

Improving the accuracy the forecasting volumes of electric grid construction in energy systems based on economic and statistical dynamic models

N S Morozova, E G Andreeva

Omsk State Technical University, Prospect Mira, 11, Omsk, 644050, Russia

E-mail: lenandr02@yandex.ru, morozova_omsk@mail.ru

Abstract. The article deals with the issues related to increasing the accuracy of forecasting the volume of power grid construction (the total length of power lines and transformer substation capacity) for power systems. A set of parameters that affects the volume of electric grid construction and reflects the level of development and structure of power systems has been identified. The parameters of a higher level of the hierarchy of forecasting the development of electric grids are selected, and, therefore, having more stable characteristics of changes in time in comparison with electric grid indicators. To consider the set of parameters in detail, a special method of multivariate statistical analysis – factor analysis, which allows relating the electric grid indicators and a large set of influencing parameters, was investigated. The novelty of the work lies in the proposed method, which allows to develop multifactor dynamic models of forecasting of power grid indicators (total length of power lines and power of transformer substations), in which the coefficients are functions of time and are based on extensive information on the development of electric grids, as well as including an additional independent variable that considers the previous value of the modeled indicator (autoregressive component). These models can be called dynamic factor-autoregressive ones, and they are more effective than the methods of direct optimization of network configuration, geometric modeling and extrapolation due to the lack of reliable information on the specific territorial characteristics of newly introduced electric grids for the long term (more than 5–7 years).

1. Introduction

When solving the complex problem of optimal development of power systems, there is a need to determine the volume of electric grid construction. Forecasts of electric grid indicators are important not only for determining optimal plans for the development of the power system, but also in developing material balances considering the sectoral structure of the country's economy.

The difficulty in determining these indicators is primarily due to the fact that the specific territorial characteristics (the size of the power systems territory, population, etc.) of newly introduced higher voltage networks (110 kV and more) are not known for the long term or are known very roughly. Therefore, methods based on direct optimization of the network configuration [1, 2] are ineffective here.

The use of only extrapolation methods for forecasting power network indicators, based only on the study of the development history, the mean-value method, moving averages, or exponential averages [3] is not recommended due to the incompleteness of their results. However, due to well-established



algorithms, they can be used only as additional techniques while reference information preparing and processing.

Several references can be noted [4, 5, 6, 7], there were attempts made to use neural networks and GZ analysis for forecasting electricity demand. However, the use of this approach to determine the volume of electric grid construction is not universal, because these methods do not reflect a significant number of parameters affecting the electric grid indicators in the models.

2. The problem statement

The article offers a methodology for constructing dynamic factor-autoregressive models to forecast electric grid indicators - the total lengths of power lines (L) and capacities of transformer substations (S_{tr}) for two stages with voltages of 220 and 110 kV. The dynamics of such models is that the coefficients of the models are functions of time; auto-regressiveness is the inclusion of an additional independent variable into the model, considering the previous value of the simulated indicator itself.

Previously, it was assumed as influencing parameters, that the volumes of the power grid construction are most influenced with the parameters characterizing the size of the territory of power systems, the population, the number of large cities, installed capacity of power plants and their structure, power units, levels of energy generation and consumption, maximum electrical loads. Such parameters can be defined at a higher level in the hierarchy of forecasting electric grids. So more stable characteristics of their changes over time (they are less susceptible to influence from random circumstances) can be obtained and can be forecasted with greater accuracy compared to electric grid indicators. Therefore, it is of interest to construct electric grid models as functions of these parameters.

Thus, to obtain and study economic and statistical models, a comprehensive approach was used, including a combination of methods of correlation, factor and regression analyzes.

Since any approach to forecasting is probabilistic in nature, therefore, to manage the risk of its reduction in the proposed models introduced a time parameter and autoregressive component, allowing to take into account changes in the previous value of the simulated indicator. This allows you to reflect the trends occurring in the market.

3. Theory

The sequence of constructing electric grid models for forecasting the total lengths of power lines and the capacities of transformer substations in power systems is considered.

As reference information for constructing and studying economic and statistical models for forecasting volumes of electric grid construction, we used data on four relatively "large" power systems (PS-1, PS-2, PS-3, PS-4), the installed capacity of power plants at the end of the study. The approximation period is in the range from 3200 to 6200 MW, and relatively "small" power systems (PS-5, PS-6, PS-7), the installed capacity of which is only 120-1300 MW. The period under review was 16 years (11 years is a training sample for constructing economic and statistical models, and 5 years is a control sample for assessing the prognostic capabilities of economic and statistical models). In addition, the energy systems of the considered set of indicators differ significantly in the average annual growth rate of the modeled indicators β , determined by the formula:

$$\beta = \frac{1}{m} \frac{(y_T - y_0)}{y_0},$$

where y_0 and y_T – are according to the values of the studied indicators at the beginning ($t = 0$) and the end ($t = T$) of the studied period. The characteristics of the considered power systems are given in table 1.

Table 1. Characteristics of the power systems.

Power systems	Installed capacity of power plants, MW	Average annual growth rates of simulated indicators β :
---------------	--	---

	Beginning of period (t = 0)	End of period (t = T)	total length of power lines		transformer substations power	
			110 kW	220 kW	110 kW	220 kW
PS-1	2900	6200	0.157	0.085	0.295	0.570
PS-2	2100	4500	0.070	0.033	0.404	0.186
PS-3	1020	3300	0.065	0.249	0.522	0.260
PS-4	1300	3200	0.083	0.040	0.158	0.440
PS-5	300	1300	0.404	0.060	1.50	0.083
PS-6	220	750	0.304	0.468	0.313	0.136
PS-7	95	120	0.120	1.100	0.680	0.500

At the first stage, such parameters as the area of the power system territory (TS, km^2), population (H , $th.p.$) in the context of urban (H_u , *thousand people*) and rural (H_r , *thousand people*), the number of cities with a population of more than 15 thousand people. ($n_{15,pc.}$), as well as with a population of more than 100 thousand people ($n_{100,pc.}$), total installed capacity of power plants (N_{inst} , MW) in the context of the structure of generating capacities (hydroelectric power plant (HEP), heating power plant (HPP), state district power plant (SDPP)), most power plant (N_{pp}), highest unit power (N_h , MW), maximum electric load (P_{max} , MW), production levels (W_p , million kWh) and consumption (W_{cons} , million kWh) of electric power were taken into consideration.

The formation of such models was preceded by a correlation analysis [8], a correlation matrix was obtained on its basis. Histograms were constructed using this correlation matrix (figure 1), which indicate that there are strong interconnections of the initial set of parameters that affect the electric grid parameters.

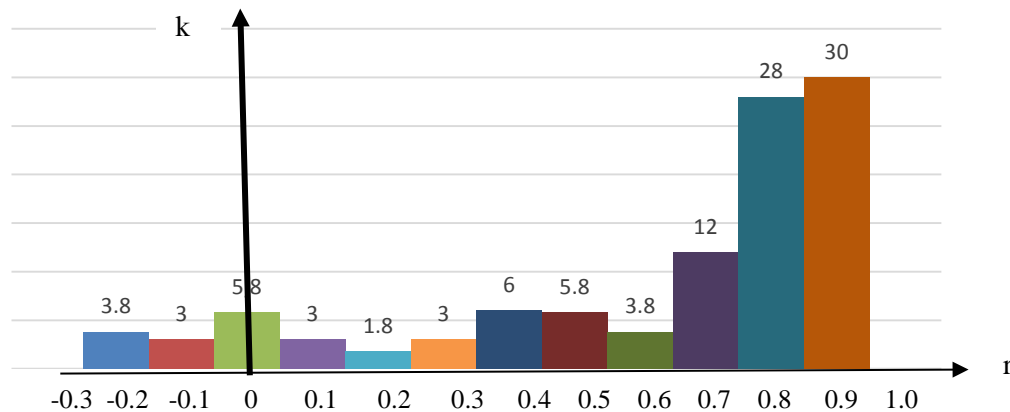


Figure 1. Cross-correlation coefficient histograms (r) of set of parameters affecting electric grid performance of power systems (k – the number of correlation coefficients at intervals).

High specific weights of strong correlation connections cause difficulties in eliminating the phenomenon of multicollinearity of variables [9]. In such cases, the accuracy of mathematical models decreases sharply. For this reason, it is not possible to consider more than three or four of these influencing parameters in the economic and statistical models for forecasting electric grid indicators, which leads to a significant loss of information.

To consider the totality of multicollinear parameters affecting the development of electric grid facilities of power systems in detail, it is recommended to use special methods, multivariate statistical analysis, and, in particular, factor analysis [10].

An analysis of the correlation matrix showed that some of the parameters selected for the study were excluded: H_r , W_{cons} – as excessive (having a strong cross-correlation – $r_{H_r-H} = 0.994$ and $r_{W_{cons}-W_p}$

$= 0.992$) and N_{HEP} , – as uninformative (having a negative and less 0.2 cross-correlation). For further factorization of the indicated set of parameters, only 12 parameters were left: TS, H, H_r , n_{15} , n_{100} , N_{inst} , N_{HPP} , N_{SDPP} , N_{pp} , N_h , P_{max} , W_p .

Factorization of the correlation matrix of the selected parameters was carried out by the principal factors method. Hypothesis testing about a sufficient number of factors was performed according to the Neumann-Pearson criterion, which showed that at a significance level $\alpha = 0.05$ for power systems, two factors F_1 and F_2 proved to be sufficient and they consider 90% of the total variance σ^2 of the factorizable set of parameters (figure 2). Here, the dispersion σ^2 is the arithmetic mean of the squares of the deviations of each parameter value from the total mean.

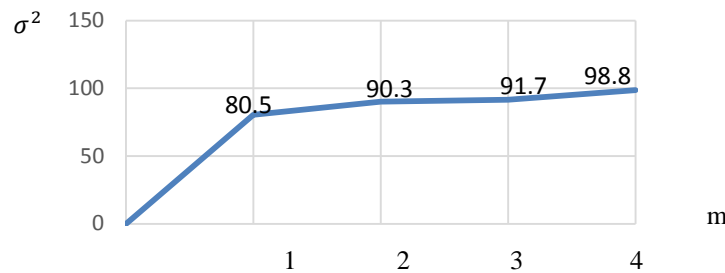


Figure 2. Dependence of contributions to the total variance of variables (σ^2) from the number of allocated factors (m).

The inclusion of factor F_3 and the following factors practically does not increase the percentage of total dispersion. Such high efficiency is explained with strong correlation between the parameters (60-70%). Moreover, the cross-correlation coefficients of the parameters have values from 0.5 to 0.99.

The final stage is the construction and analysis of economic and statistical models, and factors will act as variables there - some conditional parameters that consider the correlation relationships of the initial factorizable parameters. The article considers multifactor dynamic factor-autoregressive models of the form

$$y_t = a_{0t} + \sum_{i=1}^m a_{it} F_{it} + a_{m+1.0} y_{t-1} + b_{m+1.1} y_{t-1} t, \quad (1)$$

where y_t – is simulated indicator (total length of power lines L and power of transformer substations S_{tr}); F_{it} is factors obtained on the basis of factorization of the set of parameters affecting the modeled indicator y_t and forecasted at a higher level of the information hierarchy of power systems; m is number of factors highlighted; a_{0t} , a_{it} – model coefficients ($i = 1, \dots, m$), considering time factor t .

It is accepted as a working hypothesis:

$$\left. \begin{aligned} a_{0t} &= a_{00} + b_{01}t, \\ a_{it} &= a_{i0} + b_{i1}t, (i = 1, \dots, m). \end{aligned} \right\} \quad (2)$$

Then equation (1) can be written as follows

$$y_t = (a_{00} + b_{01}t) + \sum_{i=1}^m a_{i0} F_{it} + \sum_{i=1}^m b_{i1} F_{it} t + a_{m+1.0} y_{t-1} + b_{m+1.1} y_{t-1} t. \quad (3)$$

Such dynamic models considers trends in the time series of model coefficients, which are formed as a result of a change in time as the influence of some factors-arguments on the studied electric grid indicators, and previous values of the modeled indicator.

The quality of approximation of the factor-autoregressive models of power systems was assessed using the following statistical characteristics: multiple correlation coefficient R showing the degree of relationship between the considered parameters, multiple determination coefficient R^2 , which is often used as a measure of the effectiveness of models, the average relative error ε_{av} , characterizing the average deviation of the simulated indicator from its calculated value according to the model, and the

standard deviation σ , expressed as a percentage of the mathematical expectation of the studied indicator.

4. Results of the experiment

As a result of the study, factor-autoregressive models were obtained to forecast the lengths of power lines and the capacities of transformer substations at voltage levels of 110 kV and 220 kV according to equation (3). As noted above, two factors F_1 and F_2 , were identified as a result of factorization which is over 90%, which describe the total variance of the factorizable set of initial parameters. The simulation results are given in table 2.

Table 2. Factor-auto-regression models of electric grid indicators of power systems development.

Indicators	Free component	Coefficients of models with variables						
		t	F_1	F_2	$F_1 t$	$F_2 t$	y_{t-1}	$y_{t-1} t$
Total length of power lines (L)								
voltage:								
220 kV	140.9	-16.7	-71.9	9.8	-27.4	2.5	1.04	0.0003
110 kV	338.7	-9.7	100.1	7.2	9.6	-6.7	0.98	0.0001
Power of transformer substations (S_{tr}) voltage:								
220 kV	173.8	-15.8	115.9	-8.0	-48.9	2.5	0.9	0.02
110 kV	-318.3	46.9	-233.8	27.7	64.3	-7.5	1.2	-0.05

For the obtained models, the approximating ability was verified. The approximating ability is estimated by the above four statistical characteristics R , R^2 , σ and ε_{av} . The results are summarized in table 3.

Table 3. Comparative assessment of the approximating ability of factor-auto-regressive models of forecasting for power grid indicators.

Indicators	R , rel. units.	R^2 , rel. units	σ , %	Average approximation error, %						
				PS-1	PS-2	PS-3	PS-4	PS-5	PS-6	PS-7
L_{220}	0.986	0.972	12.19	5.54	6.93	17.20	5.90	19.30	14.21	9.43
L_{110}	0.995	0.990	6.9	3.21	3.94	2.86	5.8	11.83	9.37	10.50
$S_{tr 220}$	0.989	0.978	18.42	9.68	9.05	2.10	9.83	21.05	21.55	25.10
$S_{tr 110}$	0.996	0.995	7.41	2.33	6.23	8.30	7.31	19.52	14.95	26.09

The factor-auto-regressive models have high correlation coefficients R (table 3). Therefore, the identified factors reveal the pattern of changes in power supply indicators.

Acceptable accuracy results were obtained for relatively "large" power systems (PS-1, PS-2, PS-3, PS-4). The exception is parameter modeling L_{220} for PS-3: its average approximation error is 17.2%. The latter can be explained with significant average annual growth rates of L_{220} for PS-3. Average annual growth rate of L_{220} for PS-is 0.249, while the growth rate of this indicator for the remaining energy systems – PS-1, PS-2 and PS-4 (only relatively large power systems are compared) are in the range from 0.03 to 0.085 (table 1).

There is also a large error (up to 26%) for relatively "small" power systems – PS-5, PS-6 and PS-7, their installed capacity is much less than PS-1, PS-2, PS-3 and PS-4 (table 1). Apparently, the grid indicators of their development are more susceptible to the influence of random factors in such "small" power systems.

5. Discussion of the results

The article offers dynamic factor-auto-regressive modeling of volumes of electric grid development of power systems. According to the statistical characteristics of the approximating ability assessment, it

is proved that such models work quite well for relatively “large” power systems; the installed capacity of their power plants is in the range from 3200 to 6200 MW. Moreover, the error increases for those power systems that have high growth rates of simulated indicators.

Therefore, considering the time factor t in the indicated form of model (3) is the most effective with a more stable development of power systems. Otherwise, it is recommended to use simplified factor-autoregressive models for modeling, i.e. models without a time factor, namely:

$$y_t = a_0 + b_1 t + \sum_{i=1}^m a_i F_{it} + a_{m+1.0} y_{t-1}.$$

It should be noted that forecasting on models with an auto-regressive component assumes the availability of information on previous values of the simulated indicator. This imposes some restrictions on the scope of factor-autoregressive models.

In this regard, further research and construction of models for forecasting power grid indicators, taking into account the growth of the latter, are required.

6. Conclusion

1. One of the main stages in solving the problem of optimal development of power systems is forecasting the volume of electric grid construction.

In order to predict the volumetric indicators of electric grid construction, namely, the total lengths of power lines and the capacities of transformer substations, the possibility of using economic and statistical methods (correlation, regression, and factor analysis) was considered above.

2. A set of parameters has been identified that affect the size of electric grid construction and reflect the level of development and structure of power systems. A correlation analysis of this set of parameters was carried out considering the delay effect of some parameters and their low efficiency.

3. For the simulation of electric grid indicators, economic-statistical models are offered, and the coefficients for the argument parameters are functions of time, and also considering the autoregressive component of the modeled indicator itself, i.e. dynamic factor-autoregressive models are offered that work reasonably well for relatively “large” power systems, as well as systems with stable growth rates of simulated indicators

7. References

- [1] Rozhkov P P and Polyakov T V 2006 Linear model of electrical network configuration optimization *Power industry* **7** pp 96-101
- [2] Voropai N I 2008 System research in energy: from L. A. Melentyev to the present day in ISEM SB RAS *Proceedings of the RAS. Energetics* **6** pp 30-34
- [3] Ismagilov T S 2010 Methods for solving the forecasting problem in energy *Bull. of Ufa State Aviation. Tech. Univ.* vol 14 **4**(39) pp 93-96
- [4] Gordeev A S and Chuvilkin A V 2008 Forecasting of power consumption of objects using artificial neural networks *Issues of mad. science and practice. Univ. named after V.I. Vernadsky* vol 2 **2** (12) pp 31-36
- [5] Gnatyuk V I 2009 Forecasting of power consumption on the basis of GZ-analysis *Energy resource saving and energy efficiency* **1** (25) pp 22-25
- [6] Shashikala L Karunathilake and Harsh and Nagahamulla R. K. Artificial neural networks for daily electricity demand prediction of Sri Lanka 2017 *17th Int. Conf. on Advances in ICT for Emerging Reg.(ICTer)* pp 1-6
- [7] Saravanan S and Kannan S and Thangaraj C 2017 Forecasting India's electricity demand using Artificial Neural Network *IEEE-Int.Conf. On Advances In Eng., Science And Management (ICAESM -2012)* pp 79-83
- [8] Sis'kov V I 1975 *Correlation Analysis in Economic Research* (Moscow: Statistika) p 168

- [9] Lavrentik O A and Lavrentik A I and Simchenko O I et al. 2008 Construction of mathematical models of econometric systems based on the phenomenon of linearity *Bulletin of the Azov state technical University. Series: Technical Sciences* **18** pp 292-296
- [10] Ovsyannikov G N 2013 *Factor analysis available. The study of multi-parameter systems and processes* (Moscow: Librokom) p 174