moving average, exponential smoothing, moving average with autoregression and integration) and deep learning models (artificial neural network, recurrent neural network, long short-term memory, transformer), indicating their advantages and disadvantages, and choosing the best of them. The experimental results of a comparative analysis of power consumption forecasting models are presented, which showed that the transformer model was 1.5% - 2% more effective in power consumption forecasting according to various metrics. Its higher level of accuracy, reflected in low error values and high coefficient of determination, indicates its high adaptability to the dynamics of electricity consumption.

Keywords

Forecasting, time series, power consumption, model, neural network, deep learning, accuracy

1. Introduction

The energy sector is critical to economic development and social well-being as it provides the energy required for various activities. However, power supply is often unstable and there is a need for accurate power consumption forecasting to balance the power system [1]. There is no efficient way to store large amounts of electrical energy. Therefore, the total amount of consumed electricity must be balanced with the generated. In industrial enterprises that use electricity as the main raw material, there may be a shortage of capacity if the consumption of electricity exceeds the established norms. On the other hand, when the electricity consumption is less than the established norms, there may be a waste of money.

The task of planning and forecasting electricity consumption is quite significant in the power industry. Timely receipt of information about the future load allows you to choose the optimal operating mode of the system. Forecasting is an important factor in drawing up the electricity balance in the power system, influencing the choice of mode parameters and estimated electrical loads. The balance of electricity is necessary to ensure the stable operation of the power system. If the balance is not maintained, the quality of electricity suffers (the frequency and voltage deviate from the required values). The accuracy of forecasting allows for the optimization of the operation of the electrical system. Forecasting electricity consumption is a complex task that is influenced by many factors. Short-term, medium-term, and long-term forecasting of electricity consumption are distinguished. The work deals with short-term forecasting, which predicts the amount of energy that will be used in a short period, from several hours to several days in advance. The main advantage of short-term forecasting is that it can help optimize power generation, transmission, and consumption

in real-time. On the other hand, short-term forecasting of electricity consumption has some limitations. It is susceptible to sudden changes in weather, human behavior, and other external factors that can lead to inaccurate forecasts. One of the main requirements for forecasting methods in the power industry is the calculation of electricity consumption volumes in different time intervals. If until now it was possible to get by with a simple method of linear regression or a method of daily comparison of indicators, now there is a need to take into account the non-linear effects of external factors, which requires the use of intelligent information processing methods. The accuracy of forecast calculations is determined by the correspondence of the mathematical models of the process of power consumption fluctuations. These fluctuations represent a complex non-stationary random process that has certain cycles [2, 3]. When applying mathematical models and software, company specialists are usually limited to values averaged over past periods ("manual forecast"). Simplified, "manual" forecasting of consumption can give quite high values of the mathematical expectation of errors, and a wide confidence interval and is practically not used for quick operational calculations at the pace of the process. When solving the problem of forecasting electricity consumption, the question of choosing a mathematical forecasting model arises. The adequacy of this model affects the accuracy of determining the planned electricity consumption during the formation of a price request for the purchase and sale of electricity [2, 4]. The error of forecast estimates determines the adequacy of the used mathematical models for the process of fluctuation of electricity consumption. The purpose of this study is to conduct a comparative analysis of various models of short-term forecasting of electricity consumption and to determine the best of them, which contributes to balancing and optimizing the use of the energy system, which is an urgent task.

2. Analysis of models for short-term forecasting of electricity consumption

For short-term forecasting of electricity consumption, a large number of approaches and methods can be applied using different technologies, such as statistical approach, machine and deep learning, expert systems, etc. Accordingly, there are various short-term forecasting models such as statistical models (autoregressive model, moving average, exponential smoothing, moving average with autoregression and integration), deep learning models (artificial neural network, recurrent neural network, long short-term memory, transformer), etc. It is important to reasonably choose a model that provides analysis and effective forecasting of electricity consumption in order to optimize the use of energy resources. We will analyze the main forecasting models that can be used to forecast electricity consumption.

2.1. Statistical models

2.1.1. Auto regression (AR) model

It assumes that there is a linear relationship between energy consumption and the independent variables used in the analysis, which are described by the expression [5-7]:

$$Y = b_1 \times X + c ,\qquad\qquad(1)$$

where b_1 is the regression coefficient; X is the value of the feature factor; c is a free term, a constant.

The AR model also assumes that the historical data used in the analysis are representative of future consumption patterns. The accuracy of forecasts depends on the reliability of historical data and the extent to which the relationships between variables remain stable over time. One of the main advantages of the autoregressive model is that it provides a clear and quantitative understanding of the factors that affect electricity consumption.

By identifying and quantifying the relationships between various factors, it enables energy companies to make informed decisions about supply and demand. An autoregression model can also help identify trends and patterns in energy consumption that can be used for long-term planning and investment decisions.

One of the main limitations of the autoregressive model is that it assumes a linear relationship between electricity consumption and independent variables. In reality, the relationship between

power consumption and variables can be non-linear or complex, which can lead to inaccuracies in forecasts. Another limitation is that the autoregressive model is based on historical data, which may not accurately reflect future consumption patterns. This may lead to forecast errors and inaccuracies, particularly if there are changes in the market or regulatory environment. Therefore, the AR model is useful for forecasting electricity consumption based on historical data. However, its accuracy depends on the reliability of historical data and the stability of relationships between variables over time.

2.1.2. Moving Average (MA) model

This model is based on the average value of previous electricity consumption and assumes that future electricity consumption will be the same as in the past. In this model, the moving average is calculated as the average of a fixed number of consecutive historical data points. The resulting value is then used as a forecast for the next period and is described by the expression [5-7]:

$$y_t = \sum_{r=-q}^{+s} \alpha_r x_{t+r},$$ (2)

where x_t is a time series; α_r is the sum of weights.

However, the MA model also has limitations. One of them is that it strongly depends on the length of the moving average interval. If it is too short, the forecast may be too volatile and not reflect long-term trends. On the other hand, if the interval is too long, the forecast may be too smooth and ignore short-term fluctuations. Therefore, determining the appropriate period size for a given data set can be challenging. Another limitation of the moving average model is that it assumes a constant nature of electricity consumption over time. This may not be true in cases where there are significant changes in the structure of electricity consumption. In such cases, the MA method may be inaccurate and must be combined with other forecasting methods.

Thus, the MA method is a simple and useful tool for forecasting electricity consumption, but it has its limitations.

It is important to carefully consider the appropriate size of the window and take into account all factors that can affect the structure of electricity consumption over time. By combining the moving average method with other forecasting methods, you can develop a more accurate and reliable forecast of electricity consumption.

2.1.3. Exponential smoothing

Assumes that future values of a series are based on past observations and that recent observations are more important than previous ones [6, 7]. This means that the weights assigned to the prior values decrease exponentially. Exponential smoothing works by calculating a forecast based on previous observations and an exponentially weighted average of past observations. The weight assigned to each observation decreases exponentially as the observations age. Exponential smoothing is described as follows [6, 7]:

$$y_{t+1} = \alpha x_t + (1 - \alpha) y_t,$$ (3)

where y_{t+1} is the forecast for the next period; α is the smoothing constant; x_t is the observed value of the series for period t; y_t is the old forecast for period t.

One of the main advantages of exponential smoothing is its simplicity, and it also does not require a large amount of historical data, making it useful for short-term forecasting. Furthermore, it is flexible enough to be adapted to a wide range of time series data. Exponential smoothing has several limitations, including that it is best suited for data with a smooth trend, seasonal patterns or cycles, and limited random fluctuations. It also assumes that forecast errors are normally distributed and independent of each other, which may not always be the case in practice. Finally, it can be sensitive to outliers, so it is important to remove them or adjust the weights accordingly.

Hence, exponential smoothing is a popular time series forecasting method that is easy to use and adaptable to a wide range of data. Although it has its limitations, it can be a powerful tool for short-term forecasting when used correctly.

2.1.4. Autoregressive Integrated Moving Average (ARIMA) model

Is an extension of the autoregressive moving average (ARMA) models for non-stationary time series, which can be made stationary using the operation of taking differences of a certain order d from the original time series. In the ARIMA (d, p, q) model, the future value of the process is a finite linear combination of its previous values and errors, and can be written as [7-9]:

$$y_t = \alpha_1 \gamma_{t-1} + \alpha_2 \gamma_{t-2} + \cdots + \alpha_p \gamma_{t-p} + \cdots + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \cdots - \beta_q \varepsilon_{t-q}, \qquad (4)$$

where y_t is the current value of the process; ε_t – random error at time t; α_i, β_j – coefficients; p, q are integers corresponding to the orders of autoregression and moving average, respectively.

Using the lag shift operator L, the general form of the model can be written as [7]:

$$\beta(L)y_t = \beta(L)\nabla^d y_t = \alpha_0 + \alpha(L)x_t,$$
$$\beta(L) = 1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_p L^p, \qquad (5)$$
$$\alpha(L) = 1 - \alpha_1 L - \alpha_2 L^2 - \cdots - \alpha_q L^q,$$

here $\alpha(L) = \nabla^d A(L)$ is an autoregression operator, which is a non-stationary operator for which d roots of the equation $\alpha(L) = 0$ are equal to one; $\beta(L)$ is a moving average operator, that is, the roots of the equation $\beta(L) = 0$ are located outside the unit circle.

In general, among the statistical models, the ARIMA model received the greatest distribution. It has demonstrated an effective ability to generate short-term forecasts and often outperforms complex structural models in short-term forecasting results.

2.2. Deep learning models in electricity consumption forecasting

One of the main approaches that can be used to implement a short-term forecast of electricity consumption is based on the use of artificial neural network models [10], which include a multilayer perceptron, recurrent neural network (RNN), long-short-term memory (LSTM), convolutional neural networks (CNN), transformers (autoencoders), etc. Let's analyze the most important models.

2.2.1. Artificial neural network (ANN)

It is a powerful model for predicting energy consumption. An ANN is a model (multilayer perceptron) that learns relationships in data without taking into account time dependencies. ANNs consist of several layers of interconnected nodes, or neurons, that process information and learn patterns from historical data to make predictions about future electricity consumption [11].

One of the key advantages of ANN is its ability to handle non-linear relationships and complex patterns in data. It can capture subtle and complex relationships that may be missed by traditional forecasting methods such as moving averages and exponential smoothing. The ANN is also highly adaptable and can be easily customized to meet specific electricity forecasting needs.

2.2.2. Recurrent neural network (RNN)

It is a deep learning model that is trained to process and transform a sequential set of input data into a sequential set of output data. In other words, RNN is an architecture that can work with sequential data. It uses a re-entry mechanism that allows it to take into account previous states and use them when processing input data (Figure 1) [12, 13]. RNN is called recurrent because it performs the same task for each element of the sequence, and the output depends on previous calculations. RNN is a network that has a "memory" that takes into account previous information. Theoretically, RNNs can use information in arbitrarily long sequences, but in practice, they are limited to only a few steps.

Unlike a traditional deep neural network, which uses different parameters in each layer, an RNN has the same parameters (U, V, W) at all stages. This means that the same task is performed at each step, using only different inputs. This significantly reduces the number of parameters that need to be fitted. The main feature of RNNs is the hidden state, which contains some information about the sequence.

Although the RNN should work with the entire sequence, unfortunately, there is a "vanishing

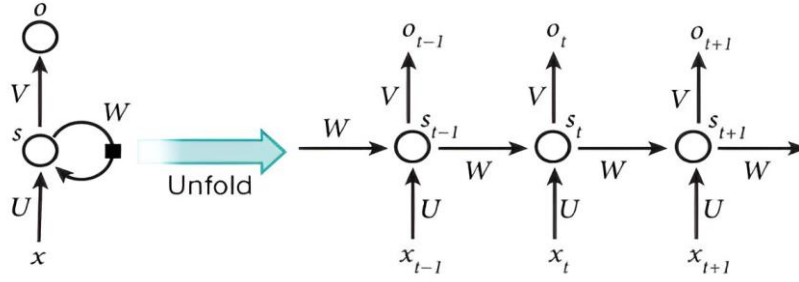gradient problem". This means that older inputs do not affect the output.

Figure 1: Recurrent Neural Network Architecture and its unfolding [12]

2.2.3. Long short-term memory (LSTM) model

The LSTM model solves the "vanishing gradient problem" by adding additional parameters (separate memory). LSTM is not fundamentally different from RNN, but it uses a different function to calculate the hidden state (Figure 2) [14-17].

LSTM is an extension of RNNs designed to overcome the problems of loss and gradient shifts. It uses special memory blocks that allow storing and updating information for a long time. The LSTM model is described by equations [14]:

$$f_t = \sigma(W_f[\,h_{t-1}, x_t] + b_f)\,,$$
$$i_t = \sigma(W_i[\,h_{t-1}, x_t] + b_i)\,,$$
$$\tilde{C}_t = tanh(W_C[\,h_{t-1}, x_t] + b_C)\,,$$
$$C_t = f_t \times C_{t-1} + i_t \times \tilde{C}_t\,,$$
$$y_t = \sigma(W_y[\,h_{t-1}, x_t] + b_y)\,,$$
$$h_t = y_t \times tanh(C_t)\,.$$

(6)

where x_t is an input vector; h_t is an output vector; C_t is a vector of states; W_f, W_i, W_C, W_y are parameter matrices; b_f, b_i, b_C, b_y are parameter vectors; f_t, i_t are valve vectors, σ is a sigmoidal activation function; $tanh$ is a hyperbolic tangent activation function.

The memory in LSTM is represented by cells that can be thought of as black boxes that take the previous state h_{t-1} as input and the current input parameter x_t. Inside, these cells decide which memory to keep and which to erase.

Then they combine the previous state, the current memory, and the input parameter. It turns out that these types of units are very effective in capturing (storing) long-term dependencies. LSTM models read the input data sequentially. If you need an architecture in which the sequence is processed simultaneously so that there is no loss of information, then such an architecture is implemented in the transformer model encoder [18, 19], which allows you to study the context of a variable taking into account its environment. In addition, it is often faster than RNNs.
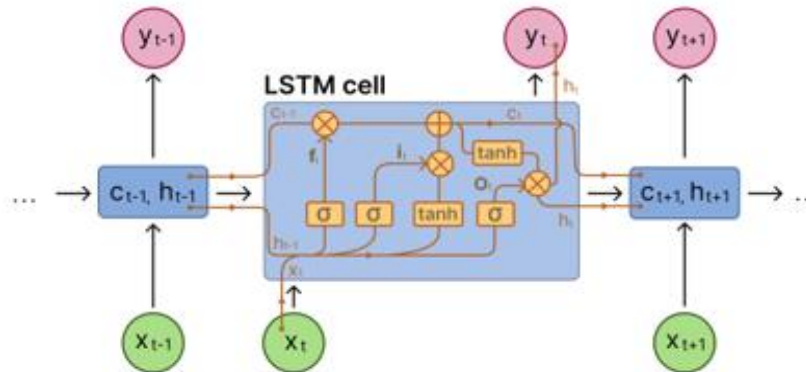

Figure 2: Long Short-Term Memory (LSTM) Architecture and LSTM cell (unit) [14]

2.2.4. Transformer (autoencoder)

The original architecture of the transformer is an autoencoder. The encoder receives as input a

sequence with positional information. The decoder receives as input a part of this sequence and the output of the encoder (Figure 3) [20].

On the left in Figure 3, the encoder processes the input sequence to create a hidden representation. On the right in Figure 3, the decoder uses the output of the encoder to generate the output sequence. In this case, the decoder works as an autoregressive model, using previously generated samples as additional input to generate the next output sample. Thus, the transformer model consists of input vector transformation, positional coding, normalization, forward propagation layers, linear layers, and attention layers.
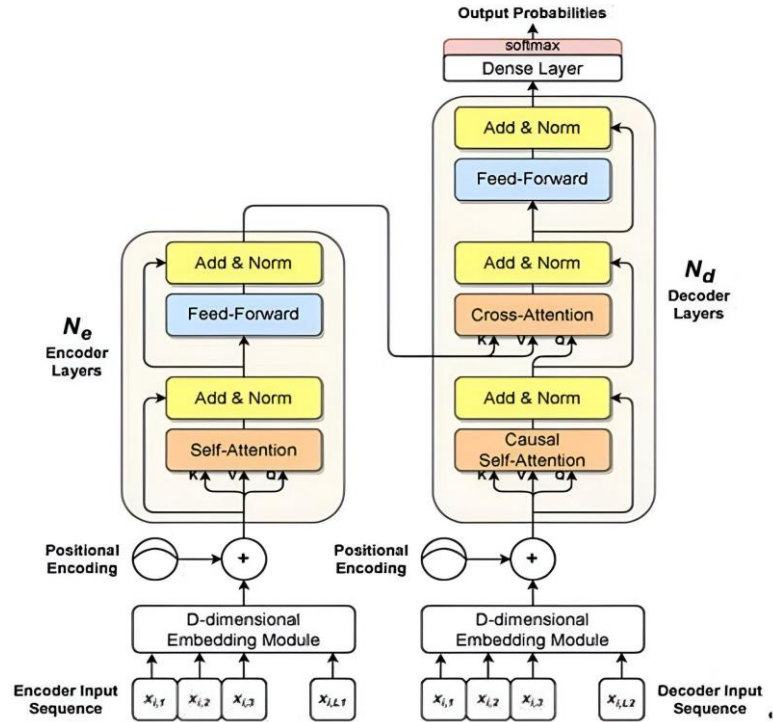

Figure 3: Transformer (Autoencoder) Architecture [20]

The most important computational blocks of a transformer are the attention mechanisms that allow the model to focus its attention on certain parts of the input data, depending on the information being processed.

An attention layer in a transformer is a mathematical model that allows you to estimate the relationship between sequence values. That is, the transformer maintains direct connections to all previous values in the time series, allowing information to spread over much longer sequences.

At the same time, transformers have the following disadvantages. In the transformer, the scaled scalar product of the attention mechanism is insensitive to the local context, which can make the model prone to anomalies in time series prediction.

Transformers are characterized by memory bottlenecks, which lead to a large complexity of the transformer space when processing long sequences. Transformers have a high temporal complexity, which limits their use for long-term forecasting. In recent years, many variants of transformers [21-25] have been proposed for time series prediction, addressing issues related to the level of application, attention mechanisms, and encoder-decoder structure. In general, transformers are a powerful tool for predicting electricity consumption.

3. Results of comparative analysis of electricity consumption models

To perform a comparative analysis of electricity consumption forecasting models and determine the best system was created, consisting of the following main modules:

1. Data set module. Collects the necessary information for training and testing forecasting models.
2. Data preparation and analysis module. The collected data is normalized, formatted, reconciled, processed, and prepared for analysis and modeling.

3. Forecasting module. Based on artificial intelligence methods, electricity consumption forecasting models are analyzed, the best one is selected, and the forecast is executed.
4. Results analysis module. The trained prediction model is evaluated on a test dataset to measure its performance.
5. Module of forecasting results. The forecast results allow managers and utilities to optimize operations, resources, and system reliability.

To forecast electricity consumption, we used the "Hourly Energy Consumption" dataset from the Kaggle online resource, which consists of data from PJM Interconnection, a regional electricity transmission organization that is part of the Eastern Interconnection network that manages the electricity transmission system in the United States.

Hourly data on electricity consumption are shown in Figure 4 and are indicated in megawatts (MW). The system's data preparation and analysis module uses a systematic approach to data processing and analysis, creating a solid foundation for further study and understanding of electricity consumption dynamics in the broader context of the energy sector. The initial stage of data preparation involves examining the dataset to provide an initial overview of the structure.
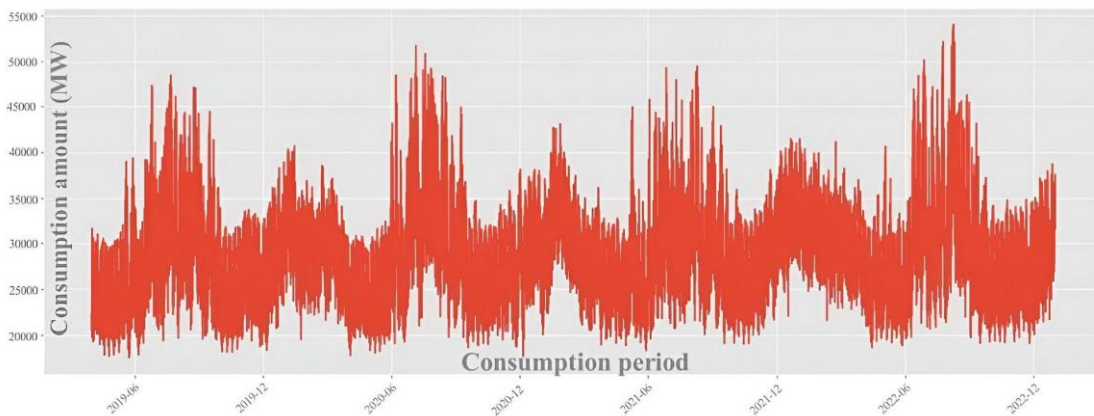

Figure 4: Hourly electricity consumption data in megawatts (MW)

Next, to enable effective time-based analysis, the data is converted to a single structure where a column over time is set as an index. The resulting dataset with a time index is fundamental for studying and analyzing the dynamics of electricity consumption over time. This makes it possible to examine long-term trends in detail and identify factors that affect energy performance.

The next step in data processing is to use methods to ensure consistency and homogeneity. An important element of this approach is normalization, which allows you to create a standardized data format, facilitating further comparison and analysis. At the end of this stage, the importance of visual interpretation of the data was taken into account. Graphs are used not only to illustrate changes in electricity consumption but also to highlight key patterns and trends. This contributes to a deeper understanding of the dynamics of energy consumption.

One of the main modules of the system is the forecasting module, the quality of the results of electricity consumption forecasting depends on the efficiency of its functioning. This system module conducts a practical analysis of various machine and deep learning models on specific data sets to select the best one for forecasting.

With the constant evolution of technology and the demands of the modern world, determining the accuracy of a model becomes an important task. To quantify the errors of forecasting models, various accuracy metrics have been calculated, namely:
1. Mean square error (MSE):

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2. \tag{7}$$

2. Mean absolute error (MAE):

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

3. Mean absolute percentage error (MAPE):

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^{n} | \frac{y_i - \hat{y}_i}{y_i} | \times 100, \tag{9}$$

where n is the number of observations; y_i is the actual value of the i-th observation; $\hat{y}_i$ is the predicted value of the i-th observation.

1. Coefficient of determination R^2:

$$R^2 = 1 - \frac{D(y/x)}{D(y)}, \tag{10}$$

where $D(y) = \sigma_y^2$ is the variance of the random variable y; $D(y/x) = \sigma^2$ is the conditional variance of the dependent variable (variance of the model error).

This indicator, which is used in statistical models, measures the extent to which changes in the independent variables affect the dependent variable. That is, it shows how accurately the model explains the variation in the dependent variable.

The coefficient of determination R^2 can take values from 1 to 0 in a classical linear multiple regression, where a higher value of the coefficient indicates a better fit of the model to the observations.

All these indicators allow us to analyze the accuracy of models in the face of constant technological evolution.

Taking into account both absolute accuracy and percentage deviations, you can get a complete picture of forecasting performance.

The use of MSE, MAE, and MAPE allows for a deeper study of various aspects of errors and their impact on the model.

The coefficient of determination R^2, in turn, becomes a key indicator in determining how well the model adapts to the data. Overall, by taking these metrics into account, an objective analysis can be made and an informed decision can be made about the effectiveness and suitability of the model in question.

The last module of the system highlights the forecasting results that have been obtained by the system.

They are presented in the form of graphs and tables that are used by managers to optimize operations, resources, and reliability of the electricity system.

When performing experimental research, the main tool for software development was the Python programming language, its NumPy and Pandas libraries, as well as the Scikitlearn and TensorFlow frameworks.

To conduct a comparative analysis of electricity consumption forecasting models, the best models from the respective groups were selected: from statistical models – ARIMA model, from recurrent models – LSTM model, from deep learning models – transformer model.

The results of predicting electricity consumption at different time intervals using these models are given in Table 1, and the results of their comparative analysis according to the selected indicators are given in Table 2.

Table 2 shows that the use of all the above models allows us to obtain a forecast with the required accuracy.

At the same time, when analyzing and comparing the results of this study, it was found that the transformer model was the most effective in predicting energy consumption. Its high level of accuracy, reflected in the low values of the mean square error, the mean absolute error, and the high coefficient of determination, indicates its high adaptability to the dynamics of energy consumption.

For a better visual analysis and comparison of energy consumption forecasting results, several detailed graphs were created for different parts of the time series, shown in Figure 5 and Figure 6, respectively.

They show the actual energy consumption data along with the predicted values obtained using the ARIMA, LSTM, and transformer models.

Table 1
The results of forecasting electricity consumption at different time intervals using these models

Date	Actual data	ARIMA model	LSTM model	Transformer
2022-02-12 00:00:00	11494.0	11196.5	11509.6	11366.0
2022-02-12 01:00:00	8728.0	9983.2	8691.0	10398.2
2022-02-12 02:00:00	8390.0	8393.5	8394.4	7870.7
2022-02-12 03:00:00	8283.0	7972.0	8279.3	8363.8
2022-02-12 04:00:00	8195.0	8243.5	8222.8	8601.4
2022-02-12 05:00:00	8150.0	8365.9	8312.2	8587.6
2022-02-12 06:00:00	8308.0	8628.6	8351.3	8640.2
2022-02-12 07:00:00	8588.0	9061.6	8671.2	8962.3
2022-02-12 08:00:00	9000.0	9138.3	8969.4	9261.0
2022-02-12 09:00:00	9290.0	9580.7	9385.9	9520.2

Table 2
Results of comparative analysis of models by different metrics

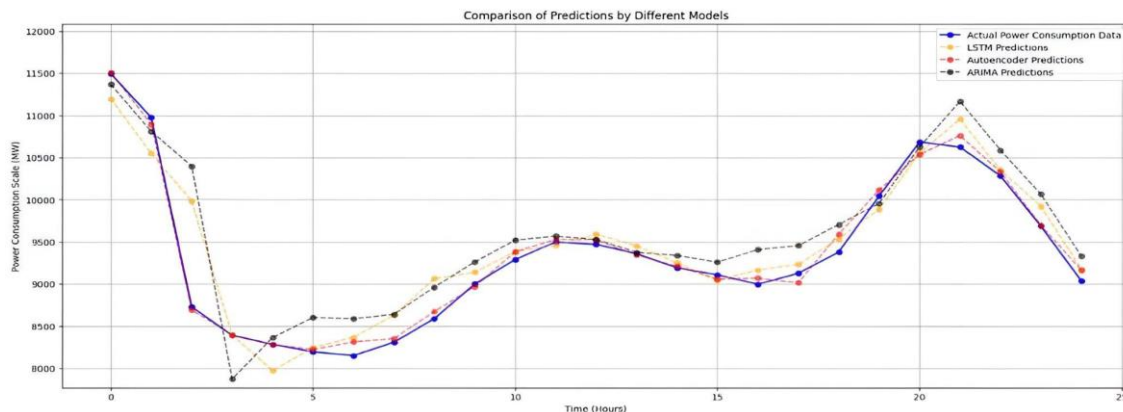| Model	MSE	MAE	MAPE	R^2
ARIMA	0,153	0,324	3,1%	0,954
LSTM	0,151	0,289	2,5%	0,965
Transformer	0,119	0,209	1,5%	0,985

Figure 5: Detailed graphs of electricity consumption forecasting for one section of the time series

Each figure allows you to perform a visual comparative analysis, where you can see the deviation between the actual and predicted values for each model. This allows for a visual assessment of the accuracy and performance of each model at different points in the time series. This visual approach contributes to a better understanding of trends and the overall adaptability of the models to changes in time-based energy consumption. These graphs are an important tool for making informed conclusions and determining the most effective model to use in forecasting.

Figures 5 and 6 show that the predicted values using the Transformer model are almost always near the line of actual electricity consumption, or even these values overlap.

Based on careful comparisons with other models, the transformer model appears to be not only the most accurate (by all metrics it is 1.5% - 2% better than other models), but also the most versatile model in different conditions. Its ability to adapt to changes in the time series and its high accuracy

make it the most effective for accurate forecasting of electricity consumption. This conclusion is supported by both quantitative data from the tables and graphs presented and conclusions drawn from the visual analysis. All these factors make the transformer model the most promising choice for further applications in the field of electricity consumption forecasting.
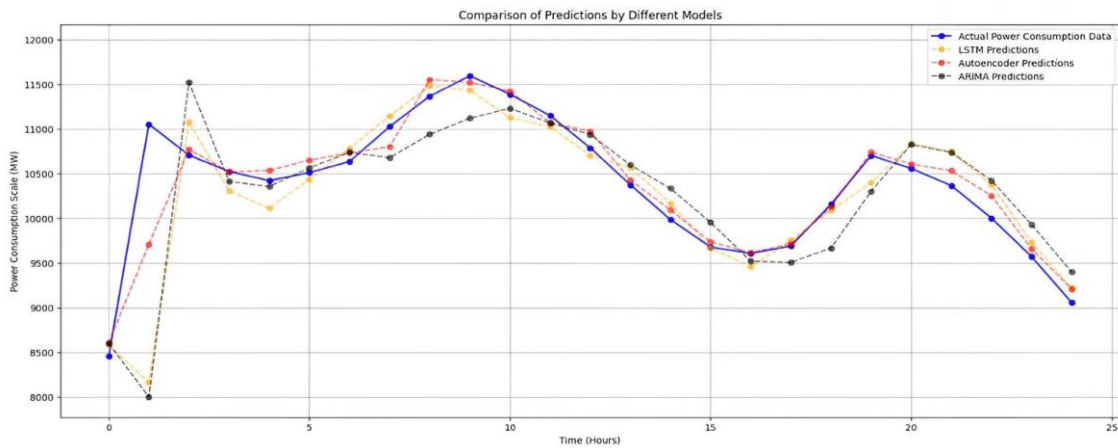

Figure 6: Detailed graphs of electricity consumption forecasting for another part of the time series

Prospects for further research include hybrid models that combine statistical methods, machine learning methods, and deep neural networks aimed at improving the reliability and accuracy of forecasting, as well as the use of graph neural networks (GNN) for multivariate time series forecasting.

4. Conclusions

Planning and forecasting of electricity consumption is quite important in the power industry. Timely receipt of information about the future load allows choosing the optimal system operation mode. There are short-term, medium-term, and long-term forecasting of electricity consumption. This paper considers short-term forecasting, which predicts the amount of energy that will be used in a short period, from several hours to several days in advance. The main advantage of short-term forecasting is that it can help optimize the production, transmission, and consumption of electricity in real time.

When solving the problem of forecasting electricity consumption, the question arises of choosing a mathematical forecasting model, the adequacy of which affects the accuracy of determining the planned electricity consumption when formulating pricing policy. There are various short-term forecasting models, so it is important to reasonably choose a model that provides analysis and effective forecasting of electricity consumption to optimize the use of energy resources. The article analyzes the main forecasting models, namely statistical models (autoregressive model, moving average, exponential smoothing, moving average with autoregression and integration) and deep learning models (artificial neural network, recurrent neural network, long short-term memory, transformer), indicating their advantages and disadvantages.

To perform a comparative analysis of electricity consumption forecasting models and determine the best one, a corresponding system was created, consisting of the following main modules: a data set module, a data preparation and analysis module, a forecasting module, a results analysis module, and a forecasting results module. To forecast electricity consumption, we used the Hourly Energy Consumption dataset from the Kaggle online resource, which consists of data from PJM Interconnection, a regional electricity transmission organization that is part of the Eastern Interconnection network that manages the electricity transmission system in the United States.

For a comparative analysis of electricity consumption forecasting models, the best models from the respective groups were selected: from statistical models - the ARIMA model, from recurrent models - the LSTM model, from deep learning models - the transformer model. The experimental results of the comparative analysis of these models by various metrics are presented, which show that the transformer model proved to be the most effective in predicting energy consumption (by all metrics it is 1.5% - 2% better than other models). Its high level of accuracy, reflected in low error

values and a high coefficient of determination, indicates its exceptional adaptability to the dynamics of electricity consumption. The transformer model appears to be not only the most accurate but also the most versatile model in different conditions. Its ability to adapt to changes in time series and high accuracy make it the most effective for accurate forecasting of electricity consumption.

Prospects for further research include hybrid models that combine statistical methods, machine learning methods, and deep neural networks aimed at improving the reliability and accuracy of forecasting, as well as the use of graph neural networks (GNN) for multivariate time series forecasting.

References

[1] R. Hafezi, M. Alipour, Energy Security and Sustainable Development. In the book: Affordable and Clean Energy. Publisher: Springer, Cham, 2020. DOI: 10.1007/978-3-319-71057-0_103-1.

[2] Y. W. Lee, T. K. Gaik, C. Y. Yee, Forecasting Electricity Consumption Using Time Series Model. International Journal of Engineering & Technology 7 4 (2018) 218-223. DOI: 10.14419/ijet.v7i4.30.22124.

[3] R. Adhikari, R. K. Agrawal, An Introductory Study on Time series Modeling and Forecasting. Publisher: LAP Lambert Academic. 2013. DOI: 10.13140/2.1.2771.8084.

[4] P. Malik, A. S. Dangi, A. S. Thakur, A. P. S. Parihar, U. Sharma, L. Mishra, An Analysis of Time Series Analysis and Forecasting Techniques. International Journal of Advance Research and Innovative Ideas in Education 9 5 (2023). DOI: 16.0415/IJARIIE-21608.

[5] A. A. Mutairi, Time-series forecasting for some statistical models. Advances and Applications in Statistics 78 (2022) 83-92. DOI:10.17654/0972361722051.

[6] J. Kaur, K. S. Parmar, S. Singh, Autoregressive models in environmental forecasting time series: a theoretical and application review. Environmental Science Pollution Research 30 (2023) 19617-19641. DOI: 10.1007/s11356-023-25148-9.

[7] M. Zhang, Time Series: Autoregressive models AR, MA, ARMA, ARIMA. University of Pittsburgh, 2018.

[8] V. I. Kontopoulou, A. D. Panagopoulos, I. Kakkos, G. K. Matsopoulos, A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data-Driven Networks. Future Internet 15 8 (2023) 255. DOI:10.3390/fi15080255.

[9] A. Petrova, M. Deyneka. ARIMA models: modeling and forecasting prices of stocks. International Scientific Journal "Internauka". Series: "Economic Sciences", 2022, https://doi.org/10.25313/2520-2294-2022-2.

[10] C. Lu, S. Li, Z. Lu. Building energy prediction using artificial neural networks: A literature survey. Energy and Buildings, Vol. 262, 2021. DOI: 10.1016/j.enbuild.2021.111718.

[11] E. Chianese, F. Camastra, A. Ciaramella, T. C. Landi, A. Staiano, A. Riccio, Spatio-temporal learning in predicting ambient particulate matter concentration by multi-layer perceptron. Ecological Informatics 49 (2019) 54–61. DOI:10.1016/j.ecoinf.2018.12.001.

[12] U. Ugurlu, I. Oksuz, O. Tas, Electricity price forecasting using recurrent neural networks. Energies 11 (2018) 1255. DOI:10.3390/en11051255.

[13] L. G. B. Ruiz, R. Rueda, M. P. Cuéllar, M. Pegalajar, Energy consumption forecasting based on Elman neural networks with evaluative optimization. Expert Systems with Applications. 92 (2018) 380–389. DOI:10.1016/j.eswa.2017.09.059.

[14] S. Hochreiter, J. Schmidhuber, Long Short-Term Memory. Neural Computation, 9 8 (1997) 1735-1780.

[15] G. V. Houdt, C. Mosquera, G. Nápoles, A Review on the Long Short-Term Memory Model. Artificial Intelligence Review 53 1 (2020). DOI: 10.1007/s10462-020-09838-1.

[16] S. Arifin, A. K. Wijaya, R. Nariswari, A. Yudistira, F. Suwarno, D. Wihardini, Long Short-Term Memory (LSTM): Trends and Future Research Potential. International Journal of Emerging Technology and Advanced Engineering (2023). DOI: 10.46338/ijetae0523_04.

[17] M. Korablyov, O. Fomichov, D. Antonov, S. Dykyi, I. Ivanisenko, S. Lutskyy, Hybrid stock analysis model for financial market forecasting, in: Proceedings of the 18th International

Conference on Computer Science and Information Technologies (2023) 1-4. DOI:10.1109/CSIT61576.2023.10324069.

[18] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is All You Need. Advances in Neural Information Processing Systems (2017).

[19] R. E. Turner. An Introduction to Transformers, 2023. DOI:10.48550/arXiv. 2304.10557.

[20] X. Amatriain. Transformer models: an introduction and catalog. A Preprint, 2023. DOI: 10.48550/arXiv.2302.07730.

[21] H. Zhou, S. Zhang, J. Peng, S. Zhang, J. Li, H. Xiong, W. Zhang, Informer: Beyond efficient transformer for long sequence time series forecasting, in: Proceedings of the AAAI Conference on Artificial Intelligence 35 (2021) 11106–11115. DOI:10.1609/aaai.v35i12.17325.

[22] A. Casolaro, V. Capone, G. Iannuzzo, F. Camastra, Deep Learning for Time Series Forecasting: Advances and Open Problems. Information 14 11 (2023) 598. DOI:10.3390/info14110598.

[23] S. Li, X. Jin, Y. Xuan, X. Zhou, W. Chen, Y. X. Wang, X. Yan, Enhancing the locality and breaking the memory bottleneck of transformer on time series forecasting. Neural Information Processing. 2019, 32.

[24] H. Wu, J. Xu, J. Wang, M. Long, Autoformer: Decomposition transformers with auto-correlation for long-term series forecasting. Neural Information Processing 34 (2021) 22419–22430.

[25] P. Delgado-Santos, R. Tolosana, R. Guest, F. Deravi, R. Vera-Rodriguez, Exploring transformers for behavioral biometrics: A case study in gait recognition, Pattern Recognition 143 (2023) 109798. DOI:10.1016/j.patcog.2023.109798.