UDC 621.311(574.3)
DOI: 10.15587/1729-4061.2025.345073

DEVELOPMENT OF A FORECASTING MODEL FOR OPTIMIZING ENERGY CONSUMPTION AT COAL ENTERPRISES

Shynar Telbayeva
Corresponding author
Doctoral Student*
E-mail: tshz@mail.ru
Leonid Avdeyev
Candidate of Technical Sciences*
Vladimir Kaverin
PhD, Acting Professor*
Dinara Zhumagulova
Senior Lecturer*
*Department of Energy, Automation and Telecommunications
Abylkas Saginov Karaganda Technical University
N. Nazarbayev ave., 56, Karaganda, Republic of Kazakhstan, 100027

The study object is daily data on electricity consumption of one of the coal mines in the Karaganda basin for 2024. This article solves the problem of the lack of accurate tools that can predict complex and variable modes of energy consumption in a coal mine and thereby ensure more efficient management of energy-intensive installations.

This article presents a comparative analysis of three electricity demand forecasting models using data from a coal mine in the Karaganda basin for 2024. The study explores the effectiveness of both classical approaches (seasonal ARIMA model and simple exponential smoothing) and an LSTM neural network model. To handle non-stationary data, the first difference method was applied, allowing the time series to be stationary. The forecast was generated for 7 days in advance. A comparative analysis of the models' accuracy was conducted using the MAPE metric on both the training and test sets. The study found that the LSTM model demonstrated the best results with a MAPE of 5.37% on the test set demonstrating its superior ability to capture complex data dynamics compared to ARIMA and simple exponential smoothing.

The developed predictive LSTM model can be effectively used in automated energy monitoring and management systems, providing accurate short-term load forecasts for coal mines and other mining and metallurgical enterprises with complex and volatile energy structures, provided the initial data is highly reliable and complete.

Keywords: power consumption mode, coal mine, time series, ARIMA, exponential smoothing, neural network model, LSTM, MAPE, test sample, stationarity

Received 16.09.2025
Received in revised form 18.11.2025
Accepted 27.11.2025
Published 17.12.2025

How to Cite: *Telbayeva, S., Avdeyev, L., Kaverin, V., Zhumagulova, D. (2025). Development of a forecasting model for optimizing energy consumption at coal enterprises. Eastern-European Journal of Enterprise Technologies, 6 (4 (138)), 26–35. https://doi.org/10.15587/1729-4061.2025.345073*

1. Introduction

The modern energy industry is facing increasing demands for the accuracy of forecasting power consumption parameters, since the efficiency of energy system management directly depends on the ability to anticipate load changes in conditions of high uncertainty. The growth of energy-intensive industries, the introduction of digital technologies, and the development of automated control systems have led to the fact that the task of accurately predicting energy consumption has become global and critically important. Worldwide, there is an increasing interest in intelligent forecasting methods that can take into account the nonlinear structure of data, seasonal fluctuations, the effects of stochastic factors and complex relationships between technological processes.

Industrial enterprises, including the mining sector, play a key role in the structure of global energy consumption. At the same time, the daily schedules of electricity consumption are characterized by high dynamics, pronounced seasonality and sensitivity to technological modes. In such circumstances, traditional forecasting methods based primarily on linear models often turn out to be insufficiently accurate, and emphasize the need to move to adaptive, computationally stable and intelligent forecasting models capable of operating in real time and providing high-quality forecasting in conditions of uncertainty.

This problem is of particular importance for energy-intensive industries, including the coal industry. Coal enterprises all over the world have a complex architecture of energy supply: long electrical networks, many heterogeneous energy consumers, irregular equipment operating hours and the stochastic nature of the electrical load. These features make the forecasting task particularly difficult and require the use of methods that take into account the nonlinear dynamics and internal patterns of time series [1].

Despite a significant amount of global research in the field of energy consumption forecasting, existing approaches often prove insufficient for industrial facilities with a complex load structure. This necessitates the development and implementation of modern intelligent models capable of providing higher accuracy of short-term forecasting and improving the efficiency of energy management at such enterprises [1].

2. Literature review and problem statement

In the work [2], an approach to optimizing the operation of an integrated coal mine energy system is considered, characterized by a high level of energy consumption and a complex structure of energy sources. To reduce the impact of uncertainties related to load changes, renewable energy generation, and forecasting errors, the authors proposed a multiscale interval strategy for optimal dispatching control, providing more accurate and flexible management of energy flows. It should be noted that electricity generation and consumption may vary due to weather, forecast errors, or unstable equipment operation, while in underground coal mining, the temperature is constant and the operating mode is round-the clock.

In this work [3], the authors propose a new model called BayesMAR for time series forecasting. BayesMAR combines the advantages of median regression and the Bayes approach allowing it to be robust to outliers and effectively account for uncertainty in the model. It should be noted that the proposed BayesMAR is very challenging for a forecasting model in coal mines since coal mines require the fastest possible forecast data acquisition.

Paper [4] present a method that uses Digital Twin technology and deep learning for highly accurate and adaptive forecasting of electricity consumption in large industrial furnaces for aluminum annealing. Traditional forecasting approaches are ineffective because multi-day furnace production cycles, interrelated operating modes, and complex thermal process dynamics create insurmountable obstacles to accurate modeling. It should be noted that aluminum annealing furnaces, as large-scale energy-intensive industrial equipment, operate at a single voltage level, whereas a coal mine is a highly complex facility with multiple network voltage levels.

Paper [5] propose a model based on the least-squares support vector machine (LS-SVM) for optimizing electricity consumption in Turkiye. The authors [6], in turn, examine and compare the effectiveness of both traditional statistical and modern machine learning algorithms (XGBoost, Linear Trees, Prophet) to forecast electricity consumption in the UK. The article [7] presents a method for short-term seasonal forecasting of hourly electricity demand in New England. However, it should be noted that the proposed models are acceptable for a specific country with specific parameters, whereas the nature of formation of electric load patterns in coal mines is stochastic, determined by the number of power consumers and a variety of technological characteristics and operating modes of electrical equipment [8].

The AECF-UC method proposed in paper [9] is designed to overcome the shortcomings of traditional energy consumption forecasting models. It adapts to gradual changes in data and accounts for user diversity. The model uses an innovative joint loss function and dynamic weight adjustment to cope with changing data patterns. A significant disadvantage of this method is its reliance on universal environments, which leads to the neglect of critical peak loads.

The paper [10] presents an adaptive method (AECF-UC) that addresses the problem of variability in data patterns and consumer heterogeneity in electricity consumption forecasting in Vietnam. It should be noted that the model uses a joint loss function and dynamic weight adjustment, which allows it to adapt to new conditions while the quantitative characteristics of primary information sources mainly depend on the size of the mine and the applied power supply scheme.

The works [11, 12] analyze the issues of energy infrastructure planning and grid management in the Kingdom of Bahrain, as well as develop approaches aimed at assisting the Ministry of Energy of the Philippines in optimizing electricity consumption and implementing effective strategies to respond to fluctuations in consumer demand. In particular, the study [11] uses the ARIMA model to analyze and predict electricity consumption based on data from the Ministry of Energy of the Philippines. The use of the K-means algorithm makes it possible to differentiate periods of high, medium and low consumption, which ensures the identification of months of peak demand. For future planning in the Kingdom of Bahrain, paper [12] proposes a new hybrid forecasting model based on a combination of ARIMA models (for linear patterns) and artificial neural networks to account for nonlinear dependencies.

It should be noted that the authors' study is mainly focused on minimizing total electricity consumption across the national grid (Philippines) and the Kingdom of Bahrain, while our study focuses on optimizing the power consumption modes of specific energy-intensive installations in a mining enterprise. This approach requires taking into account the specifics of operational modes and a multi-level network architecture.

Thus, in the reviewed studies, the existing universal models do not always effectively predict coal mine electricity consumption modes as they fail to account for the facility specifics, namely 24-hour operation, constant temperature, and the highly complex, stochastic nature of the load caused by multiple technological processes and a multi-level network. As a result, these methods are too complex for operational application, use irrelevant external factors, and fail to account for critical peak loads the analysis of which is key to this study.

The above justifies the need to develop a forecasting model capable of taking into account the complex dynamic characteristics of energy consumption in coal mines to optimize energy consumption for coal enterprises.

3. The aim and objectives of the study

The work's aim is to develop forecasting models for optimizing energy consumption at coal enterprises the results of which will enable implementation of optimal energy consumption modes and ensure effective management of underground coal mining.

To achieve this aim, the following objectives were accomplished:

to perform the necessary transformations to bring the time series to a stationary form;

– to conduct a study of modern time series forecasting models and determine the optimal ones for modeling the energy consumption of coal enterprises;

– to perform a comparative analysis of the accuracy of forecast models on a test sample using the MAPE metric and to determine the most effective model for forecasting the dynamics of electricity consumption.

4. Materials and methods

The study object is daily data on electricity consumption of one of the coal mines in the Karaganda basin for 2024.

The main hypothesis of the study is that the use of deep learning models, in particular the architecture of recurrent neural networks with long short-term memory (LSTM), provides significantly higher accuracy in short-term forecasting of electricity consumption at coal enterprises compared to classical time series models (ARIMA, SES), which is a critical condition for effective operational optimization of energy load.

It was assumed that the emissions detected in the time series of electricity consumption were seasonal fluctuations and did not need to be eliminated.

The following assumptions were made when conducting the study:

– stationarity of the time series: for the correct application of a number of classical analytical methods, it is necessary that the time series under study be stationary;

– unchanged technological scheme: within the study period, the technological scheme for coal mining and ventilation at the enterprise is considered unchanged, excluding major reconstructions or the introduction of new energy-intensive units capable of causing a structural shift in the data;

– quality of input data: the collected data on electricity consumption must be reliable, complete, and reflect actual consumption without significant measurement errors.

Time series reflecting energy consumption dynamics have a number of features, such as seasonality, trend, and random fluctuations, which require the use of various modeling methods to improve the accuracy of forecasts. The graph of electricity consumption for 2024 is shown in Fig. 1.

Fig. 1 shows that the initial time series has a pronounced seasonality. There are regular peaks and troughs in electricity consumption throughout the year. Peaks tend to occur in the winter months, while declines occur in the summer months. The data has high volatility, i.e. there are significant fluctuations in electricity consumption from one day to the next. They may indicate atypical failures in the accounting system, which will later require additional analysis. In the study, these emissions were considered seasonal, and the possibility of eliminating them was not considered.

The time series shown in Fig. 1 is not stationary. Forecasting non-stationary time series is a difficult task due to the presence of trends, seasonal fluctuations and variability of statistical properties over time.

One of the key problems is the presence of a deterministic or stochastic trend, which leads to a change in the average level of the series. If the trend is not eliminated or not taken into account in the model, this will cause a systematic bias in forecasts. Failure to account for seasonality leads to underestimation or overestimation of future values, especially during peak periods. Therefore, for the methodologically correct application of a wide range of classical analytical approaches, it is required that the time series possess the properties of stationarity.

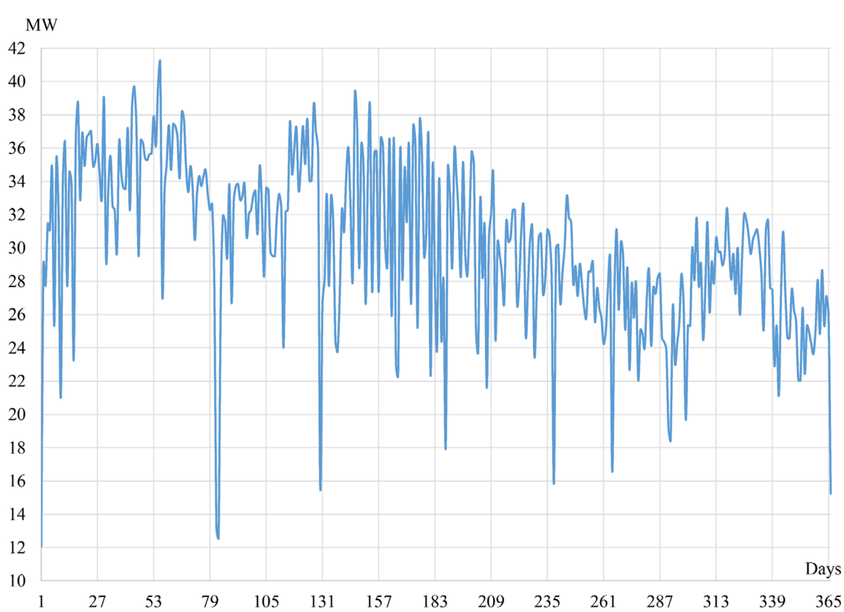

Fig. 1. Electricity consumption chart for 2024

To bring the studied time series of power consumption to a stationary form, standard data preprocessing methods were used: logarithm of the dependent variable and the method of first differences. The stationarity of the time series was evaluated using the extended Dickey-Fuller test.

The experiment plan includes the following steps:

– data preparation: the selected series will be divided into training and test samples. The training sample will contain values from January 1, 2024, to December 24, 2024. The last 7 values (from December 25, 2024, to December 31, 2024) of the total time series are selected for the test sample in order to compare the forecast values obtained for the models built based on the training sample with the existing data and, based on this, evaluate the quality of the forecast;

– building forecast models: each of the three models (ARIMA, SES, LSTM) will be trained on the training sample, and forecast values for 7 days ahead will be calculated for each model;

– comparative analysis: the obtained forecast values will be compared with the actual data from the test sample, and the accuracy will be assessed using the MAPE metric.

The research methods used include analysis of existing forecasting models:

– a seasonal ARIMA model, it allows to take into account the autocorrelation structure of the data and the trend due to the integration parameter, and also takes into account the seasonal component;

– simple exponential smoothing (SES) model;

– neural network modeling based on the LSTM model.

The seasonal ARIMA model is an extension of the classical ARIMA (Autoregressive Integrated Moving Average) model designed for analyzing and forecasting time series with seasonal fluctuations. Seasonal ARIMA is a powerful tool for seasonal time series, but requires careful tuning. The general form of the model is [13]

$$\Phi_P\left(B^s\right)\phi_p\left(B\right)\nabla_s^D\nabla^d X_t = \Theta_Q\left(B^s\right)O_q\left(B\right)\varepsilon_t, \quad (1)$$

where X_t – time series; ε_t – white noise; B – shift operator ($BX_t = X_{t-1}$); s – seasonal period; ∇^d – non-seasonal differentiation operator of d order; ∇_s^D – seasonal differentiation operator of D order; $\Phi_P(B)$ – non-seasonal AR polynomial of p order; $O_q(B)$ – non-seasonal MA polynomial of q order; $\Phi_P(B^s)$ – seasonal AR polynomial of P order; $\Theta_Q(B^s)$ – seasonal MA polynomial of Q order.

The correlogram consisting of the autocorrelation function.

Simple exponential smoothing (SES) is a time series forecasting method that assigns exponentially decreasing weights to past observations. To forecast future values using the simple exponential smoothing method, the forecast is based on a weighted average of past data, where more recent observations receive higher weights. The formula for simple exponential smoothing is given as follows [14]

$$\hat{y}_{t+1} = \alpha * y_t + \left(1-\alpha\right) * \hat{y}_t, \quad (2)$$

where $\hat{y}_{t+1}$ – forecasted value of a series in the period $t+1$; y_t – actual value of the series level in the period t; $\hat{y}_t$ – forecasted value of a series in the period t; α – smoothing parameter.

A long short-term memory (LSTM) neural network is a deep learning model specifically designed for analyzing sequential data with long-term dependencies. Each subsequent computational step in an LSTM model is described by the following equations [15]:

$$f_t = \sigma\left(W_f * \left[h_{t-1}, x_t\right] + b_f\right), \quad (3)$$

$$i_t = \sigma\left(W_i * \left[h_{t-1}, x_t\right] + b_i\right), \quad (4)$$

$$\tilde{C}_t = \tanh\left(W_C * \left[h_{t-1}, x_t\right] + b_C\right), \quad (5)$$

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t, \quad (6)$$

$$o_t = \sigma\left(W_o * \left[h_{t-1}, x_t\right] + b_o\right), \quad (7)$$

$$h_t = \sigma_t \odot \tanh\left(C_t\right), \quad (8)$$

where σ – sigmoid activation function; $\odot$ – element-wise multiplication; W, b – trainable parameters; h_t, C_t – hidden state and cell state at step t.

The procedure for processing experimental data consists of a comparative quantitative assessment of the accuracy of forecasts obtained by each of the three developed models by comparing them with the actual data of the test sample. The mean relative approximation error (MAPE) will be used as a metric for assessing the quality of the constructed forecast, calculated using the following formula [16]

$$MAPE = \frac{1}{n}\sum\frac{\left|y_t - \hat{y}_t\right|}{y_t} * 100\%, \quad (9)$$

where y_t – initial value of a time series; $\hat{y}_t$ – modal value of a time series; n – number of observations.

5. Results of development of a forecasting model for optimizing energy consumption at coal enterprises

5. 1. Transformation of time series into stationary form

The logarithm method of the dependent variable was chosen to transform the time series to stationarity. The Dickey-Fuller test was used to test the series for stationarity. Its result is presented in Fig. 2 [17].

The extended Dickey-Fuller test for the l_Qt
test. starting with 30 lags, the AIC criterion
is the sample size of 350
the null hypothesis of the single root: a = 1
  the consanta test
  including 8 lags for (1-L)l_Qt
  model: (1-L)y = b0 +(a-1)*y(-1) + ... + e
  score for (a-1): -0.196167
  test statistic: tau_c(1) = - 2.72789
  asymptotic p-value 0.06927
  1st order autocorrelation coefficient for e: 0.016
  lag for differences: F(8, 340) = 6.783 [0.0000]

Fig. 2. Result of the Dickey-Fuller test

As Fig. 2 shows, the asymptotic p-value for the constant test is 0.06927, which exceeds the test's specified value of 0.05. Thus, it can be concluded that logarithmic transformation did not help to overcome the non-stationarity of the series. The distribution of the logarithm does not differ from the distribution of the original electricity consumption values. The above graph still does not exhibit constant variance, which confirms the lack of stationarity in the data. To eliminate the trend in the time series, let's take the first differences

$$\Delta y = y_1 - y_0, \quad (10)$$

where y_0 – first value of the time series; y_1 – subsequent value of the dependent variable.

Fig. 3 shows a graph of the distribution of first differences for visual analysis of the stationarity of the series.

Analysis of Fig. 3 shows that the original time series became stationary, thus eliminating the trend using first-order differences. The result of the repeated Dickey-Fuller test is presented in Fig. 4.

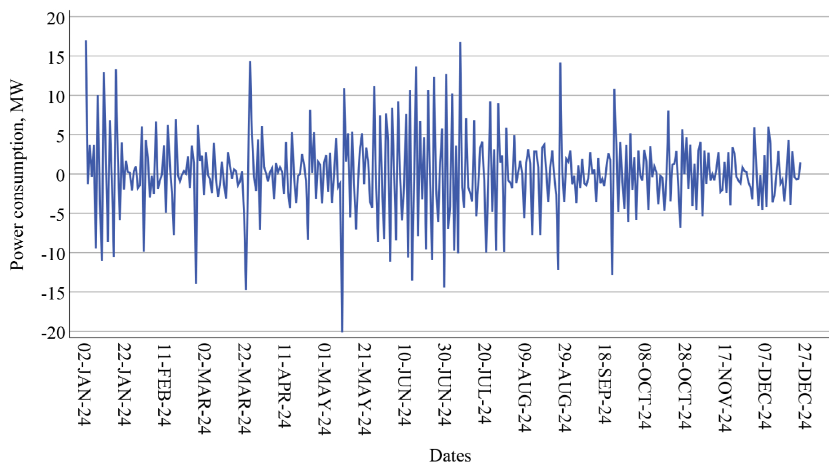

Fig. 3. Graph of first differences in electricity consumption for 2024

The extended Dickey-Fuller test for the d_Qt
test. starting with 30 lags, the AIC criterion
is the sample size of 348
the null hypothesis of the single root: a = 1
  the consanta test
  including 9 lags for (1-L)d_Qt
  model: (1-L)y = b0 +(a-1)*y(-1) + ... + e
  score for (a-1): -4.01349
  test statistic: tau_c(1) = - 8.35826
  asymptotic p-value 4.332e-14
  1st order autocorrelation coefficient for e: -0.011
  lag for differences: F(9, 337) = 12.961 [0.0000]

Fig. 4. Result of the Dickey-Fuller test

The obtained p-value when conducting the test is (4.332e-14) < 0.05, which means that the test confirms the stationarity of the time series.

5. 2. Development of forecasting models for optimizing energy consumption of coal enterprises

5. 2. 1. Autoregressive integrated moving average model

The correlogram consisting of the autocorrelation function (ACF) and partial autocorrelation function (PACF) graphs serves as a key diagnostic tool in time series analysis. It provides both visual and statistical assessments of the internal autocorrelation structure of the series. Interpretation of ACF and PACF is critically important for the proper specification of the ARIMA model parameters (p and q) as it allows one to empirically determine the orders of the autoregressive and moving average components.

The integrating component in time series modeling was defined as $d = 1$, which corresponds to linearizing the data using first-order differences to stabilize the first-order (mean) moments. Achieving stationarity allows for the correct estimation of covariance functions. A graphical representation of the ACF of this stationary time series, which is necessary for subsequent identification of the autoregressive and moving average orders is shown in Fig. 5.

Since its distribution has fluctuations, and the largest of them is at the first lag, this means that the parameter q in the further constructed model will be equal to 1.

Examination of the empirical autocorrelation function (ACF) revealed a statistically significant coefficient exclusively at lag $p = 1$, followed by a sharp drop to insignificance. This structure is diagnostic of a first-order moving average process, AR(1). Therefore, the order of the moving average (q) in the specified model is set to one ($q = 1$).

Next, to determine the autoregressive order (p), the partial autocorrelation function (PACF) was analyzed. The PACF graph exhibits a sinusoidal decay with decreasing amplitude without an abrupt break. According to the Box-Jenkins model identification criteria, a decaying PACF with a simultaneous ACF break indicates the presence of a first-order autoregressive component, AR(1). Therefore, the autoregressive order (p) is assumed to be equal to one ($p = 1$). The PACF graph confirming this identification is shown in Fig. 6.

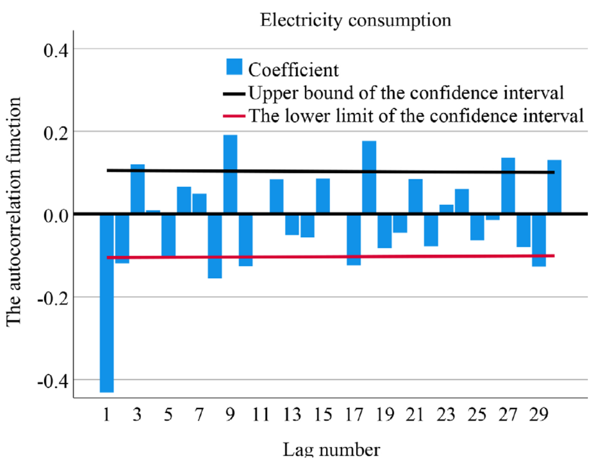

Fig. 5. The distribution graph of the autocorrelation function, for $q = 1$

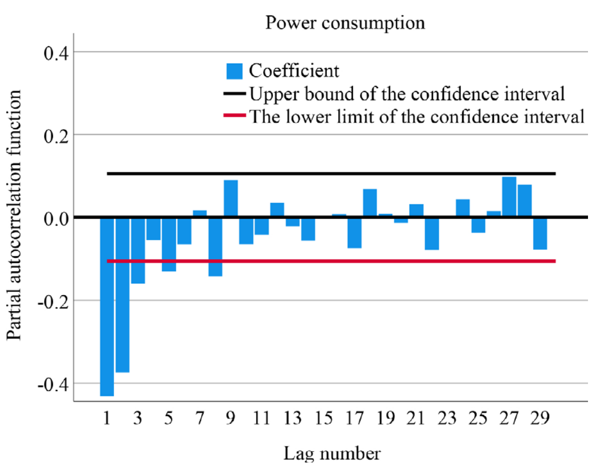

Fig. 6. Partial autocorrelation function distribution graph

To construct the ARIMA model, the following values were selected with parameters $p = 0$, $d = 1$, $q = 1$, as well as seasonal components $P = 1$, $D = 0$, $Q = 1$. Table 1 presents the results and statistics of the resulting model.

Table 1

Parameters of the ARIMA model

Parameters	Estimate	Standard error	Student's t-distribution	Significance
Difference	1	0	–	–
Moving average	0.825	0.031	26.543	0.000
AR (seasonality)	–0.770	0.213	–3.622	0.000
Moving average (seasonality)	–0.831	0.187	–4.454	0.000

As a result of the model construction, the R-squared value is 0.263, which means that 26.3% of the model is explained by the included factors, while the remaining 73.7% are explained by error and factors not included in the model. The mean absolute percentage error (MAPE) of the training dataset is 11.27%, which means that 88.73% of the model is explained by the included factors. A seven-day forecast was generated using the developed ARIMA model. The adequacy of the selected model was assessed by comparing the forecasting data with the actual values from the test dataset, the details of which are provided in Table 2.

Table 2

Comparison of forecasting values for the ARIMA model with a test dataset on actual data

Day number	Actual data	Forecasting values
360	28.08	24.73
361	24.84	24.63
362	28.68	24.69
363	25.32	24.78
364	27.12	24.76
365	25.92	24.93
366	15.24	25.12

Analysis of Table 2 shows that the actual data fluctuate within the range of 15.24–28.68 demonstrating variability. The ARIMA forecast provides more stable values within 24.63–25.12 indicating a possible underestimation of volatility by the model. The key discrepancies lie in the significant underestimation of the forecast and the failure to capture sharp fluctuations in the actual t rend – its peaks and drops. The developed seasonal ARIMA model failed to account for external factors, such as sharp change on the 366^{th} day of 2024 (with values of 15.25 and 25.12), which could have been caused by seasonality. The mean absolute percentage error (MAPE) of the forecast was calculated using formula (9)

$$\text{MAPE}\left(\text{ARIMA}\right) = \frac{0.412}{6} * 100\% = 6.87\%.$$

A MAPE of 6.87% means that, on average, forecasting values deviate from actual data by approximately 7%. In the context of economics, business, or other socioeconomic processes, this MAPE level indicates a very good level of model accuracy.

Fig. 7 shows the graphs of the autocorrelation function (ACF) and partial autocorrelation function (PACF) for the residuals of the ARIMA model, which can be used to check the adequacy of the constructed model.

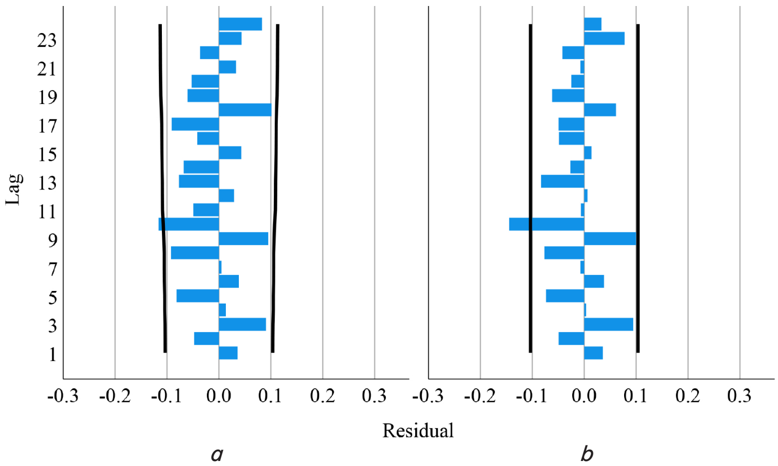

Fig. 7. Graphs residuals for the autoregressive integrated moving average model: *a* – autocorrelation function; *b* – partial autocorrelation function

Validation analysis of the residual series was performed using a correlogram (ACF and PACF). According to the diagnostic test criterion, the white noise hypothesis is accepted if all empirical autocorrelation coefficients (ACF) lie within their statistical confidence intervals. Since no statistically significant spikes are observed on the ACF graph of the residuals, it can be concluded that the residuals represent white noise, which verifies the adequacy and completeness of the constructed model. Additionally, in the original series represented by daily data, a pronounced weekly seasonality was identified, manifested by periodic peaks on the ACF graph at lags that are multiples of seven, with a similar but less pronounced pattern on the PACF graph. This indicates the need to account for the seasonal component in subsequent stages of modeling.

5. 2. 2. Simple exponential smoothing method

Based on the developed simple exponential smoothing model, the parameter α is set to 0.174. To assess the adequacy of the forecast and the relative effectiveness of the SES and ARIMA models, a comparison of their seven-day forecasted values with the actual data was conducted, and the results are presented in Table 3.

Table 3

Comparison of the forecast results of the SES and ARIMA models with actual data

Day number	Actual data	Forecast values, ARIMA model	Forecast values, SES
360	28.08	24.73	24.66
361	24.84	24.63	24.66
362	28.68	24.69	24.66
363	25.32	24.78	24.66
364	27.12	24.76	24.66
365	25.92	24.93	24.66
366	15.24	25.12	24.66

Analysis of Table 3 shows that the forecast based on the Simple Exponential Smoothing (SES) method produced identical results for all seven points, since the smoothing coefficient was set to a value close to zero, i.e., $\alpha = 0.174$. A small α value indicates that the model almost ignores recent observations relying primarily on the previous forecast. The calculation of the mean absolute percentage error (MAPE) for the SES model was performed using formula (9)

$$MAPE(SES) = \frac{0.435}{6} * 100\% = 7.25\%.$$

Since MAPE = 7.25%, this means that, on average, the forecast deviates by 7.25% from the actual electricity consumption data. This is a low error, which falls within the range of "very good" forecast accuracy. The ARIMA model forecast (6.87%) turned out to be slightly more accurate than the SES model forecast (7.25%), although the difference is insignificant. Compared to the ARIMA model's forecast values, it can be noted that the ARIMA model adapts to changes, so its predictions vary (24.63–25.12), while the SES model's forecast values remain static.

It should also be noted that the ARIMA model partially captures the trend (its forecasts increase from 24.63 to 25.12), but still fails on the 366th day, whereas the SES model completely ignores changes in the time series.

As with ARIMA model diagnostics, to validate the alternative SES model, it is possible to analyze the autocorrelation function (ACF) and partial autocorrelation function (PACF) of its residuals (forecast errors). PACF analysis allows to estimate the conditional correlation between current and past error values excluding the linear influence of intermediate time lags. In this case, since the ACF graph of the residuals lacks statistically significant outliers (i.e., all coefficients lie within the confidence interval), it is possible to conclude that the residual series conforms to the properties of white noise. This confirms the adequacy and completeness of the model specification: all autocorrelations present in the original series were effectively extracted and accounted for by the model.

The autocorrelation functions of the residuals of the ARIMA model and the SES model show high similarity and almost identical behavioral parameters.

In addition, when analyzing the ACF residuals, as with the seasonal ARIMA model, periodic peaks with lags of seven days (lags of 7, 14, 21, etc.) are observed. This feature is due to the presence of pronounced weekly seasonality in the original daily time series. Presence of these seasonal peaks in the residuals may indicate that the seasonal component was not fully or adequately accounted for in the structure of the analyzed model.

Visual analysis is used to further evaluate the model's predictive ability. The graph of predicted values obtained using the SES method is presented in Fig. 8 and serves to visually assess the adequacy of the model relative to actual data and confidence intervals.

As shown in Fig. 8, the confidence intervals of the forecast obtained using the simple exponential smoothing method do not capture most of the sharp peaks and drops, which can be explained by the choice of a small value for the smoothing parameter α.

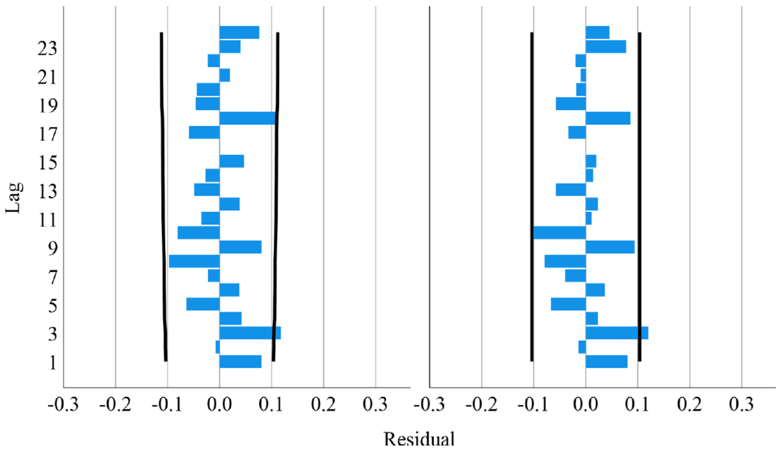

Fig. 8. Graphs residuals for the simple exponential smoothing method: *a* – autocorrelation function; *b* – partial autocorrelation function

5. 2. 3. Neural network model of long short-term memory

When building an LSTM neural model using Python code in Google.Colab, it is possible to split the dataset (previously the training set) into a test and a training dataset. The test dataset will be 80% of the data while the training dataset will be 20%. It is possible to base the forecast on the test dataset (80% of the data from the initial training dataset), and use the same data to compare the forecast results. The forecast result based on the LSTM model is shown in Fig. 9.

```
MAPE  on the training sample: 10.1194%
1/1━━━━━━━━━━━━━━━━━━━━ 0s 45ms/step
1/1━━━━━━━━━━━━━━━━━━━━ 0s 46ms/step
1/1━━━━━━━━━━━━━━━━━━━━ 0s 43ms/step
1/1━━━━━━━━━━━━━━━━━━━━ 0s 46ms/step
1/1━━━━━━━━━━━━━━━━━━━━ 0s 44ms/step
1/1━━━━━━━━━━━━━━━━━━━━ 0s 50ms/step
1/1━━━━━━━━━━━━━━━━━━━━ 0s 49ms/step
7-day forecast: [ 27.372805  26.19292   24.96334   26.387417   27.883404
26.96911   25.87351 ]
```

Fig. 9. Forecast results based on the model of long short-term memory

Analysis of Fig. 9 shows that the MAPE of the LSTM model on the training dataset is 10.1194%, which is at the boundary between "very good" and "satisfactory". This indicates that the model already demonstrates high quality, but still has some potential for optimization.

The average relative error modulus of the forecast values obtained by the LSTM neural network without taking into account the 7th forecast point (366th day) was obtained using formula (9)

$$\text{MAPE}(\text{LSTM}) = \frac{0.322}{6} * 100\% = 5.37\%.$$

A MAPE value of 5.37% means that, on average, the LSTM model's forecast deviates from the actual data by about 5%, which corresponds to "very high forecast accuracy" for highly variable time series. Fig. 9 presents a graph of all forecasted values for visual comparison of the results.

5. 3. Comparative analysis of the forecasting accuracy of models

For making informed decisions in the field of energy planning, it is critically important not only to be able to forecast but also to compare the effectiveness of different forecasting models.

Table 4 presents the actual data and seven-day forecast values obtained using three different approaches: the classical statistical method ARIMA, simple exponential smoothing, and the long short-term memory (LSTM) neural network. A detailed analysis of these data will make it possible to assess their accuracy and identify the most suitable model for solving the given task.

Fig. 10 shows a graph of all forecasting values for visual comparison of results.

Table 5 shows comparisons of the MAPE of all constructed models, including those on the training dataset and on the forecasting values.

Table 4

Comparison of the forecast results of SES, ARIMA and LSTM models with actual data

Day number	Actual data	Forecasting values, seas, ARIMA	Forecasting values, SES	Forecasting values, LSTM
360	28.08	24.73	24.66	27.37
361	24.84	24.63	24.66	26.19
362	28.68	24.69	24.66	24.96
363	25.32	24.78	24.66	26.39
364	27.12	24.76	24.66	27.88
365	25.92	24.93	24.66	26.97
366	15.24	25.12	24.66	25.87

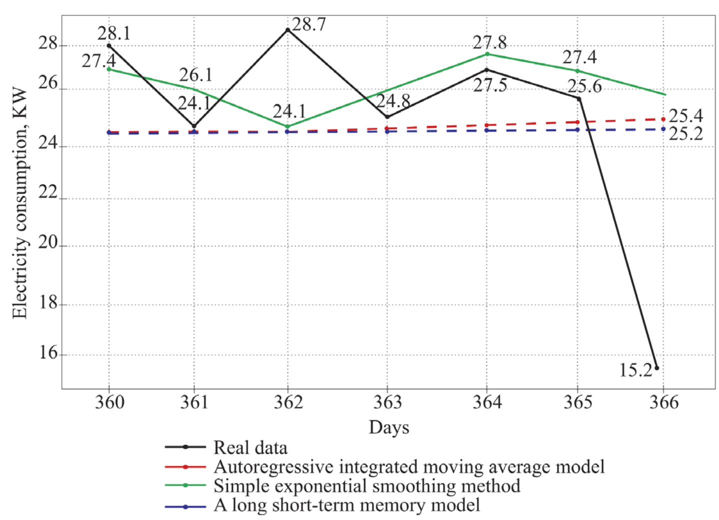

Fig. 10. Comparison of forecasts with actual data from models: an autoregressive integrated moving average model, a simple exponential smoothing method and model of long short-term memory

Table 5

Comparison of MAPE forecasts of SES, ARIMA and LSTM models

Model	MAPE on training dataset (1.01.2024–24.12.2024), %	MAPE on forecasting values of training dataset (25.12.2024–31.12.2024), %
ARIMA	11.269	6.87
SES	11.426	7.25
LSTM	10.1194	5.37

When analyzing errors on the training dataset, the LSTM model (10.12%) shows the best result and the highest ability to learn and adapt to historical data. The MAPEs of the ARIMA and SES models on the training dataset are similar, but worse than those of the LSTM model, which is typical for linear modeling methods when working with rapidly changing data.

6. Comparative analysis and evaluation of the effectiveness of the developed predictive models for predicting energy consumption in a coal mine

As part of the study, three forecasting models were developed and tested: a seasonal ARIMA model, a simple exponential smoothing (SES), and a long short-term memory (LSTM) neural network.

The analysis of the initial time series revealed a pronounced trend and instability of variance. The applied logarithmic transformation did not achieve the required stationarity, which is confirmed by the Dickey-Fuller test result in Fig. 2, where the p-value of 0.06927 exceeds the significance threshold of 0.05, indicating that the series remains non-stationary. Thus, logarithmic transformation did not eliminate the trend component.

After applying the transformation of the first differences (formula (10)), the structure of the series has changed significantly, which is graphically shown in Fig. 3. Visually, there is a disappearance of the trend and stabilization of fluctuations around the zero level. The repeated Dickey-Fuller test shown in Fig. 4 shows a p-value of 4.332e–14 < 0.05, finally confirming the stationarity of the transformed series. Thus, the transformation of the first differences is correct and sufficient to prepare the data for modeling.

The components of the model were determined based on the analysis of ACF and PACF graphs. Fig. 5 shows the values of the ACF of the stationary series: the dominant peak in the first lag indicates the presence of the MA(1) component, which allows to assume $q = 1$. The PACF in Fig. 6 demonstrates a gradual attenuation without a sharp break, a diagnostic sign of AR(1), which justifies the choice of $p = 1$.

The parameters of the final seasonal ARIMA model (0,1,1)(1,0,1) are presented in Table 1. All coefficients are significant ($p < 0.001$), which confirms the correctness of the model. However, the coefficient of determination R2 = 0.263 indicates the limited ability of the model to explain the variations in the series, despite the acceptable value of MAPE for the training sample (11.27%).

According to Table 2, actual values range from 15.24 to 28.68, while the ARIMA forecast ranges from 24.63 to 25.12, indicating that the model underestimates the degree of variability. Significant discrepancies are associated with the omission of peaks and troughs, including a sharp change on day 366 of 2024, probably caused by seasonal fluctuations.

The residue estimates shown in Fig. 7, *a*, *b* demonstrate the absence of significant autocorrelations, which confirms the hypothesis of "white noise" and, consequently, the adequacy of the model. The observed seasonal peaks at lags of multiples of 7 indicate the presence of weekly seasonality, which is only partially accounted for by the model.

Analysis of the data presented in Table 3 shows that the forecast made using the simple exponential smoothing method was the same for all seven observation points. This behavior is due to the small smoothing coefficient ($\alpha = 0.174$), which indicates that the model practically does not take into account the latest observations and is mainly based on previous forecast values.

The analysis of the remnants of the SES model shown in Fig. 8 shows the absence of significant autocorrelations, however, there are still signs of weekly seasonality in the data. The results obtained (MAPE = 7.25%) confirm that SES is less sensitive to the dynamics of the series under consideration and is less suitable for predicting complex non-stationary processes.

The forecast results for the LSTM model, shown in Fig. 9, demonstrate improved forecasting quality due to the model's ability to capture nonlinear dependencies. Its MAPE in the training sample was 10.1194%, which corresponds to the boundary between "very good" and "satisfactory". However, when analyzing the average relative errors of the forecast (excluding the anomalous 366th day), MAPE = 5.37%, which is the best result among all the models under consideration.

In Table 4, the actual and forecast data for 7 days shows strong variability (values ranging from 15.24 to 26.68). The ARIMA and SES models produce almost constant forecasts in the range of 24.6–25.1 failing to capture sharp fluctuations. The LSTM neural network model, on the other hand, provides a more flexible forecast (values ranging from 24.96 to 27.88), which follows the trend of the actual data.

Fig. 10 confirms that the LSTM model is the most adequate and accurate for predicting electricity compared to ARIMA and SES, as it visually matches the real graph better and shows fewer deviations.

Table 5 shows that the MAPE values decreased on the test dataset for all models: ARIMA (–4.399%), SES (–4.176%), LSTM (–4.749%). This change may be due to the training dataset being more "noisy" than the test dataset. The MAPE decrease on the test dataset indicates that there is no overfitting in the forecast models.

Within the framework of this study, the LSTM neural network model was proposed and tested, which, as the discussion showed, made it possible to adequately take into account the complex dynamics of actual energy consumption modes and demonstrated superiority over classical methods (ARIMA, SES). Its best predictive indicator MAPE (5.37%) on the test sample confirms that the model effectively copes with significant variability and stochasticity of the load schedule.

Existing approaches to forecasting electricity consumption do not adequately take into account the complex, stochastic nature and multi-level network architecture of the electric load of a coal mine and are either too complex for operational use or are focused on universal environments that ignore critical peak loads.

The developed LSTM forecast model can be effectively used in automated systems for operational control and management of electricity consumption and for short-term forecasting of electrical load at coal mines and other mining and metallurgical enterprises with a complex and stochastic energy consumption profile.

The conditions for applying the research results are that the developed model demonstrates high efficiency with significant variability and stochasticity of the load graph, which limits the application of classical linear methods. The optimal area of application for the model is short-term planning (up to 7 days). To ensure high reliability of forecasts in areas of abnormal emissions, additional verification and preliminary filtering of the initial data is necessary.

The use of the LSTM model makes it possible to increase economic efficiency by optimizing electricity consumption, improve the manageability of production processes, minimize the risks of technological disruptions, and ensure a technological advantage over traditional forecasting methods.

A limitation of this study is presence of anomalous outliers in the original time series (a sharp drop in consumption to 15–20 units), which were not fully processed or explained. Forecasting such critical drops remains a challenging task for all developed models, reducing the overall reliability of the forecast at points of extreme load decline.

A disadvantage of this study is that the training dataset turned out to be noisier than the test dataset as evidenced by the decrease in MAPE on the test dataset for all models (for example, for LSTM by 4.749%).

Further development of the study involves considering the possibility of ensemble of several models to increase the reliability of forecasts in critical areas, and it is also necessary to conduct additional verification and filtering of historical data to reduce the "noisiness" of the training dataset.

7. Conclusion

1. The initial data for the electricity consumption time series were analyzed and converted to stationarity using the first-difference method. This was confirmed by the Dickey-Fuller test, which showed a p-value of 4.332e-14 providing a valid basis for further modeling.

2. To study the complex, stochastic dynamics of energy consumption at a coal enterprise, three forecasting models were developed and tested, representing both classical linear approaches (seasonal ARIMA and PES) and modern nonlinear deep learning (LSTM).

3. A comparative analysis of the developed forecasting models was conducted using the MAPE metric, which revealed that the LSTM neural network model provides the most accurate forecasts. Its MAPE is 5.37% on the test dataset. This result quantitatively demonstrates the superiority of LSTM in capturing the complex dynamics of the data making it the most effective tool for optimizing energy consumption in coal enterprises.

Conflict of interest

The authors declare that they have no conflict of interest in relation to this study, whether financial, personal, authorship or otherwise, that could affect the study and its results presented in this paper.

Financing

The study was performed without financial support.

Data availability

Data will be made available on reasonable request.

Use of artificial intelligence tools

The authors confirm that they did not use artificial intelligence technologies in creating the submitted work.

Authors' contributions

Shynar Telbayeva: conceptualization; methodology; visualization; **Leonid Avdeyev**: validation; writing – review & editing**; Vladimir Kaverin**: writing – review & editing; supervision**; Dinara Zhumagulova**: writing – original draft; writing – review & editing; visualization.

References

1. Kazakhstan Electric Power Industry Key Factors. Available at: https://www.kegoc.kz/en/electric-power/elektroenergetika-kazakhstana/
2. Hu, H.-J., Sun, X., Zeng, B., Gong, D.-W., Zhang, Y. (2024). Multi-time-scale interval optimal dispatch of coal mine integrated energy system considering source-load uncertainty. Control and Decision (Kongzhi yu Juece), 39 (3), 827–835. https://doi.org/10.13195/j.kzyjc.2022.1507
3. Zeng, Z., Li, M. (2021). Bayesian median autoregression for robust time series forecasting. International Journal of Forecasting, 37 (2), 1000–1010. https://doi.org/10.1016/j.ijforecast.2020.11.002
4. Xiao, H., Wang, B., Zhou, H., Hu, W., Liu, G.-Ping. (2026). Digital twin-empowered power consumption prediction for energy-intensive aluminum annealing furnaces. Expert Systems with Applications, 296, 129079. https://doi.org/10.1016/j.eswa.2025.129079
5. Kaytez, F., Taplamacioglu, M. C., Cam, E., Hardalac, F. (2015). Forecasting electricity consumption: A comparison of regression analysis, neural networks and least squares support vector machines. International Journal of Electrical Power & Energy Systems, 67, 431–438. https://doi.org/10.1016/j.ijepes.2014.12.036
6. De Silva, S. N., Mishra, B. K., Sayers, W., Loukil, Z. (2025). Predicting Long-Term Electricity Consumption Using Time Series Data: Use Case of the UK Electricity Data. Intelligent Systems with Applications in Communications, Computing and IoT, 37–58. https://doi.org/10.1007/978-3-031-92614-3_3
7. El-Azab, H.-A. I., Swief, R. A., El-Amary, N. H., Temraz, H. K. (2025). Seasonal forecasting of the hourly electricity demand applying machine and deep learning algorithms impact analysis of different factors. Scientific Reports, 15 (1). https://doi.org/10.1038/s41598-025-91878-0

8. Telbayeva, S., Nurmaganbetova, G., Avdeyev, L., Kaverin, V., Issenov, S., Janiszewski, D. et al. (2024). Development of mathematical models of power consumption at coal plants. Eastern-European Journal of Enterprise Technologies, 5 (8 (131)), 22–32. https://doi.org/10.15587/1729-4061.2024.313932
9. Zhou, S., Ni, S., Han, Y., Dong, Z., Lai, C. S. (2025). Adaptive electricity consumption forecasting approach for universal environments. Scientific Reports, 15 (1). https://doi.org/10.1038/s41598-025-10147-2
10. Bui, T. H., Lee, K. (2025). Forecasting annual electricity consumption in Vietnam using radial basis function neural network. Energy, 334, 137762. https://doi.org/10.1016/j.energy.2025.137762
11. Tolentino, J. A. (2025). Forecasting Electricity Consumption Using ARIMA Model. Smart Trends in Computing and Communications, 41–51. https://doi.org/10.1007/978-981-96-7517-3_4
12. Al-Dahhan, I. A. H., Ashour, M. A. H. (2025). A Hybrid ARIMA-ANN Model for Enhanced Electricity Consumption Forecasting in Bahrain. Integrating Big Data and IoT for Enhanced Decision-Making Systems in Business, 399–407. https://doi.org/10.1007/978-3-031-97609-4_34
13. ARIMA model. Available at: https://docs.exponenta.ru/econ/arima-model.html
14. Exponential smoothing. Available at: https://help.fsight.ru/ru/mergedProjects/lib/02_time_series_analysis/uimodelling_expsmooth.htm
15. LSTM – long-term short-term memory networks. Available at: https://habr.com/ru/companies/wunderfund/articles/331310/
16. Fundamentals of forecasting theory. Available at: https://openforecast.org/ru/etextbook/about/
17. Methods and formulas for Augmented Dickey-Fuller Test. Available at: https://support.minitab.com/en-us/minitab/help-and-how-to/statistical-modeling/time-series/how-to/augmented-dickey-fuller-test/methods-and-formulas/methods-and-formulas/?utm_source=chatgpt.com