ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2026, Vol. 87, No. 5, pp. 404–417.

SPECIAL ISSUE

Using Machine Learning Methods for Analyzing and Forecasting of Small Samples of Macroeconomic Indicators in the Energy Sector of the Russian Federation

V. A. Ivanyuk

Financial University under the Government of the Russian Federation, Moscow, Russia
e-mail: VAIvanyuk@fa.ru

Received July 8, 2025
Revised October 14, 2025
Accepted November 20, 2025

Abstract—Forecasting methodologies are widely applied in the analysis of socio-economic systems. Employing robust forecasting techniques enables organizations to anticipate future developments, optimize resource allocation, and mitigate potential risks. In the energy sector, accurate forecasting of supply and demand is essential for maintaining grid stability, reducing operational costs, and enhancing reliability. This study aims to assess the effectiveness of statistical and neural network modeling methods in forecasting macroeconomic indicators within the energy sector of the Russian Federation.

Keywords: machine learning methods, ensemble forecast, energy balance, statistics

DOI: 10.7868/S1608303226050025

1. INTRODUCTION

Amid significant transformations in the energy sector and increasing emphasis on sustainable development, examining the challenges of forecasting the energy balance is essential. The energy balance represents the interdependence between production capacity and consumption within the energy system. Effective analysis of energy balance fluctuations requires consideration of a comprehensive range of economic, environmental, and technical factors.

The energy balance encompasses more than energy production and consumption. It also requires consideration of both renewable and non-renewable resource utilization, seasonal fluctuations in demand, and the influence of industrial and other economic sectors. These factors are essential for a comprehensive analysis.

The relevance of the chosen topic is dictated by the need to overcome the general economic problem of electricity shortage.

The article has the following structure. Section 2 describes the collection and processing of data for making a forecast. Section 3 contains an analysis of forecasting methods. Section 4 presents the results of the ensemble forecast. The final results are presented in Section 5.

2. DATA COLLECTION AND PROCESSING FOR FORECASTING

In order to obtain up-to-date data, annual macro-economic reports for the period 2005–2020 from the Unified Interdepartmental Information and Statistical System of the Russian Federation (UIISS) were studied [1, 2]. The data summary sheets had different formats, encoding, and indicator names. Hence, manual transposition was performed, the data were reduced to a common format, and the names of the indicators were unified.

For the 2021–2022 data collection, annual summary statistical reports on the energy balance of the Russian Federation were obtained from the official website of the Ministry of Energy [3]. Since the data in these reports were based on the year-to-year relative change (percentage change) in indicators, a consistent conversion of relative values to natural indicators (in million tons of reference fuel) was carried out. At the next stage, an in-depth analysis of data sources was performed, which proved the absence of freely available energy balance data for the Russian Federation in the period 2023–2024. After receiving confirmation that the highest possible completeness of the macroeconomic data is reached, the 2005–2020 and 2021–2022 data were combined and reduced to a single format. As a result, a table of indicators has been formed, which contains data sorted by date, type of energy, and consumption sector.

An array of data consisting of 29 blocks on 8 time series (sampled values) was obtained, where each series had a length of 18 successive values. In this case, two values in each block are aggregators, so that the total number of input data in the table can be defined as 4032 values. The fragments of the input data are presented in Table 1.

Table 1. Fragment of the input data.

Supply				Ext. demand		Demanc																								
In mln ton ref. fuel.		Production (generation)	Imports	Exports	Total consumption, RF	Conversion into other forms of energy	Processing into fuel	Processing into non-fuel products	As non-fuel resources	Agriculture and forestry	Mining and quarrying	Manufacturing	Manufacture of food products	Manufacture of textiles	Manufacture of footwear and leather	Manufacture of products of wood	Manufacture of paper, printing trade	Manufacture of coke and refined petroleum products	Chemical industry	Manufacture of rubber and plastic	Manufacture of mineral products	Manufacture of metals	Machine tool manufacture	Manufacture of electronics	Manufacture of motor vehicles	Generation and distribution of electric power, gas and water	Construction	Logistics	Population	Other
Oil and natural gas liquids	2021	754,3	0,0	349,5	404,8	0,7	370,3	33,3	0,2	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Gas		856,0	11,4	257,1	610,4	316,6	38,3	62,3	12,4	2,1	29,0	50,7	2,0	0,1	0,0	0,3	0,8	5,8	3,5	0,1	15,1	17,8	2,6	0,3	0,8	1,7	4,6	37,2	6,2	0,1
Coal		250,5	20,1	156,1	114,5	76,0	27,9	0,1	0,3	0,1	0,3	2,8	0,1	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,8	1,6	0,0	0,0	0,0	2,8	0,0	0,2	1,0	0,0
Refined fuels		411,8	7,0	168,8	250,0	87,0	96,6	21,2	2,9	0,5	6,5	11,9	0,5	0,0	0,0	0,1	0,2	1,3	0,8	0,0	3,5	4,3	0,6	0,1	0,2	1,0	1,0	8,3	1,6	0,0
Combustible energy waste		22,8	0,0	0,0	22,8	5,8	0,0	0,1	0,1	0,1	0,0	8,3	0,2	0,0	0,0	0,6	0,3	1,2	0,2	0,0	0,1	5,5	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0
Electrical energy		417,2	1,1	5,1	413,2	0,9	0,0	0,0	0,0	7,2	49,3	114,4	6,0	0,5	0,1	1,7	6,6	10,2	16,9	1,8	6,0	55,8	1,2	1,8	4,3	44,8	4,3	30,9	45,9	2,5
Thermal energy		172,9	0,0	0,0	172,9	0,0	0,0	0,0	0,0	4,4	5,8	67,0	6,3	0,3	0,1	1,8	6,8	12,8	20,1	0,7	3,2	9,0	1,5	1,1	2,9	12,9	0,5	3,0	12,5	0,3
Renewable sources		1,2	0,0	0,0	1,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,2	0,0
		2886,8	39,6	936,6	1989,9	487,1	533,0	116,9	15,9	14,3	91,2	255,1	15,1	1,0	0,2	4,5	14,7	31,3	41,9	2,7	28,7	94,1	5,9	3,3	8,1	63,4	10,5	79,6	68,4	3,0
Oil and natural gas liquids	2022	770,2	0,0	356,9	413,3	0,7	378,0	34,0	0,2	0,0	0,2	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0
Gas		827,4	9,9	222,6	614,7	318,8	38,5	62,7	12,5	2,1	29,2	51,1	2,0	0,1	0,0	0,3	0,8	5,9	3,5	0,1	15,2	17,9	2,7	0,3	0,8	1,7	4,7	37,4	6,2	0,1
Coal		255,6	21,5	161,8	115,3	76,6	28,1	0,1	0,3	0,1	0,3	2,9	0,1	0,0	0,0	0,0	0,0	0,0	0,3	0,0	0,8	1,6	0,0	0,0	0,0	2,9	0,0	0,2	1,0	0,0
Refined fuels		410,1	6,9	164,0	253,0	87,7	98,4	21,4	2,9	0,5	6,6	11,9	0,5	0,0	0,0	0,1	0,2	1,3	0,9	0,0	3,5	4,3	0,6	0,1	0,2	1,0	1,0	8,3	1,6	0,0
Combustible energy waste		21,9	0,0	0,0	21,9	5,6	0,0	0,1	0,1	0,1	0,0	8,0	0,2	0,0	0,0	0,6	0,3	1,2	0,2	0,0	0,1	5,3	0,0	0,0	0,0	0,2	0,0	0,0	0,0	0,0
Electrical energy		423,1	1,1	5,1	419,0	1,0	0,0	0,0	0,0	7,3	50,0	116,0	6,1	0,5	0,1	1,7	6,7	10,4	17,1	1,8	6,1	56,6	1,2	1,8	4,4	45,5	4,4	31,3	46,5	2,6
Thermal energy		166,0	0,0	0,0	166,0	0,0	0,0	0,0	0,0	4,2	5,6	64,3	6,0	0,3	0,1	1,7	6,5	12,2	19,3	0,7	3,1	8,7	1,4	1,0	2,8	12,3	0,5	2,9	12,0	0,3
Renewable sources		1,3	0,0	0,0	1,3	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	0,0	1,3	0,0
		2875,6	39,4	910,4	2004,6	490,3	543,1	118,3	16,0	14,3	91,9	254,1	14,9	1,0	0,2	4,5	14,5	30,9	41,3	2,7	28,8	94,5	5,9	3,2	8,1	63,5	10,6	80,2	68,7	3,0

Based on machine learning methods and statistics, we will create forecasts for the next 7 years (2023–2030).

3. FORECASTING SMALL SAMPLES OF MACROECONOMIC INDICATORS OF THE RUSSIAN ENERGY SECTOR

3.1. Machine-Learning-Based Linear Forecast

Linear regression estimation and forecasting imply a number of conditions, such as the normality of the sample distribution, the robustness of the predicted trend, a small quantity of outliers, and the homoscedasticity of the series. Despite the fact that the linear method is the fastest, it is extremely sensitive (not robust) to outliers and produces the largest confidence interval (forecast error). This method should be used only in cases where the immediacy is much more important than the accuracy of the result [4–6].

Methodology. To use a machine-learning-based linear forecasting method, it is necessary to determine the permissible limits of the values of the predicted variable. It should be noted that a preliminary determination of the permissible limits of the values of functions and their arguments is mandatory in almost all machine learning methods using gradient or stochastic optimization. Usually, the limits are determined by judgmental methods (so, for example, the annual gas production limit depends not so much on planning as on the availability of actual natural, material, and human resources). If it is impossible to obtain the values collected from experts, the limits are defined as the upper and lower likelihood bounds corresponding to the standard data spread being two interquartile ranges of sample variance, or as logically reasonable values corresponding to theoretically achievable maximum and minimum.

It should also be noted that in addition to the boundaries that ensure values are within the likelihood range, the Tikhonov regularization method is also applied to linear forecasts, which consists of eliminating sample outliers based on the assumption of a normal input data distribution, followed by linear imputation of the average values to replace identified anomalies. Another widely used method is linear forecasting with scheduling imputation, where outliers are determined not by higher-order statistical moments, but rather based on the hypothesis that they are related to the known crisis events.

In this case, the values corresponding to crisis moments are imputed using the smart method (less commonly, the linear method).

1. The limits of permissible forecast values are defined.
2. If necessary, scheduling or statistical regularization is carried out.
3. The linear regression equation is calculated.
4. After checking the results, forecast values that exceed the limits are replaced.

For the mathematical implementation of the logical component of the fourth algorithm step, piecewise constant functions are usually used, such as:

$$\text{Kronecker function: } \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases}$$

$$\text{Sign function: } \operatorname{sgn} i = \begin{cases} +1, & i > 0, \\ 0, & i = 0, \\ -1, & i < 0, \end{cases}$$

$$\text{Heaviside function: } \mathrm{H} i = \begin{cases} 1, & i \geqslant 0, \\ 0, & i \leqslant 0. \end{cases}$$

Using these functions as product elements that take values of either 1 or 0, they can be employed to include or exclude blocks of values in additive equations.

In general, the linear forecast equation looks as follows:

$$\begin{aligned} y_{n+f} = {} & H(k_1)H\big(Y_{+0} - (k_1(n+f) + k_2)\big)\big(k_1(n+f) + k_2\big) \\ & + H(-k_1)H\big(-Y_{-0} + (k_1(n+f) + k_2)\big)\big(k_1(n+f) + k_2\big) \\ & + H(k_1)H\big(-Y_{+0} + (k_1(n+f) + k_2)\big)Y_{+0} \\ & + H(-k_1)H\big(Y_{-0} - (k_1(n+f) + k_2)\big)Y_{-0}, \end{aligned}$$

where

$$k_1 \text{ (slope coefficient) } = \frac{n\Sigma_{i=1}^n iY_i - \Sigma_{i=1}^n i\Sigma_{i=1}^n Y_i}{\Sigma_{i=1}^n i^2 - (\Sigma_{i=1}^n i)^2},$$

$$k_2 \text{ (base coefficient) } = \frac{\Sigma_{i=1}^n Y_i - \left(\frac{\Sigma_{i=1}^n iY_i - \Sigma_{i=1}^n i\Sigma_{i=1}^n Y_i}{\Sigma_{i=1}^n i^2 - \left(\Sigma_{i=1}^n i\right)^2}\right)\Sigma_{i=1}^n i}{n},$$

y_{n+f}—forecast value for a given date, Y_i—sample elements, Y_{+0}—upper limit of permissible forecast values, Y_{-0}—lower limit of permissible forecast values, n—sample size, f—prediction interval, $y_i = k_1 i + k_2$—linear regression equation.

Explanation of the logical components: $H(k_1) = 1$, if the function increases; $H(-k_1) = 1$, if the function decreases; $H\big(Y_{+0} - (k_1(n+f) + k_2)\big) = 1$, if the function did not cross the upper limit; $H\big(-Y_{-0} + (k_1(n+f) + k_2)\big) = 1$, if the function did not cross the lower limit; $H\big(-Y_{+0} + (k_1(n+f) + k_2)\big) = 1$, if the function crossed the upper limit; $H\big(Y_{-0} - (k_1(n+f) + k_2)\big) = 1$, if the function crossed the lower limit.

3.2. ETS Exponential Smoothing Forecasting Model

The abbreviation ETS (error/trend/seasonality) means that the following predictive factors are taken into account when making a forecast:

1. The general tendency of growth or decline (trend).
2. Seasonal fluctuations (periodic/harmonic motion).
3. The error significance level increases as we approach the starting point of the forecast.

The ETS forecast is also referred to as the exponential smoothing forecast, which is related to the method of calculating the significance of the error (exponential amplification).

As in linear forecasting, the main approximation accuracy criterion in the ETS method is the squared deviation, but in this case, instead of calculating the sum of squares, we obtain the sum of the products of the squares by their weights, which increase exponentially as they approach the forecast point [7–9].

The result of the forecast is an additive function consisting of linear and periodic components [10–14].

The forecast algorithm:

Step 1. Definition of the permissible forecast value limits.
Step 2. Identification of the linear trend using the classic diagonal method.
Step 3. Approximation of the trend residuals by a periodic function.
Step 4. Expert definition of the "data inertia" value.
Step 5. Sample approximation by a linear function, taking into account the error weights.
Step 6. Summation of linear and periodic functions.
Step 7. Forecast generation based on the overall trend.
Step 8. Verification and replacement of the forecast values that exceed the limits.

Let us formalize the implementation of the ETS forecast as part of the algorithm steps mathematically:

Step 2. Identification of the linear trend using the classic diagonal method:

$$y_{i_1} = \frac{n\left(n\Sigma_{i=1}^n iy_i - \Sigma_{i=1}^n i\Sigma_{i=1}^n y_i\right)}{\Sigma_{i=1}^n i^2 - (\Sigma_{i=1}^n i)^2} + \frac{\Sigma_{i=1}^n y_i - \left(\frac{\Sigma_{i=1}^n iy_i - \Sigma_{i=1}^n i\Sigma_{i=1}^n y_i}{\Sigma_{i=1}^n i^2 - \left(\Sigma_{i=1}^n i\right)^2}\right)\Sigma_{i=1}^n i}{n},$$

where y_{i_1}—initial linear trend for obtaining seasonal residuals, y_i—sample elements, i—element indices, n—sample size,

$$k_1 \text{ (slope coefficient) } = \frac{n\Sigma_{i=1}^n iY_i - \Sigma_{i=1}^n i \Sigma_{i=1}^n Y_i}{\Sigma_{i=1}^n i^2 - (\Sigma_{i=1}^n i)^2},$$

$$k_2 \text{ (base coefficient) } = \frac{\Sigma_{i=1}^n Y_i - \left(\frac{\Sigma_{i=1}^n iY_i - \Sigma_{i=1}^n i \Sigma_{i=1}^n Y_i}{\Sigma_{i=1}^n i^2 - \left(\Sigma_{i=1}^n i\right)^2}\right) \Sigma_{i=1}^n i}{n},$$

where Y_i—sample elements.

Step 3. Approximation of the trend residuals by a periodic function:

$$\sum_{i=1}^{n} \left(k_3 \varphi(k_4 i) - (k_1 i + k_2 - Y_i)\right)^2 \to \min,$$

where k_1—slope coefficient, k_2—base coefficient, k_3—selected amplitude coefficient of the periodic function, k_4—selected frequency coefficient of the periodic function.

Step 5. Sample approximation by a linear function with regard to the error weights:

The value of the data inertia coefficient $\alpha \in]0;1]$ is usually determined according to the expert opinions; in the absence of such, it is recommended to take a value within 0.25–0.30. The significance of up-to-date data grows with the increase of the coefficient, while the significance of "stale" data decreases.

Step 6. Summation of linear and periodic functions:

$$\sum_{i=1}^{n} \left(\frac{e^{\alpha i}(k_5 i + k_6 - Y_i)}{e^{\alpha n}}\right)^2 \to \min,$$

where α—coefficient of "data inertia" (staleness), k_5—selected slope coefficient, k_6—selected base coefficient, Y_i—sample elements.

Step 7. The final prediction equation without limits looks as follows:

$$y_{n+f} = k_5(n+f) + k_6 + k_3\varphi\big(k_4(n+f)\big).$$

Step 8. Next, we check the results and replace forecast values that exceed the limits.

Thus, we obtain the forecast with limits:

$$\begin{aligned} y_{n+f} = {} & H(k_5) H\Big(Y_{+0} - \Big(k_5(n+f) + k_6 + k_3\varphi\big(k_4(n+f)\big)\Big)\Big)\big(k_5(n+f) + k_6 + k_3\varphi\big(k_4(n+f)\big)\big) \\ & + H(-k_5) H\Big(-Y_{-0} + \Big(k_5(n+f) + k_6 + k_3\varphi\big(k_4(n+f)\big)\Big)\Big)\big(k_5(n+f) + k_6 \\ & + k_3\varphi\big(k_4(n+f)\big)\big) + H(k_5) H\Big(-Y_{+0} + \Big(k_5(n+f) + k_6 + k_3\varphi\big(k_4(n+f)\big)\Big)\Big) Y_{+0} \\ & + H(-k_5) H\Big(Y_{-0} - \Big(k_5(n+f) + k_6 + k_3\varphi\big(k_4(n+f)\big)\Big)\Big) Y_{-0}. \end{aligned}$$

3.3. Neural-Network-Based Forecast

A neural network functions through interconnected nodes called neurons. These nodes are organized into a multi-layered structure that resembles the structure of the human brain.

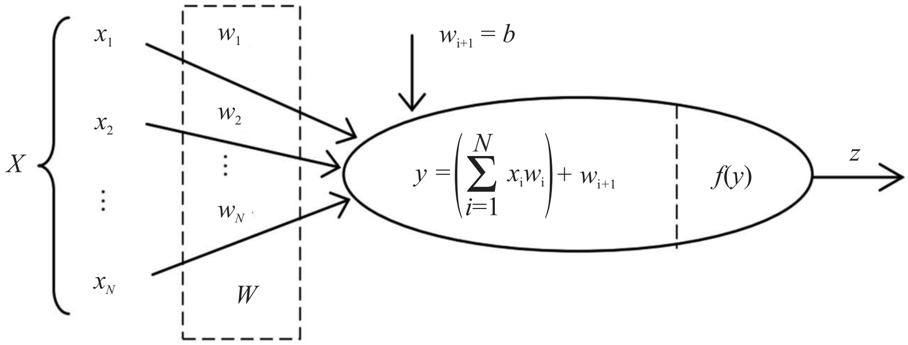

Fig. 1. Schematic drawing of an artificial neural network.

The most elementary neural network, called a perceptron, is a simplified simulation of the activity of a biological network consisting of neurons and connections between them.

The structure of a formal neuron is shown in Fig. 1.

A set of signals is coming to the input. Each signal is subjected to a weighting procedure, i.e., multiplied by a preset coefficient:

$$net_j = \sum_{i=1}^{N} x_i w_{ij},$$

where net_j—total value that would result from processing all incoming signals of neuron j (neuron's summed input), N—total number of elements whose input parameters affect the input signal j; w_{ij}—weight determining the strength of the connection linking neuron i with neuron j.

Summing up all the input signals, previously multiplied by the corresponding weights, we get the total input signal of the neuron.

Each neuron functions according to its own internal algorithm that determines the input signal. This algorithm calculates the result based on the total value received by the neuron. This rule is called the activation function $f(x)$.

Let us turn to the Widrow–Hoff learning rule, also known as the delta rule. Its objective is to minimize the root-mean-square error that has arisen in the neural network. This error is calculated for the input data using the formula:

$$E = \frac{1}{2}(Y - d)^2,$$

where d—desired output value.

Each neuron performs weighted sum calculations using the following formula:

$$S = w_1 X_1 + w_2 X_2 - b,$$

where b—defines the critical level (threshold).

For example, consider the linear activation function given as $Y = x$. In this case, the functional of the error will be determined by the following formula:

$$\frac{\partial E}{\partial w_1} = (X - d)X_1,$$
$$\frac{\partial E}{\partial w_2} = (X - d)X_2,$$
$$\frac{\partial E}{\partial b} = -(X - d).$$

The weights and biases of the neuron are determined by the following formulas:

$$w_i(t+1) = w_i(t) - \alpha(Y - d)X_i,$$
$$b(t+1) = b(t) + \alpha(Y - d).$$

In this particular scenario, the error functional is defined as:

$$E = \frac{1}{2}\Sigma_j (Y_j - d_j)^2.$$

The coefficients determining the weights and biases of neurons are calculated using the following formulas:

$$w_{ij}(t+1) = w_{ij}(t) - \alpha(Y_j - d_j)X_i,$$
$$b_j(t+1) = b_j(t) - \alpha(Y_j - d_j).$$

The error back-propagation algorithm serves as the core tool for training multilayer neural networks based on the principle of the forward signal propagation. Training essentially boils down to two consecutive stages covering each layer of the network: forward and backward passes. Synaptic weights, denoted as $w_{i,j}^k$ (i is the weight number, j is the neuron number, k is the layer number), are adjusted to achieve the best match between the actual signal output and the desired output. The output value of the jth neuron located in the kth layer is determined as follows:

$$Y_j^k = F\left(\Sigma w_{i,j}^k Y_i^{k-1} - b_j^k\right).$$

The input signal for the jth neuron of the last layer is calculated as follows:

$$Y_j = F\left(\Sigma w_{i,j} Y_i^{n-1} - b_j\right).$$

The error functional of the neural network is defined as:

$$E = \frac{1}{2}\Sigma_j (\gamma_j)^2,$$

where $\gamma_j = Y_j - d_j$—error of the jth neuron located in the output layer.

Error of the jth element located on the kth hidden layer:

$$\gamma_j^k = \frac{\partial E}{\partial Y_j^k} = \Sigma_j \frac{\partial E}{\partial Y_j}\frac{\partial Y_j}{\partial S_j}\frac{\partial S_j}{\partial Y_j^k} = \Sigma_j \frac{\partial E}{\partial Y_j}\frac{\partial Y_j}{\partial S_j} w_{i,j} = \Sigma_j (Y_j - d_j) F'(S_j) w_{i,j} = \Sigma_j \gamma_j F'(S_j) w_{i,j}.$$

Error gradients:

$$\frac{\partial E}{\partial w_{i,j}} = \frac{\partial E}{\partial Y_j}\frac{\partial Y_j}{\partial S_j}\frac{\partial S_j}{\partial w_{i,j}} = \gamma_j F'(S_j) Y_j^k,$$
$$\frac{\partial E}{\partial b_j} = \frac{\partial E}{\partial Y_j}\frac{\partial Y_j}{\partial S_j}\frac{\partial S_j}{\partial b_j} = -\gamma_j F'(S_j),$$
$$\frac{\partial E}{\partial w_{i,j}^k} = \Sigma_j \frac{\partial E}{\partial Y_j}\frac{\partial Y_j}{\partial S_j}\frac{\partial S_j}{\partial Y_j^{k-1}}\frac{\partial Y_j^{k-1}}{\partial S_j^{k-1}}\frac{\partial S_j^{k-1}}{\partial w_{i,j}^k} = \gamma_j F'\left(S_j^k\right) Y_j^k.$$

The weights and biases of the neurons are determined by the following formulas:

$$w_{ij}^k(t+1) = w_{ij}^k - \alpha\gamma_j^k F'\left(S_j^k\right) Y_j^k,$$
$$b_j^k(t+1) = b_j^k - \alpha\gamma_j^k F'\left(S_j^k\right),$$

where $\alpha (0 < \alpha < 1)$—learning rate.

The algorithm will be executed until $E > E_m$, E_m, where Em is the desired value of the root mean square error of the neural network.

Let us create a simple, non-recurrent, perceptron neural network for the autoregression of each data series:

$$y_{n+f} = k_{n0}\tanh k_{n1}\tanh\left(\sum_{i=1}^{i=m} k_{ni}\tanh\left(\sum_{j=i}^{i=1} k_{nij}\tanh(Y_{ij})\right)\right),$$

where y_{n+f}—forecast at time $n+f$, Y_{ij}—sample elements, $\tanh(x)$—hyperbolic tangent function, k_{nij}—coefficients of the neuron input weights calculated during training, f—forecast lead time, m—number of input neurons, n—number of inputs of each first-layer neuron.

Let us train the neural network using the standard stochastic Hopfield method or the gradient descent method (EBP), minimizing the sum of squared deviations.

3.4. The Ensemble Forecast

Figure 2 shows graphs for all three of the above types of forecasts (linear forecast, ETS forecast, neural forecast).

As one can see, the forecasts are considerably different:

1. The *linear* forecast shows a significant decrease.
2. The *ETS* forecast shows a decrease, an increase, and then a decrease again.
3. The *neural* forecast shows an increase.

Such significant discrepancies in forecasts usually appear when predicting small and ultra-small time series (less than 30 elements), since their statistical deviations are very large and, for example, in a sample of 20 elements, may be up to 2σ.

In such cases, a predictive ensemble (combined forecast) is created from a set of disparate forecasts, the main task of which is to give a consensual prediction depending on the degree of

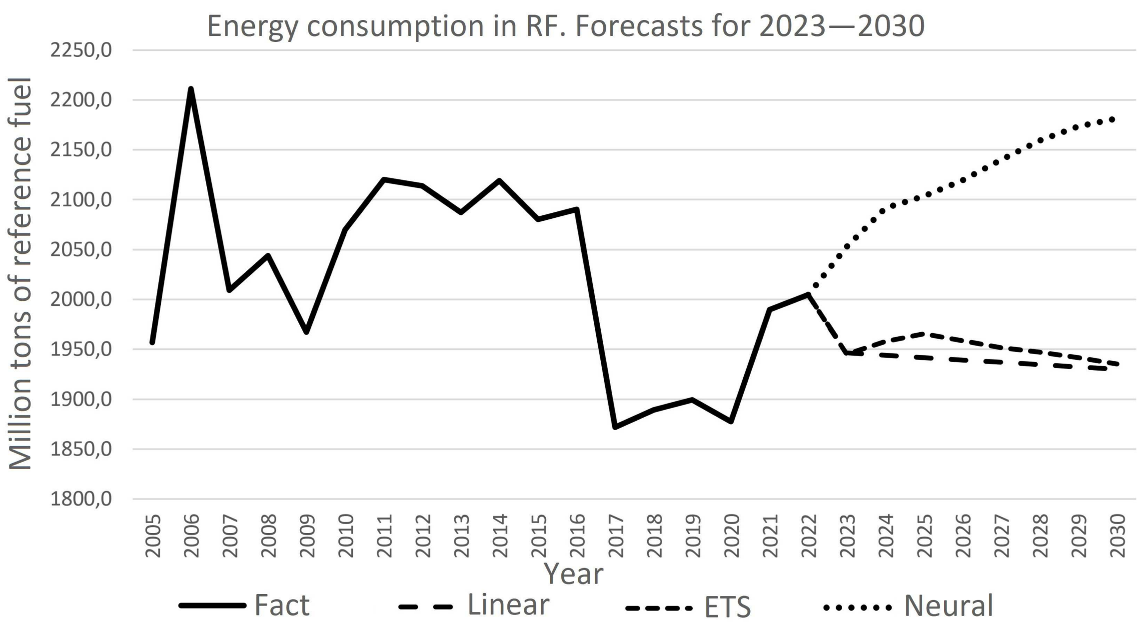

Fig. 2. Linear forecast, ETS forecast, neural forecast.

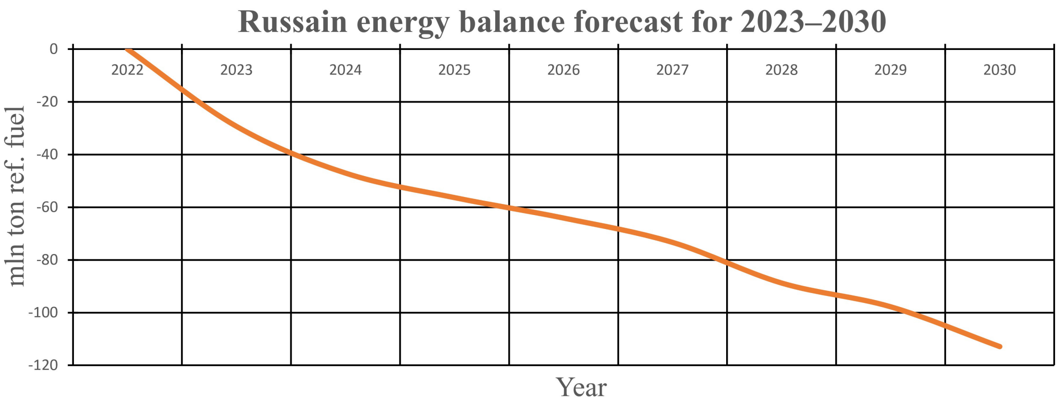

Fig. 3. Russian energy balance forecast.

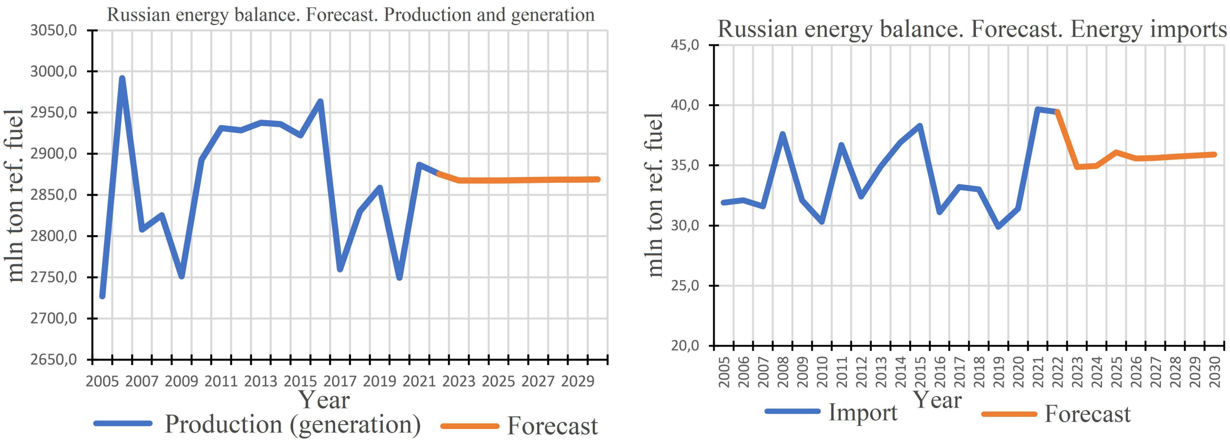

Fig. 4. Forecast. Production and generation. Energy imports.

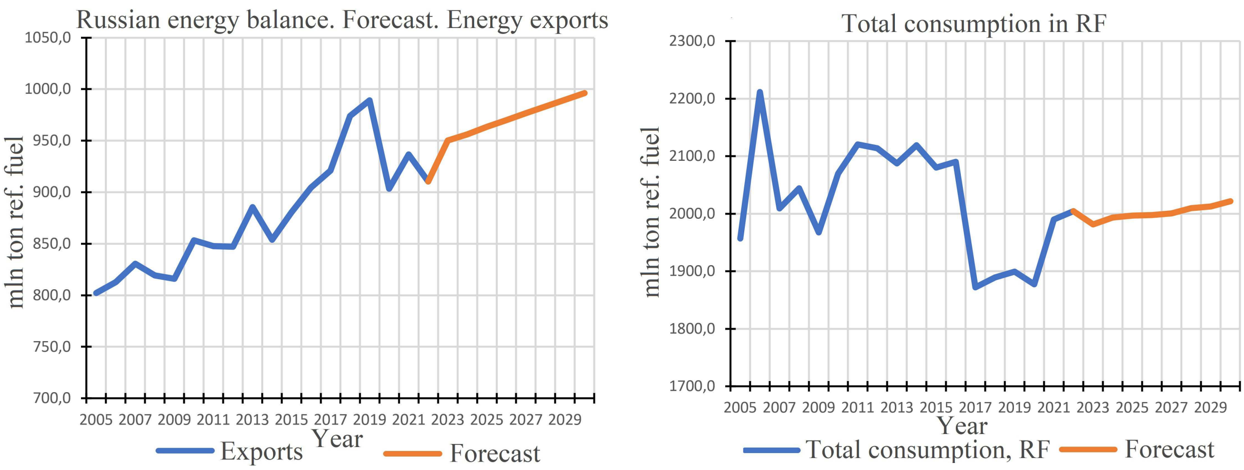

Fig. 5. Forecast. Energy exports. Total consumption in RF.

confidence in a particular methodology:

$$y_{n+f} = \sum_{i=1}^{m} d_i y_{n+f_i}; \quad \sum_{i=1}^{m} d_i = 1,$$

where y_{n+f}—forecast at a time $n + f$, y_{n+f_i}—forecast according ith method, d_i—confidence coefficient of the ith method, m—number of forecasting methods.

The simplest ensembling method is the Bayesian method, which assumes an equal degree of confidence in all forecasting methods [15–17]:

$$y_{n+f} = \sum_{i=1}^{m} \frac{y_{n+f_i}}{m}.$$

Using this method, we will calculate ensemble forecasts and plot such forecasts both for individual indicators and the energy balance as a whole.

Let us consider the general ensemble forecast of the energy balance (Fig. 3).

As we can see, if current trends continue in the period 2023–2030, an energy deficit in the amount of 20 to 120 million tons of reference fuel is expected, which, with the total volume of current production, ranges from 0.7% to 4.2%. Such a deficit can be easily covered by a planned increase in production by 0.6–0.7% per year.

Now let us find out the causes of a possible shortage. Figure 4 shows that total production will remain unchanged, while energy imports will increase slightly.

Figure 5 shows that exports and domestic consumption will increase significantly.

Thus, even an intuitive analysis of the graphs let us conclude that the growth of domestic consumption and exports of energy resources will be the main causes of a possible energy shortage.

4. RESULTS. ASSESSMENT OF FORECAST QUALITY

The estimate of the possible deviation of the predicted value represents the estimated limits of fluctuations in the predicted value, defined as the confidence interval of the forecast. They are usually designated as "pessimistic" and "optimistic" limits of the forecast. The value of the confidence interval is based on the assumption that the predicted value has a Gaussian normal distribution, as well as on the amount of uncertainty allowed in the forecast.

The value of the confidence interval is calculated for each forecast period using the formula:

$$\Delta X_n = nZ_\alpha RMSE,$$

where n—prediction interval, Z_α—multiplier of the significance level, $RMSE$—standard deviation of the trend.

Thus, each subsequent value of the predicted variable expands its probabilistic bounds by at least one standard deviation, and the maximum significance level of the forecast corresponds to a deviation of 68%.

Let us estimate the confidence interval of the components of linear, exponential, neural, and ensemble forecasts of the energy balance, which include exports, imports, production, and consumption of energy resources.

For this estimation, it is first necessary to find the standard sample approximation error for the linear, exponential, and neural trends of these components. In all three cases, it can be estimated as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (X_i - x_i)^2}{n-1}},$$

where X_i—actual values of the predicted variable, x_i—values of the predictive trend, n—sample size of the actual values.

In this case, the limits of the confidence interval for the predicted values are defined as:

$$D_{i>n} = x_i \pm Z_\alpha (i-n) RMSE,$$

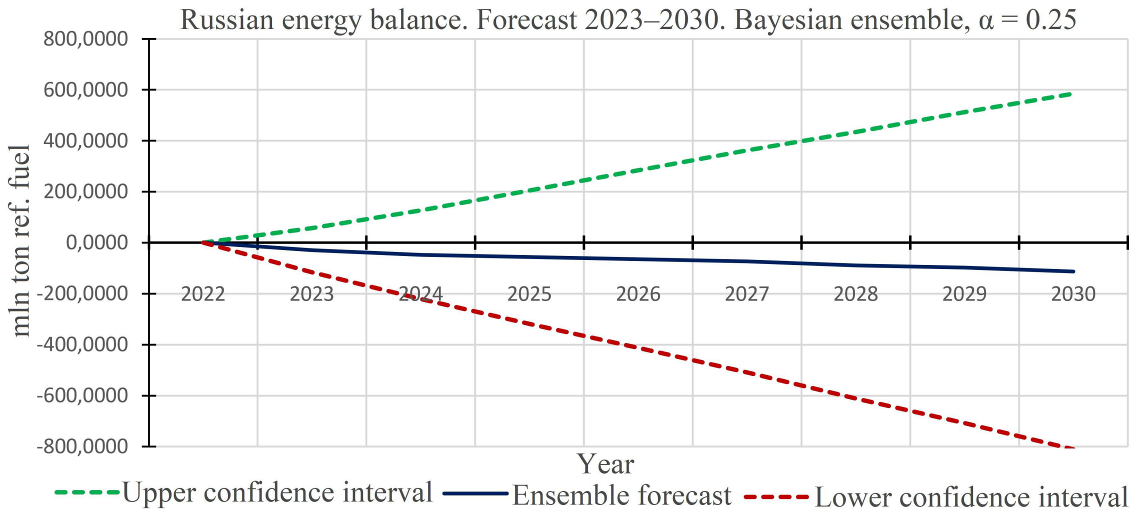

Fig. 6. Ensemble forecast of the Russian energy balance with confidence interval.

where x_i—values of the predictive trend, n—sample size of the actual values, Z_α—multiplier of the significance level, $RMSE$—standard deviation of the trend.

Since the actual energy balance is a value reduced to zero, it is impossible to estimate its deviation from any predictive trends. In this case, calculated estimates of the error in energy balance components can be used, and regardless of whether the component has a negative or positive effect on the energy balance, the fraction of error it introduces should be accounted for as increasing the total error.

For each type of energy balance forecasts, we obtain the values of the cumulative RMSE using Bayesian ensembling and estimate the error of the ensemble forecast as an arithmetic mean (see Table 2).

Table 2. Benchmarking of forecasts

Energy balance forecast	Absolute cumulative RMSE	Relative cumulative RMSE
Linear	74.5223	5.14%
ETS (exponential)	75.4225	5.22%
Neural	77.5736	5,67%
Ensemble	75.8395	5.23%

Let us plot the resulting ensemble forecast with confidence intervals (see Fig. 6).

It should be noted that the confidence interval reflects outer limits beyond which the predicted value will not exceed at the specified significance level [18–20]. However, the display of the confidence interval, especially for long-term prognostic periods, suggests a low forecast accuracy and a significant spread of possible values. Hence, for graphs, it is recommended to use a predictive corridor, which is a visual representation of the standard deviation or the results of frequency analysis. In this case, the display of the forecast limits will be smoother, narrower, and non-expansive. Let us consider the limits of the predictive corridor for displaying the ensemble forecast of the energy balance.

In this case, the limits of the predictive corridor for the predicted values are defined as:

$$F_{i>n} = x_i \pm Z_\alpha RMSE,$$

where x_i—values of the predictive trend, n—sample size of the actual values, Z_α—multiplier of the significance level, $RMSE$—standard deviation of the trend.

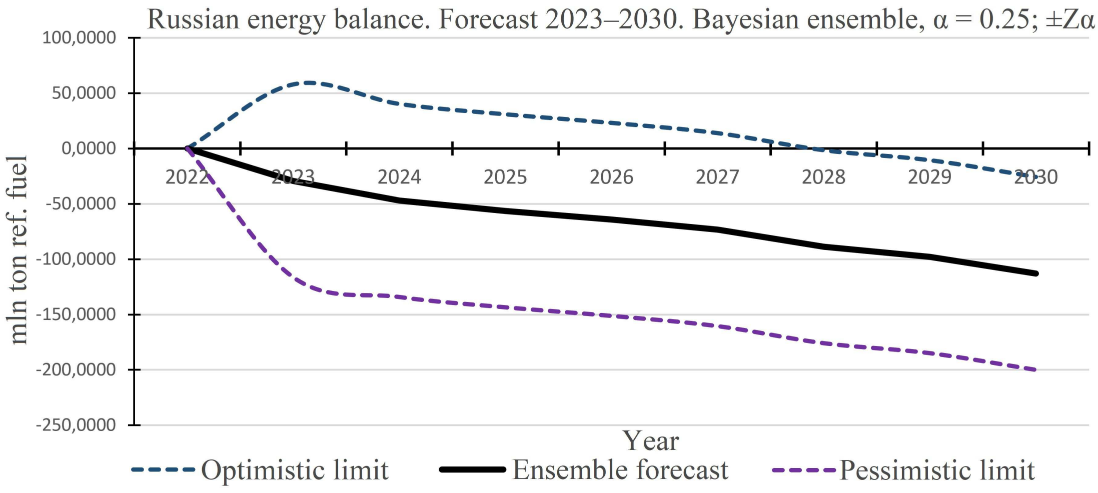

Fig. 7. Ensemble forecast of the Russian energy balance with predictive corridor.

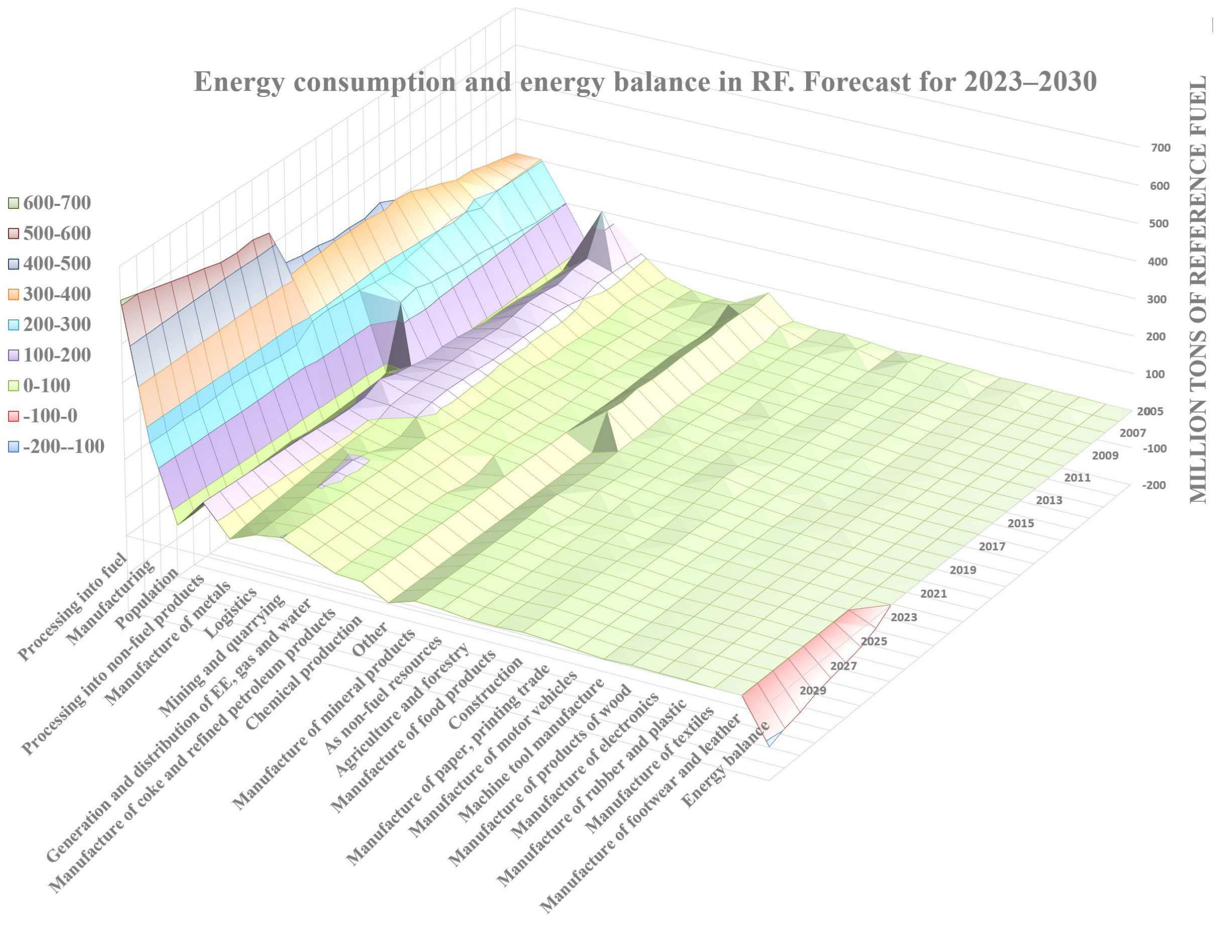

Fig. 8. Three-dimensional graph of sectoral forecasts.

So, for $\alpha = 0.25$, the expected deviation will be 1.15σ, which is reflected in the graph in Fig. 7. As the figure shows, this representation of the forecast is much more favorable in terms of psychological perception.

As a conclusive outcome, a three-dimensional graph of sectoral forecasts is constructed (Fig. 8) in order to identify industries where energy supply requires special attention.

From the graph (the green area of near-zero values on the right), it immediately becomes clear which industries in the Russian Federation are not receiving enough attention. In order to better diversify the economy, these industries need to be stimulated.

5. CONCLUSION

Based on the time series analysis of indicators of the energy balance of the Russian Federation, both classical statistical models and deep learning methods, as well as their ensemble combination, were implemented and investigated.

The results revealed the advantages and limitations of each forecasting model. It is shown that the use of an ensemble model combining the predictions of various methods provides higher stability and accuracy of the results compared with individual models.

The economic value of the developed forecasting system is confirmed by the possibility of its application for strategic and operational planning, risk management, and the formation of competitive advantages in the energy market. The practical implementation of the models and the constructed ensemble based on real data has shown their applicability to the tasks facing energy companies, traders, and government regulators.

REFERENCES

1. UIISS. State Statistics. https://www.fedstat.ru
2. Rosstat. Federal State Statistics Service. https://www.rosstat.gov.ru
3. Ministry of Energy. https://www.minenergo.gov.ru
4. Vasil'ev, M.V. and Dranko, O.I., Two-level revenue forecasting model for a large-scale energy system, *Sensors and systems*, 2023, vol. 267, no. 2, pp. 71–78.
5. Dranko, O.I., On Forecasting Enterprise Conversion Financing, *South Ural State University Bulletin. Series: Computer technology, management, radio electronics*, 2020, vol. 20, no. 4, pp. 74–82.
6. Dranko, O.I. and Blagodarnyj, E.V., Modeling the Destruction of the Value of Russian Energy Companies, *Information and mathematical technologies in science and management*, 2022, vol. 27, no. 3, pp. 104–112.
7. Dranko, O.I., Rezchikova, A.F., Stepanovskaya, I.A., et al., Scenario Modeling of the Country's Development Based on Indicative Planning, *Problems of Management*, 2024, no. 5, pp. 25–41.
8. Dranko, O.I. and Taroyan, K.K., Revenue Forecasting for a Fast-Growing Company Using a Logistic Curve, *Automation and Modeling in Design and Management*, 2024, vol. 24, no. 2, pp. 84–92.
9. Dranko, O.I. and Taroyan, K.K., On a Revenue Forecasting Model for a Fast-Growing Enterprise, *South Ural State University Bulletin. Series: Computer Technology, Management, Radio Electronics*, 2023, vol. 23, no. 4, pp. 66–75.
10. Ivanyuk, V.A., A Long-Term Forecasting Technique Based on a Multi-Trend Forecast, *Soft Measurements and Calculations*, 2023, vol. 73, no. 12–1, pp. 128–138.
11. Brockwell, P.J. and Davis, R.A., *Introduction to Time Series and Forecasting*, New York: Springer, 3rd ed., 2016.
12. Brown, R.G., *Smoothing Forecasting and Prediction of Discrete Time Series*, New York: Prentice Hall, 1963.
13. Chen, W., Xu, H., Liu, Z., et al., Hybrid Modeling of Energy Price Forecasting Combining Temporal Features and Economic Indicators, *Energy*, 2022, vol. 243, pp. 123–135.
14. Hamid, A., Islam, M.S., and Hasan, M.R., A Comprehensive Review of Deep Recurrent Neural Networks for Energy Price Forecasting , *Renewable and Sustainable Energy Reviews*, 2020, vol. 135, art. no. 110354.

15. Ivanyuk, V., The Method of Residual-Based Bootstrap Averaging of the Forecast Ensemble, *Financial Innovation*, 2023, vol. 9, no. 1, pp. 37.

16. Ivanyuk, V., Forecasting of Digital Financial Crimes in Russia Based on Machine Learning Methods, *Journal of Computer Virology and Hacking Techniques*, 2024, vol. 20, no. 3, pp. 349–362.

17. Li, K., Liu, X., Zhu, J., and He, Z., Electricity Price Forecasting Using Gradient Boosting and Random Forests, *Applied Energy*, 2019, vol. 253, pp. 113–122.

18. Smith, J. and Reeve, D., Forecasting Oil and Natural Gas Prices Using ARIMA and Exponential Smoothing Models, *Energy Economics*, 2021, vol. 93, pp. 105–120.

19. Wang, L., Zhou, Y., and Yang, G., Ensemble Forecasting Model Combining ARIMA, XGBoost, and LSTM for Short-Term Natural Gas Price Prediction, *Energy Reports*, 2022, vol. 8, pp. 1430–1440.

20. Zhang, G.P., Time Series Forecasting Using Hybrid ARIMA and Neural Network Models, *Neurocomputing*, 2003, vol. 50, pp. 159–175.

This paper was recommended for publication by A.A. Galyaev, a member of the Editorial Board