MSC 97M40 DOI: 10.14529/jcem170401

ANALYSIS OF FORMALIZED METHODS FOR FORECASTING THE VOLUME OF ELECTRICITY CONSUMPTION

***V. G. Mokhov*[1]**, mokhov50@mail.ru,
***T. S. Demyanenko*[1]**, oop_flp@mail.ru,
***K. V. Demyanenko*[1]**, gelu@mail.ru.
[1] South Ural State University, Chelyabinsk, Russian Federation.

We consider the construction of basic formalized forecasting methods. The methods are widely applied in the world markets of electric power industry. The models are tested on the actual hourly data of the United energy system of the Wholesale Electricity and Capacity Market of Russia. Each model is tested for adequacy by means of Fisher's F-test, t-statistics, mean error of approximation and determination coefficient. We construct a correlogram in order to prove a cyclicity of the time series of electricity consumption, where periods are 1 week and 1 year. An autoregression of the 7th order, that is, with cyclicity in a week, shows the highest efficiency among the examined models. A multifactorial linear regression, taking into account 3 external factors (market rate per day forward; average daily ambient temperature; qualitative factor of working and non-working days), has the least efficiency. The proposed methods are recommended for the operations of subjects of electric power industry to forecast the main parameters of energy market in order to reduce the penalties by improving the forecast accuracy.

Keywords: WECM forecasting, regression, formalized method of forecasting, model, evaluation of significance, volume of electricity consumption.

Introduction

Electricity is one of the most significant products of intermediate consumption of the country and makes a weighty share in the costs of practically all sectors of the economy. The electric power industry is leading infrastructure one, which determines the limits of economic development possibilities practically without alternative. Therefore, the electric power industry needs to ensure rapid growth of generating capacities, as well as profound changes in their qualitative characteristics [1]. Electricity deficit in regions and especially in the country as a whole inevitably leads to the limitation of economic growth. In these conditions, there is a problem of providing as much as possible high accuracy of short-term forecasts of energy consumption, because the work of the domestic energy market is organized such that an error in the forecast leads to a significant increase in costs due to the current system of hourly penalties of the balancing market. The article considers models based on the most widespread formalized methods of forecasting, i.e. methods that use a mathematical description of the identified regularities in the object development to obtain a forecast.

1. Identifying Periodic Oscillations

In order to identify the presence or absence of periodic oscillations we determine the degree of tightness of the autocorrelation relation between the series levels. The degree of tightness can be determined using autocorrelation coefficients, i.e. coefficients of linear correlation between the levels of the original time series and the levels of the series, which is obtained from the original one by shifting the levels a few steps back in time.

$$r_\tau = \frac{\sum_{t=\tau+1}^{n}(y_t - \bar{y}_{1\tau}) \cdot (y_{t-\tau} - \bar{y}_{2\tau})}{\sqrt{\sum_{t=\tau+1}^{n}(y_t - \bar{y}_{1\tau})^2 \cdot \sum_{t=\tau+1}^{n}(y_{t-\tau} - \bar{y}_{2\tau})^2}}, \tag{1}$$

where τ is a shift value, called *a lag*, defines an order of the autocorrelation coefficient,

$$\bar{y}_{1\tau} = \frac{\sum_{t=\tau+1}^{n} y_t}{n-\tau}, \tag{2}$$

$$\bar{y}_{2\tau} = \frac{\sum_{t=\tau+1}^{n} y_{t-\tau}}{n-\tau}. \tag{3}$$

An analysis of the autocorrelation function and correlogram allows to identify a structure of the series, i.e. to identify the presence of one or another component.

A correlogram of electricity consumption for $\tau \in [1, 401]$ is shown in Fig. 1.

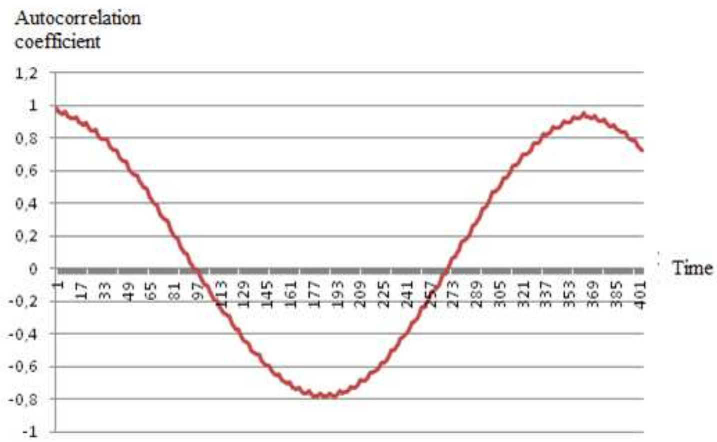

Fig. 1. A correlogram of electricity consumption for lag range from 1 to 401 for URES data from 2009 to 2016

The correlogram clearly shows weak oscillations having a small period. In order to confirm the presence of the oscillations, we construct a correlogram for a month period, see Fig. 2.

According to the autocorrelation study, we conclude the following.

1. The first-order autocorrelation coefficient is the highest, therefore there is a trend in the time series.

2. An autocorrelation coefficient for $\tau = 364$, $\tau = 729$, $\tau = 1091$ turned out to be insignificantly different from the first-order autocorrelation coefficient. Therefore,

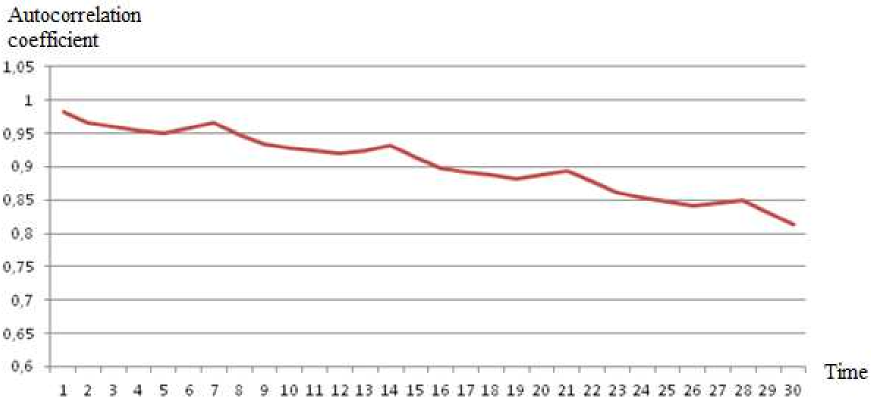

Fig. 2. A correlogram of electricity consumption for lag range from 1 to 30 for URES data from 2009 to 2016

the time series of the electricity consumption volume contains cyclical fluctuations with a period of 1 year.

3. In addition to annual fluctuations, there is a cyclic component with a lag of 7 days.

2. Regression Models

Regression models allow to investigate the relations between two or more variables. In many sources the classical regression model is the following:

$$Y = X\beta + \varepsilon, \tag{4}$$

where Y is $n \times 1$-vector of dependent variables, X is $n \times k$-matrix of independent variables, ε is $n \times 1$-vector of errors, β is $k \times 1$-vector of parameters.

Regression models are applied when there is a clear dependence of the time series on external factors.

Significant factors of forecasting the consumption volume are the market rate per day forward, average daily ambient temperature and working (non-working) days of the week, see [2]. Taking into account these factors in the model, we increase the coefficient of determination from 0.017 to 0.89 in contrast with the paired regression.

The multifactorial regression equation is the following:

$$\hat{y} = 594561.7 + 100.9 \cdot x_1 - 4948.9 \cdot x_2 + 18826.8 z_1, \tag{5}$$

where x_1 is a market rate per day forward; x_2 is an average daily ambient temperature; z_1 is a qualitative factor of working and non-working days.

An increase in the market rate per day forward by unit leads to an increase in the volume of electricity consumption by 100.9 units, when the rest factors are fixed. This fact is interesting for the economic analysis, because there is an obvious direct dependence of demand on price. An increase in temperature by unit leads to a decrease in consumption by 4948.9 MW*h, and in working days the volume of electricity consumption increases on average by 18826.8 MW*h. The Student's t-test analysis shows that the coefficients of the regression equation are significant, i.e. the coefficients are not accidental. The determination coefficient estimates a fraction of the result variation due to the factors presented in the equation. For the obtained model $R^2 = 0.89$, i.e. there is a very close relation between the factors and result. The average partial elasticity coefficients prove an absence of significant elasticity between factors and result feature, but the market rate per day forward has a stronger influence on the volume of electricity consumption than the other two factors. The average partial coefficient of elasticity for this factor is 0.14. Therefore, although the elasticity is very weak, but the dependence is positive, i.e. the consumption volume increases, when the rate is increased.

The mean error of approximation is 2.71%. Therefore, the obtained multiple regression equation is valid for forecasting the volume of electricity consumption.

3. Autoregressive Models

Autoregressive model implies that a future value of the process linearly depends on several previous values of the same process.

The model of forecasting is the following:

$$Z_t = c + \sum_{i=1}^{n} a_i Z_{t-i} + \varepsilon_t, \tag{6}$$

where $a_1 ... a_n$ are model parameters (autoregression coefficients); c is a constant; ε_t is a model error (white noise).

According to the results of correlogram, we construct the autoregression equation based on weekly cyclicity:

$$Z_t = 24065.43 + 0.9653 Z_{t-7}. \tag{7}$$

The coefficients of equation and the equation as a whole are statistically significant. The determination coefficient is higher than the coefficient of multifactor regression $R^2 = 0.93$. Mean error of approximation is 1.96%.

4. Moving Average Method

According to the moving average method, the forecast is based on the the last k values of the time series, which is convenient for large series. The moving average of the order k is an average value of k consecutive observations.

$$\bar{y}_t = \frac{(y_{t-1} + y_{t-2} + ... + y_{t-k})}{k}, \tag{8}$$

where $\bar{y}_t$ is a forecasted value of the time series; y_i are previous values of the series; t is a length of the series; k is a number of observations in the moving average.

The purpose of smoothing time series by the moving average method is to obtain a series with a smaller spread of levels in order to construct the equations of tendencies.

We smooth the time series of consumption volume for the following intervals of smoothing:

1. week, g=7;
2. month, g=30;

3. year, g=365.

We have the time series without any fluctuations, when the smoothing interval is equal to 365, i.e. the period is 1 year. The smoothing results for different intervals are presented in Fig. 3.

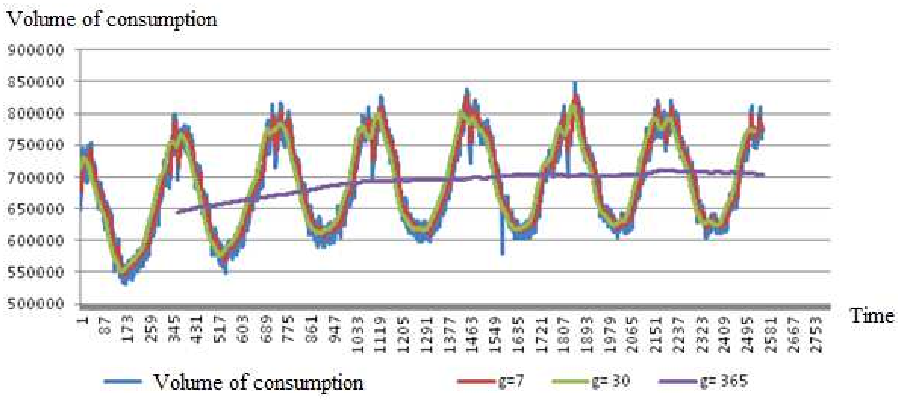

Fig. 3. A smoothing of the electricity consumption volume by the moving average method

We construct different models describing the obtained smoothed time series. Summary results are presented in Table 1.

Table 1

Trend model equations and their characteristics

Model	Equation	Fisher Criterion	Error of approximation, %	Determination Coefficient
Second-order polynomial	$y = -0.0198x^2 + 82.53x + 621322$	84371.26	0.33	0.975
Linear	$y = 24.834x + 664577$	9366.73	0.98	0.810
Exponential	$y = 664603 \cdot e^{0.00004x}$	8943.71	1.89	0.803
Power	$y = 585841 \cdot x^{0.0248}$	25601.88	1.22	0.921
Logarithmic	$y = 16862 \ln(x) + 578918$	24681.17	1.21	0.919

According to the results of Table 1, the second-order polynomial model is the most effective, because the model has the highest determination coefficient and the smallest mean error of approximation, i.e. the model best describes the original data. The obtained results can not be taken as the most accurate among others formalized forecasting models considered in the paper, since the obtained error of approximation shows accuracy of forecasting only for the trend component of the time series.

5. Models Represented by a Periodic Fourier Series

This class of models is applied, if the presence of periodic oscillations (seasonal fluctuations) in the investigated process is proved. In this case, the original process is considered as a harmonic oscillatory process.

For each point the following holds:

$$y_t = f(t) + \sum_{k=1}(a_k \cos(kt\frac{2\pi}{n}) + b_k \sin(kt\frac{2\pi}{n})), \tag{9}$$

where y_t is the actual level of the series at the time (interval of time) t; $f(t)$ is the aligned level of the series at the same time, a_k, b_k are parameters of the oscillatory process (harmonics) with number k, collectively estimating the range (amplitude) of deviation from the general trend and a shift of oscillations with respect to the starting point.

In the general case, we can choose $n/2$ oscillatory processes for a series consisting of n levels. The formulas for calculating the oscillatory process are the following:

$$\varepsilon_t = y_t - f(t), \tag{10}$$

$$a_k = \frac{2}{n}\sum_{t=1}^{n} \varepsilon_t \cdot \cos(kt\frac{2\pi}{n}), (k = 1, 2, ..., \frac{n}{k} - 1), \tag{11}$$

$$b_k = \frac{2}{n}\sum_{t=1}^{n} \varepsilon_t \cdot \sin(kt\frac{2\pi}{n}), (k = 1, 2, ..., \frac{n}{k} - 1), \tag{12}$$

$$a_{n/2} = \frac{1}{n}\sum_{t=1}^{n} \varepsilon_t \cdot \cos(\pi t), \quad b_{n/2} = 0, \tag{13}$$

where k is a parameter of the harmonic.

For the time series of electricity consumption volume, the model of periodic oscillations takes the form:

$$y_t = a_0 + a_1 \cdot \cos(t\frac{2\pi}{T_1}) + b_1 \sin(t\frac{2\pi}{T_1}) + a_2 \cdot \cos(t\frac{2\pi}{T_2}) + b_2 \sin(t\frac{2\pi}{T_2}); \tag{14}$$

$$a_0 = \bar{y}; \quad a_1 = \tfrac{2}{n_1}\sum_{t=1}^{n} y_t \cdot \cos(t\tfrac{2\pi}{n_1}); \quad b_1 = \tfrac{2}{n_1}\sum_{t=1}^{n} y_t \cdot \sin(t\tfrac{2\pi}{n_1}); \tag{15}$$

$$a_2 = \tfrac{2}{n_2}\sum_{t=1}^{n} y_t \cdot \cos(t\tfrac{2\pi}{n_2}); \quad b_1 = \tfrac{2}{n_2}\sum_{t=1}^{n} y_t \cdot \sin(t\tfrac{2\pi}{n_2}), \tag{16}$$

where T is a length of the period. In our case $T_1=365$, $T_2=7$, which is confirmed by the autocorrelation study.

As a rule, in practice, the time series forecasting models, which are represented exclusively by decomposition into harmonics, are not applied widely, in contrast with the additive models consisting of both the model of trend component and the model of seasonal fluctuations.

After the calculations, the equation takes the form:

$$\begin{aligned} y_t = {} & 68766.6 + 92183.7\cos(t\tfrac{2\pi}{365}) + 16740.9\sin(t\tfrac{2\pi}{365}) + 8931.8\cos(t\tfrac{2\pi}{7}) + \\ & +4503.5\sin(t\tfrac{2\pi}{7}). \end{aligned} \tag{17}$$

The determination coefficient is 0.985. The mean approximation error of the obtained model is 3.32%.

According to Fisher's F-criterion, the obtained harmonic model is statistically significant, and Student's t-test confirms the significance of the coefficients obtained.

Therefore, we combine the obtained models of the trend with the seasonal component and have the following equation to forecast the volume of electricity consumption:

$$\begin{aligned} y_t = &-0.0198x^2 + 82.53x + 621322 + 92183.7\cos(t\tfrac{2\pi}{365}) + 16740.9\sin(t\tfrac{2\pi}{365}) + \\ &+8931.8\cos(t\tfrac{2\pi}{7}) + 4503.5\sin(t\tfrac{2\pi}{7}). \end{aligned} \tag{18}$$

The mean approximation error of the obtained model is 2.08%. The determination coefficient is 0.989.

6. Neural Network Models

In order to forecast a time series, the network of direct propagation, where a signal moves only in one direction (from input to output), is usually used, see [3]. In the literature, depending on a way of neurons communication, the network is divided into three main types [4, 5, 6, 7, 8]:

- single-layer networks of direct propagation,
- multilayered networks of direct propagation,
- recurrent networks.

A multilayer perceptron is a subclass of networks of direct propagation and is a set of several elements (Fig. 4):

- input layer (a set of input nodes),
- n hidden layers,
- output layer.

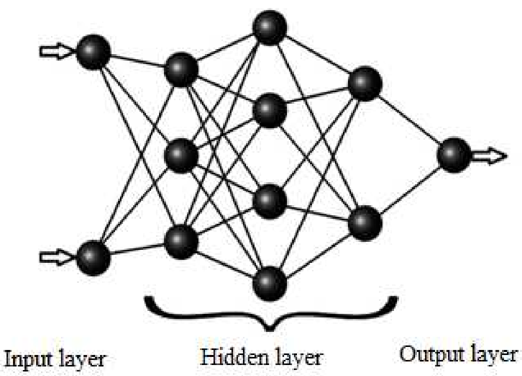

Fig. 4. Three-layer perceptron architectur.

The output of each network neuron is processed by a nonlinear activation function. In order to apply the learning algorithm of backward propagation of the error, the activation function must be continuous, differentiable, and monotonically nondecreasing. Also, for more effective calculations, it is desirable that the activation function is easy differentiable.

There are three main types of the activation functions [4]:

- function of the unit jump;

- piecewise-linear function;

- sigmoid function.

We use hourly electricity consumption data in order to create a training sample. The training sample is a set of N pairs (input, output), where the input is a vector of (48, 1) size, and the output is a vector of (24, 1) size.

The input vector includes electric consumption data for two days. The output vector includes electric consumption data for one day. In the paper we use a three-layer perceptron to forecast an electricity consumption. The perceptron contains 48 neurons at the input layer, 72 neurons at the hidden layer and 24 neurons at the output layer, see Fig. 5.

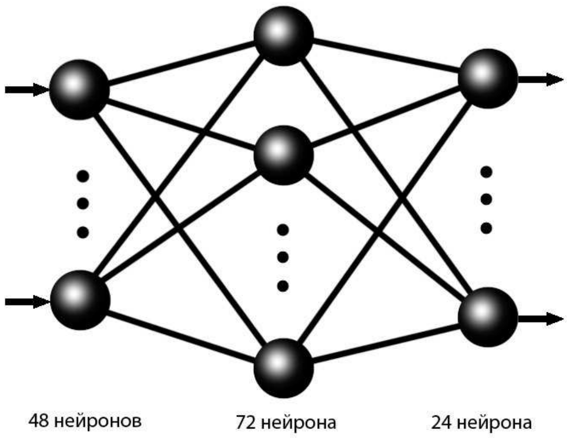

Fig. 5. Multilayer perceptron

In the neural network in the input and hidden layers we use an activation function

$$tansig(x) = \frac{2}{1 + e^{-2x}} - 1. \tag{19}$$

The activation function does not apply to the output layer.

In order to calculate an output error of neural network we use a mean square error

$$E(y, \hat{y}) = \frac{\sum_{i=1}^{k} (y_k - \hat{y_k})^2}{k}. \tag{20}$$

For training we use the hourly data of electricity consumption in Russia from 2009 to 2015. The data is taken from the official web-site of the System Operator of the United Energy system.

In order to evaluate the accuracy of forecast we use data of December 2015.

A parameter of model evaluating is MAPE (mean absolute percentage error)

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}_i}{y_i}\right|, \tag{21}$$

where y_i is a theoretical value, $\hat{y}_i$ is a practical value.

Also, in order to find the optimal parameters of neural network, the tests with different number of neurons in the hidden layer are conducted. Namely, we test neural network having 24, 48, 72, 96 neurons in the hidden layer. The results are shown in Table 2.

Table 2

The results of training a neural network having different number of neurons in the hidden layer

Number of neurons	Error, %
24	11.83
48	4.52
72	2.13
96	2.24

Based on the training accuracy, we choose the network having 72 neurons in hidden layer.

In order to test the forecast we use hourly data of electricity consumption in Russia for 2016. Mean forecast error is 2.29 %. Graphical forecast image is shown in Figure 6.

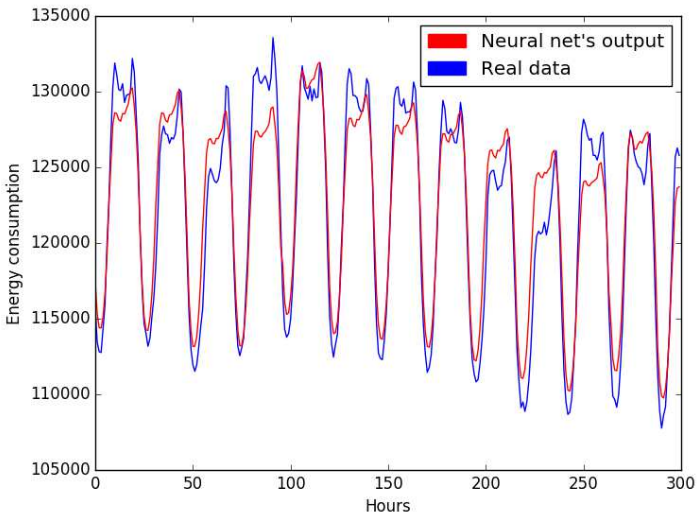

Fig. 6. Forecast of neural network

Conclusion

Moving average model can be used for short-term analysis of trends and to identify the seasonal trends in the electricity market. Therefore, the model is advisable to use in combination with other forecasting models.

The time series of the electricity consumption volume is not linear process, as well as external factors (regressors). Therefore, forecasting consumption volumes with the help of regression models gives a low forecasting accuracy, but an application of the regression models is possible to select external factors for other forecasting models.

For autoregressive models, an acceptable accuracy of forecasting, as a rule, occurs only for one period in advance. For longer periods in advance, the forecast is based on the obtained model values, which leads to a significant decrease of the forecasting accuracy. Therefore, this class of models is advisable to apply for short-term forecasting (1 period) and to identify seasonality (autocorrelation function).

As a separate type of model, the periodic Fourier series does not apply to forecast the WECM, because the series does not take into account the trend (constant component). Nevertheless, in combination with regression and autoregressive models, periodic Fourier series is used in forecasting the WECM of Russia.

Neural network models are widely used to forecast the electricity consumption volume. The models give a sufficiently high forecasting accuracy, but have high resource cost and complexity of the algorithms. Therefore, the search for the forecasting methods, which are more simple to implement and understand, continues.

References

1. Domnikov A. Yu. *Konkurentnoe razvitie sistem kogeneratsii energii v usloviiakh reformirovaniia elektroenergetiki.* [*Competitive Development of Systems to Cogenerate Energy in the Conditions of Reforming the Electric Power Industry.* The Dissertation for Scientific Degree of the Doctor of Economy Sciences]. Ekaterinburg, Ural State Technical University named after the first President of Russia B.N. Yeltsin, 2009.

2. Mokhov V.G., Demyanenko T.S. The Formation of Prices in the Wholesale Electricity Market and Power Russia. *Journal of Economy and Entrepreneurship*, 2015, no. 12-1 (65-1), pp. 1073–1082. (in Russian)

3. Kriesel D. *A Brief Introduction to Neural Networks*. Germany, 2007.

4. Haykin S. *Neural Networks: A full course*. Moscow, Publishing house Villiams, 2006. (in Russian)

5. Pradhan R.P., Kumar R. Forecasting Exchange Rate in India: An Application of Artificial Neural Network Model. *Journal of Mathematics Research*, 2010, vol. 2, no. 4, pp. 111–117. doi: 10.5539/jmr.v2n4p111.

6. Yildiz B., Yalama A., Coskun M. Forecasting the Istanbul Stock Exchange National 100 Index Using an Artificial Neural Network. *An International Journal of Science, Engineering and Technology*, 2008, vol. 46, pp. 36–39.

7. Catalão J., Mariano S. An Artificial Neural Network Approach for Day-Ahead Electricity Prices Forecasting. *6th WSEAS International Conference on Neural Networks*. USA, Stevens Point, 2005, pp. 80–83.

8. Kumar M. *Short-term Load Forecasting Using Artificial Neural Network Techniques.* Thesis for Master of Science degree in Electrical Engineering. India, Rourkela, National Institute of Technology, 2009.

Veniamin G. Mokhov, DSc(Economics), Professor, Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), mokhov50@mail.ru.

Tatyana S. Demyanenko, Department of Mathematical and Computer Modeling, South Ural State University (Chelyabinsk, Russian Federation), oop_flp@mail.ru.

Konstantin V. Demyanenko, Undergraduate, Department of Mathematical Analysis and Methods of Teaching Mathematics, South Ural State University (Chelyabinsk, Russian Federation), gelu@mail.ru.

Received October 31, 2017

УДК 621.31:339.1 + 519.86 + 621.31(470.5) **DOI: 10.14529/jcem170401**

АНАЛИЗ ФОРМАЛИЗОВАННЫХ МЕТОДОВ ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПОТРЕБЛЕНИЯ ЭЛЕКТРОЭНЕРГИИ

В.Г. Мохов, Т.С. Демьяненко, К.В. Демьяненко

В статье рассмотрено построение основных формализованных методов прогнозирования, которые широко применяются на мировых рынках электроэнергетики. Модели протестированы на фактических почасовых данных Объединенной энергосистемы Оптового рынка электроэнергии и мощности России. Каждая модель проверена на адекватность с помощью F-критерия Фишера, t-статистик, средней ошибки аппроксимации и коэффициента детерминации. С помощью построения коррелограммы доказана цикличность временного ряда потребления электроэнергии с периодами в 1 неделю и 1 год. Наибольшую эффективность среди рассмотренных моделей показала авторегрессия 7 порядка, то есть с цикличностью в неделю. Наименьшую эффективность имееи многофактораная линейная регрессия, учитывающая 3 внешних фактора (тариф рынка на сутки вперед; среднесуточная температура окружающей среды; качественный фактор рабочих и нерабочих дней). Разработанный научный инструментарий рекомендуется в операционной деятельности субъектов электроэнергетики при прогнозировании основных параметров энергетического рынка для снижения штрафных санкций за счет повышения точности прогнозов.

Ключевые слова: прогнозирование ОРЭМ, регрессия, формализованный метод прогнозирования, модель, оценка значимости, объем потребления электроэнергии.

Литература

1. Домников, А. Ю. Конкурентное развитие систем когенерации энергии в условиях реформирования электроэнергетики: дис. ... докт. экон. наук / А.Ю. Домников. – Екатеринбург: Уральский государственный технический университет имени первого Президента России Б.Н. Ельцина, 2009.

2. Мохов, В.Г. Формирование цен на оптовом рынке электрической энергии и мощности России / В.Г. Мохов, Т.С. Демьяненко // Экономика и предпринимательство. – 2015. – № 12-1 (65-1), С. 1073–1082.

3. Kriesel, D. A Brief Introduction to Neural Networks / D. Kriesel. – Germany, 2007.

4. Хайкин, С. Нейронные сети: Полный курс / С. Хайкин. – М.: Издательский дом Вильямс, 2006.

5. Pradhan, R.P. Forecasting Exchange Rate in India: An Application of Artificial Neural Network Model / R.P. Pradhan, R. Kumar // Journal of Mathematics Research. – 2010. – V. 2, № 4. – C. 111–117.

6. Yildiz, B. Forecasting the Istanbul Stock Exchange National 100 Index Using an Artificial Neural Network / B. Yildiz, A. Yalama, M. Coskun // An International Journal of Science, Engineering and Technology. – 2008. – V. 46. – P. 36–39.

7. Catalão J. An Artificial Neural Network Approach for Day-Ahead Electricity Prices Forecasting / J. Catalão, S. Mariano // 6th WSEAS International Conference on Neural Networks. – USA: Stevens Point, 2005. – P. 80–83.

8. Kumar, M. Short-term Load Forecasting Using Artificial Neural Network Techniques: Thesis for Master of Science degree in Electrical Engineering / M. Kumar. – India, Rourkela, National Institute of Technology, 2009.

Мохов Вениамин Геннадьевич, доктор экономических наук, профессор, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), mokhov50@mail.ru.

Демьяненко Татьяна Сергеевна, кафедра математического и компьютерного моделирования, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), oop_flp@mail.ru.

Демьяненко Константин Викторович, магистрант, Математический анализ и методика преподавания математики, Южно-Уральский государственный университет (г. Челябинск, Российская Федерация), gelu@mail.ru.

Поступила в редакцию 31 октября 2017 г.