Paper accepted for presentation at 2009 IEEE Bucharest Power Tech Conference, June 28th - July 2nd, Bucharest, Romania

Use of the "ANAPRO" software to analyze and forecast operating parameters and technological characteristics on the basis of macro applications

V. G. Kurbatsky and N.V. Tomin

***Abstract*— A progressive trend in creating advanced specialized software packages for processing considerable information arrays, identifying and classifying states and parameters is related to the integration of different artificial intelligence technologies which aims to combine their best features. The paper presents the intelligent ANAPRO software. It implements approaches to the analysis and forecasting of state parameters and process characteristics in the electric networks on the basis of modern methods of nonlinear data analysis and artificial intelligence technologies. The findings of experimental studies are given for different sections of electric network in East Siberia on the basis of the proposed ANAPRO software.**

***Index Terms* — software package, artificial intelligence technologies, macro, process characteristics, analysis, forecasting.**

I. Introduction

Specific features of electric networks in Russia are related to huge volumes of data and uncertainty of initial information. They are particularly conspicuous at the low level of power system hierarchy and complicate essentially the qualitative analysis and forecasting of operating parameters (load flow, power flows, voltage levels, etc.) and process characteristics (power and electricity losses, fuel prices, etc.) (OPPC). However, the existing specialized software packages designed to solve the problems of electric power industry employ a rather limited number of mathematical approaches and methods that do not allow the specific features of power systems to be taken into account properly.

II. Structure of the ANAPRO software on the basis of macros

A certain way out of the situation is to use a specialized software on the basis of application macros. According to this approach the procedures and blocks of integrated software system (ISS) that have already been created and debugged are used as basic computation modules. The presented concept of developing intelligent software on the basis of combined involvement of modern statistic methods and AIT underlay the ANAPRO software [5] (Fig.1). The software includes three subsystems implemented in STATISTICA 6.0 [4] (produced by "StatSoft") in Windows environment:

- Subsystem of OPPC analysis;
- Subsystem of OPPC forecasting;
- Subsystem of joint OPPC processing;

The subsystems of OPPC analysis and forecasting can operate independently, i.e. as separate programs or jointly, when the subsystem of joint OPPC processing starts to operate.

A. Subsystem of OPPC analysis

A specific feature of the subsystem is joint application of cluster analysis on the basis of self-organizing Kohonen maps (SOM) (block 3) and factor analysis – the principal component analysis (PCA) (block 4). In order to decrease the dimensionality of initial arrays and determine the most informative OPPC the input data are analyzed in block 2. The analysis is based on the algorithm of nonlinear principal component analysis (NPCA).

Neural network procedure of the OPPC analysis and visualization on the basis of SOM (block 3) makes it possible to effectively determine the features and differences in the electric network operation conditions and timely detect abnormal conditions [3]. In order to improve the quality of cluster analysis of OPPC the ANAPRO software employs a two-stage adaptive classifier [1] that includes subsequent learning within SOM and *Learning Vector Quantization* (LVQ) algorithms. PCA (block 4) should be applied to reduce after clustering the number of factors in the arrays of studied OPPC with the view to efficiently represent the data with minimal loss of information.

The results of SOM-based cluster analysis are visualized in block 5 with the help of Hinton diagrams. Hinton diagrams are formed for each time interval, for example, a month, and reflect its typical features within the clusters obtained for different time intervals.

The decisions made in block 6 are based on the Hinton diagram analysis for different time periods (for example at the beginning and at the end of a year) and on the basis of results obtained in block 4 with the help of PCA.

V. G. Kurbatsky is with the Department of Electric Power Systems, Energy Systems Institute, Irkutsk, Russia (e-mail: kurbatsky@isem.sei.irk.ru).

N. V. Tomin is with the Department of Electric Power Systems, Energy Systems Institute, Irkutsk, Russia (e-mail: tomin@isem.sei.irk.ru).

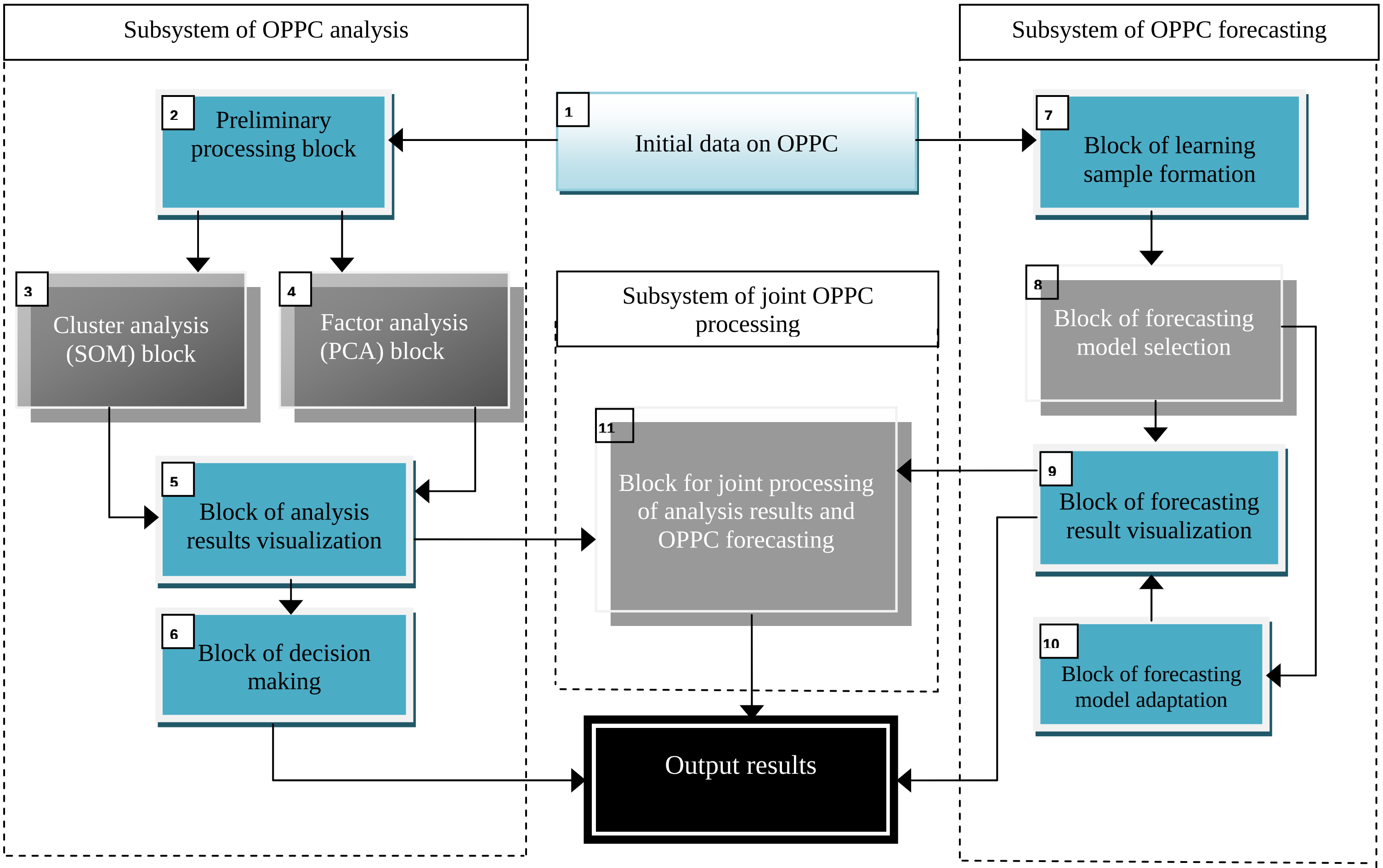


Fig.1. A general structure of the ANAPRO software

B. *Subsystem of OPPC forecasting*

The subsystem that contains blocks 7-10 implements an intelligent approach to forecasting OPPC on the basis of neural network technologies. This approach makes it possible to effectively solve forecasting problems when requirements for accuracy of calculations are rather strict and the dynamics of parameters is non-stationary which is characteristic of the Russian electric networks.

The set of learning data (block 7) may vary depending on what OPPC are forecasted. The initial data can be represented by the parameter of the forecasted OPPC (for example, only electric load data) and the additional parameters that affect the parameter to be predicted. The studies show that the use of additional parameters can lead to increase in the error of the calculation. Further in the analysis of learning sample in block 8 on the basis of the neural genetic input selection (NGIS) algorithm certain input data can be rejected as less informative.

The use of nonlinear optimization algorithms in block 8, and first of all the algorithms of simulated annealing (SA) [6] and NGIS provides the procedure for selecting the best forecasting models for each specific sample. The SA algorithm allows one to analyze the properties of the initial sample and organize a "Competition-based" system among different neural network forecasting models when the best forecasting model is selected in the process of nonlinear optimization. In the neural network forecasting these models are the structures of artificial neural networks (ANN).

In the event that the computational forecasting problem is rather complex (for example, when learning sample contains many additional factors) it is sensible to use the committee machine (CM) principle (Fig.2). The calculations show that under different combinations of learning and test samples the SA algorithm can be started 3-6 times to form neural networks-experts.

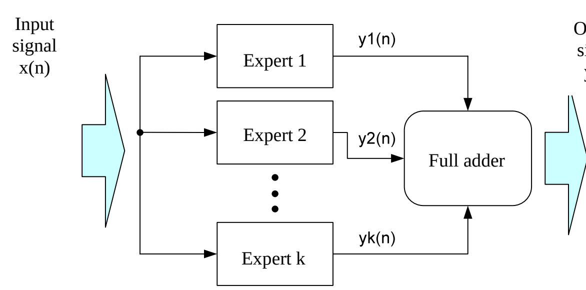


Fig.2. Committee machine for forecasting problem

C. *Subsystem of joint OPPC processing*

Based on the operation results of the OPPC analysis and forecasting subsystems it is possible to arrange joint processing of OPPC in the subsystem of joint OPPC processing (Fig.1). The idea of joint processing is to specify the results of OPPC forecasting in block 11 and correct the decisions made on perspective planning in the considered electric networks

D. Macros of user applications in the ANAPRO software

As was said the ANAPRO software was developed on the basis of macros of user applications [2, 9]. The core of the software is a system implemented on the basis of the software product STATISTICA (Fig.3).

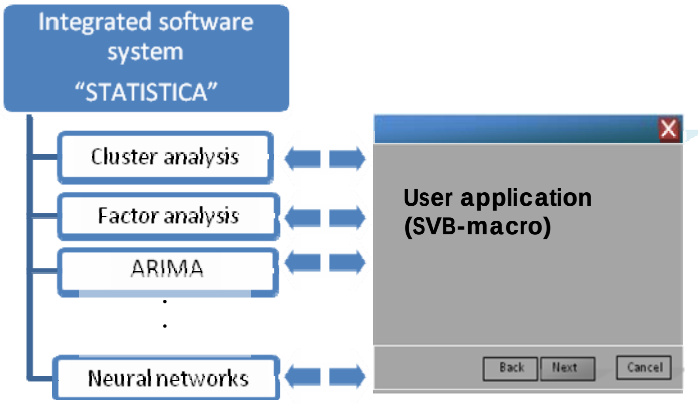


Fig.3. Operation of the ANAPRO software in terms of macro conception.

The advantages of this approach are related to the possibilities to choose the efficient models for solving a specific electric power problem. The object-oriented programming language Statistica Visual Basic (SVB) is applied as the main language of the macro-program.

The developed "intelligent" approaches to the analysis and forecasting of OPPC are experimentally checked in the ISS "STATISTICA" and automatically recorded in the software background mode in the SVB language in the form of a program code. Thus obtained user macros representing the basis of the ANAPRO software subsystems can be edited. It is also possible to change the settings of analysis procedures, used variables and their types, files with data, and elements of user interface, etc.

Further these "construction blocks" were used by the authors to create their own specialized applications (Figs.4 and 5) that underlay the ANAPRO software.

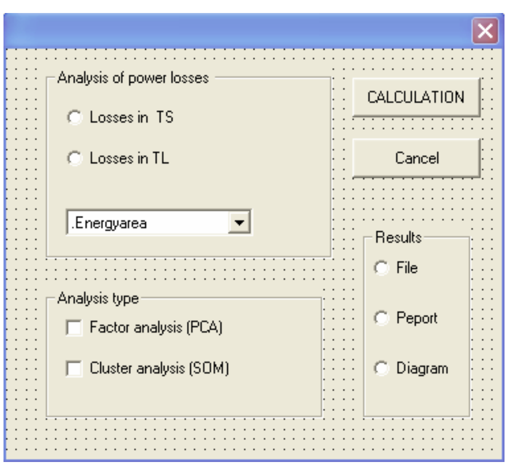


Fig. 4. View of application macro for the analysis of power losses in the Subsystem of OPPC analysis

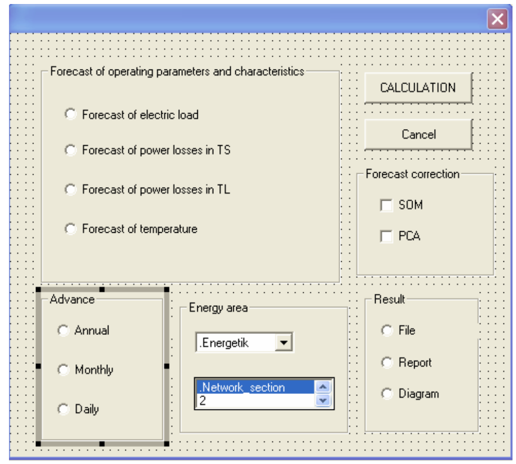


Fig.5. View of application macro for forecasting state variables in the Subsystem of OPPC forecasting.

Depending on the problem specifications the macros obtained for analysis and forecasting of OPPC (Figs. 4 and 5) can work autonomously within the respective subsystems of the ANAPRO software.

If it is necessary to specify the results of calculation in order to increase the accuracy of forecasting the studied OPPC the macro of joint processing of OPPC starts to operate (Fig.6.)

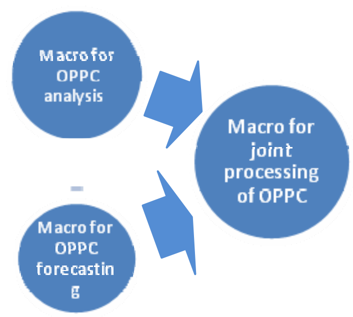


Fig.6. Macro for joint processing of OPPC

III. Experimental Calculations

The proposed "intelligent" approaches and methods that were implemented in the ANAPRO software were tested for efficiency on real electric networks in East Siberia within the problems [5,7-9] reflected in Figure 7.

Let us illustrate the ANARPO software operation by the example of calculations within the analysis and forecasting of power losses in transmission lines for the Bratsk network (Fig.8a) For initial information in block 1 of the ANAPRO software we used an annual data array on calculated power losses in transmission lines in several Bratsk network sections. The results of SOM-based cluster analysis (block 3) and PCA (block 4) for these sections are presented in Figs 8b and 8c.

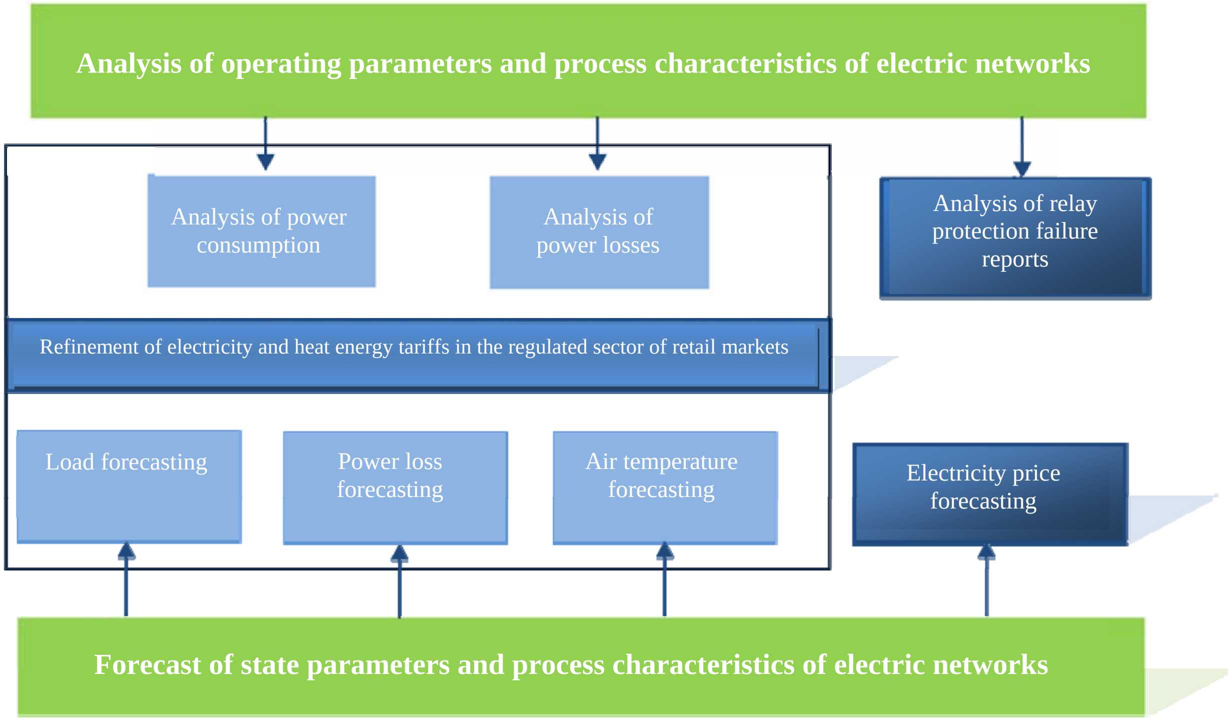


Fig.7. Structure of the OPPC analysis and forecasting problems for electric networks

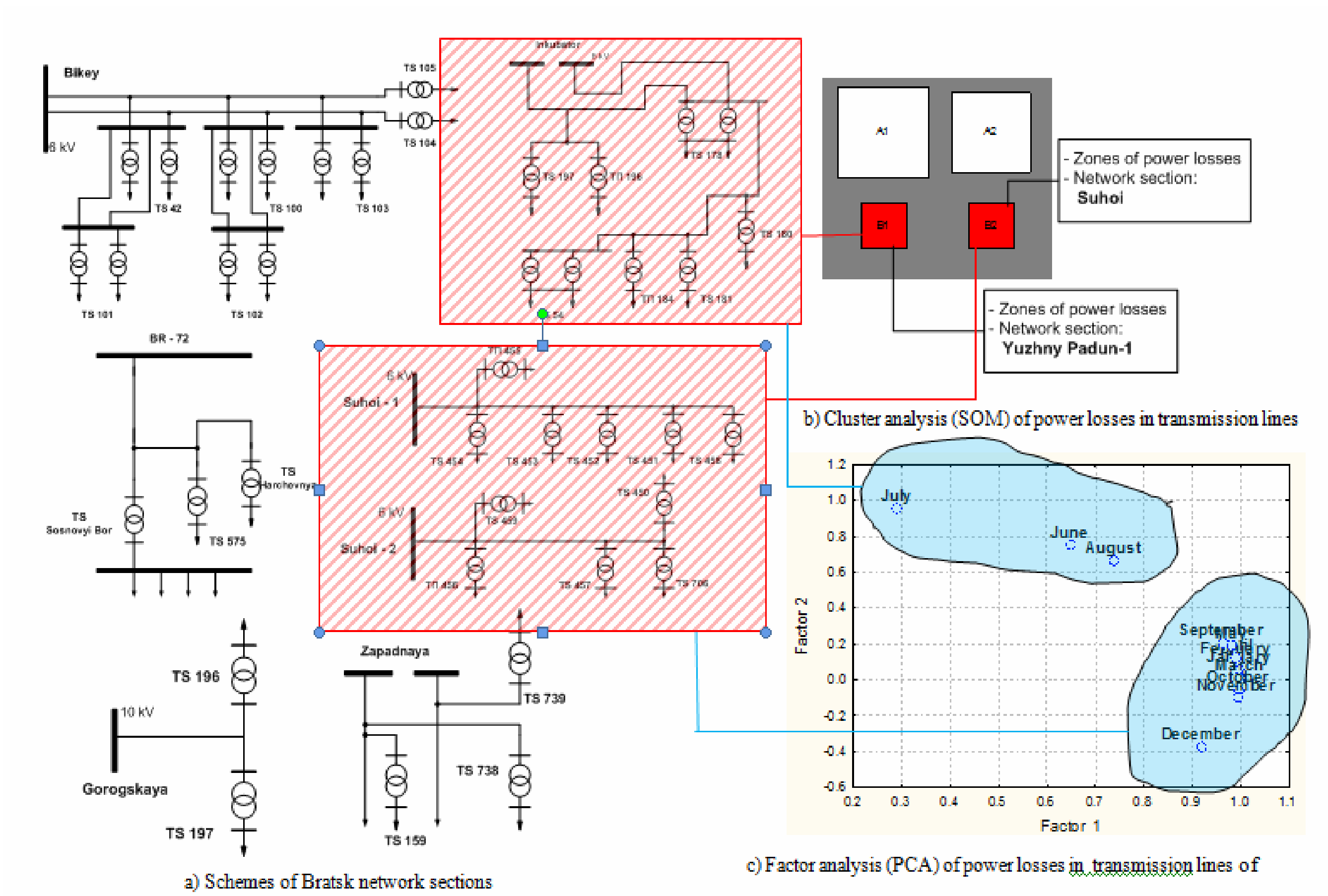


Fig.8. Analysis of power losses in Bratsk distribution networks

With the SOM-based cluster analysis the entire Bratsk section was divided into compact groups with similar features according to the criterion of power loss distribution during a year. It also helped detect the “maximal power loss zones” in transmission lines of the studied network (Fig. 9, Table 1).

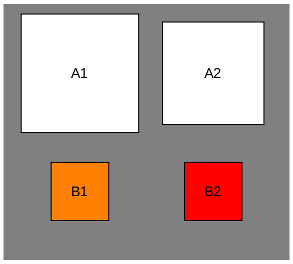


Fig.9 Hinton diagram for analysis of power loss distribution in transmission lines for June 2006

TABLE 1
THE RESULTS OF CLUSTER ANALYSIS (SOM) FOR BRATSK NETWORK SECTIONS

Cluster	Network sections included in the cluster	Months of the largest power losses
A1	Lesokhimik, Nadezhda, Sosnovy Bor, Yuzhny Padun-2	-
A2	Bikey, Severny Artek, Stenikha	-
B1	Yuzhny Padun-1	All months but June and July
B2	Sukhoi	Summer months

The factor analysis (PCA) allows one to determine the main factors in the studied array of change in power losses, including the components of above-standard losses (ASL) (Fig. 8c). The cluster analysis made it possible to identify the factors highlighted in block 4 by the PCA procedure:

- Factor 1 – power losses in the network section “Sukhoi” (cluster 1)
- Factor 2 – power losses in the network section “Yuzhny Padun – 1” (cluster B2)

The analysis of Fig. 8 and Table 1 shows that summer months stand out due to their high values of factor 2 (0.65-0.95). In these months power losses in the network section Sukhoi are high.

On the other hand in December factor 2 has a minimal value – 0.35, which is indicative of low losses. Similarly, the value of factor 1 that varies in the range of 0.92-0.99 determines the place of all months on the PCA plot except for June and July. In these months high power losses are observed in the network section Yuzhny Padun-1.

The subsystem for OPPC analysis has determined that the maximum power losses occur in the network section Sukhoi. Therefore, for this section the forecast of ASL was made for a year interval on the basis of retrospective sampling for the previous 4 years for some indices of power consumption including the ASL component.

In order to carry out a comparative analysis for solving the problem of ASL forecasting several forecasting models were applied [2]:

1. ARIMA (2,0,0), (2,2,0). Due to limited number of input parameters in the regression models the data on the above-standard power losses of the previous years were used as the input data.
2. ANN of Multilayer Perceptron type (5 input - 5 hidden – 1 output neuron). It was chosen by the SA algorithm (Table 2). The following input parameters were used for learning: the volume of electric energy, process power losses, planned power sales, real power sales and number of month. The output parameter was the above-standard power losses.
3. he committee machine structure. It consisted of 3 neural networks-experts of the MLP type of different architecture. The set of input and output values was similar to the set in the previous point.

TABLE 2
“COMPETITION-BASED” SELECTION USING SA ALGORITHM FOR ASL FORECASTING

Number of SA stages	Type of forecasting model	Absolute error, kWh	Number of input neurons	Number of hidden neurons	ANN performance
1	MLP	238882.2	2	1	0.367
2	GRNN	140589.3	6	37	0.366
3	GRNN	140580.9	6	37	0.366
4	GRNN	140580.9	6	37	0.366
5	MLP	105530.2	2	6	0.268
6	RBF	43894.11	6	19	0.115
7	RBF	32463.19	6	22	0.082
8	RBF	28295.02	6	29	0.076
9	MLP	8707.817	3	1	0.022
10	**MLP**	**6565.176**	**3**	**5**	**0.015**

The results of calculations within the forecasting models applied (Table 3) show high forecasting error in the ARIMA model. At the same time the forecast made by CM implemented in the Subsystem of OPPC forecasting is practically twice as accurate as the forecast of the single ANN. The use of the Subsystem of joint OPPC processing in the ANAPRO software essentially increases the forecasting accuracy. As was revealed earlier by PCA the high values of losses for the network section “Sukhoi” are observed in summer months (Fig. 8 c). This affects the accuracy of the ASL forecast in summer months (Table 3) according to models ANN and CM, particularly for July (according to ANN – 30%, CM model – 16.5%) which has the greatest deviation on the PCA diagram (Fig 8 c).

TABLE 3
ASL FORECAST FOR THE NETWORK SECTION "SUKHOI" BY DIFFERENT FORECASTING MODELS

Forecasting model	Relative error for months, %												Mean error, %
	1	2	3	4	5	6	7	8	9	10	11	12	
ARIMA	43.9	43.5	49.1	42.9	45.8	35.9	35.6	43.8	40.1	44.1	40.1	49.3	**42.8**
MLP	2.3	0.4	2.5	4.9	8.3	8.3	30.0	12.5	1.2	0.5	1.4	1.3	**6.1**
CM	1.0	1.5	2.8	3.9	4.3	4.6	16.5	8.1	0.6	0.3	0.2	1.2	**3.7**

Thus, to increase the accuracy of the CM forecast within the Subsystem of joint processing an additional input parameter was included into the learning sampling. This was a correlation coefficient r of factor 1 (networks section Sukhoi), that was calculated in the analysis of power losses (Table 4).

TABLE 4
CORRELATION COEFFICIENT OF FACTOR 1 FOR DIFFERENT MONTHS OF THE NETWORK SECTION "SUKHOI"

Index	Months											
	1	2	3	4	5	6	7	8	9	10	11	12
r	0.996	0.991	0.999	0.989	0.980	0.648	0.289	0.738	0.964	0.997	0.995	0.920

TABLE 5
ASL FORECAST FOR THE NETWORK SECTION SUKHOI ON THE BASIS OF CM(FORECASTING AND JOINT PROCESSING MACROS)

Forecasting model	Relative error for months, %												Mean error, %
	1	2	3	4	5	6	7	8	9	10	11	12	
CM (forecasting macro)	1.0	1.5	2.8	3.9	4.3	**4.6**	**16.5**	**8.1**	0.6	0.3	0.2	1.2	**3.7**
CM (joint processing macro)	0.5	0.1	0.4	1.7	2.5	**2.3**	**2.4**	**0.1**	0.6	0.4	0.5	2.9	**1.2**

The results of the ASL forecasting taking into account correction by PCA and SOM are given in Table 5. As is seen from the Table, the forecast made on the basis of the previously made analysis gave more accurate results and, first of all, a considerable decrease in the forecasting error for the summer months that are most problematic for this section of the network.

A. Analysis of failure reports of relay protection devices and automatic systems

Cluster analysis on the basis of SOM was carried out in the Subsystem of OPPC analysis to estimate the failure reports of relay protection and automatic systems [7]. Two schemes of the SOM-based cluster analysis of failure reports were used (Figs. 10 and 11):

SOM learning resulted in division of the entire array of input data into four clusters with general properties.

Application of cluster analysis within the Subsystem of OPPC analysis made it possible in 2003-2006 to vividly and efficiently determine the specific features of faults of relay protection and automatic systems in different months of a year and carry out a fault diagnosis depending on input data.

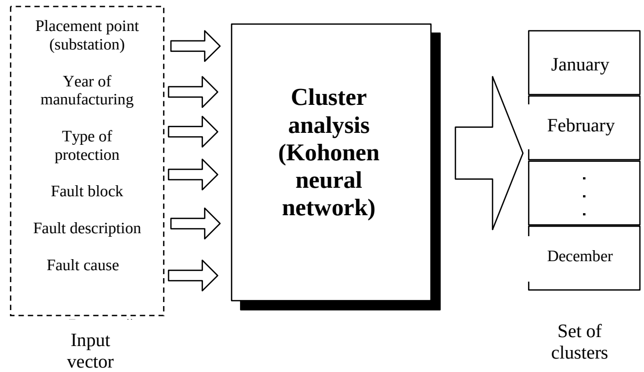


Fig.10. Clustering scheme to detect specific features of faults in relay protection and automatic systems in different months

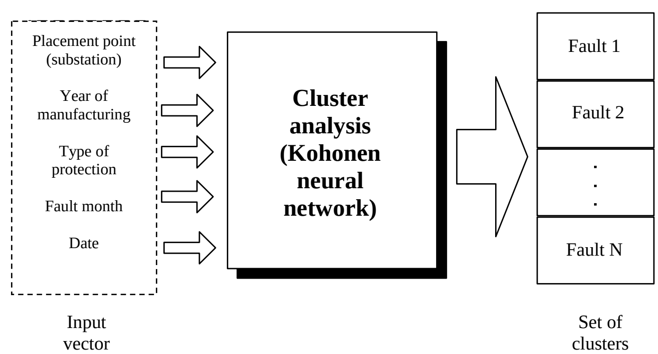


Fig.11. Clustering scheme to estimate the fault probability of relay protection and automatic systems

It is necessary to note that the analysis of obtained clusters proves the correct operation of the chosen neural network model (SOM) that reflected a well-known situation:

- Breaking and flaws of design typical of relay protection and automatic systems that were put into operation in 1965-1990.
- Thunderstorm as one of the fault causes of relay protection devices that takes place normally in summer months.

B. Forecasting electricity prices

Forecast of nodal and spot electricity prices for different lead intervals was made within the Subsystem for OPPC forecasting of the ANAPRO software. The objective was to forecast prices in several price zones (Price zone 1 and Price zone 2) for the following time intervals:

- A day ahead;
- 14-day ahead average hourly

– 4-week ahead average weekly

Let us illustrate solving the problem of price forecasting by the example of the 14-day ahead forecast.

The initial sampling (block 1) represented an array of retrospective hourly electricity price values for the period of 05.11.2007 – 24.11.2008 for two price zones.

Based on the SA algorithm operation the best neural network model of RBF type was chosen in block 8 for Price zone 1 and the neural network model of type GRNN for Price zone 2.

The results of price forecasting for the two price zones on the basis of EBF and GRNN models are illustrated in Fig.12 and presented in Table 6. The Table shows that the maximum errors of the ANN model forecast were made for 24.11.2008. On this day hourly prices changed sharply which caused the largest error in the forecast of hourly price values for both Price zones.

TABLE 6
ERROR OF THE 14-DAY AHEAD HOURLY FORECAST OF ELECTRICITY PRICE FOR PRICE ZONE 1 AND PRICE ZONE 2

Days	Forecast error, %	
	Price zone 1	Price zone 2
11.11.2008	4,39	5,09
12.11.2008	2,82	8,10
13.11.2008	3,62	2,17
14.11.2008	5,09	1,87
15.11.2008	2,30	3,37
16.11.2008	3,22	4,29
17.11.2008	6,55	3,24
18.11.2008	4,30	4,05
19.11.2008	3,44	5,80
20.11.2008	3,05	7,35
21.11.2008	5,13	7,28
22.11.2008	6,31	10,23
23.11.2008	5,07	4,94
24.11.2008	19,73	13,74

The use of the ANAPRO software to forecast prices in two price zones for other lead-time intervals allowed one to obtain

– The forecast error of 2-3.5% in the day-ahead forecast;
– The errors no higher than 6% in the forecast of average weekly value for a 4-week ahead time interval.

IV. CONCLUSIONS

The paper presents the results of experimental studies conducted on the basis of the intelligent software ANAPRO. The studies concerned the analysis and forecasting of state parameters and process characteristics of the electric networks in East Siberia.

V. REFERENCES

Books:

[1] Haykin S. Neural networks. A comprehensive foundation. Second edition / S. Haykin. – Williams Publishing House, 2006. – 1104 p.
[2] Ossovsky S. Neural networks for information processing / S. Ossovsky. - transl. from Polish by I.D.Rudinsky. – M.: Finansy i statistika, 2004. – 344 p.
[3] Kohonen T. Self-organizing maps. / T. Kohonen. – Berlin ets: Spzinger, 1995. XV. 362 p.
[4] Borovikov V.P. Forecasting in STATISTICA system in Windows environment: Theory and intensive practice on computer: Tutorial/V.P.Borovikov, G.I.Ivchenko. – 2nd edit., revised and enlarged. – M.: Finansy i statistika, 2006. – 368 p.

Papers from Conference Proceedings (Published):

[5] Kurbatsky V.G. Application of ANAPRO software for analysis and forecasting of state parameters and process characteristics in electric power systems / V.G.Kurbatsky, N.V.Tomin. Proceedings of the 8th Baikal All-Russian Conf. “Information and mathematical technologies in science and management.” Part 1. – Irkutsk: SEI SB RAS, 2008. – P.91-99.
[6] Kikpatrick S., Gelatt C.D. – Veccht M. P. Optimization by simulated annealing // Science, 1983. – Vol. 220. Pp. 671-680

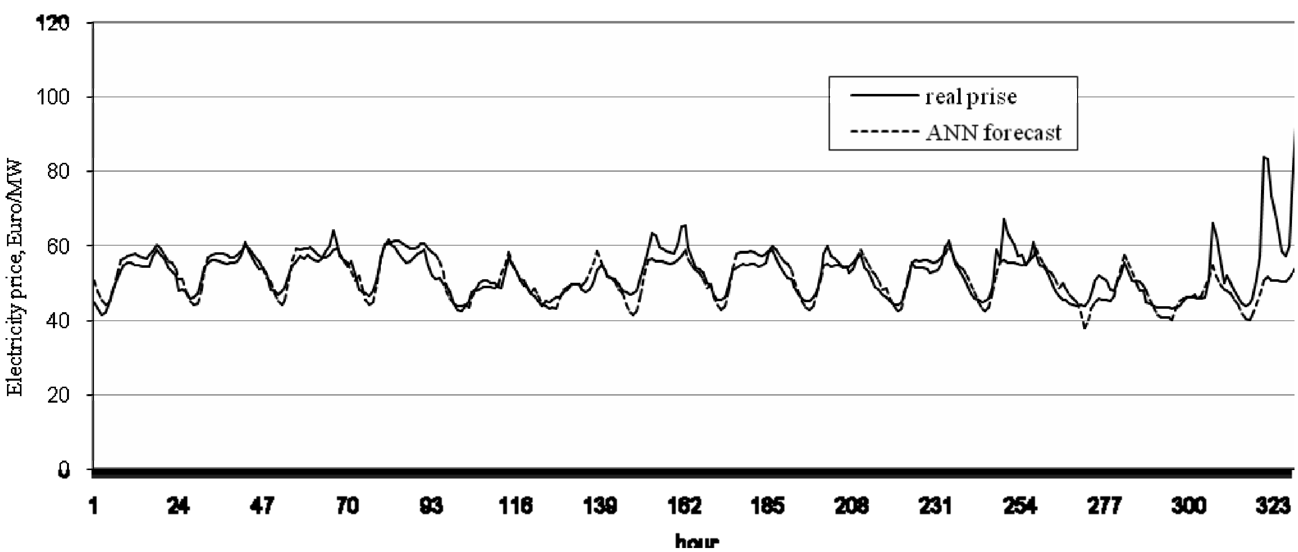


Fig.12. 14-day ahead hourly forecasts of electricity price for Price zone 1.

[7] Tomin N.V. Analysis of failure reports of relay protection and automatic systems on the basis of neural network approach/ N.V.Tomin. Energy: control, quality and efficiency of using energy resources: Collected papers of the 5-th All-Russian scientific and technical conference with international participation. Blagoveshchensk, Publishing Comp. of Am SU, 2008. – P. 265-269.

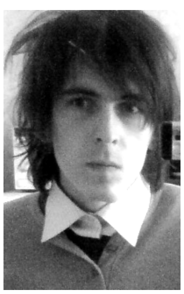
[8] Kurbatsky V.G. Application of neural approach for prognostication of parameters of regime and technological descriptions in electric networks / V.G. Kur-batsky, N.V. Tomin // Applications of optimization methods: Proceedings of XIV Baikal International School-seminar "Optimization methods and their applications», Irkutsk, Baikal, 2008. Vol.4. Irkutsk: Melentiev Energy Systems Institute SB RAS. - 2008. - Pp. 73-79

[9] Kurbatsky V.G. Software for electric power industry problems on the basis of user application macros conception / V.G.Kurbatsky, N.V.Tomin. Proceedings of the 8th Baikal All-Russian Conf. “Information and mathematical technologies in science and management.” Part 1. – Irkutsk: SEI SB RAS, 2008. – P.206-212

VI. BIOGRAPHIES

Kurbatsky G. Victor was born on May, 27th, 1949 in Komsomolsk-on-Amur (Russia), PhD, Professor, Doctor of Science. In 1997 Prof. Kurbatsky defended doctor’s thesis "Monitoring of quality of the electric power in electric networks of Russia for a choice of actions on maintenance of electromagnetic compatibility" at the Energy Systems Institute of the Russian Academy of Science (Irkutsk). He is Corresponding member of the International Research Institute of Electrical Engineers (New York). Professor Kurbatsky is the author of several monographs and manuals and more than 270 scientific papers.

Professor Kurbatsky is a Leading researcher at the Energy Systems Institute, SB RAS.



Tomin V. Nikita was born on 18th in December, 1982., in Bratsk (Russia), Ph.D. In 2007 He defended his PhD thesis “Analysis and forecasting of operating parameters and characteristics based on artificial intelligence technologies in electric power networks” at the Energy Systems Institute of the Russian Academy of Science (Irkutsk).

Dr. Tomin specializes in the field of artificial intelligence technologies in electric power systems. He is the author and co-author of more than 45 scientific papers.

Dr. Tomin is a Senior researcher at the Energy Systems Institute. SB RAS.