VESTNIK

scientific journal
of Siberian State Aerospace University
named after academician M. F. Reshetnev

Issue 7 (33)

Krasnoyarsk 2010

VESTNIK

scientific journal
of Siberian State Aerospace University
named after academician M. F. Reshetnev

Issue 7(33)

Founder
Siberian
State
Aerospace
University
named after
academician
M. F. Reshetnev

TO THE READERS' AWARENESS

"Vestnik of Siberian State Aerospace University named after academician M. F. Reshetnev" (Vestnik of SibSAU) is the inter-regional scientific journal, comprising the results of scientific research in the sphere of natural, technical and applied sciences, which concern the design, production and running of aviation and space and rocket machinery, as well as economics

Articles, submitted to the journal are published after required reviewing and under their compliance with editorial requirements (www.vestnik.sibsau.ru)

Frequency – 4 times a year
Special issue – once a year
English version – once a year

Electronic version of the journal is available on the site of the Scientific on-line library – http: //www.elibrary.ru and journal site (www.vestnik.sibsau.ru)

The journal is included in the catalogues:
Russian Press "Rosspechat" – 11399;
Obyedinenniy "Pressa Rossii" – 39263

EDITORIAL BOARD ADDRESS:

Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, 660014, Krasnoyarsk, Krasnoyarskyi rabochyi prospect, 31, P-415
Tel./ fax (391)291-90-19
E-mail: vestnik@sibsau.ru

The journal is registered by the Federal service supervising adherence to laws in the sphere of mass communications and cultural heritage protection, certificate PI № ФС77-22189 dd 27.10.2005.

The journal is included by VAK of Russia in the "list of leading reviewed scientific journals and publications, issued in Russian Federation, where the main scientific work results on doctor dissertations are to be published"

Circulation 1000 copies.
Editors: A. E. Goncharov, A. I. Kuklina,
M. V. Savelyeva, T. V. Strekaleva, N. Y. Popova,
E. G. Pylaeva, T. V. Ukhova, N. B. Kokurina,
S. S. Vershnyak, O. I. Suslova
Layout and make-up M. A. Belousova
Passed for printing 19.09.2011. Format 70×108/16. Paper poster. Flat printing. Conventionally printed sheets 31,3.
Publisher's signature 43,6. Order 87/106. C 89/11.
Publishing department
of Siberian State Aerospace University.
660014, Krasnoyarsk, Krasnoyarskiy rabochiy prospect, 31.
Printed in Yu. P. Sukhovolskaya's printing house.
660125, Krasnoyarsk, Vodopianov st., 2.

Part 1

MATHEMATICS, MECHANICS, COMPUTER SCIENCE

E. I. Algazin
Novosibirsk State Technical University, Russia, Novosibirsk

SPECIFICATIONS OF AN INFORMATION PROCESSING INVARIANT SYSTEM IN CONDITIONS OF NONCOHERENT RECEPTION AND INACCURATE DETERMINATION OF THRESHOLDS

An information processing invariant system based on a linear detector in conditions of inaccurate determination of thresholds is considered. Quantitative estimation of noise immunity of such a system with its further comparison with noise immunity of an ordinary binary system with non-coherent reception is carried out.

Keywords: invariant system, noise immunity.

The main requirement to an information processing system is undistorted transmission through communication channels with variable parameters.

There are methods which are reduced to using of ARA, diversed reception, adaptive methods with a training signal, systems with feedback.

These methods have both positive and negative characteristics. One of the drawbacks of the methods mentioned above is a difficulty in realization of transmission algorithms of signals with multilevel amplitude modulation.

In the given paper the algorithm of multilevel amplitude modulated signals transmission through the channels with variable parameters is synthesized and quantitative estimation of the noise immunity in conditions of non-coherent reception is carried out.

There is a communication channel restricted by the frequencies f_{low} и f_{high}. The state of the communication channel is defined by the stationary interval inside which the influence of multiplicative noise is described by the stability of the transmission coefficient $k(t)$ on a certain frequency.

The algorithm of reception is defined by the carrying frequency given as an average frequency of the channel, the amplitude of which is modulated by rectangular impulses.

It is required to determine the technical characteristics of an invariant transmission system in conditions of imprecise definition of thresholds.

Each transmitted block will contain the informative part and the sequence of training signals S_{PILOT}.

On the receiving side the training signals are averaged and used for modulation of the informative part of the block.

At the same time due to the changing of communication channel parameters the information and training signals are interfered with the adaptive noise.

To decrease the influence of adaptive noise of the communication channel the operation of averaging of the training signals in each block is used [1].

Let us carry out the analysis of noise immunity of the invariant system in fig. 1, where two processing channels are used.

In the first channel, consisting of a synchronous detector (SD) and the first solving device (SD1) the estimation of the channel transmission coefficient and dispersion of normal noise is carried out. Later these data are used for calculating the threshold in conditions of invariants demodulation.

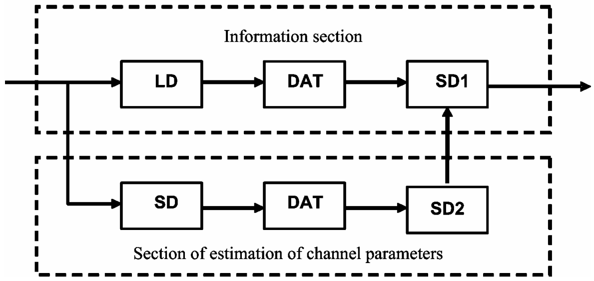


Fig. 1. Extended structural scheme of an invariant system: LD is a linear detector; DAT is a digital – analogue transducer; SD1 is a solving device 1; SD is a synchronous detector; SD2 is a solving device 2

In the second channel a non-coherent system with a linear detector (LD) and the second solving device (SD2) are used. In this channel reception signals are really demodulated.

Let us estimate the quantitative indicators of the method offered.

The principle of information section operation consists in the separation of the reception signals envelope together with the normal noise with the help of LD. The result of transformation into DAT further on is recorded in SD1.

In SD1 the decision in favour of one or another invariant is made.

As it is known [2], in the process of LD using the displacement of mathematic expectation appears. Mathematic expectation is calculated by the following formula [2]:

$$m_R = \sigma\sqrt{\frac{\pi}{2}}\left\{ I_0\left(\frac{\alpha^2}{4\sigma^2}\right) + \frac{\alpha^2}{2\sigma^2} \times \right. \\ \left. \times \left[I_0\left(\frac{\alpha^2}{4\sigma^2}\right) + I_1\left(\frac{\alpha^2}{4\sigma^2}\right)\right]\right\} e^{-\frac{\alpha^2}{4\sigma^2}}, \quad (1)$$

where m_R is the quantity of mathematic expectation; σ^2 is dispersion component of normal noise; I_0 and I_1 are modified Bessel functions of zero and first order; $\alpha = k\cdot\text{INV}_l$, where k is a coefficient of transmission of the channel; INV_l is l transmitted invariant.

The quantity of dispersion on the output of LD is calculated by the following formula [2]:

$$\sigma_R^2 = m_2 - m_R^2 = 2\sigma^2 + \alpha^2 - m_R^2 . \quad (2)$$

The variables in (2) are described above.

To decide in favour of one or another invariant it is necessary to know the values of thresholds for each pair of invariants.

To estimate thresholds it is necessary to calculate m_R and $\sigma^2{}_R$.

It can be done with the help of section of channel parameters estimation (fig. 1) where calculation of quantities k and σ^2 is made.

Joint operation of the information section and the section of channel parameters estimation consists in reception and recording of values of amplitude modulated informative and training signals in SD1 and SD2 by a non-coherent receiver and calculating of invariant estimation on their basis.

On the basis of the latter and the calculated thresholds a decision in favour of one or another invariant is made.

Let us calculate the probability of erroneous reception in case of multilevel invariant amplitude modulated transmission of signals. The well-known approach is used to do this [3]:

$$P_{tr} = P_1 \int_0^{z_p} W_2(z)\,dz + P_2 \int_{z_p}^{\infty} W_1(z)\,dz , \quad (3)$$

where P_{tr} is the probability of transition of the first invariant into the second one and vice versa; P_1 is the probability of appearing of the first invariant; P_2 is the probability of appearing of the second invariant; the first integral is the probability of appearing of the second invariant, when the first one is sent; the second integral is the probability of appearing of the first invariant, when the second one is sent; z_P is a threshold value necessary to calculate P_{tr} with known P_1 and P_2.

The quantity z_P is defined with the help of the best bias estimation by minimizing P_{tr} by z_P. With unknown P_1 and P_2 let us choose $P_1 = P_2 = 0.5$.

As we can see from the expression (3), it is necessary to know the analytical expression $W_1(z)$ and $W_2(z)$.

For coherent reception the calculation of quantities $W_1(z)$ and $W_2(z)$ is known and is shown in [1]. The same approach can also be used in case of non-coherent reception.

Thus the quantity of estimation of the invariant in such a system is calculated as follows:

$$\mathrm{INV}_l^* = \frac{\sum_{i=1}^{N}\left(k \cdot \mathrm{INV}_l + \xi(i)\right)}{\frac{1}{L}\sum_{m=1}^{L}\sum_{j=1}^{N}\left(k \cdot S_{PILOT} + \eta(m,j)\right)} S_{PILOT} ,$$

where INV_l is l transmitted invariant; $\xi(i)$ is i value of Relay noise; k is the coefficient of communication channel transmission; in the denominator: S_{PILOT} is the value of the training signal; $\eta(m, j)$ is j value of Relay noise in m realization of signal S_{PILOT}; N is the number of readings taken by the envelope INV_l or S_{PILOT}; L is the number of training signals.

Without loss of generality let us take $S_{PILOT} = 1$, as $S_{PILOT} > 0$, and we can divide the values of invariants INV_l and root-mean-square deviation into S_{PILOT}.

When $S_{PILOT} = 1$ we obtain the following analytical expression:

$$\mathrm{INV}_l^* = \frac{\sum_{i=1}^{N}\left(k \cdot \mathrm{INV}_l + \xi(i)\right)}{\frac{1}{L}\sum_{m=1}^{L}\sum_{j=1}^{N}\left(k + \eta(m,j)\right)} . \quad (4)$$

To calculate P_{tr} it is necessary to know mathematic expectations and dispersion of the numerator and the denominator of the expression (4).

To calculate it let us use the following approach.

Mathematic expectation of the numerator (4) will be:

$$m_{num} = m_R \cdot N . \quad (5)$$

Dispersion of the numerator (4) will be:

$$D_{num} = N \cdot \sigma_R^2 , \quad (6)$$

where m_R and $\sigma^2{}_R$ are calculated in accordance with the expressions (1) и (2). Mathematic expectation of the denominator (4) after transformation will be as follows:

$$m_{den} = m_{R2} \cdot N , \quad (7)$$

where m_{R2} is calculated in accordance with (1) by $\alpha = k$, as $S_{PILOT} = 1$ is used instead of INV_l.

Dispersion of the denominator (4) will be:

$$D_{den} = \frac{N \cdot \sigma_{R2}^2}{L} , \quad (8)$$

where $\sigma^2{}_{R2}$ is calculated in accordance with (2) by $\alpha = k$, where m_{R2} is used instead of m_R.

Then the expression of density of the probability of the estimation of the invariant will be [4]:

$$W(z) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(zx-m_1)^2}{2\sigma_1^2}} e^{-\frac{(x-m_2)^2}{2\sigma_2^2}} |x|\,dx , \quad (9)$$

where $\sigma_1 = \sqrt{D_{num}}$; $\sigma_2 = \sqrt{D_{den}}$; $m_1 = m_{num}$; $m_2 = m_{den}$.

The calculation of P_{tr} is carried out quantitatively by approximation of the formula (9).

In the systems with AM and non-coherent reception the analogue of the probability of the pairwise transition is the probability of error Per, which is calculated by the know formulas [3].

The probability of the pairwise transition and the probability of error are calculated for the similar h – noise-to-signal ratio which is calculated by the formula $h = k \cdot \mathrm{INV}_1 / \sigma_R$.

Threshold z_P are calculated by minimization of P_{tr} in formula (3). For $k = 1$ and $\mathrm{INV}_1 = 1$, $\mathrm{INV}_2 = 2, 3, 4, 5, 6$ the calculations result in z_P = 1,23; 1,49; 1,77; 2,07; 2,36.

For $k = 0{,}7$ and $\mathrm{INV}_1 = 1$, $\mathrm{INV}_2 = 2, 3, 4, 5, 6$ the calculations result in z_P = 1,14; 1,30; 1,50; 1,68; 1,92.

The results of modulation are shown in fig. 2 and fig. 3, from which we can see that the peculiarity of any invariant system based on the principle of invariant relative amplitude modulation is that amplitude modulated signals formed by INV_l and S_{PILOT} are transmitted through the channel.

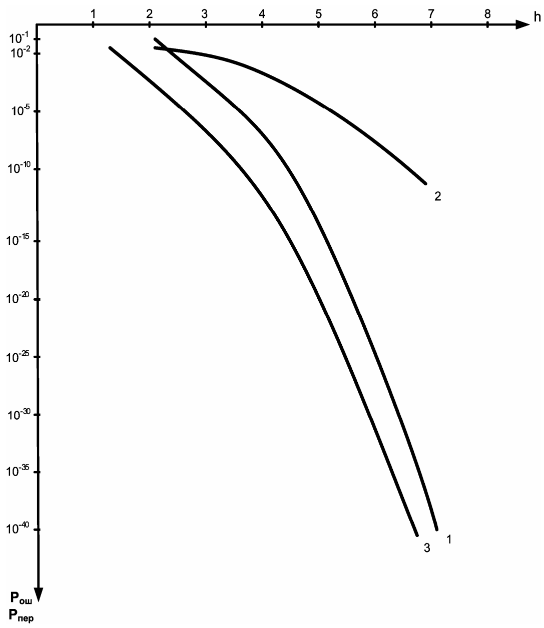


Fig. 2. Results of modulation:
1 – the probability of pairwise transition of one invariant into another under the following given conditions: $k = 1$; $\text{INV}_1 = 1$; $\text{INV}_2 = 2, 3, \ldots, 6$ and non-coherent reception; 2 – the probability of error in classical amplitude modulation and non-coherent reception; 3 – the probability of the pairwise transition of one invariant into another under the following given conditions: $k = 1$; $\text{INV}_1 = 1$; $\text{INV}_2 = 2, 3, \ldots, 11$ and coherent reception

As a rule the transmission of these signals on the basis of classical algorithms provides low noise immunity of information processing [3].

Only after processing of these signals in accordance with the quotient algorithm using expression (4), we obtain the invariant estimation which is really a number but not a signal.

As we can see from fig. 2 and fig. 3 the probability of the pairwise transition of one invariant into another in conditions of great noise-to-signal ratio is defined by the values (10^{-30}–10^{-40}). In recalculation of the shown above quantities the probability of erroneous reception of a single symbol in classical systems is within the limits (10^{-6}–10^{-10}).

However, in real situations it is impossible to determine the value of the transfer constant of communication channel accurately. The consequence of this would be inaccurate definition of the thresholds. The summand of the denominator X_j in formula (4) of evaluation of the invariant ISPR can be represented as:

$$X_j = \frac{1}{L}\sum_{m=1}^{L}\left(k + \eta(m,j)\right), \qquad (10)$$

where L is the number of averages; k is thetransfer constant of communication channel; $\eta(m, j)$ is j-th reading of additive noise in the m-th realization of the training signal.

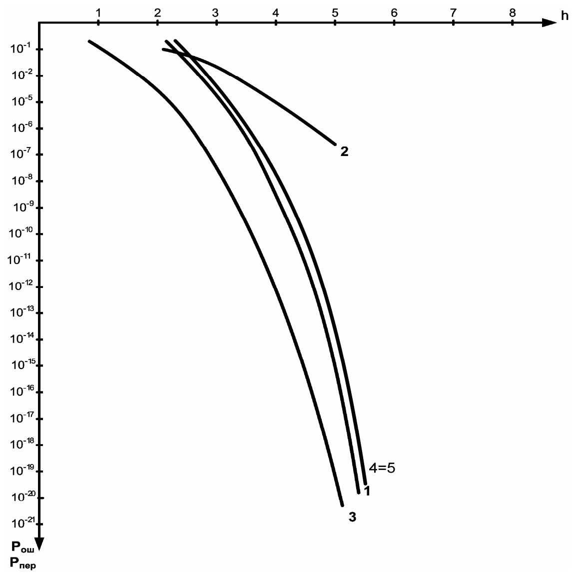


Fig. 3. Results of modulation:
1 – the probability of the pairwise transition of one invariant into another under the following given conditions: $k = 0{,}7$; $\text{INV}_1 = 1$; $\text{INV}_2 = 2, 3, \ldots, 6$ and non-coherent reception; 2 – the probability of error in classical amplitude modulation and non-coherent reception; 3 – the probability of the pairwise transition of one invariant into another under the following;given conditions: $k = 0{,}7$; $\text{INV}_1 = 1$; $\text{INV}_2 = 2, 3, \ldots, 11$ and coherent reception; 4 – the probability of pairwise transition at $k = 0{,}7$, and thresholds, calculated with k_-; 5 – the probability of pairwise transition at $k = 0{,}7$, and thresholds, calculated with k_+

Then expected X_j is equal to:

$$EX_j = E\left(k + \eta(m,j)\right) = m(k). \qquad (11)$$

In addition, we have

$$\overline{X} = \frac{1}{N}\sum_{j=1}^{N} X_j = m(\hat{k}), \qquad (12)$$

$$\hat{k} = g(\overline{X}), \qquad (13)$$

where $\hat{k}$ is the evaluation of the transfer constant of communication channel; g is the inverse of the function m;

$$D\hat{k} \approx \left(g'(m(k))\right)^2 \frac{\sigma^2}{NL} \qquad (14)$$

(according to the theorem on asymptotic normality),

$$m(k) = X, \qquad (15)$$

$$k = g(X), \qquad (16)$$

$$g'(X) = \frac{1}{\left(m(k)\right)'} = \frac{1}{m'(k)} = \frac{1}{m'\left(g(X)\right)}, \qquad (17)$$

$$m(k) = \sigma\sqrt{\frac{\pi}{2}}\, e^{-\frac{k^2}{2\sigma^2}} \sum_{i=0}^{\infty} \frac{(2i+1)!!}{(i!)^2\, 4^i}\left(\frac{k}{\sigma}\right)^{2i}. \qquad (18)$$

Then

$$m'(k) = \sigma\sqrt{\frac{\pi}{2}}\left(-\frac{k}{\sigma^2}e^{-\frac{k^2}{2\sigma^2}}\sum_{i=0}^{\infty}\frac{(2i+1)!!}{(i!)^2\,4^i}\left(\frac{k}{\sigma}\right)^{2i} + \right.$$
$$\left. + e^{-\frac{k^2}{2\sigma^2}}\sum_{i=0}^{\infty}\frac{(2i+1)!!}{(i!)^2\,4^i}\frac{2i\cdot k^{2i-1}}{\sigma^{2i}}\right), \quad (19)$$

$$D\hat{k} = \frac{\sigma^2}{(m'(k))^2\,N\cdot L}, \quad (20)$$

$$k_- = k - 3\sqrt{D\hat{k}}, \quad (21)$$

$$k_+ = k + 3\sqrt{D\hat{k}}. \quad (22)$$

Fig. 3 shows the curves 4 and 5 corresponding to the curves of noise immunity at k_- and k_+, respectively. In this case $D\hat{k} = 1.2\cdot 10^{-19}$, k_- = 6.99999988·10^{-1} and k_+ = 6,99999989·10^{-1}. As it is evident from these curves, a decrease in immunity ISPR is observed.

The invariant non-coherent system of information transmission is offered and its qualitative characteristics in conditions of inaccurate definition of thresholds are defined.

The developed method can find application in the systems of information processing.

In the author's opinion it is necessary to compare the noise immunity of the investigated invariant system with the noise immunity of similar invariant systems. That will be done in the subsequent papers.

References

1. Gonorovsky I. S. Radio Engineering circuits and signals M. : Sovetskoye Radio, 1971.
2. Invariant method of analysis of telecommunication systems of information transmission : monograph / V. B. Malinkin, E. I. Algazin, D. N. Levin, V. N. Popantonopulo. 2006.
3. Teplov N. L. The noise immunity of the systems of discrete information transmission. 1964.
4. Levin B. R. Theoretical foundation of statistic radio engineering. The third edition. M. : Radio and communication, 1989.

S. S. Aplesnin, A. I. Moskvin
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

MAGNETOELECTRIC EFFECT INDUCED BY ORBITAL ORDERING OF ELECTRONS

Relationship between orbital order and the formation of the spontaneous magnetic moment, lattice constant, correlation function of orbital and spin moments between nearest neighbors have been investigated in terms of the continuous Potts model for set of electron-phonon parameters and spin-phonon interactions. A change in the permittivity and orbital correlation functions in the external magnetic field has been found.

Keywords: permittivity, magnetoelectric effect, electron-lattice interaction, orbital and spin moment.

The study of multiferroics with the coexistence of at least two of the three order parameters (magnetic, electric, and crystallographic) [1] is an urgent problem, for it describes the possibility of controlling the magnetic properties of a material by means of an electric-field and, vice versa, magnetic-field modulation of electric properties. In the future, multiferroics may find wide technical application in sensors and recording devices, reading and storing information. While the spintronic devices transform information by changing the magnetization to electric voltage; in multiferroics the correlation between the magnetic and electric subsystems manifests itself in the magnetoelectric effect [2; 3].

The $Co_xMn_{1-x}S$ solid solutions can be attributed to the multiferroic class [4]. In the temperature ranges of $T \approx$ 110–120 K and $T \approx$ 230–260 K, the correlation between the magnetic and electric subsystems has been found [5]. The presence of this correlation is confirmed by sharp rise of the magnetization and the maximum in the relative variation of permittivity, measured in the external magnetic field and without it at a lowering temperature [6].

Electron density redistribution inside a 3d-shell arising from electron transitions from e_g to t_{2g} levels; or due to the different electro negativities of cobalt and manganese ions can lead to changes the orbital occupancy at the t_{2g} shell of Mn ions.

An important feature is that magnetic exchange interaction depends on orbital occupancy. This means that even the sign could change. Therefore, it is possible that magnetic correlation at normal can be very different from that in the ordered phase, when the orbital order is accompanied by magnetic transition. The variation of the orbital occupancy may be caused shift in polarizability and in spin state of cation.

The aim of this study is to investigate the physical properties of the $Co_xMn_{1-x}S$ solid solutions typical of multiferroics, induced by spin-charge ordering, and to establish the interrelation between the magnetic, electric, and elastic subsystems.

Interrelation model between electron and elastic subsystems. Interpretation of the results obtained requires consideration of the interrelation between the electron and crystal structures. We shall consider competition between the Coulomb and direct exchange interactions between manganese ions. The consideration of the electrons' kinetic energy induces redistribution of the electronic density between the e_g and t_{2g} levels towards the increase in electron population in the t_{2g} state. In this case, one of the t_{2g} orbitals contains two electrons and hopping between the neighboring sites and different orbitals occurs with no change in the Coulomb interactions of electrons. In addition, kinetic energy of electrons grows due to the formation of a narrow miniband by electron hopping over to the nearest manganese ions in the vicinity of the Fermi level.

For a pair of manganese ions with the half-filled orbitals the integral of electron hopping over sulfur anions can be estimated as $t^x_{\alpha\beta} = E_{x,\alpha}E_{x,\beta}/[(\varepsilon_p - \varepsilon_d) + U]$, where $E_{x,a}$ are the integrals of overlap of the sulfur p-orbitals and manganese t_{2g}-orbitals with the value $E_{x,a}$ = 1.1 eV, $\varepsilon_p - \varepsilon_d$ = 1.5 eV is the charge gap, and U = 4 eV. For two electrons at one of the five d-levels of manganese ions surrounded by sulfur ions, the overlap integral is zero, ($E_{x,a}$ = 0) in virtue of the Pauli principle and upon the lateral oscillations of sulfur ions weak overlap of the electron wave functions of manganese ions with the miniband width of $W = 2zt$ = 0.5 – 0.8 eV is possible. In other words, the transport properties are related to hole hopping over the e_g-orbitals and motion of electrons over the t_{2g}-states. The type of the carriers corresponds to the lattice polarons, i. e., the motion of electrons in the lattice induces the coupled lateral and longitudinal oscillations of sulfur ions and is implemented along one of the subbands of the t_{2g} band with the orbital magnetic moment, while in the e_g band motion of the current carriers is directly connected with the type of the magnetic structure. As a result, different population of the *dxy*-, *dxz*-, and *dyz*-orbitals is induced and orbital-charge ordering forms accompanied by ordering the orbital angular moments.

In the phenomenological representation, the Hamiltonian for two orbitals x and y and two sites can be written as:

$$H = -J_1(n_{1x}-n_{1y})(n_{2x}-n_{2y})(1-c) - g(n_{1x}+n_{2x})(1-c)x -$$
$$- g(n_{1y}+n_{2y})(1-c)y + \frac{1}{2}k(x^2+y^2) - b(x^3+y^3) -$$
$$- h(n_{1x}+n_{2x}-n_{1y}-n_{2y})(1-c) - J_2(n_{1ix}-n_{1iy})(n_{2ix}-n_{2iy})c -$$
$$- KM_1M_2 - g_s(x+y)M_1M_2 - \lambda(1-c)\left(n_{1x}-n_{1y}\right)M_1 -$$
$$- \lambda(1-c)(n_{2x}-n_{2y})M_2 - h(M_1+M_2) - h(n_{1ix}-n_{1iy} +$$
$$+ n_{2ix}-n_{2iy})c - c\lambda\left(n_{1ix}-n_{1iy}\right)M_1 - \lambda\left(n_{2ix}-n_{2iy}\right)M_2c,$$

where $n_{1,2ix}$ and $n_{1,2x,y}$ are the electron densities on the *dxz*- and *dyz*-orbitals of manganese ions surrounding a cobalt ion with concentration c and on manganese ions in the matrix with concentration (1 – c), J_1 and J_2 are the exchange interactions between the orbital magnetic moments Mn–Mn and Co-Mn, g is the parameter of the electron-lattice interaction, x and y are the displacements of ions in the directions corresponding to the square sides, k and b are the elastic constants, h is the magnetic field, a is the lattice constant, $K < 0$ is the exchange interaction between magnetic moments M_1 and M_2, g_s is the constant of the spin-lattice interaction, and λ is the parameter of the spin-orbital coupling. At the interaction of orbital and spin moments, the high-order terms (L_1L_2) (M_1M_2) appear, but for the t_{2g} electrons they are much smaller than the spin-orbital interaction.

Now, within this Hamiltonian, we will try to answer a number of questions. How do the concentration of the Mn-Co-Mn clusters and the effective orbital interaction influence the temperature of the formation of the spontaneous magnetic moment? How will the lattice parameter change at orbital-charge and magnetic ordering? Which changes in the temperature behavior of the orbital correlation functions (correspondingly, charge redistribution) will result from action of the external magnetic field?

The estimate density of the electron on the t_{2g} orbitals of manganese ions in the matrix by the value of the spin moment at site S = 4.4 μ_B for MnS ($n_{1,2}$ ~ 0.1). The electron density in the Mn-Co cluster can reach value $n_{1,2i}$ ~ 0.5.

Thermodynamic characteristics of model. We shall calculate the thermodynamic characteristics, the correlation function between the nearest neighbors for the orbital $\langle L_1 L_2\rangle$ ($L = n_{1x} - n_{1y}$) and magnetic $\langle M_1 M_2\rangle$ moments, and the average displacement along the OX $\langle x\rangle$ and OY $\langle y\rangle$ axes using a continuous Potts model, where the quantities vary within intervals $0 < n_{1,2x,y} < 0.1$, $0 < n_{1,2ix} < 0.5$, $0 < x,y < 1$, and $-1 < M_{1,2} < 1$.

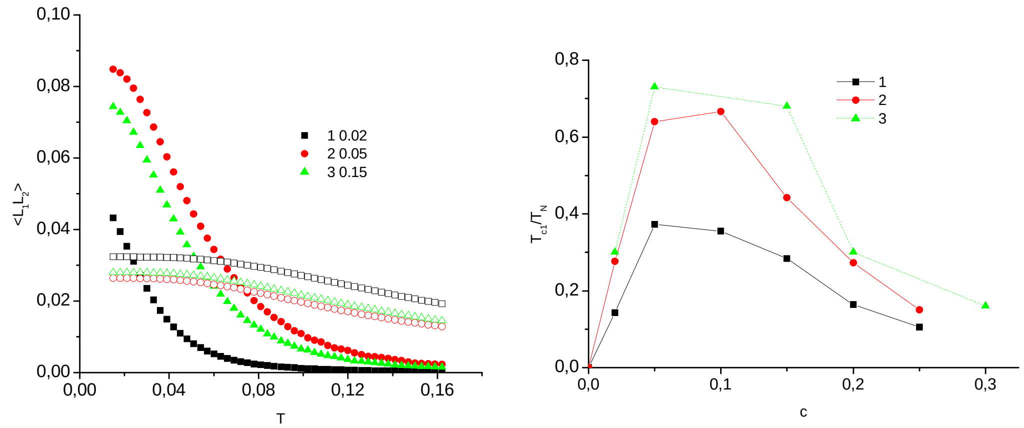
The correlator of the magnetic moments in the region of the transition temperature from the magnetically ordered to paramagnetic phase decreases by a factor of two or three, and has an inflection point. Therefore, we associate the temperature at which the long-range ferromagnetic order of the orbital moments disappears with the temperature at which the inflection in the temperature dependence of the correlation function $\langle L_1 L_2\rangle (T)$ and $\langle M_1 M_2\rangle (T)$ is observed. Fig. 1, *a* depicts temperature dependences of the correlators for several concentrations of cobalt ions. One can extract two temperatures T_c and T_{c1} at which the orbital magnetic moments of manganese ions and the orbital moments of the Mn-Co-Mn clusters are ordered. The typical concentration behavior of T_{c1} (c) is presented in fig. 1, *b* for two exchange parameters J_2/J_1. It correlates with the concentration of the Mn-Co-Mn clusters in dependence of cobalt concentration $c = zx(1-x)^{z-1}$, where z = 12 is the number of the nearest neighbors for the FCC lattice. According to the results of our calculations, the dependence of T_{c1} from the value of the exchange between the orbital magnetic moments in the Mn-Co-Mn cluster is linear with the slope depending on the concentration of cobalt ions and the electron-lattice interaction which shifts the temperature of the transition of orbital ordering towards the higher temperatures within 20 % with an increase in g/J ~1. The experimental results illustrated in

Fig. 1 are described satisfactorily within the model with orbital ordering of the angular moments.

Let us now consider the effect of the electron-lattice and spin-lattice interactions on lattice deformation and variation in the lattice constant with temperature. Fig. 2 shows temperature dependences of the average displacement of ions for several parameters of the electron-lattice and spin-lattice interactions with allowance made for unharmonism of the lattice oscillations. The interaction of electrons with allowance for the lattice degrees of freedom leads to the increase in distance between electrons, i. e., to the growth of the lattice constant. In particular, the displacement is directly proportional to the constant of the electron-phonon interaction $\langle x\rangle = gn/k$ in the systems with strong electron correlations, when the Coulomb interaction exceeds the band width and unharmonism of the oscillations is neglected. With an increase in the value of the electron-lattice interaction, the change in the slope is observed at some temperature which shifts towards higher temperatures as g grows.

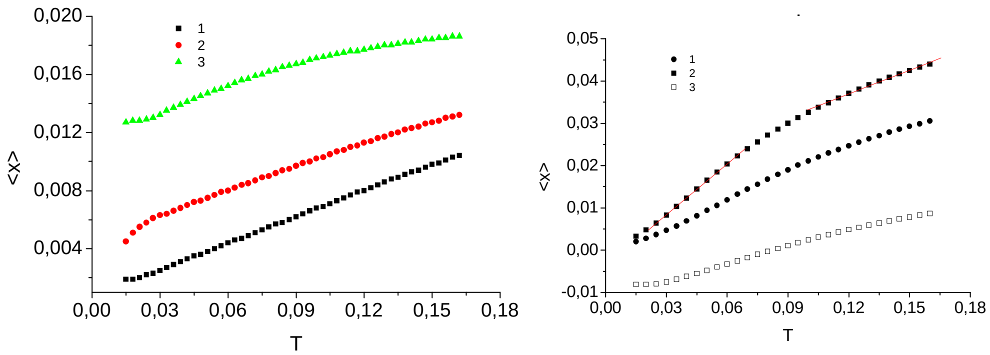
The exchange interaction between the localized spins depends exponentially on the distance; due to the lattice compression the density of overlap of the wave functions grows and the exchange energy increases. The enhancement of the spin-lattice coupling leads to a change in the sign of the average ion displacement. In fig. 2, one can see temperature dependences of $\langle x\rangle$ on two concentrations and two parameters of the spin-lattice coupling. Due to the competition of the electron-lattice and spin-lattice interactions, the lattice compresses. The change in the slope of the dependence $d\langle x\rangle/dT$ is observed in the region of the transition to the magnetically ordered state.

In the presence of the spin-orbital interaction, fluctuations of the magnetic moments in two sublattices are asymmetric; as a result, the total magnetic moment $M_1 + M_2$ with the maximum at the Neel temperature is induced (fig. 3). This is similar to the action of the external magnetic field $h_{ef} = \lambda \langle L\rangle$, when the magnetization value of one of the sublattices changes. Redistribution of the electron density on the *dxz*- and *dzy*-orbitals results in the occurrence of the tetragonal distortion illustrated in fig. 3 for several parameters of the spin-orbital coupling.



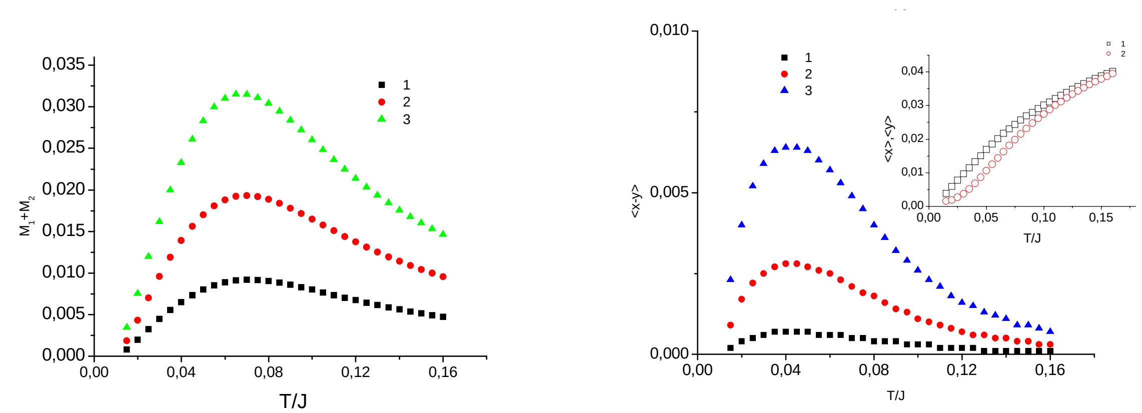
a *b*

Fig. 1. Correlation of magnetic orbital moments $\langle L_1L_2\rangle$ (*a*) in the matrix of manganese ions (light symbols) and in the Mn-Co-Mn cluster (dark symbols) versus temperature for the parameters $J_1 = 10$, $J_2 = 6$, $g = 6$, $k = 20$, $b = 3$, $K = -0.35$, $g_s = 0.1$, $\lambda = 0.1$, and $x = 0.02$ (1), 0.05 (2), and 0.15 (3) (*a*). The temperature of the orbital magnetic moment formation in the Mn-Co-Mn cluster is normalized by the Neel temperature for the parameters $J_2 = 5$ (1), 8 (2), $J_1 = 10$, $g = 6$, $k = 20$, $b = 3$, $K = -0.35$, $g_s = 0.1$, and $\lambda = 0$ versus cobalt concentration (*b*)



a *b*

Fig. 2. Average ion displacement $\langle x\rangle$ versus temperature for the parameters of the electron-phonon interaction $g = 2$ (1), 5 (2), and 8 (3) with $b = 1$, $J_2 = 6$, $J_1 = 10$, $k = 20$, $K = -0.35$, $g_s = 0.1$, $\lambda = 0$, and $x = 0.15$ (*a*) and for the two parameters of the spin-phonon interaction $g_s = 0.1$, $x = 0.02$ (1), $x = 0.05$ (2), $g_s = 0.2$, and $x = 0.05$ (3) with $J_1 = 10$, $J_2 = 6$, $g = 6$, $k = 20$, $b = 3$, $K = -0.35$, and $\lambda = 0.1$ (*b*)



a *b*

Fig. 3. Total magnetization ($M_1 + M_2$) of the spin moments (*a*) and tetragonal distortion $<x–y>$ of the lattice (*b*) versus temperature for the parameters of the spin-orbital interaction λ = 0.1 (1), 0.2 (2), and 0.3 (3) with $J_1 = 10$, $J_2 = 5$, $g = 6$, $k = 20$, $b = 3$, $K = –0.35$, $g_s = 0.1$, and $x = 0.05$. The insert shows the temperature dependence of lattice parameters $<x>$ (1) and $<y>$ (2) for λ = 0.3

The change in the electron density at e_g and t_{2g} states induces electron polarization of an ion determined by the polarizability of an atom $\alpha = \alpha_n + 2b_n\, [M^2_J – 1/3\, J(J + 1)]$, where α_n and b_n are the constants, J is the total moment of an atom, M_J is the projection of the moment onto a selected direction. Permittivity is related to polarizability as $\varepsilon = 1 + 4\pi N\alpha$ and the change in permittivity in the external magnetic and electric fields is determined as $\Delta\varepsilon \sim \Delta\alpha \sim \Delta\, M^2_J \sim \Delta < L_1L_2 >$. Reconstruction of orbital structure by the magnetic or electric fields will manifest itself in the variation of the correlation functions. In fig. 4, one can see the difference in the correlators of the orbital magnetic moments calculated with and without the magnetic field for several concentrations.

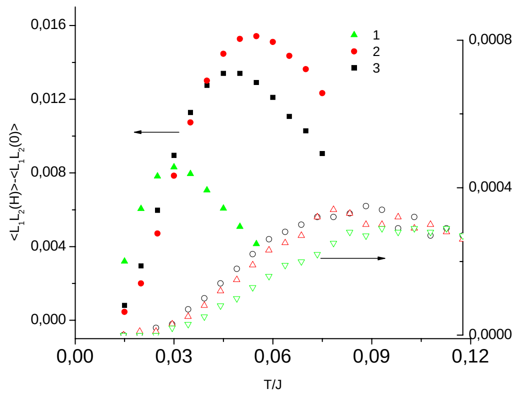


Fig. 4. Difference between the correlation functions of the orbital magnetic moments for the Mn-Co-Mn clusters (left scale) and manganese ions (right scale) calculated with and without magnetic field versus temperature for $J_1 = 10$, $J_2 = 5$, $g = 6$, $k = 20$, $b = 3$, $K = –0.35$, $g_s = 0.1$, λ = 0.1, and $x = 0.02$ (1), 0.05 (2), and 0.15 (3)

The first maximum in the low-temperature region is caused by the change in the orbital correlation function in the Mn-Co cluster; the second maximum is related to the growth of the orbital order in the manganese system. This qualitatively explains the presence of two maxima in the ME effect. A similar behavior is observed in the external electric field due to the change in the electron density on the orbitals caused by the dependence on the potential energy of an electron at a distance in the external uniform electric field. The substantial contribution to permittivity is made also by the Jan-Teller ion displacement, which is not considered in this model.

The competition of the Coulomb interactions between the electrons located on one orbital and on different orbitals along with the change in the hopping integrals causes ordering the electrons on the certain orbitals and orbital magnetism. Due to the redistribution of the electron density, the elastic energy changes and coupled modes of the ion oscillations are induced under the action of the electron-phonon interaction. The magnetoelectric effect magnitude is estimated.

References

1. Epitaxial $BiFeO_3$ Multiferroic Thin Film Heterostructures / J. Wang [et al.] // Science. 2003. Vol. 209. P. 1719–1728.

2. Zvesdin A. K., Pyatakov A. P. Phase transition and colossal magnetoelectric effect // UFN. 2004. Vol. 174. P. 465–468.

3. Smirnov A. I., Hlusticov I. N. Magnetoelectric effect and Stark effect // UFN. 1995. Vol. 165. P. 1215–1219.

4. Transport Properties and Ferromagnetism of $Co_xMn_{1–x}S$ Sulfides / S. S. Aplesnin [et al.] // JETF. 2008. Vol. 106, № 4. P. 765–772.

5. Magnetoelectric effect in $Co_xMn_{1–x}S$ / S. S. Aplesnin [et al.] // Vestnic of SibSAU. 2009. Vol. 1(22). P. 41–45.

6. Spin Glass Effects in $Co_xMn_{1–x}S$ Solid Solutions / S. S. Aplesnin [et al.] // Bulletin of the Russian Academy of Sciences: Physics. 2009. Vol. 73, № 7. P. 965–967.

O. V. Aripova, A. N. Guschin
Baltic State Technical University "Voenmeh" named after D. F. Ustinov,
Russia, Saint Petersburg

METHODS OF AGENTS CONTROL IN MULTI-AGENT EXPERT SYSTEM

In the article we discuss the user models interacting with distributed network resource, where every user has a correlated agent, as well as on the methods of multi-agent expert system control.

Keywords: expert system, the agent, interaction, behavior, control.

The administration of distributed network resources requires solving tasks, concerned with the complexity of resource-to-user interaction organization.

To solve the set task [1] the following multi-agent expert system (fig. 1) has been developed.

To organize interaction between users and distributed network resource on a basis of the developed multi-agent expert system model we will consider the following models of agent's behavior coordination [2–6]:

1. Game-theory modes – solve the tasks of selection solutions in conditions of equivocality and conflict, which if followed, allow the constructing of rule sets and conversations, permitting agents to achieve equilibrium agreements.
2. Models of collective behavior for automats – are based on constructing conversation rules and protocols in tasks, which are characterized by a large quantity of simple interactions with indeterminate characteristics.
3. Models of collective behavior planning reveal methods of agent behavior planning (centralized, partially centralized, distributed) for the purpose of making decisions regarding the selection of self-actions in the process of implementing the plans' coordination.
4. Models based on *BDI*-architecture – apply axiomatical methods of the game theory and the artificial intellect's logical paradigm. The task of the agents' coordination behavior consists in coordinating the output results in the knowledge bases of these agents, obtained for the current state of external environment.
5. Models based on competition – imply the "auction" concept as a mechanism of the agents' behavior coordination. The concept is based on a postulate regarding the possibility of undisguised transfer of "usability" from one agent to another, or to the agent-auctioneer.

On the interaction assumption with distributed network resource, the following subpopulations were segregated from the full set of users U: A – administrators subpopulation (fig. 2), E – experts subpopulation, RU – registered users (fig. 3), GU – guest users (fig. 4). Since data regarding candidate users is formed on the base of processing information received from experts, who also have a right to influence all aspects of expert system functioning, the following assumption was made: the authentication of subpopulation U by agents is unviable. We shall build the behavior models for segregated users.

The following agent types were associated with all the user classes: for each guest user GU_i – reactive agent $RAGU_i$ (fig. 5), for each registered user RU_i – reactive agent $RARU_i$ (fig. 6), for each user-administrator A_i – intellectual agent IAA_i (fig. 7).

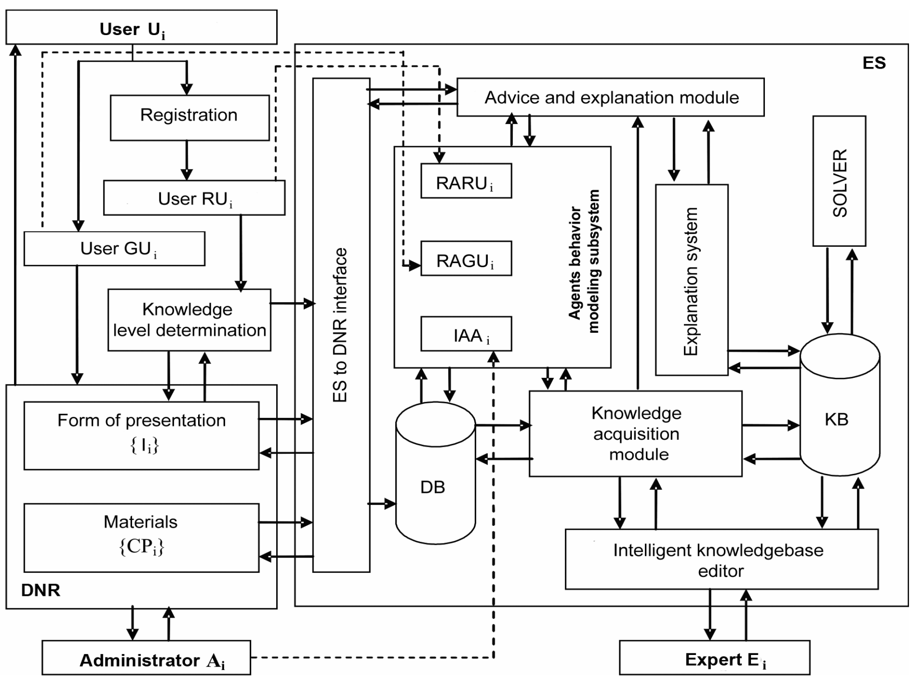


Fig. 1. Model of a multi-agent expert system

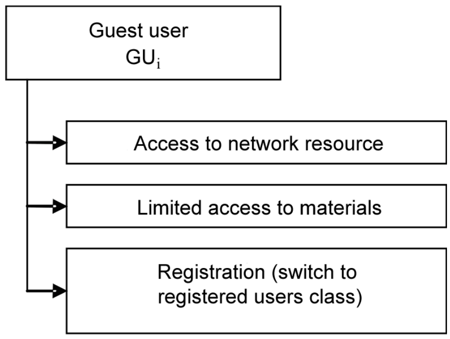


Fig. 2. Guest user behavior model

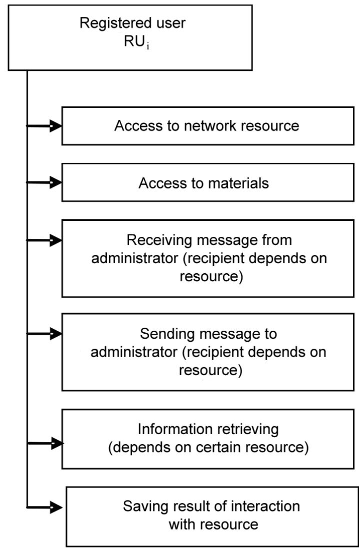


Fig. 3. Registered user behavior model

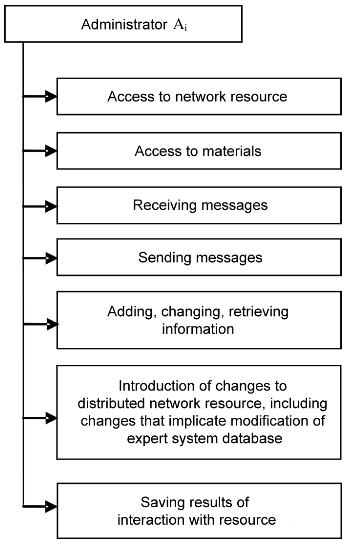


Fig. 4. Administrator behavior model

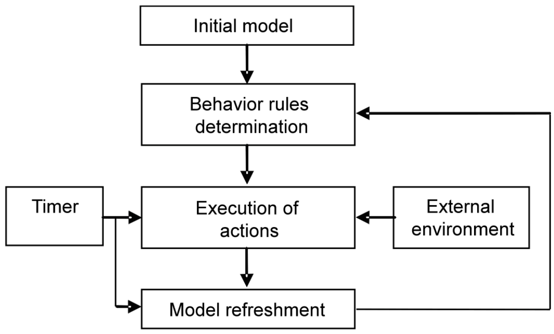


Fig. 5. Model of guest user agent “life cycle”

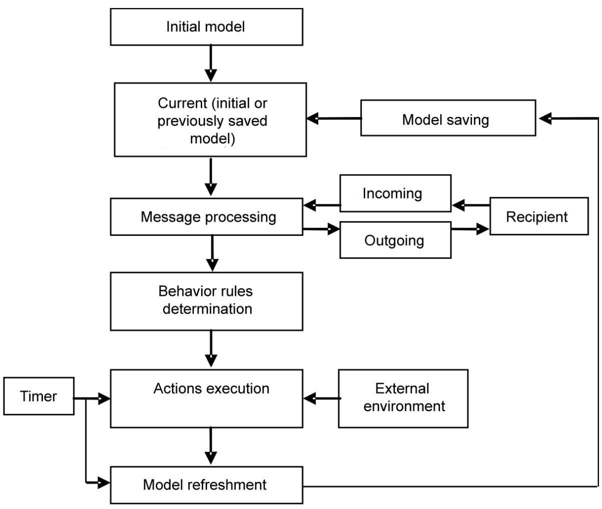


Fig. 6. Registered user agent “life cycle” model

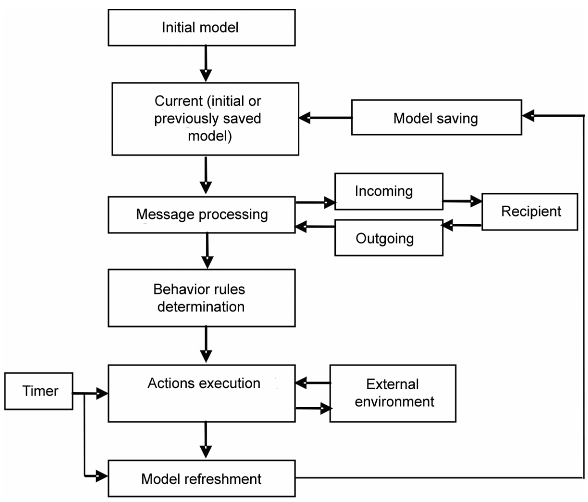


Fig. 7. Administrator agent “life cycle” model

In addition to the interaction with distributed network resource it is necessary to organize relations between existing agents, whose main features are directedness, selectiveness, intensity, and dynamism. It is essential to consider the finiteness of existence time for every agent, and the influence of the agent's behavior within the applied materials, located in a distributed network resource.

We shall consider requirements for agents [2; 3], correlated to each users subpopulation (see table).

Requirements for user agents

Characteristics	Agent type		
	$RAGU_i$	$RARU_i$	IAA_i
Independent execution	+	+	+
Interaction with other agents	+	+	+
Reactivity	+	+	+
Adaptive behavior		+	+
Education based on interaction with environment		+	+
Tolerance towards mistakes and/or improper input signals			+
Real-time functioning			+
Life cycle finiteness	+	+	+
Persistence of behavior in expert system DB		+	+

Due to the fact that reactive agents do not possess complex knowledge of the environment and are fully dependent on the objective (used to compose reactions for the designed situations and are also independent and have finitesimal life cycle), we can consider the definition of their functioning based on the production systems with an imperative kernel. In this case every agent possesses a rule set $R = \{r_1, ..., r_i, ..., r_n\}$, which has the following structure:

$$r_i : p_i; a_i \to b_i,$$

here $a \to b$ – is the kernel, which is the basic component of production and is interpreted as "if *a*, then *b*", *a* – is inference *b* existence conditions ; *i* – name, which allows to identify the present production from the production set; p_i – constraint of kernel production usability (predicate): if *p* – true, the kernel is actuated.

Let us assume that for the reactive agent:

$$a_i = \vee_j \left(\wedge_k a_{ijk} \right),$$

where a_{ijk} – explicit predicate over state *S* of agent $a_{ijk} = a_{ijk}(S)$, $S = \{s_i, ..., s_j\}$, s_i, s_j – state parameters.

Example:

– WHEN user = registered user;

– AND logon to distributed network resource;

– IF unread messages are in presence;

– THEN user is invited to review messages.

Thereat the intellectual agent requires a change in the structure of rules, videlicet the replacement of the kernel with an optional one, having an estimation of implementation on the base of fuzzy logic and introducing the following tail conditions to the structure:

$$r_i : p_i; a_i \to b_i; n_i,$$

here n_i – are tail conditions for *i* – is production.

Let us assume that for the intellectual agent:

$$a_i(S) = \vee_j \left(\wedge_k a_{jk}(S) \right) \Rightarrow f_{a_i}(S) = \max_j \left(\min_k f_{a_{ijk}}(S) \right),$$

where *f*(*S*) – is the membership function, dependent on predicable truth, as follows: $f_{a_i}(S) = 0 \to b_i$ – false, $f_{a_i}(S) = 1 \to b_i$ – true. For definiteness of change in the agent state we shall introduce $f_{cutoffi}$ for which $f_{a_i}(S) > f_{cutoff_i} \to b_i$ – is true and the tail condition n_i is fulfilled.

Example:

– IF unread messages are in presence;

– AND these are not urgent;

– AND NOT (administrator is busy);

– WHEN with f_{cutoff} = 0,75: offer to show messages to the administrator;

– TAIL CONDITION = show messages to the administrator;

– IF unread messages are in presence;

– AND these are urgent;

– AND NOT (administrator is busy);

– WHEN with f_{cutoff} = 0,1 show messages to the administrator;

– TAIL CONDITION = show messages to the administrator.

Such requirements allow coordinating of the agents' behavior using a model based on *BDI*-architecture, since such concepts as beliefs, desires, and intentions are accentuated. Logical inference in knowledge bases is done directly in the process of agent functioning.

Multi-agent expert system, based on a hybrid architecture, with the use of two types of agents: intellectual, for the administrator subpopulation and reactive, for the subpopulations of registered and guest users, should implement reflexive control over distributed a network resource, the task of which is to make users consciously fall under external influence (i. e. to externalize such desires and intentions, which correspond to the environment).

We understand reflexive control as:

– the actualization of socially significant demands;

– the generation of a set of unique administrative concepts and relations;

– the realization of an activity set, particularly by means of engaging specialists (experts) to take part in the concept interpretation.

The expert system itself is based on the production model with fuzzy logic [7], where the agents' behavior modeling subsystem is used as a potential argument for predicates in antecedent rules, (the observed agents' behavior can be used in the rules of the expert system, but it can only be influenced by environmental change).

Let us assume that S_a – is the agent's state, S_v – is the observed agents state, S_e – is the expert system state, S_r – is the distributed network resource state;

$R_e = \{r_{e_1}, ..., r_{e_i}, ..., r_{e_n}\}$ – is the set of rules possessed by an expert system of type $r_{e_i} : p_{e_i}; a_{e_i} \rightarrow b_{e_i}; n_{e_i}$, where $a_{ei}(S_e)$ – is the existence condition for inference b_{ei}, moreover S_e includes S_v ($S_v = \{S_{v_1}, S_{p_{v2}}, ..., S_{v_i}, ..., S_{v_n}\}$), $S_v \subset S_a$ ($S_v \subset S_e$), b_i – are the results of redefinition of S_e, which again can result in redefinition of S_r. As far as S_r can cause change of the user's behavior (this changes environment influencing of the agent's behavior), which, again is observed by the expert system; videlicet behavior of user at a moment of time t_1 (beginning of the expert system operation) results in changes of distributed network resource at a moment in time t_2 and in possible subsequent behavior of user at a moment in time t_3 and so on (Пд – behavior):

$$S_r^{t_n} = \left(S_r^{t_{n-1}}, S_e^{t_{n-1}} \text{Пд}^{t_{n-1}} \left(S_r^{t_{n-2}} \left(S_r^{t_{n-3}}, S_e^{t_{n-3}}, \text{Пд}^{t_{n-3}} \times \right.\right.\right.$$
$$\left.\left.\left. \times \left(S_r^{t_{n-4}} \left(... \left(S_r^{t_2}, S_e^{t_2} \text{Пд}^{t_2} \left(S_r^{t_1} \right) \right) \right) \right) \right) \right) \right).$$

A model of interaction between distributed network resource and users has now been developed. Its main peculiarity is the use of the multi-agent expert system. Methods of agent management allow the rationalization of the present interaction by means of the user's behavior reflexive control via dynamic modifications in the representation and the content of network resource.

On the base of educational the following resource (www.i5nfo.ru) the present methods are applicable for tasks of interaction between organizations, distributed network resources, and the user. They have also been applied to instruments measuring and monitoring radiation levels, which had been affirmed by implementation acts.

References

1. Aripova O. Models of Interaction between User and Distributed Network Resource: Research and information magazine. Innovations. SPb. : JSC "TRANSFER", 2009.
2. Andreychikov A., Andreychikova O. Intelligent Information Systems : the Textbook. M. : Finance and statistics, 2006. P. 424.
3. Gavrilova T., Khoroshevskiy V. Intelligent Systems Knowledgebases. SPb. : Piter, 2001.
4. Gushin A. Basic Concepts of Personal-Centered Information Systems Development / Voenmeh. Baltic State Technical University Bull. SPb. : "Sot" printing establishment, 2008. P. 34–44.
5. Laurier J.-L. Artificial Intelligence Systems: transl. from the French. M. : Mir, 1991. P. 568.
6. Rassel S., Norwig P. Artificial Intelligence: Modern Approach. M. : Williams, 2006. P. 1048.
7. Yasnitskiy L. Artificial Intelligence Guide-book : Study guide for students of inst. of tertiary education. M. : Academia, 2005. P. 176.

V. V. Bukhtoyarov, E. S. Semenkin
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

A COMPREHENSIVE EVOLUTIONARY APPROACH FOR NEURAL NETWORK ENSEMBLES AUTOMATIC DESIGN

A new comprehensive approach for neural network ensembles design is proposed. It consists of a method of neural networks automatic design and a method of automatic formation of an ensemble solution on the basis of separate neural networks solutions. It is demonstrated that the proposed approach is not less effective than a number of other approaches for neural network ensembles design.

Keywords: neural networks, ensemble, automatic design, genetic programming, probabilistic evolutionary algorithm.

At the present time data analysis systems which are based on intelligent information technologies are increasingly demanded in many fields of human activity and the scale requirements to these systems are continuously increasing. In connection with these facts the problem of developing methods for automatic design and adaptation of IIT for specific tasks is becoming more urgent. Such methods could allow to abandon the use of expensive, mostly human, resources for the design of the IIT and to reduce the time required for the development of intelligent systems.

One of the most widely used and popular intellectual technologies are artificial neural networks. The range of problems solved by using neural networks is extremely large because of many advantages of systems based on their use. Despite the fact that this information technology could be called a universal tool for solving problems of data analysis, in each case we have to create a unique neural network. One of the approaches to improve the efficiency of systems based on the use of neural networks is the use of neural networks ensembles. Problem solving with the help of neural network ensembles supposes simultaneous use of a finite number of preliminarily trained neural networks.

This approach was first proposed in [1], where it has been shown that the ability of a neural network system to generalize can be significantly improved by uniting a number of component neural networks into an ensemble.

In general, neural network ensemble formation consists of two steps. For the use in data analysis systems it is desirable to automate both of these steps. The first step is to form the structure and to train a number of component networks which will be included in an ensemble or a preliminary pool. The second step is to select the networks, the solutions of which will produce a final solution, and to define a way and parameters of a common solution formation.

The approach based on preliminary choice and fixation of neural networks' structure and their training on different learning sets formed from an original learning sample is often used to perform the first stage of a neural networks ensemble formation. In some cases in the process of training of each subsequent neural network special attention is paid to its training on those subsets of a learning set on which the mistake of all previous neural networks was big. Many other approaches for forming and training of component networks are also widespread. Neural networks with different number of hidden neurons are offered to use in work [2]. Usage of different object functions for training of each component network is advised in work [3]. The main disadvantage of such approaches is the need for a priori fixed structure of component networks which could adversely affect the adaptability of the resulting collective solution.

Other promising approaches developed in different works are based on the use of genetic algorithms (GA) [4] to form component neural networks. The main difficulty in these approaches is the necessity to tune a great number of GA parameters. Besides, it is often very difficult to choose proper settings for GA, because it requires a lot of time and computational efforts. At the same time a poorly tuned GA may fail to solve the task.

The analysis of existing methods for automatic design of neural networks (in particular the method based on the use of genetic algorithms [4]) shows that methods accumulate and process statistical information about the structure of the designed neural network. However this information is not used explicitly. Instead information is collected and processed in an implicit form by operators of a specific method, the type of which depends on the setting of relevant parameters. But if we suggest a method of processing this information in an explicit form, it would allow to avoid tuning a great number of genetic algorithm parameters, which is very difficult in solving real practical problems. In connection with this we propose a new probabilistic method for automatic generation of neural networks structures. This method uses information-processing principles firstly used in the operators of probabilistic genetic algorithm [5].

The proposed method is based on the computation and use of estimated probabilities $p^k_{i,j}$, where $i=\overline{1,N_l}$ is the number of a hidden layer, N_l is a maximum number of hidden layers, $j=\overline{1,N_n}$ is a number of a neuron on a hidden layer, N_n is a maximum number of neurons on a hidden layer, $k=\overline{0,N_F}$, where N_F is the cardinality of a set of activation functions which can be used for neural network structure formation. If $k\in[1;N_F]$ it corresponds to a number of an activation function in a neuron, if $k=0$ it means the absence of j-th neuron on i-th hidden layer.

A full multi-layer perceptron with the number of hidden layers equal to N_l and the number of neurons in each hidden layer equal to N_n is used as the most complete (in terms of a number of layers and a number of neurons in layers) neural network architecture. This architecture allows to indicate all the possible positions of a neuron in the network clearly with the help of the hidden layer number and numbers of neurons on the layer.

We propose to present solutions as a string S of integers from the interval $[0;N_F]$ with the length equal to $L=N_l\cdot N_n$.

The number of elements of this string is equal to the maximum possible number of neurons on hidden layers of a "full" multilayer perceptron. It is clear that the identification number of each element of the string can be evaluated using the following formula:

$$r=N_n\cdot(i-1)j,\quad i=\overline{1,N_l},\quad j=\overline{1,N_n} \tag{1}$$

and it defines the place of a neutron in the structure of an artificial neural network with the architecture of a "full" perceptron.

So each element of the string S can be interpreted in the following way:

– If $S_r=0$, then j-th neuron on i-th layer does not exist;

– If $S_r\neq 0$, then j-th neuron on i-th layer exists and its activation function identifier is equal to S_r.

A neural network with any configuration of layers and neurons on them, the dimensions of which do not exceed N_l and N_n mentioned for the architecture of a multilayer perceptron, can be described by such a string.

The general scheme of the proposed method for automatic formation of neural networks structure is described below:

1. Execute initialization steps.

2. Form N vectors $S^{o,i}, i=\overline{1,N}$ using random generator, set iteration counter $k=0$.

3. On k-th step of search to evaluate an objective function for each network presented by $S^{k,i}, i=\overline{1,N}$, establish an intermediate set $S'=\varnothing$.

4. Select N_{par} individuals from a current set $S^k=\{S^{k,1},S^{k,2},...,S^{k,N}\}$ by applying a rank selection procedure and place them into an intermediate set S'.

5. Using the solutions from S' calculate a set of estimated probabilities:

$$\overline{P} = \left\{ p_{i,j}^l \right\} = \left\{ p_r^l \right\}, \quad r = \overline{1, L}, \quad l = \overline{1, N_F} .$$

6. According to the estimated probabilities obtained on the previous step generate a temporary set $\overline{S^k}$, which consists of N new solutions.

7. Apply a mutation operator to each solution of $\overline{S^k}$.

8. Create a new set of solutions S^{k+1} from sets S^k and $\overline{S^k}$.

If a termination criterion is satisfied the best found solution is the result. Otherwise set $k = k + 1$, and go to step 3.

Initialization steps include determination of a maximum number of hidden layers of a neuron network N_l , determination of a maximum number of neurons on each layer of a neuron network N_n , specifying of activation functions set F and a maximum number of steps of N-step structure formation as well as a number of networks looked through at each step of formation of structure N.

The second step implies random generation of a string which represents the initial solutions. In order to generate originally simpler structures of neural networks in this method the following probabilities of the random-number generator for initialization of strings were established:

$$S_r^{1,i} = \begin{cases} 0 \text{ with probability equal to } 0.5 \\ s \in \left[1;\, N_F\right] \text{ with probability equal to } 0.5/N_F \end{cases} .$$

Here $r = \overline{1, L}$, $i = \overline{1, N}$. The key steps of our method are step number 5 and step number 6. Considering equation (1) we get the following formula for probabilities evaluation for a fully connected neural network (it is used on step 5):

$$p_i^k = \frac{\sum\limits_{r=N_n(i-1)}^{N_n i} \left| S_r^l : S_r^l = k,\ l = \overline{1, N_{par}} \right|}{N_n \cdot N_{par}},$$

$$p_{i,j}^k = p_i^k, \quad k = \overline{0, N_F}, \quad i = 1, N_l, \quad j = \overline{1, N_n} .$$

Here N_n is a maximum available number of neurons on a hidden layer, N_l is a number of hidden layers, N_{par} is a number of solutions in an intermediate parents' subset, $|\cdot|$ is a set cardinality.

An intermediate set $\overline{S^k}$ is generated according to the probabilities specified above on the sixth step. Accumulation and processing of statistical information about the structures of neural networks in an explicit form are implemented in steps 5 and 6 of the proposed method.

As it was mentioned above the comprehensive approach for neural network ensemble design consists of two main steps. The first step includes design and training of component neural networks. We propose probabilistic approach described above for this step.

The second step includes selection of networks which will be used for collective problem solving and finding out the way according to which the solutions of individuals will be taken into account in a collective solution. The most popular approaches of combining solutions of different neural networks are plurality voting or majority voting for classification problems [1] and simple or weighted averaging for regression problems [6]. The most developed methods are those with weighted averaging and majority voting. For example to evaluate the weighting coefficients of separate neural networks' contributions into the general solution Jimenez [7] uses the evaluation of the quality of their individual solutions. Zhou [8] uses genetic algorithm to find proper weights for each member of an ensemble.

To increase the efficiency of this stage implementation we developed another approach. This method allows you to automatically choose those neural networks that will participate in ensemble solution finding. It also forms an ensemble solution in the form of various transformations and combinations (linear and nonlinear) of individual neural networks solutions. We suppose that, while using the ensemble of neural networks we can find a more efficient solution with the help of formation of more complex combinations of individual neural networks solutions than the simple or weighted averaging and plural or majority voting.

The proposed approach is based on the genetic programming method [9], which is used for solving the tasks of symbolic regression. The basic idea of the proposed method is to adapt and use the operators of genetic programming to automatically generate a solution, which is a formula (a program). This program calculates a general solution of a neural networks ensemble on the basis of solutions received from its individual members.

To devise a solution the proposed method uses the elements of a terminal and a functional sets. Instead of independent variables of the problem (as in genetic programming) answers given by neural networks from a preliminary pool are used as a terminal set T . It includes therms of which solutions will be formed. A terminal set also includes numeric literals tuning of which allows to get a more adaptive general solution of a neural networks ensemble. A functional set F consists of different operations and functions which specify the dependency between an ensemble solution and solutions of individual neural networks.

The general scheme of the proposed method is presented below and basically coincides with the steps of the method of genetic programming used to approximate functions.

1. Initialization of the initial generation.

2. Execute the following steps until the termination criterion is satisfied:

2.1. Calculate fitness for each individual of the current generation.

2.2. With the probability connected with the found value of fitness select one or several individuals for crossover by applying a selection operator.

2.3. Create a new generation of individuals-programs by execution of the following genetic operators:

2.3.1. Create a new individual by applying a crossover operator.

2.3.2. Modify a new individual with a mutation operator.

2.3.3. Copy some initial individual into a new generation.

3. After the termination criterion is satisfied the best found solution is declared as a problem solution.

Solutions are encoded in the form of trees. Outer vertexes are formed from elements of the terminal set and the inside vertexes are formed from the functional elements of the set. Arithmetic operations, mathematical functions, Boolean operations can be used as the functional elements of the set.

An ensemble solution generated by our method is a function input whose parameters are individual solutions of neural networks included in an ensemble:

$$o = f\left(o_1, o_2, ..., o_n\right).$$

Where o is the ensemble solution; o_i is i-th individual neural network solution; n is the number of networks in an ensemble (or a preliminary pool).

The proposed method allows to improve the flexibility of the system based on the use of neural networks ensembles, due to the lack of firmly fixed structure of the interaction between the individual networks. The proposed method not only forms the structure of the interaction between ensemble members, but also indirectly (through inclusion or non-inclusion of the relevant arguments into the formula of general solution) selects those neural network solutions which will be most useful in terms of the solution effectiveness. The use of additional algorithms (e. g. a genetic algorithm) to adjust the parameters of interaction models can further improve the model efficiency.

A number of numerical experiments were carried out on a set of test problems in order to study the effectiveness of the proposed comprehensive approach for neural network ensembles designing. Our approach was compared with three other approaches:

1. GASEN approach [7] – an approach which uses a genetic algorithm to specify weights. Further on while forming an ensemble solution it takes into account those neural networks whose weighting coefficient is higher than some preset value.

2. An approach based on the usage of GA for selection of a fixed number of component networks from a preliminary pool for joint use and specifying their weights (it will be called GA-based 1 below).

3. An approach based on the usage of GA for selection of an arbitrary (non-fixed in advance) number of component networks from a preliminary pool (it will be called GA-based 2 below).

A standard method for neural networks design based on a genetic algorithm with preselected setup of parameters was used to generate a preliminary pool of neural networks in these methods. For neural networks the maximum number of hidden layers was set equal to three, the maximum number of neurons on each layer is equal to five. During the preliminary investigation it was found that the increase of maximum number of hidden layers of the network and the maximum number of neurons on them does not give additional benefits to the formed neural network model for the problem.

To conduct research we also used the data set "Concrete Slump Data Set" from UCI Machine Learning Repository [10; 11]. The data set includes 103 data points describing the dependence between concrete composition and measured figures which characterize deformation and strength of products made of it. There are 7 input variables and 3 output variables in this data set. Examples of other test problems used in comparative studies are listed below (tab. 1).

Table 1

Examples of test problems

Problem	Function	Range of input variables	Sample size
1	$y = \sin x$	$x \in [-4, 3]$	150
2	$y = x_1^2 \sin x_1 + x_2^2 \sin x_2$	$x_i \in [-4, 3]$	150
3	$y = \frac{x_1 \cdot x_2}{x_3^2}$	$x_i \in [1, 20]$	200
4	$y = 100\left(x_2 - x_1^2\right)^2 - \left(1 - x_1\right)^2$	$x_i \in [-2, 3]$	200

The main efficiency criterion are estimation of expectation and estimated variance of error calculated after 50 independent runs of algorithms. We use the following formula to calculate an approximation error in each run:

$$\text{Error} = \frac{100\ \%}{s\left(y^{\max} - y^{\min}\right)} \sum_{i=1}^{s} \left|o_i - y_i\right|.$$

Where i is the record number in the sampling; o_i is the output of an ensemble or a single network; y^i is a real value of an output variable in the sampling; $y^{\max}$ and $y^{\min}$ are maximum and minimum values of an output variable, s is a number of instances in the sampling.

The results of a comparative study of the proposed probabilistic method for automatic design of the structure of neural networks and the method using the genetic algorithm (GA) to solve this problem are presented in tab. 2.

The results of a comparative study of the proposed integrated approach for neural networks ensemble design and other methods described above are presented in tab. 3–4.

An example of a formula describing the method of calculating the ensemble solution for Concrete Slump Test problem on the basis of solutions of individual neural networks is given below:

$$o = o_4 - \frac{\left(o_{20} - 1.953 o_4 + o_{18}\right)}{\left(o_9 / \left(o_4 + o_{18}\right) - 2.824\right)},$$

Where o is a solution of a neural networks ensemble; o_4, o_9, o_{18}, o_{20} are escapes of the fourth, ninth, eighteenth and twentieth networks from a pool. These networks were automatically chosen to form an ensemble. Network 18 allows to achieve a minimum error (4.92 %). A maximum error (4.98) corresponds to network 20. An error of a neural networks ensemble is 3.57 %.

Table 2

The results of a comarative study of methods for neural networks automatic design

Problem	Probabilistic method Estimation of an error expectation, %	GA Estimation of an error expectation, %
1	1,880	1,857
2	4,355	4,428
3	0,736	0,780
4	6,750	6,643

Let us analyze the characteristics of a neural networks ensemble design for Concrete Slump Test problem. For comparison we analyze the characteristics of the networks included in the preliminary pool and networks selected for an ensemble. Statistical data were obtained by the results of twenty runs of a software system which implements the proposed method for neural networks automatic design.

There were two hidden layers with two neurons on each layer in the minimum size network included in the preliminary pool. A maximum size network included in the preliminary pool was characterized by the structure of hidden layers 5–4–4, that is, there were 13 neurons on the hidden layers of the network: 5 neurons on the first hidden layer, 4 neurons on the second and third hidden layers. The average number of neurons on the hidden layers of networks included in the preliminary pool is 9.

The average number of networks selected for an ensemble from the 20 networks of a preliminary pool is equal to 5 (the exact value is equal to 4.75). The average number of neurons in such networks is equal to 7.

For the problem of predicting the strength characteristics of concrete the mean of error obtained for neural networks in the preliminary pool is equal to 4,95 %. The estimated mean of error for the averaged output of all the networks from a preliminary pool is equal to 4,58 %. The average error of a neural networks ensemble designed with the proposed method is equal to 3,52 %. It is important to note that an ensemble usually includes not only networks with the lowest individual modeling error but also those neural networks which allow to minimize the total value of errors which characterizes the quality of an ensemble general solution.

The results of statistical studies show that the proposed comprehensive evolutionary approach for neural network ensembles design demonstrates high efficiency in all used test problems. As far as the first phase of the neural network ensemble design based on the results of comparative research is concerned we can conclude that the probabilistic method of forming neural networks structures is not less effective than the commonly used method based on a genetic algorithm. An advantage of this method is the smaller number of algorithm parameters to be set up for its effective work. It is very important for complex tasks due to saving significant time and computational resources necessary to adjust the settings.

The results shows that the proposed approach for neural networks automatic design consisting of a probabilistic method for individual neural networks design and a method for forming an ensemble solution demonstrates high efficiency and it is not inferior to other approaches on the used test problems. The relative superiority of the proposed method over other methods of forming a common solution in neural networks ensembles in terms of an average modeling error is within 50 % for test problem 1 to 13 % per cent for test problem 2. The average superiority of this indicator is about 25 %. For the problem of predicting the concrete strength characteristics relative superiority of the evolutionary method was about 15 %, which can be considered a fairly good result in the processing of real data.

Table 3

Results of comparative study on test problems

Approach	Test problem 1		Test problem 2	
	Estimation of an error expectation, %	Estimated error variance, %	Estimation of an error expectation, %	Estimated error variance, %
Single neural network	1.880	0.510	4.355	0.736
GASEN	1.444	0.112	3.479	0.026
GA based 1	1.335	0.223	3.486	0.028
GA based 2	1.302	0.162	3.482	0.024
Proposed approach	0.855	0.065	3.037	0.003

Approach	Test problem 3		Test problem 4	
	Estimation of an error expectation, %	Estimated error variance, %	Estimation of an error expectation, %	Estimated error variance,%
Single neural network	2.537	0.245	6.643	1.447
GASEN	1.679	0.016	6.192	0.163
GA based 1	1.651	0.019	6.147	0.178
GA based 2	1.639	0.020	6.100	0.209
Proposed approach	1.389	0.035	5.036	0.115

Table 4

Results of comparative study on the problem of concrete strength characteristics prediction

Approach	Estimation of an error expectation, %	Estimated error variance, %
GASEN	4.119	0.040
GA based 1	4.113	0.047
GA based 2	4.012	0.036
Proposed approach	3.521	0.028

The results allow us to draw a conclusion that the proposed approach can effectively automatically generate neural networks ensembles and calculate the final solution of the ensemble. Thus, this integrated approach should be used to improve the effectiveness of the solution of complex practical problems which were traditionally solved by systems using a single artificial neural network, for example, complex problems of approximation and prediction.

References

1. Hansen L. K., Salamon P. Neural network ensembles // IEEE Trans. Pattern Analysis and Machine Intelligence. 1990. № 12 (10). P. 993–1001.

2. Cherkauer K. J. Human expert level performance on a scientific image analysis task by a system using combined artificial neural networks // Proc. AAAI-96 Workshop on Integrating Multiple Learned Models for Improving and Scaling Machine Learning Algorithms : P. Chan, S. Stolfo, D. Wolpert (eds.). Portland, OR : AAAI Press ; Menlo Park, CA. 1996. P. 15–21.

3. Hampshire J., Waibel A. A novel objective function for improved phoneme recognition using timedelay neural networks // IEEE Transactions on Neural Networks. 1990. № 1 (2). P. 216–228.

4. Goldberg D. E. Genetic algorithms in search, optimization and machine learning. Reading. MA: Addison-Wesley, 1989.

5. Semenkin E. S., Sopov E. A. Probabilities-based evolutionary algorithms of complex systems optimization. In Intelligent Systems (AIS'05) and Intelligent CAD (CAD-2005). M. : FIZMATLIT, 2005. Vol. 1. P. 77–79.

6. Perrone M. P., Cooper L. N. When networks disagree: ensemble method for neural networks // Artificial Neural Networks for Speech and Vision ; R. J. Mammone (ed.). N. Y. : Chapman & Hall, 1993. P. 126–142.

7. Jimenez D. Dynamically weighted ensemble neural networks for classification // Proc. IJCNN-98. Vol. 1. Anchorage: AK. IEEE Computer Society Press; Los Alamitos, CA. 1998. P. 753–756.

8. Zhou Z. H., Wu J., Tang W. Ensembling neural networks: Many could be better than all // Artif. Intell. 2002. Vol. 137, № 1–2. P. 239–263.

9. Koza J. R. The Genetic Programming Paradigm: Genetically Breeding Populations of Computer Programs to Solve Problems. Cambridge, MA : MIT Press, 1992.

10. Yeh I.-C. Modeling slump flow of concrete using second-order regressions and artificial neural networks // Cement and Concrete Composites. 2007. Vol. 29. № 6. P. 474–480.

11. URL: http://www.archive.ics.uci.edu/ml/datasets/Concrete+Slump+Test

M. V. Damov
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

BACKGROUND RESTORATION IN FRAME AREAS WITH SMALL-SIZE OBJECTS IN VIDEO SEQUENCES

A general concept of removal of artificially overlaid images, natural damages of video images and other small-size objects is presented. The classification of artificially overlaid images is developed. The algorithms of feature points detection and feature points tracking used in video sequence restoration are considered.

Keywords: movement tracking, video sequence, feature points, texture, texture filling.

The task of restoration of video sequences is getting more and more actual as a result of computer engineering development. The restoration of the original images under artificially overlaid object (TV channel logotypes, subtitles etc.) and other small-size objects such as a man, a tree, a stone etc. on a certain background as well as the images distorted as a result of damage of information carrier (scratches on the film etc.) is of primary importance. The solution of this problem in general will result in reduction of costs of video reutilization such as old film remastering, original video forwarding by different TV channels with removal of earlier overlaid, but irrelevant now, computer graphics, and accidentally shot objects, for example, advertising structures.

Overlaid computer graphic images occurred in video can be divided into several groups. There are TV channels logotypes that can be defined as small-size images arranged in one or several frame corners or frame borders; the second group is titles, that is, text areas with information about film makers arranged in any place of a frame; the third group is subtitles, which can be defined as text areas near the top or bottom frame borders with periodically changed static text; and finally a creeping line that can be defined as a text area at the top or bottom frame borders with the text moving according to generally accepted rules of reading and writing.

All variety of overlaid computer graphic images can be classified according to different criteria. Let's consider the most frequently used classifications. They are the following: according to their size: small (less than 5 % of the frame), middle (less than 20 % of the frame), huge (less than 35 % of the frame); according to their position: corner; stretched along the horizontal border of the frame; stretched along the vertical border of the frame, in accordance with Substation Alpha standard; or some other kind; according to their action: static (image is always constant), moderately changing (image size is constant), fully dynamic (image changes its size and some other video sequences may be overlaid within this size); according to their duration: always constant in video sequence, periodically absent in video sequence; according to their color: monochrome; black and white; gradient; with the limited number of colours; full-colour; according to their transparency: transparent and opaque; according to the presence of a border line: bordered, borderless; according to the presence of their own background: with background, without background [1].

Usually images distortions as a result of damage of information carrier have stretched geometric structure and can appear in any place of the frame with different angles of inclination. Their presence in several consecutive frames unassociated with the change of scene foreshortening is a unique feature of such structures. Several similar structures can be present in video sequence and each of them is characterized by its own behavior and can overlay others. Indeterminacy and unpredictability of damages appearance make the problem difficult to solve automatically. Only the hypothesis about the stretch of geometric structure, its small size relative to full frame, homogeneous brightness and stability of existence in frame sequence allows to devise a method of allocation and restoration of similar images.

Accidentally shot and unnecessary objects in the frame must be characterized by small size (less than 10 % of the image) relative to the size of frame, also they can be characterized by static position on dynamic background; by dynamic position on static background; by dynamic position on dynamic background.

The right classification of objects to be removed allows to choose the right complex of algorithms for restoration of original video sequence. The general order of restoration of original video sequence is presented below with the detailed description of every step:

Step 1. Determination of characteristics of a video sequence (feature points of a frame, movement vectors in a frame, object and texture movement in a frame).

One of the technologies used to extract structured and intelligent information from video sequence is feature points tracking in image sequence. A feature point is defined as a point in the scene situated on the plane area of the scene surface. Meanwhile, the depiction of environment of this point could be distinguished from the depictions of environment of other points by some particular environment of this point.

Harris detector that computes the value of special corner response function for every pixel is often used for the detection of feature points of an image or a frame. This function evaluates the degree of similarity between the depiction of point environment and the corner. To evaluate this degree matrix is calculated the following way:

$$\mathbf{M} = \begin{bmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{bmatrix},$$

where $I(x, y)$ is image brightness in the point with the coordinates (x, y).

If both eigenvalues are large then even small displacement of the point (x, y) provokes considerable changes in brightness that corresponds to the feature point of the image and corner response function is written as follows:

$$R = \det \mathbf{M} - k(\text{trace }(\mathbf{M}))^2,$$

where $k = 0.04$ (coefficient suggested by Harris); trace ($\mathbf{M}$) is the sum function calculation of matrix elements on the principal diagonal.

The points of image corresponding to local maximum of corner response function are considered to be the feature points.

Let us consider a simple scheme of feature points tracking [2]:

1. Detection and evaluation.

1.1. Detection of a set of feature points {F} in terms of characteristics of feature points such as the degree of corner response function extremity, the position of feature point (in the image center, at the image borders, at the image corners), the position of feature point relative to other feature points or compactness of feature points in some area of the image.

1.2. Detection of quality of all feature points – Q{F}. The feature points with large degree of corner response function extremity, sufficiently remote from the frame border, with low compactness of feature points in the area of the frame of interest are considered to be the most qualitative feature points. Methods of multi-attributive decision making, for example, the method of ordered preference by way of similarity with the ideal decision, can be used for evaluating of the quality of feature points.

1.3. Choice of feature points with the quality higher than some earlier or dynamically defined threshold, and generation of the set of feature points {G}.

2. Tracking and evaluation.

For each next frame:

2.1. Finding of a new position of all feature points from set {G} – tracking in the current frame;

2.2. Estimation of current quality of all the elements of set {G};

2.3. Choice of such feature points only whose quality satisfies a certain criterion. As a rule, this is an integral criterion or a degree of corner response function extremity.

2.4. If the number of tracking feature points becomes less than the required number then a detector of feature points is applied to the current image and a number of new points is added to set {G}.

Modification of Lucas-Kanade algorithms is applied for tracking of feature points coordinates changes [3]. The last modification of Lucas-Kanade algorithms is Favaro-Soatto algorithm that takes into consideration the displacement of feature points, affine distortions of feature points, affine changes of brightness of feature points. The problem of feature points tracking is reduced to the problem of definition of movement parameters and distortion of feature point space. Then the difference is minimized:

$$c = \iint_W \left(J\left(Ax+d\right) - I\left(x\right)\right)^2 w(x)dx,$$

where W is feature point space, and $w(x)$ is weight function which can be equal to one in the whole space, $J(x)$ и $I(x)$ are two images; $Ax + d$ is the displacement of points.

The expression is differentiated relative to the movement parameters, and derivative function is equal to zero. Then the system is linearized by means of image function expansion in Taylor's series:

$$J\left(Ax+d\right) = J(x) + g^{\mathbf{T}}(u).$$

It gives us a linear system of six equations with six unknown quantities:

$$\mathbf{Tz} = \mathbf{a},$$

where all required parameters are merged in vector $\mathbf{z}$:

$$\mathbf{z}^{\mathbf{T}} = [d_{xx} \quad d_{yx} \quad d_{xy} \quad d_{yy} \quad d_x \quad d_y].$$

Error vector **a** is written as:

$$\mathbf{a} = \iint_W \left(I(x) - J(x)\right) \begin{bmatrix} xg_x \\ xg_y \\ yg_x \\ yg_y \\ g_x \\ g_y \end{bmatrix} wdx,$$

and matrix **T** can be presented as:

$$\mathbf{T} = \iint_W \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}^{\mathbf{T}} & \mathbf{Z} \end{bmatrix} wdx,$$

$$\mathbf{U} = \begin{bmatrix} x^2 g_x^2 & x^2 g_x g_y & xyg_x^2 & xyg_x g_y \\ x^2 g_x g_y & x^2 g_y^2 & xyg_x g_y & xyg_y^2 \\ xyg_x^2 & xyg_x g_y & y^2 g_x^2 & y^2 g_x g_y \\ xyg_x g_y & xyg_y^2 & y^2 g_x g_y & y^2 g_y^2 \end{bmatrix},$$

$$\mathbf{V}^{\mathrm{T}} = \begin{bmatrix} xg_x^2 & xg_x g_y & yg_x^2 & y \\ xg_x g_y & xg_y^2 & yg_x g_y & yg_y^2 \end{bmatrix},$$

$$\mathbf{Z} = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix}.$$

The obtained system of equations is resolved also iteratively with the help of Newton–Raphson method.

If movement is considered to be not affine, and just a displacement, then first four elements of the required vector **z** become zero and only the last two remain meaningful. The algorithm is turned into Tomashi-Kanade algorithm.

Let us supplement the algorithm for variable brightness case.

Let the scene surface for which the scene feature point is detected will be Lambert one. Then the intensity of point brightness $x = P(X)$, where X is scene point, P is projection operator, x is image point, can be described as:

$$I(x,0) = \upsilon_E E(X) + \xi_E \ \forall_X \in W_U,$$

where $E(X)$ is albedo (reflection power) of scene point; U is environment of scene point; υ and E are permanent parameters which present the changes of contrast and brightness respectively. While a camera is moving the parameters are changing, that is, they depend on time. Changes of brightness in time can be written as:

$$I(x,0) = \upsilon(t) I(x,t) + \xi(t) \ \forall_X \in W_U,$$

where

$$\upsilon(t) = \frac{\upsilon_E(t)}{\upsilon_E(t)}, \ \xi(t) = \xi_E(t) - \frac{\upsilon_E(t)}{\upsilon_E(t)} \xi_E(0),$$

if $t > 0$. Having merged affine movement of environment of feature point with changes of brightness we get the following expression:

$$I(x,0) = \upsilon(t) I(Ax+d,t) + \xi(t) \ \forall_X \in W.$$

This equation will never be realized in reality because of noise in the image, approximate modeling of movement and changes of brightness. That's why the task of feature points tracking is reduced to minimization of difference between the environments of current and new feature point position:

$$c = \int_W \left(I(x,0) - \upsilon I(Ax+d,t) + \xi\right)^2 w(x)dx,$$

where $w(x)$ is weight function. With the help of expansion in Taylor's series in the environment $d = 0$, $\upsilon = 1$, $\xi = 0$ we get:

$$\upsilon I(y,t) + \xi \approx \upsilon I(x,t) + \xi + \nabla I \frac{\partial y}{\partial u}(u - u_0),$$

$$y = Ax + d \ A = \{d_{ij}\},$$

$$A = \{d_{ij}\},$$

$$\mathbf{d} = \begin{bmatrix} d_1 & d_2 \end{bmatrix}^T,$$

$$\mathbf{u} = \begin{bmatrix} d_{11} & d_{12} & d_{21} & d_{22} & d_1 & d_2 \end{bmatrix}^T,$$

$$\mathbf{u_0} = \begin{bmatrix} 1 & 0 & 0 & 1 & 0 & 0 \end{bmatrix}.$$

Having rewritten the obtained equality in matrix form we get the following equation:

$$I(x,0) = F(x,t)^T z ,$$

where

$$\mathbf{F}(x,t) = \begin{bmatrix} xI_x & yI_x & yI_y & I_x & I_y & I & 1 \end{bmatrix},$$

$$\mathbf{z} = \begin{bmatrix} d_{11} & d_{12} & d_{21} & d_{22} & d_1 & d_2 & \upsilon & \xi \end{bmatrix}^T .$$

Having multiplied by $F(x,t)^T$ and having integrated over all space of feature point W with weight function w, we get a system of eight equations with eight unknown quantities:

$$\mathbf{S}\mathbf{z} = a,$$

$$a = \int_W F(x,t)^T I(x,0) w(x) dx,$$

$$S = \int_W F(x,t)^T F(x,t) w(x) dx .$$

Having changed integration into the sum of all pixels in space W we get the following system of linear equations:

$$\mathbf{S}\mathbf{z} = a ,$$

$$\mathbf{S} = \sum_{x \in W} \begin{bmatrix} \mathbf{T} & \mathbf{U} \\ \mathbf{U}^{\mathrm{T}} & \mathbf{V} \end{bmatrix},$$

$$\mathbf{T} = \begin{bmatrix} x^2 I_x^2 & xyI_x^2 & x^2 I_x I_y & xyI_x I_y & xI_x^2 & xI_x I_y \\ xyI_x^2 & y^2 I_x^2 & xyI_x I_y & y^2 I_x I_y & yI_x^2 & yI_x I_y \\ x^2 I_x I_y & xyI_x I_y & x^2 I_y^2 & xyI_y^2 & xI_x I_y & xI_y^2 \\ xyI_x I_y & y^2 I_x I_y & xyI_y^2 & y^2 I_y^2 & yI_x I_y & yI_y^2 \\ xI_x^2 & yI_x I_y & xI_x I_y & yI_x I_y & I_x^2 & I_x I_y \\ xI_x I_y & yI_y^2 & xI_y^2 & yI_y^2 & I_x I_y & I_y^2 \end{bmatrix},$$

$$\mathbf{U}^{\mathrm{T}} = \begin{bmatrix} xI_x I & yI_x I & xI_y I & yI_y I & I_x I & I_y I \\ xI_x & yI_x & xI_y & yI_y & I_x & I_y \end{bmatrix},$$

$$\mathbf{V} = \begin{bmatrix} I^2 & I \\ I & 1 \end{bmatrix}.$$

If matrix **S** is invertible then the solution of the system of linear equations can be written as follows:

$$\mathbf{z} = \mathbf{S}^{-1} a .$$

As in all tracking algorithms the system is solved iteratively with the help of Newton–Raphson method. Iterations are carried out until the solution changes become negligibly small.

Step 2. Division of a video sequence into scenes.

To increase the quality of restoration of video sequence it is necessary to divide the video sequence into scenes. The division occurs according to the following algorithm:

1. Calculation of distance from every feature point to the central point of frame:

$$R_{ij} = \sqrt{(x_{G_i} - x_c)^2 + (y_{G_i} - y_c)^2} ,$$

where x_{G_i}, y_{G_i} are coordinates of the i-th feature point; x_c, y_c are coordinates of the central point of the frame.

2. Calculation of point displacement:

$$\left| R_{ij} - R_{ij-1} \right| < e,$$

where e is the threshold of point displacement in the frame.

3. Calculation of a number of strongly displaced points in j-th frame:

$$f(R, e, j) = \mathrm{count}(e > e_n),$$

where e_n is the general threshold of displacement. If function f obtains local maximum in the current frame j then current and next frames of video sequence are the borders of the scene. The quality of detection of the scene border can be described by the following parameters:

– accuracy, that is the probability that the detected scene border is not wrong:

$$P = \frac{C}{C+F} ;$$

– threshold signal that is the probability that the expected scene border will be found:

$$V = \frac{C}{C+M} ;$$

– synthetic measure of accuracy based on accuracy and threshold signal:

$$F1 = \frac{2PV}{P+V} ,$$

where C is the number of right actuations; M is the number of missed scenes; F is the number of wrong actuations.

Step 3. Detection of a scene type (with movement, without movement).

If feature points movement vectors were detected successfully at the stage of tracking of feature points of video sequences then let us consider that this video sequence and scene as its part have movement. Hence at the stage of restoration of the original video sequence let us use the spatial-temporal algorithm of obtaining information from the neighboring frame. When movement vectors can not be detected successfully at the stage of tracking of feature points of video sequences then let us consider that the scene does not have movement and use spatial algorithm of restoration of video sequences based on obtaining the information from neighboring area in the current frame.

Step 4. Detection of borders of artificially overlaid computer graphics areas in case of restoration of video sequence of such type.

The localization of text areas with artificially overlaid computer graphics is based on modification of Rares–Reinders–Biemond spatial algorithm [4]. This algorithm is constructed on the principle of detection of extreme brightness areas on the basis of soft and hard dynamic thresholds. To detect the areas of extreme brightness we must define some thresholds for searching and localization of bright and dim pixels. However, the use of fixed thresholds is not desirable as brightness changes from frame to frame. A hard threshold is a good solution for detection of such areas. A soft threshold results in a large quantity of wrong actuations. To avoid these problems the dynamic threshold is used to detect extreme brightness areas. The algorithm works in our case rather efficiently. The main idea of choosing the dynamic threshold is that firstly a hard threshold is chosen. Then only the areas with brightness values higher than the initial threshold are detected. The areas obtained at this stage are expanded by the neighboring areas satisfying the soft threshold.

Step 5. Detection of characteristics of areas to be restored, choice of the complex of algorithms for restoration of video sequence.

For video sequence with movement signs in the frame the structure of several previous frames of video sequence and the changes of the obtained structure of the previous frames in comparison with the current frame are analyzed. The decision about modification of the current frame with the use of information obtained from the previous frames taking into consideration the changes of the structure of the previous frame is made on the basis of the obtained data. For video sequence without movement signs the texture of the neighboring area is analyzed. Then the structure and the probability of its changes are defined. Texture area analysis with the help of dynamic sized space and comparison of the image elements on space borders can be a good alternative. It can be suggested that when the main elements of the image coincide on the space border then the image inside the space is a desired texton and on the basis of this image it can be acceptable to generate the texture for filling of the object to be removed. The filling of restoration area is carried out taking into consideration the data obtained.

For a video sequence scene with movement signs let us suggest that the position presented in rectangular coordinates relative to the top left corner of the frame ($x1_a$, $y1_a$, $x2_a$, $y2_a$) and linear sizes ($d_x = x2_a - x1_a$; $d_y = y2_a - y1_a$) of the restoration area are well-known, the frame moves in one direction in general, its motion being uniform and linear. After finishing of Lucas–Kanade algorithm's work we know the position of the set of feature points G_i for each frame, the previous position of the set of feature points G_{i-1} in one of previous frames and vector (x_v, y_v) or direction and movement value of each feature point between a couple of related frames. Having this information we get the possibility to calculate the frame number relative to the current frame from which we shall obtain the information for restoration. The description of work of this algorithm is given below.

In general the restoration area can be rectangular in shape that is why the frame number n is computed as minimal so that the restoration point is beyond the restoration area $n = \min(d_x/x_v;\ d_y/y_v)$ meanwhile the displacement of the replacement point with respect to the i-th frame will be $x_{i-n} = n \cdot x_v$; $y_{i-n} = n \cdot y_v$, and the coordinates will be the following:

$$\begin{bmatrix} x'_i \\ y'_i \end{bmatrix} = \begin{bmatrix} x_{i-n} \\ y_{i-n} \end{bmatrix} = \begin{bmatrix} x_i - n \cdot x_v \\ y_i - n \cdot y_v \end{bmatrix},$$

there i is the current frame; n is the frame displacement; $i–n$ is the previous frame containing information for restoration; $[x_i, y_i]$ is the restoration point; $[x'_i, y'_i]$ is the point to be reconstructed, coordinates value contains colour; $[x_{i-n}, y_{i-n}]$ is the point on the previous frame used for restoration.

Restoration process is repeated for each point of the restoration area for each frame of the restoration scene of video sequence in order to restore the whole scene of video sequence. We can use the reconstructed points for the restoration of other points of the same frame or the reconstructed frames for the restoration of other frames of the scene to be reconstructed.

This algorithm is applicable for the restoration of areas of any foreground objects mentioned above, but constraints are imposed on the size and position of the restoration area. In case if the restoration area is not situated near the frame border it can occupy up to 90 % of the frame on its larger side and 10 % on its smaller side. In case if the restoration area is situated near the frame border then the side length of the restoration area cannot exceed 10 % of the frame border length.

For the scene of video sequence without movement signs for small-sized restoration areas we can use a modified algorithm introduced by Ponce and Forsyth [5]:

1. Choice of a texture fragment in a desired located area in terms of the hypothesis about the continuation of similar texture in the restoration area.

2. Insertion of a texture fragment into the restoration area of image (until the restoration area is filled) in the cycle.

3. Selection of the environment of the position after the example of image, ignoring the estimation of the similarity of position with undefined values in process of calculation; choice of the value for this position from the set of relevant values of the chosen environments in a random and equiprobable way (until all the points values on the borders of the restoration area are selected in the cycle).

4. End of the cycle, step 3 and step 2.

A disadvantage of the presented algorithm is its computational complexity and high dependence on random values. Advantages are the solution of the problem of pouring the structure of areas with the uncertain form, and joining of the generated and the initial image. The results of the algorithm can also be improved by using a median filter. The desirable video sequence should be the result of algorithms work, but the main drawback is that the results can be estimated only subjectively or by a number of experts.

References

1. Damov M. V. Spatial method of localization of logotypes images in video sequences // Science. Technology. Innovations. NTI-2008 : Materials of the All-Russian scientific Conference of young scientists. P. 1. Novosibirsk, 2008. P. 191–193.

2. Lucas B. D., Kanade T. An iterative image registration technique with an application to stereo

vision // Proc. of Imaging understanding workshop. 1981. P. 121–133.

3. Making good features to track better / T. Tommasini [et al.] // Proc/ IEEE Computer Society Conference on Computer Vision Pattern Recognition. 1998. P. 145–149.

4. Rares A., Reinders M. J. T., Biemond J. Recovery of partially degraded colors in old movie // Proc. of EUSIPCO-2002. Toulouse, 2003.

5. Forsyth D. A., Ponce J. Computer vision: modern approach : transl. from English. M. : Williams, 2004. P. 928.

M. V. Damov, A. G. Zotin
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

SCENE IMAGE CONSTRUCTION BY WAY OF SEQUENCED IMAGES SUPERPOSITION

A concept of scene image construction by way of video sequence frames or photo shot series superposition is presented in the paper. Scene construction is performed on the basis of displacement map which reflects vectors of blocks shifts.

Keywords: video sequence reconstruction, image superposition, scene construction, displacement map.

The level of development of modern computing technologies allows to solve problems of great computational complexity including processing of video sequences. Construction of an image of a video sequence scene is a problem necessary to be solved in reconstruction of video sequence. Reconstruction of video sequences is an important area of work in connection with increasing requirements of potential customers such as experts working with video archives, experts in the field of production and restoration of films, TV broadcast experts as well as experts in the analysis of visual data received by various methods namely: air photography, satellite photography, laser location and other systems of sensors. But generally a video sequence scene is understood as a part of a film or a sequence of images shot from one foreshortening for some period of time. In paper [1] one of algorithms of video sequence division into scenes is presented.

A displacement map is a two-dimensional data array whose elements are two-dimensional vectors. Each vector sets a shift from a point on the first image to a corresponding point on the second image. This information is used for construction of a scene image on the basis of several neighboring frames. Thus, the algorithm of construction belongs to a spatial-temporal type. There are three basic approaches to the definition of parameters of global movement: the approach with the use of the instrument of feature points; the approach with the use of vectors of blocks movement; global search. In this article the approach with the use of vectors of blocks movement or neighborhoods is considered. The advantage of the offered approach is the use of an image pyramid. At first the displacement map is searched for greatly reduced copies of images. The found values are initial ones for displacement maps for more detailed copies etc. Thus, at each level of detail it is required only to update a displacement map which considerably reduces time of calculation and probability of finding false values. At the same time the algorithm assumes the presence of large enough objects on a scene i. e. a piecewise-smooth displacement map. Let's present an algorithm of search of a displacement map.

There is a pair of images for which it is required to construct a displacement map. Two pyramids of detail (for each image) are constructed. A pyramid in the area of images processing is the representation of a source image in a set of images of the smaller resolution additionally processed by one of smoothing filters. Pyramid formation occurs by image smoothing at the previous level and by a choice of points with a step more than one pixel with the help of bilinear interpolation. A half-width σ in Gaussian function is connected with the relation k ($k > 1$) of sizes of pyramid images at the neighboring levels:

$$k = \sigma\pi/2, \quad (1)$$

where a half-width of Gaussian function is a distance between two extreme values of the independent variable for which the value of the function is equal to half of its maximum value.

On the one hand such choice leaves in the smoothed image only those frequencies which a reduced image will contain, and on the other hand it does not lead to loss of details. At the same time the formation of a detail pyramid for a displacement map takes place. The process of search of a displacement map occurs gradually beginning from the top of a detail pyramid. Processing of images near the top of a pyramid and near its base is different. For images near the tops of pyramids (strongly reduced images) we should find geometrical transformation (affine, projective) of images in the first detail pyramid combining it as a whole with an image at the same level of the second detail pyramid . Transformation search represents the updating of an algorithm presented in work [2] and is described below. The found transformation between images at the set level of a detail pyramid allows to calculate a displacement map at the same level. In transition to the next level of detail resolution the image in the first pyramid will be transformed according to a displacement map for current level, thus, only the specification of a displacement map is searched at the next level.

Beginning from some level it is no longer possible to find simple geometrical transformation combining the images in a detail pyramid. Images break into squares of a predetermined size (from 8 to 16 pixels), and for each of them geometrical transformation (usually a shift) and a displacement map is searched the same way which was applied to the whole image. It is obvious that transformation search is appropriate only in the case when one square does not grasp objects at different distances from a camera, otherwise different parts of a square should undergo different shifts. For the purpose of reduction of displacement jumps on borders of squares it is necessary to perform averaging of displacements. The larger the degree of squares overlapping the more smoothed the displacement map is, but at the same time the operating time increases.

A pair of images $I_1(\vec{x})$, $I_2(\vec{x})$ and a displacement map $\vec{d}(\vec{x},\vec{\omega})$ are considered as input data. Vector $\vec{\omega}$ is a vector of parameters setting a model of a displacement map. Two models of a displacement map – a shift and similarity transformation – are considered below. Shift

$$\vec{\omega} = (t_x, t_y)\,, \tag{2}$$

$$\vec{d}(\vec{x},\vec{\omega}) = l\vec{t}\,, \tag{3}$$

where $\vec{x}$ is some known coordinates of a pixel in an image; t_x, t_y are pixel displacement between images on coordinates.

Similarity transformation

$$\vec{\omega} = (t_x, t_y, \lambda, \varphi)\,, \tag{4}$$

$$\vec{d}(\vec{x},\vec{\omega}) = (1+\lambda)R_\varphi \vec{x} + l\vec{t} - \vec{x}\,, \tag{5}$$

$$R_\varphi = \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\phi & \cos\varphi \end{pmatrix}, \tag{6}$$

where λ is a coefficient of image scaling; φ is an angle of image turning .

The parameter l is introduced for nondimensionalization of components $\vec{\omega}$ and reduction of their values to one order of magnitude so that each component does not exceed several tenths. Its numerical value is necessarily equal to the size of a diagonal of an image:

$$l = \sqrt{w^2 + h^2}\,, \tag{7}$$

where w and h are width and height of an image accordingly. In both cases:

$$\vec{d}(\vec{x},0) = 0\,. \tag{8}$$

The search of a displacement map is reduced to the search of $\vec{\omega}$. Let $\varepsilon(\vec{\omega})$ denote an integral from a square of difference of the second and the transformed first images:

$$\varepsilon(\vec{\omega}) = \int_\Omega \left(I_1\left(\vec{x} + \vec{d}(\vec{x},\vec{\omega})\right) - I_2\left(\vec{x}\right)\right)^2 d\vec{x}\,,$$

$$\varepsilon(\vec{\omega}) = \int_\Omega \left(I_1\left(\vec{x} + \vec{d}(\vec{x},\vec{\omega})\right) - I_2\left(\vec{x}\right)\right)^2 d\vec{x}\,, \tag{9}$$

where Ω is image sizes.

The following designation is used below:

$$\vec{y} = \vec{x} + \vec{d}(\vec{x},\vec{\omega})\,. \tag{10}$$

It is required to find $\vec{\omega}$, minimizing $\varepsilon(\vec{\omega})$:

$$\frac{\partial\varepsilon}{\partial\vec{\omega}} = 2\int_\Omega \left(I_1(\vec{y}) - I_2(\vec{x})\right) \frac{\partial I_1(\vec{y})}{\partial\vec{y}} \frac{\partial\vec{y}}{\partial\vec{\omega}} d\vec{x} = 0\,. \tag{11}$$

The following designations are used below:

$$\vec{g}^T \equiv \frac{\partial I_1(\vec{y})}{\partial\vec{y}}\,, \tag{12}$$

$$Y \equiv \frac{\partial\vec{y}}{\partial\vec{\omega}} = \frac{\partial\vec{d}(\vec{x},\vec{\omega})}{\partial\vec{\omega}}\,. \tag{13}$$

In the assumption of smoothness $I_1(\vec{x})$, small difference of $\vec{x}$ from $\vec{y}$ and taking into account (8) correct in the first order we get:

$$\vec{g}(\vec{y}) = \vec{g}(\vec{x})\,, \tag{14}$$

$$I_1(\vec{y}) = I_1(\vec{x}) + (\vec{g}, \vec{y} - \vec{x})\,, \tag{15}$$

$$\vec{y} - \vec{x} = Y\vec{\omega}\,. \tag{16}$$

Using (15, 16), the equation (11) is rewritten as follows:

$$\int_\Omega \left(I_1(\vec{x}) - I_2(\vec{x}) + \vec{g}^T Y\vec{\omega}\right)\vec{g}^T Y d\vec{x} = 0\,. \tag{17}$$

From (17) we get a linear system relative to $\vec{\omega}$:

$$A\vec{\omega} = \vec{b}\,, \tag{18}$$

where

$$A = \int_\Omega Y^T \vec{g}\vec{g}^T Y^T d\vec{x}\,, \tag{19}$$

$$\vec{b} = \int_\Omega \left(I_2(\vec{x}) - I_1(\vec{x})\right) Y^T \vec{g} d\vec{x}\,. \tag{20}$$

The search of the solution of system (18) demands accuracy as an image (or its fragment) may not possess the texture open enough for reliable determination of $\vec{\omega}$. The idea consists in estimating an error of matrix **A** according to a priori value of noise n on an image and smoothing radius σ of original images. Further system (18) is solved with the help of *SVD*-decomposition, and singular values of the smaller threshold are nulled [2]. To estimate a threshold singular value of matrix **A** we will consider a random image with dispersion of pixels intensity $\langle n^2 \rangle$ and zero average value: $\langle n \rangle = 0$. In case of absence of borders in a local area for σ→1 the estimation for dispersion of the considered image derivatives [2] is carried out:

$$\mu^2 \equiv \frac{\langle |\vec{g}|^2 \rangle}{2} = \frac{\langle n^2 \rangle}{8\pi\sigma^4}\,. \tag{21}$$

Estimation (21) can be easily received by transition into a spectral area. A considered image can be presented in the form of a spectrum of a homogeneous and stochastic function whose dispersion can be calculated from the theorem of integral equality from a square of a signal module to an integral from a square of its spectrum. Image differentiation is reduced to multiplication of its spectrum by a corresponding derivative of Gaussian function which allows to estimate μ^2. Knowing the model of required transformation and estimation (21) it is possible to estimate a minimum singular value of matrix **A**:

$$S_{\min} = \mu^2 l^2 wh . \quad (22)$$

Singular values of matrix **A** smaller than $S_{\min}$, should be nulled in calculation of a pseudo inverse matrix. In case of estimation of a minimum singular value during the work with a fragment variables ω, h, l the sizes of a fragment are designated in expression (22).

Let's present concrete models of displacement maps and an algorithm of superposition of the images received on the basis of analytical calculations.

Shift:

$$Y = \begin{pmatrix} l & 0 \\ 0 & l \end{pmatrix}. \quad (23)$$

Similarity transformation:

$$Y = \begin{pmatrix} l & 0 & x & -y \\ 0 & l & y & -x \end{pmatrix}. \quad (24)$$

An algorithm of images superposition:

Input: the first image I_1, the second image I_2.

Output: the transformation converting the first image into the second image.

Step 1. To calculate derivatives of the first image by convolution with Gaussian function. A half-width of Gaussian function should be of an order of expected values in a displacement map.

Step 2. To make matrix **A** of a systems according to expression (19).

Step 3. To make right part $\vec{b}$ according to expression (20) and to solve system of equations (18).

Step 4. To transform the second image according to the reversed transformation constructed from $\vec{\omega}$. It excludes the necessity to recalculate **A** on each iteration.

Step 5. If the module of vector $\vec{\omega}$ does not exceed the set value then pass to step 7.

Step 6. Pass to step 3.

Step 7. The algorithm ends.

As one can see from tests (sce figure) for piecewise-smooth surfaces the algorithm works well. Incorrect values are received near borders of objects. The algorithm operating time is proportional to a number of pixels on images.

The considered algorithm has two drawbacks: firstly, there are wrong values of displacement on borders of objects; secondly, the offered approach is in principle inapplicable in case of presence of small objects located close to a camera. It is connected with the fact that at low resolution the object disappears, and at high resolution the object shift turns out to be too big for the offered algorithm to be able to fix it. However the analysis of a difference frame can reveal similar problem objects and for them the application of other methods for post-processing is possible.

It is necessary to note also that as a result of algorithm work one-to-one correspondences between pairs of pixels of the first and second image which are true for a significant amount of points of an image (except occlusions) are received. It creates the possibility to use stable methods to find a fundamental matrix in epipolar geometry. Thus, the second pass of the algorithm which is carried out at imposed epipolar restrictions is possible.

It is necessary to note that data received by means of an instrument of feature points or an instrument of global search can be also used as initial data for the construction of a displacement map. A shortcoming of algorithms using special points as a base is low productivity, which is caused by high computing complexity of search operations (Feature Selection) and of feature points tracking (Feature Tracking) [3]. The advantage of similar algorithms consists in high reliability of definition of local movement. Algorithms of global search define optimum transformation without use of information on local movement [4]. Calculating complexity of these algorithms is so high that it practically excludes the possibility of their application for the solution of real problems.

a *b*

The image of a test video sequence scene and vectors of movement after the algorithm work (*a*); one of frames of a test video sequence with vectors of movement before the algorithm work (*b*)

Thus, at this stage there are two partially overlapping images I_1 and I_2; a displacement map representing an array of displacement vectors (Δx, Δy); the transformation converting the first image into the second (similarity or shift transformation).

The array of vectors found by an algorithm of movement definition has some characteristic features, namely, a movement vector is highly probably found incorrectly, if:

– the error of definition of a movement vector is great (it means that there is little similarity between compared blocks);

– the movement vector differs from the neighboring vectors considerably (it follows from smoothness of a field of movement vectors); the neighborhood dispersion is small (as a rule, in this case there is little difference between one block and neighboring ones). Hence it is necessary to introduce the quality or trust estimation taking into account the features of a field of vectors.

In paper [5] for the definition of movement vectors reliability it is offered to introduce the function of trust of the following kind:

$$fb_{x,y} = \left(a \cdot e_{x,y} + \frac{b}{d_{x,y}^{\ 2}} + c \cdot \sigma_{x,y} \right)^{-1}, \qquad (25)$$

where a, b, c are parameters; $e_{x,y}$ is an error of a movement vector definition (a sum of absolute differences of pixels brightness in the neighboring area of a vector in a positions (x, y) from a current frame and a compared neighboring area from the previous frame; $d_{x,y}$ is a neighboring area dispersion in a position (x, y) from a current frame; $\sigma_{x,y}$ is a standard deviation of a movement vector in a position (x, y) from the neighboring vectors

$$\sigma_{x,y} = \frac{1}{4} \sum_{i=-1}^{1} \sum_{j=-1}^{1} \left(\left(p_{x,y}^{X} - p_{x+i,y+j}^{X} \right)^2 + \left(p_{x,y}^{Y} - p_{x+i,y+j}^{Y} \right)^2 \right), \qquad (26)$$

where $p_{x,y}^{X}$, $p_{x,y}^{Y}$ is a projection of a movement vector in a position (x, y) on the abscissa axis and ordinate axis accordingly. In paper [5] the following values of parameters of trust function received by an experimental way are given: a = 0,25; b = 32; c = 1. Movement vectors with small value of trust function are rejected; the other vectors form a set of reliable vectors.

In the process of formation of the scene I_s image it is necessary to take into consideration features of a displacement map. A displacement map displays only vectors for similar points located on overlapping images I_1 and I_2. In this connection for the points belonging to objects which appeared or disappeared again vectors of displacement will be absent. On a displacement map the display of such points will look like a vector of a zero size.

For generation of a scene on the basis of a displacement map one can use two approaches. The essence of the first approach is that a completion of an initial frame takes place. During this completion there occurs an expansion of borders and filling of new blocks with information according to the data from the next frame which underwent certain transformations. The second approach is that for each new frame the change of the size occurs by addition of a so-called loop received as a result of transformation of the boundary information from the previous frames and comparison.

As the area on frame borders is of greatest interest for us during the construction of the full image of a scene so for speed increase it is expedient to use boundary data for an analyzed frame. Taking into account the possibility of non-uniform movement on frame borders it is necessary to carry out calculation of displacement parameters for analyzed regions, however inside the region there can also be a non-uniformity. On the basis of information given above it was decided to apply the information about vectors of local movement for segments and data about similarity and shift transformations as a basis for formation of a scene image .

The general algorithm for a scene image formation on the basis of a displacement map will consist of the following steps:

Step 1. Formation of preliminary boundary segments. On this step boundary segments with width and height $W_\omega \times W_h$ accordingly are formed depending on the size of an analyzed area. We calculate them from dimension of a shot dividing it by 8. Then the generation of a card of segments M_{segm} taking into consideration their spatial position in conformity with initial division takes place.

Step 2. Calculation of vectors of local movement from frame I_1 to frame I_2. During this step calculation of average, maximum and minimum vectors of movement for each segment occurs. To calculate an average vector of movement one can choose the following ways to receive value VSs_i as an arithmetical mean or median value from corresponding elements of a displacement array .

Step 3. The analysis of local movement vectors and formation of displacement vectors . On this step local vectors of movement from frame I_1 to frame I_2 for each segment are analyzed and distances from a comparable segment for frame I_2 with borders of shot I_2 are calculated.

Step 4. The analysis of the received vectors of displacement for all segments and formation of an expansion map M_{scale}. The given map is formed as a rectangular area taking into consideration the greatest possible displacement for each segment. Unused areas, i. e. those areas in which movement is not defined unequivocally are marked on completion of an expansion map generation.

Step 5. Filling of a scene image on the basis of an expansion map and vectors of displacement taking into account possible transformation. For each vector of displacement (Δx, Δy) of vicinities (x, y) not equal to zero this vicinity is transferred into a scene image. For each vector of displacement (Δx, Δy) of vicinities (x, y) equal to zero the neighboring vectors are analyzed. The location of such vicinity in the scene image is calculated using data of the neighboring vectors. Then a vicinity is placed into a scene image using the calculated location.

After the work of an algorithm of scene formation there is an image of a scene of some height and width in the form of pixels array. The reliability of a scene image constructed on the basis of several neighboring shots of one scene of a video sequence can be estimated visually by means of an expert estimation. In fig. 1 the result of work of an algorithm after processing a scene of test video sequence as well as the found field of movement vectors are presented.

The result of research work is the algorithm of scene image formation on the basis of the neighboring frames of a video sequence scene. The offered algorithm is applicable for the solution of many important problems, for example, for construction of a panoramic image, video sequence stabilization and compression, video sequence reconstruction in systems of video editing and electronic video archives.

References

1. Damov M. V. Restoration of background in parts of a frame with small-size objects in videosequence // Vestnik. Scientific Journal of Siberian state aerospace university named after academician M. F. Reshetnev. Issue 1(27). 2010. P. 52–56.

2. Mowler K., Forsyte G., Malcolm M. Machine methods of mathematical calculations, transl : from English. Mosckov : World (Mir), 1980.

3. Making good features to track better / T. Tommasini [et al.] // Proc. IEEE Computer Society Conference on Computer Vision Pattern Recognition. 1998. P. 145–149.

4. Litvin A., Konrad J., Karl W. C. Probabilistic Video Stabilization Using Kalman Filtering and Mosaicking // Proc. of SPIE Conference on Electronic Imaging. 2003.

5. Soldatov S. A., Strelnikov K. N., Vatolin D. S. Quick and reliable definition of global motion in videosequences. // Materials of 16-th International conf. on computer graphics and its applications. 2006. P. 430 – 437.

I. M. Danilin
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

A. I. Danilin
Institute of Forest named after V. N. Sukachev, Siberian department of Russian Academy of Sciences, Russia, Krasnoyarsk

D. I. Svischev
Krasnoyarsk State Agrarian University, Russia, Krasnoyarsk

LASER LOCATION AND DIGITAL AERIAL SURVEY AS A SUBSATELLITE COMPONENT IN THE SYSTEM OF INFORMATION SUPPORT OF INVENTORY, MONITORING AND CADASTRE OF FOREST LANDS

Approaches and solutions in the area of forest remote sensing methods for the purposes of information support of inventory, monitoring and cadastre of forest lands with the use of innovation methods and high-end technologies of laser location, digital aerial survey and global satellite positioning are considered in the paper.

Keywords: laser location, digital aerial survey, satellite positioning, inventory, monitoring and cadastre of forest lands.

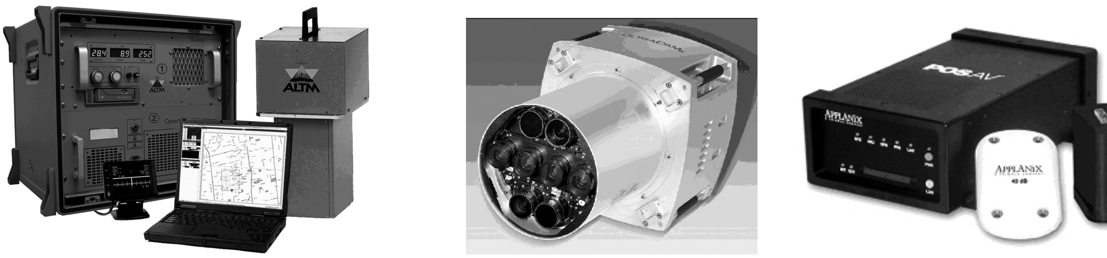
In the contemporary practice of forest use, monitoring and cadastre to receive reliable and efficient information about state and dynamics of forest lands is an urgent problem concerning both natural resources and environment and nature protection.

Nowadays laser location and digital aerial survey, which are major constituents of geomatics – the newest integral direction in the development of remote sensing methods of the Earth (aerial- and space survey), geoinformation technologies, digital photogrammetry and cartography, satellite geopositioning and telecommunications are more and more widely used in the solution of this problem in many countries of the world as well as in Russia.

Those advanced and highly effective methods today find wide application in many branches of economy. In fact they are an information basis of natural resources cadastres, land- and forest planning, ecological monitoring, systems of data gathering, processing and analysis. Judging by their accuracy and economical efficiency they exceed other methods of studying and measuring parameters of the Earth cover and natural systems [1–7].

Modern high-end aviation laser location systems develop quickly and today they have a scanning frequency of more than 200 thousand pulses (measurements) per second (fig. 1).

The highest density of scanning pulses is 1 pulse per 5–7 cm of the surface, the accuracy of measuring geometric parameters of the ground objects and vegetation morphostructural elements in plan and profile projections is about ± (5–10) cm. The accuracy of satellite positioning of forest line contours and boundaries, sample plots, separate trees and morphostructural elements of their stems and crowns, including under crown space, is practically unlimited and is determined by technical parameters of the GPS instruments [1; 4].



a *b* *c*

Fig. 1. Instrument and technological components of a laser location method:
a – universal aerial surveying laser topographic system Optech ALTM 3100 supplies laser-location data (purpose: Digital Terrain Models, contours allocation, data decoding); *b* – large-format digital photogrammetric aerial camera Vexcel UltraCamXp supplies digital aerial photographs of ultrahigh (centimeter) resolution (purpose: traditional); *c* – system of direct geopositioning and orientation of aerial photography sensors Applanix POSAV supplies elements of external orientation of digital aerial photographs and laser-location data (purpose: direct geopositioning (geoaffixment))

Satellite snapshots got by modern optical-electronic systems such as Landsat, Resurs-DK, Ikonos, OrbView-3, WorldView-2, GeoEye-1 or other systems of high- and ultrahigh resolution and decoded by main parameters and characteristics of vegetation cover [5] may also be both the tools of spatial and detailed display of contours and relief of the Earth surface together with its vegetation and the base for preliminary routes tracing for laser and digital aerial survey.

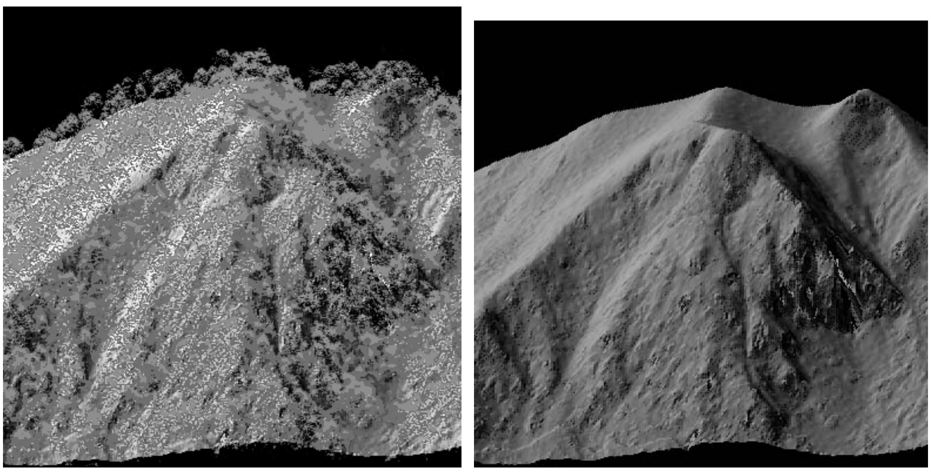
Therewith the structure, volumetric trees and tree stands indices, their phytomass are determined more reliably and precisely by laser-location data ("laser portraits"), integrated with digital geotransformed aerial photographs on the basis of DTM and forest canopy distribution field, which are generated from the initial laser location data by filtering (separating) laser pulses, reflected from the Earth surface and vegetation, by means of the ground pulses interpolation with the following triangulation of pulses reflected from vegetation in systems of differential satellite positioning GPS, GLONASS [3; 5].

Mathematic morphology methods operating with the concept of set and fuzzy set theory [8] are used in processing and analyzing laser-location data and digital aerial photographs.

Digital (laser-location) terrain and vegetation model allows to get detailed coordinates and morphostructural characteristics of relief and tree stands by means of three-dimensional computer graphics and visualization with the use of Altexis 2,0, ArcView Spatial & 3D Analyst software, or other known software [5] (fig. 2–4).

In some works implemented earlier in Russia and in other countries [1; 2; 5–7], it had been shown that accuracy of timber stock and forest biomass estimation, as well as by aviation sensing methods, can be increased up to 5–7 % with the use of morphological classification and allometric correlations among tree sizes.

Our research conducted in Krasnoyarsk region shows that the structure of the Earth's surface and vegetation cover elements with the use of DTM, generated by laser location and digital aerial photography are more adequately and effectively determined by characteristics of trees distribution lines according to the main morphometric indices – diameter and height, vertical and horizontal crown extent, which in their turn are closely intercorrelated. Therewith volumetric and weight trees and tree stands parameters are approximated with high accuracy by allometric functions through their morphometric indices (fig. 5, 6; table) [1].



a *b*

Fig. 2. Three-dimensional visualization of forest vegetation
(*a*) and land relief in under crown space; (*b*) by laser-location data

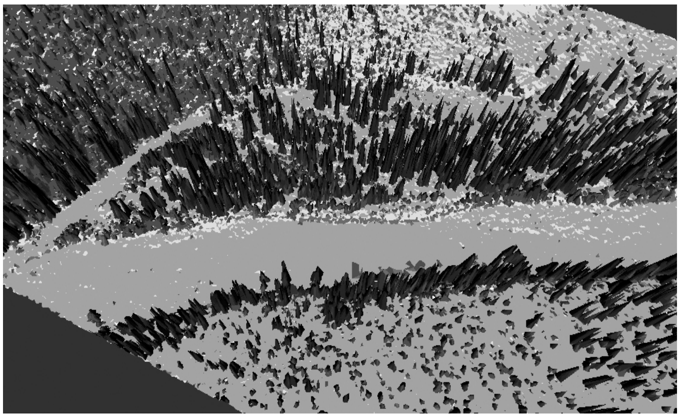


Fig. 3. Digital polygonal model of a larch stand, generated by laser-location data

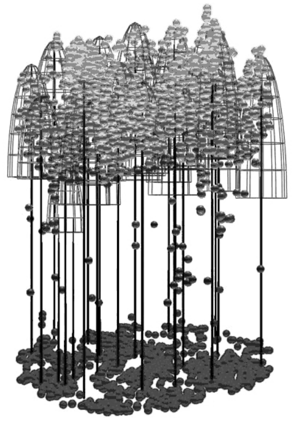
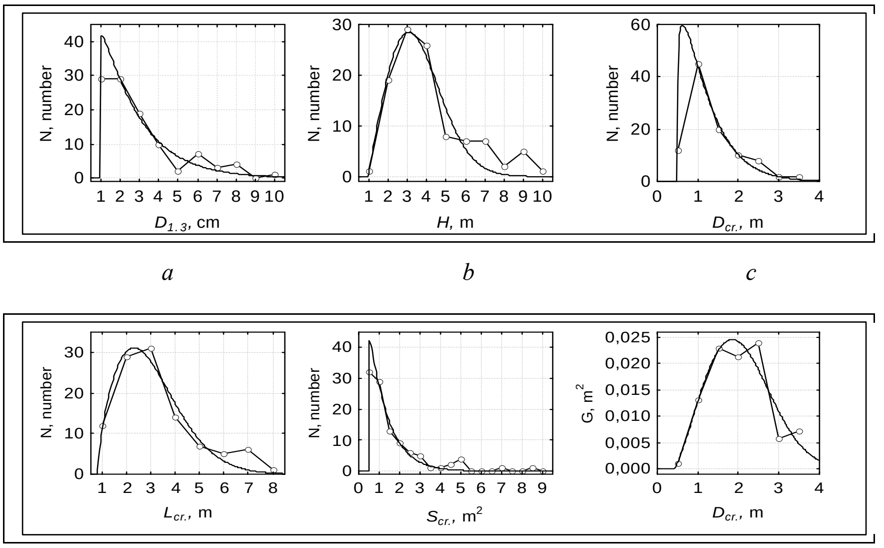


Fig. 4. Digital reconstruction of a larch stand morphological structure, implemented by laser location data



a *b* *c*

d *e* *f*

Fig. 5. Distribution of larch trees (N) by morphometric indices of stems and crowns, approximated by Weibull function: *a* – $D_{1.3}$ is a tree stem diameter 1,3 m above its base, cm; *b* – H is a tree height, m; *c* – D_{cr} is a crown diameter, m; *d* – L_{cr} is a crown length, m; *e* – S_{cr} is a crown area, m^2; *f* – G is a sum of tree stems cross sectional areas 1.3 m above their bases, m^2 ($G\,f\,(D_{cr})$)

Regression coefficients of morphometric indices and phytomass of larch trees

Model of approximation	$P = aD^2_{1.3}H$*			$P = aD^2_{cr}H$		
Parameters of the equation:	a*	S*	R^2*	a*	S*	R^2*
Above ground part of a tree	0.029	0.505	0.996	0.266	2.122	0.964
Stem	0.0203	0.055	0.999	0.187	1.750	0.951
Timber	0.017	0.037	0.999	0.153	1.452	0.950
Bark	0.004	0.008	0.996	0.034	0.307	0.951
Crown	0.008	0.258	0.976	0.079	0.427	0.983
Branches ∅ > 1 cm	0.003	0.068	0.940	0.024	0.217	0.959
Branches ∅ < 1 cm	0.002	0.021	0.969	0.020	0.272	0.892
Shoots of a current year	0.0001	0.000	0.873	0.0001	0.005	0.966
Needle	0.003	0.089	0.917	0.024	0.074	0.995
Dead branches	0.001	0.002	0.987	0.009	0.110	0.918

*P is a weight of a tree fraction in absolutely dry state, kg; $D_{1.3}$ is a tree stem diameter 1.3 m above its base, cm; H is a tree height, m; D_{cr} is a crown diameter, m; a is a constant of the equation; S is a standard error of the equation; R^2 is an index of determination.

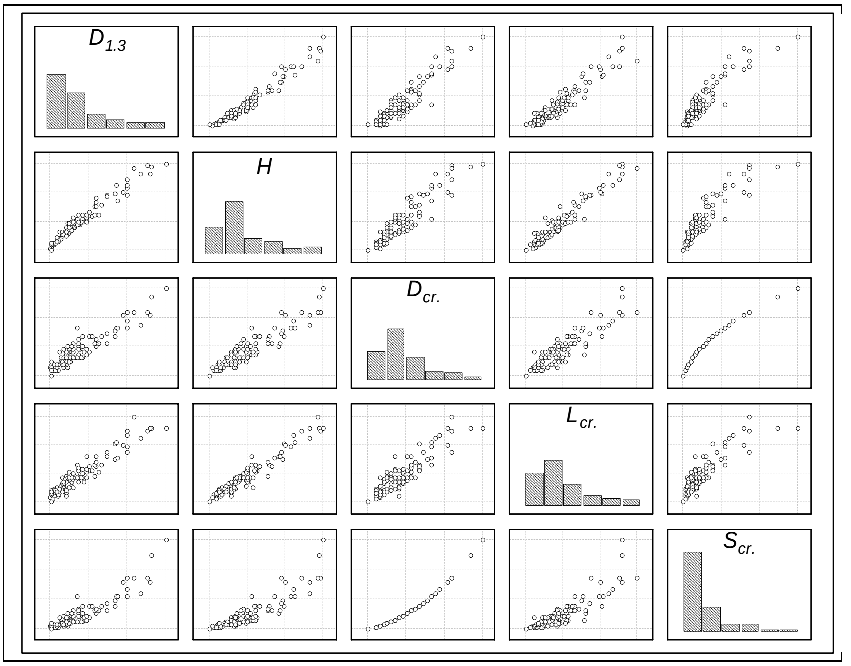


Fig. 6. An overlap matrix of distribution histograms and correlated scattering fields of the main morphometric indices of a larch tree stand (Central Evenkia)

It's generally known, that the construction of trees distribution lines by their morphometric indices traditionally supposes completing time- and labor-consuming ground biometric procedures, measuring operations and trees recalculations (continuous or selective), which require substantial financial expenditures. At the same time laser location method integrated with digital high- and ultrahigh (centimeter) resolution aerial survey allows to perform pixel instrumental-measuring forest inventory on the basis of precision satellite geodesy and detailed topographic survey, to study forest cover dynamics, horizontal and vertical structure of tree stands, to reconstruct trees distribution lines by any morphometric indicator, and to calculate required forest inventory parameters and forest biomass automatically, with high accuracy and on sufficiently large areas (up to 500–600 square km per one working day).

Assessment of timber stock and forest phytomass by laser location and digital aerial survey data in every case is reduced to identification of basic regulations of an investigated object and to determination of correlations between tree stems volumes, tree and crown heights and diameters, and phytomass, which in their turn constitute 87–99 % of explained variability of different phytomass fractions – stems, crowns skeletons and needles [1].

The results of practical approbation of an aviation laser location method combined with digital aerial survey and satellite geopositioning, integrated into geoinformation systems, give us an evidence of its high potential in use for the purposes of forest cover and disturbance analysis and simulation, inventory, on-line ecological monitoring, information support of forest lands cadastre and forest use control. This highly effective method provides remote sensing evaluation of forest resources, their state and dynamics, with minimum field works and significant saving of time and financial resources.

Cost efficiency of this method is ensured by principal increase of accuracy of the measurement results and possibility of their repetition (verification), as well as by significant reduction of labour-intensiveness and work complexity (of both field and laboratory kinds of work), at the expense of high level of automation in getting and processing laser location and digital aerial survey data. Thanks to this method the amount of field works is minimized and is necessary only for calibration of laser location results and support of aerial survey data interactive decoding.

References

1. Danilin I. M. Morphological structure, productivity and remote methods of inventory of Siberian forest stands : abstract of a thesis. … Doctor of agricultural sciences: 06.03.02. Krasnoyarsk : Institute of forest named after V. N. Sukhachev, Russian Academy of sciences, 2003.

2. High technology of XXI century for aerospace monitoring and forest inventory. Purposes of investigation and perspectives of usage / I. M. Danilin, E. M. Medvedev, N. I. Abe [et. al.] // Forest inventory and forest management. 2005. № 1 (34). P. 28–38.

3. Danilin I. M., Sveda T. Laser profiling of forest canopy // Silvics. 2001. № 6. P. 64–69.

4. Medvedev, E. M., Grigoryev A. V. Laser scanning forever // Geoprofi. 2003. № 1. P. 5–10.

5. Medvedev E. M., Danilin I. M., Melnikov S. R. Laser location of ground and forest // Tutorial, 2-nd revised and corrected edition. M. : Geocosmos.

Krasnoyarsk : Institute of forest named after V. N. Sukhachev, Russian Academy of sciences, 2007.
6. Holmgren J., Persson A. Identifying species of individual trees using airborne laser scanner // Rem. Sens. Environ. 2004. Vol. 90. № 4. P. 415–423.
7. Laser scanning of forest resources: the Nordic experience / E. Nasset [et al.] // Scand. J. For. Res. 2004. Vol. 19. № 6. P. 482–499.
8. Soille P. Morphological Image Analysis: Principles and Applications. 2nd ed. Berlin : Springer-Verlag, 2003.

M. N. Favorskaya, A. M. Popov
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

CONTEXT-DEPENDENT GRAMMAR DESIGN FOR DESCRIBING COMLEX SCENES WITH MULTI-LEVEL OBJECT MOTION

The problems of context-dependent grammars formation which describes structural information about patterns and pattern interaction in complex scenes are discussed in this article. The application of three-level grammar based on the task of an image sequences syntactic analysis (with extended contents of main and auxiliary dictionaries) and the task of scene syntactic analysis with multi-level object motion is suggested.

Keywords: context-dependent grammar, syntactic analysis, multi-level motion.

Initially, the structural or linguistic approach had been based upon using different linguistic structures, consisting of a dictionary and rules of sentence building from a specified dictionary. Such structural description of images permits to make an analogy between image structure and language syntax of formal grammars. Notice that this line of development appeared in the 1960’s as one of the first approaches for image describing and recognition. Structural approaches not only permit the reference of supervise static objects to a definite pattern, but also describe some object properties that exclude its referring to another pattern.

Traditional methods of structural approach are based on syntax description of complex image sets with limited sets of primitives and grammatical rules. It is considered that these images are formed from elements which are connected in a variety of ways just as phrases and sentences of languages are built by connecting words, and the words are composed from letters. The simplest elements, from which words and then sentences are built, are called primitives. Designing rules of composing primitives are usually assigned with special grammars of images description. Grammar rules (rules of substitution) may be applied any number of times, which allows for a compact and sufficient definition of primary structural characteristics for a sentences infinite set. The language for image structural description in terms of primitives sets and designs such element compositions; it is called the image description language. During identification, the recognition of primitives and image description in terms of special language is realized. Essentially, pattern recognition consists of a syntax analysis (or grammar analysis) and a “sentence” which describes some image. Recognition maintains the syntax correspondence between the analyzable “sentence” or image description and special grammar [1].

The system of pattern syntax recognition includes three main modules: the pre-processing module, the description module, and the syntax analysis module. The pre-processing module realizes coding, approximation, filtration, reconstruction, and the improvement of the image. The description module includes the primitives’ segmentation and allocation based on predetermined syntax operations. Each allocated part of the image is identified relatively to a special primitives set, and the whole image is characterized by a set of primitives’ sequences as the structures of language types. The syntax analysis module checks the accuracy of the sets in the context of predetermined grammars. The predetermined grammar corresponds to each pattern, and if the description of analyzable image is syntactically correct in the context of such grammar, then the image is related to the pattern for which this grammar corresponds.

The development of grammar describing both structural pattern information and patterns interaction is connected to the necessity of designing grammar reconstruction (or conclusion) algorithms according to a defined set of dynamic images that present the learning sampling. Such algorithms accomplish the learning of the recognition system. In result, pattern structural descriptions and their relationship descriptions are formed; then they are used for a syntax analysis of events and the genre of a complex scene. Basically the learning process isn’t executed; the choice of grammar and primitives set are realized by a tutor. Since the dynamic scene with multi-level motion has a very complicated and time-dependent structure, it’s necessary to use context grammar rules, which form a multi-level context grammar.

Let’s consider some main regulations which are peculiar to the structural methods of scene describing or recognition. The generative grammar is a well-ordered set of parameters **GR** = (V_T, V_N, P, S), where V_T – is a finite alphabet, defining the set of terminal symbols; V_N – is an alphabet, defining the set of nonterminal symbols; P – is a finite set of conclusion rules, i. e. a set of following pairs $u \to v$, where $u, v \in (V_T \cup V_N)^*$; S – is an initial symbol (grammar axiom), $S \in V_N$. The sequences of grammar generative language consist of terminal symbols. The symbol in left part of the first rule of grammar conclusion is an axiom. In grammar **GR** the sequence x directly generates the sequence y, if $x = \alpha u \beta$, $y = \alpha v \beta$, and $u \to v \in P$, i. e. the sequence y directly concludes from the sequence x, that denotes $x \Rightarrow y$. The language which is generated from grammar **GR** = (V_T, V_N, P, S) is called the set of terminal sequences, concluding in grammar **GR** from axiom: $L(\mathbf{GR}) = \{x \mid x \in V_T^*;\ S \Rightarrow {}^*x\}$, where symbol $\Rightarrow^*$ – is deducibility.

Rules generated by grammars permit string transformations. Constraints on the rule type determine grammar classes. The classification which was proposed by N. Kholmskyi, defines four grammar types:

– grammars type 0 – grammars, which don't have any constrains on the conclusion rules;

– grammars type 1 (context grammars) – grammars, the rules of which have the following view: $x\mathrm{A}y \to x\varphi y$, where $A \in V_N$, $x, y, \varphi \in (V_N \cup V_T)^+$;

– grammars type 2 (context-free grammars – CF-grammars). Conclusion rules in these grammars have the following view: $A \to \varphi$, where $A \in V_N$, $\varphi \in (V_N \cup V_T)^*$;

– grammars type 3 – finite state grammars which are divided into two types:

a) left linear (left recursive) grammars, conclusion rules for which have the following view: $A \to Aa \mid a$, где $A \in V_N$;

b) right linear (right recursive) grammars, conclusion rules for which have the following view: $A \to Aa \mid a$.

Language L is called i type language if the grammar of type i exists and generates language L. A conclusion tree often called the tree of grammar analysis or syntax tree, and the building process of the conclusion tree called – the grammatical analysis (syntax analysis). For one language sequence more than one tree can correspond, because this sequence can have different conclusions which are generated by various trees. For example, CF-grammar **GR** = (V_T, V_N, P, S) is called an ambiguous (indefinite) grammar, if the sequence $x \in L(\mathbf{GR})$ exists, and has two or more conclusion trees. It should be remembered that the tree of syntax analysis isn't grammar in the form of graph. Grammar graphs contain sentential forms (any sequences which are generated from an axiom) as nodes.

The principal disadvantages of the mentioned grammars are connected to their suitability in a greater extent for description scenes than for their recognition. This disadvantage has ben avoided due to investigations which had been carried out under M. I. Shlezinger who used two-dimensional programming method. The suggested two-dimensional grammar $\mathbf{GR}_S$ in [2] is a function of six parameters:

$$\mathbf{GR}_S = \langle \mathbf{V}_O, \mathbf{S}, \mathbf{T}_V, \mathbf{T}_S, R, \{\mathbf{Z}, \mathbf{Z}(t, t'); (t, t') \in \mathbf{N}\}\rangle .$$

Let object recognizing images be situated in images $\mathbf{T}_V$. Each of these images is a function given in images $\mathbf{T}_V$, possessing values of the object alphabet $\mathbf{V}_O$, which corresponds to the primary alphabet in one-dimensional grammars. The elementary images of this alphabet are used for the composition of more complex images. Besides the signal alphabet there is the alphabet of **S** structural elements which corresponds to auxiliary alphabets in one-dimensional grammars. On one hand, the structural elements define possible values of corresponding signals. On the other hand, they maintain constraints on image structures as local constraints. Structural elements make up an image description which is defined as function $\bar{S}$ on finite set $\mathbf{T}_S$ (description) and possessing values from set **S**. Generally these descriptions are not clear isomorphic images. An element of sets $\mathbf{Z} = \mathbf{V}_O \cup \mathbf{S}$ called "a symbol", and denoted as z. The set **T** is a combination of image and description. This element is called "a cell", and is denoted as t. Two cells t and t' are regarded as adjacent cells if some fixed for this grammar symmetric predicate $R(t, t')$ is equal "1". At that, **N** is the set of adjacent cells.

A pair $\left(\bar{V}, \bar{S}\right)$ image – description is called a variant $\bar{Z}$. This means that the variant is a function, specified on a set $\mathbf{T} = \mathbf{T}_V \cup \mathbf{T}_S$ and assumed values from set **Z**, such that $\mathbf{Z}(t) \in \mathbf{V}_O$, if $t \in \mathbf{T}_V$, and $\mathbf{Z}(t) \in \mathbf{S}$, if $t \in \mathbf{T}_S$. Sets **Z**, $\mathbf{Z}(t, t')$ of allowable pairs ($\mathbf{Z}$, $\mathbf{Z}'$) of symbols $\mathbf{Z}, \mathbf{Z}' \in \mathbf{Z}$ are defined for each pair of adjacent cells t and t'. Variant $\bar{Z}$ is called an allowable variant, if for each pair $(t, t') \in \mathbf{N}$ ratio $(\mathbf{Z}(t), \mathbf{Z}(t')) \in \mathbf{Z}, \mathbf{Z}(t, t')$ is executed. An image $\bar{V}^*$ is called an allowable image, if the allowable variant $\bar{Z}\left(\bar{V}^*, \mathbf{S}\right)$ exists. If variant $\bar{Z}\left(\bar{V}^*, \bar{S}\right)$ is an allowable variant, then description $\bar{S}$ is called a possible description of image $\bar{V}^*$.

Let the assign with two-dimensional grammar $\mathbf{GR}_S$ not be a whole set of images $\mathbf{X}^*$, which are concerned to one visual pattern, but if it's a small part, it is called a set $\mathbf{V}^*(\mathbf{GR}_S)$ of ideal or pattern images. Any pattern image $\bar{V} \in \mathbf{V}^*\left(\mathbf{GR}_S\right)$ corresponds to some set of real images which are similar to pattern image $\bar{V}$. The membership function $f_V(\mathbf{X})$ of recognizing image to set $\mathbf{V}^*(\mathbf{GR}_S)$, is called a similarity, possessing variable values (not only equal "0" or "1"). The syntax analysis task of image **X** is in the pattern definition of the image, which is generated by grammar $\mathbf{GR}_S$ and maximized by a similar function:

$$\mathbf{V}^*\left(\mathbf{X}\right) = \arg \max_{\bar{V} \in \mathbf{V}^*\left(\mathbf{GR}_S\right)} f_V\left(\mathbf{X}\right). \qquad (1)$$

In paper [2] a solution of this task was proposed: it was called a method of two-dimensional programming. This method permits to simultaneously receive a description with an optimal image $\mathbf{V}^*(\mathbf{X})$ $\mathbf{S}^*$ which corresponds to this image, i.e. to find a possible optimal variant $B^* = \left(\bar{V}^*, \mathbf{S}^*\right)$. The most essential peculiarities of the two-dimensional grammars are their universality (any set of images can be specified to the corresponding two-dimensional grammar) and construction (effective algorithms for finding possible optimal variant B^*). Another advantage of such algorithms is that they work directly with visual signals; that are why they permit the use of different methods for image pre-processing. Note that elementary images of grammars, describing complex pattern images can a have constant size (that decreases possibilities of two-dimensional programming) and a different size. In the latter case, such elementary images are related to so-called block two-dimensional grammars.

The solution for equation (1) with a larger noise level requires considerable computer calculations for practical problems. However the possibility of computer process paralleling usually exists during the realization of algorithms of two-dimensional programming. With noise reduction, such algorithms are not more complex that another algorithms of images analysis, but they provide a higher result reliability.

However, the two-dimensional grammar of M. I. Shlezinger is meant for the recognition of simplest binary graphical primitives in static scenes. For dynamic scenes with a multi-level motion the system of syntax pattern recognition is more complex. Temporal relationships appear between objects for which the describing of the patterns relationships recognition system design is required. The patterns relationships recognition system realizes four main principles of dynamic object recognition: the recognition aim on initial stages of video sequences processing; the recognition of behavioral situations for dynamic objects; the prehistory estimation of dynamic objects; changeable supervising object numbers in complex scenes.

A context grammar of complex scenes recognition with multi-level objects motion realizes the following procedures:

1. Pre-segmentation of the scene.
2. Description of regions with local features of motion.
3. Grouping of regions with local features of motion according to neighborhood.
4. Video objects recognition.
5. Grouping of video objects with global features of motion according to levels.
6. Description of multi-level motion in scenes.
7. Temporal events recognition in scenes.
8. Scene genre recognition (for multi-media libraries).

Analysis of these procedures shows that in case of recognition complex dynamic scenes with multi-level motion should use the following tree-level grammar:

$$\mathbf{GR}_D = \langle \mathbf{V}_{O,E,G}, \mathbf{S}_{S,LM,GM}, \mathbf{T}_V, \mathbf{T}_S, \mathbf{T}_E, R_E, \{\{\mathbf{E}, \mathbf{E}(a, a'); (a, a') \in \mathbf{M}\}, R_O, \{R_R, \{\mathbf{Z}, \mathbf{Z}(t, t'); (t, t') \in \mathbf{N}\}\}\} \rangle .$$

where $\mathbf{V}_{O,E,G}$ – is the main vocabulary of the objects, temporal events, scene genres; $\mathbf{S}_{S,LM,GM}$ – is the additional vocabulary of structural elements, local features of motion and global features of motion; R_R – is the predicate of region building; R_O – is the predicate of object building; R_E – is the predicate of temporal events. A set element $\mathbf{E} = \mathbf{V}_{O,E,G} \cup \mathbf{S}_{S,LM,GM}$ is called the event. A set $\mathbf{T}_E$ describes event sequence. A set $\mathbf{T} = \mathbf{T}_V \cup \mathbf{T}_S \cup \mathbf{T}_E$ in this case is the association of the event and description.

A context grammar of complex scenes recognition with multi-level objects motion realizes two tasks: the task of syntax analysis of image sequence $\mathbf{X}$ (with extended contents of main $\mathbf{V}_{O,E,G}$ and additional $\mathbf{S}_{S,LM,GM}$ vocabularies) according to equation (1), and the task of scene syntax analysis **SC**. Let's consider them in detail.

The aim of the syntax analysis of the image sequence $\mathbf{X}$ is the recognition of dynamic objects, which are classified into two large groups:

– objects originating from regions with continual colors, texture features during determined light conditions, and having a fixed set of projections in the frontal plane; contours of region changes in compliance with affine or projective groups of transformations (man-made items);

– objects originating from regions with continual colors, texture features during determined light conditions, and having a random set of projections in the frontal plane; contours of region changes arbitrarily (anthropometrical items). These regions are characterized by constant relative directions and speed values in some temporal interval.

These groups are characterized by a few different features, at which projection scattering of anthropometrical items is compensated by local motion features of separate statistically homogeneous regions. The methodology of the object recognition with restricted possible numbers of projections is well designed; we can propose the following formal scheme for the recognition of such objects. Let's assume that each pattern is represented by only one image. We shall call it the initial pattern template and state it as v^j. Let's also set the possible transformations G_b of initial template parameterized by blending parameter b as given. The result of using the transformation G_b for template v^j is a transformed template:

$$V(j, b) = G_b v^j.$$

The set of values at which template $V(j, b)$ accepts the fixed value j and possible values $b \in B$, is assumed an area of pattern template j. Observable images are the realization of multi-dimensional random quantity with known probability distribution $P(\mathbf{X}/V(j, b))$, depending on the transformed template $V(j, b)$ as well as the multi-dimensional parameter. Value $V(j, b)$ is the expectation value or mode of this distribution.

Such formal schemes permit solving the recognition task for the known distribution $P(\mathbf{X}/V)$ and known functional dependence of transformed template from parameters j and b. Here we can use the method of maximum likelihood. For solving the value parameter j it's necessary to identify the maximum of the likelihood function for parameters j and b, and accept for a solution d such value j for which this maximum is achieved:

$$\mathbf{V}^*(\mathbf{X}) = \arg\max_j \max_{b \in B} P\left(\mathbf{X}/V(j,b)\right) \quad . \qquad (2)$$

Solution d doesn't change if the likelihood function will be replace by any other function of parameter, values, which are connected with values $P(\mathbf{X}/V(j,b))$ by steadily increasing function, i. e. if $g(\mathbf{X}, V)$ – any function of parameter **X** and V, and satisfies condition:

$$P(\mathbf{X}/V) = f(g(\mathbf{X}, V)), \tag{3}$$

where $f(\cdot)$ – is a steadily increasing function, then rule (2) may be changed by the following expression:

$$\mathbf{V}^*(\mathbf{X}) = \arg\max_j \max_{b\in B} g(\mathbf{X}, V(j,b)). \tag{4}$$

The state would not in be principally changed, if function $f(\cdot)$ will be a monotonically decreasing function. However, the maximization in expressions (2) and (4) ought to be replaced by minimization. Because we can understand (4) as the solution rule for finding such values j and b for which similarity of observed image **X** with transformed template $V(j, b)$ is the maximum, then values of any function $g(\mathbf{X}, V)$, satisfactory condition (3), will be a measure of similarity template V with image **X**. Value j, found from expression (4), will be our solution.

Syntax analysis of the image sequences with the objects having a random set of projections is a more complex process. It is impossible to predetermine the set of available transformations; it requires a recurrent procedure for tracing regions with local features of motion for their following grouping in a common video object. In this case a template structure $\mathbf{V}^*(\mathbf{X})$ consisting of region sets, each of which has its own set of local motion features. Hence, a recurrent procedure of finding image template $\mathbf{V}_i^*(\mathbf{X})$ at step i, generating by grammar $\mathbf{GR}_D$, maximizes the similarity function the following way:

$$\mathbf{V}_i^*(\mathbf{X}) = \mathbf{V}_{i-1}^*(\mathbf{X}) + \gamma_i f_V(\mathbf{X}_{i-1}, \Delta_i), \tag{5}$$

where i, i–1 – are approximation steps; γ_i – is some function depending from approximation step (for example, a sequence of positive numbers); Δ_i – is changing quantity during the process of image sequence analysis; $\mathbf{V}_i{}^*(\mathbf{X}) = \arg\max_j P(\mathbf{X}/V(j))$ – is the similarity function on step i.

Function (5) is a variation of the stochastic approximation method for solving the task of pattern recognition learning. It is important to choose a function loss for the organization of the recurrent procedure (5). For example, we can use the following rule: if an image is classified correctly by some separating function, then the penalty equals “0”. If the classification was realized incorrectly, then the penalty is assigned to the value of proportional distance between the vector corresponding to recognizing images and separating hyperplane.

The target of scene **SC** syntax analysis is the recognition of events produced by single objects, interactive objects, and also the definition of the dynamic scene genre. These questions are covered for the problems of image understanding and scene analysis. In the case of complex scenes, before events recognition, it is essential to create a model of multi-level motion, i. e. to define the number of significant levels (in the simplest case to reach a decision on the existence of two levels – foreground and background), and to relate each recognized video object to one level or another. Such a task is most essential for virtual 3D-reconstruction in cartography, navigation system, and in cases when the video sensor is maintained on a moving platform, and the relative motion of all scene objects occurs. It gives an impression that the objects which are nearer to the video camera “move” faster than remote objects. In this case the model of global motion is similar to the model of multi-level motion which is defined by the set of various but internal similar motion levels, associating with solids located on various distances from the moving camera and k segments (image [3]). Let’s propose that the motion levels assign in a parametric form, and there are h motion levels. For image sequence it is necessary to define: a) the motion level for which each video object is associated; b) the parameter values of each level. For a known motion level, parameter values of the level are determined, and vise versa if we know the parameter values then we may determine to which motion level the video object is associated.

During the scene syntax analysis temporal events accumulating may be classified into motion classes and their interpretation on a conceptual level. The motion in image sequences, accounting their repeatability in time and space, can be classified into three classes: temporal textures, active actions and events. Temporal textures are determined as statistical regularities in space and time (sea waves, movement of clouds, leaves, birds, etc.). Active actions are interpreted as some are repeatable in time (but not in space) structures (walking, dances, separate movements of animals, insects, etc.). Events consist from isolated simple movements but do not repeat in time and in space (expressions, coming into a room, ball casting, etc.).

For dynamic changing scenes there are additional characteristics for objects such as the prehistory of the object motion and the procedure of associations, used for the final concept of forming and interpretation in terms of them. A prehistory of the object’s motion as a coordinate function from time to time may be roughly approximated because it is required to define the properties of motion but not its concrete characteristics. Then the prehistory is interpreted since some event of the object’s motion at the concept level – an analysis of temporal relationships between the objects play an essential role (handshake, discussion, aggressive action, etc.) [4].

For scene interpretation, let’s use a procedure of association which is described by two characteristics [5]:

– association value as measures of similarity calculating the nearness of the vector’s features for the scene objects, a nearness of relative transformations of these objects, and the objects’ significance;

– association similarity is subset of the objects spanned by the global motion event.

For association forming, each object of the knowledge base O_j, besides its direct description, has a set of additional characteristics the values of which are calculated in accordance with common scene dynamics:

– nearness $r_j^t = r(O_j, O^t)$ to interest object determining membership of the trace;

– nearness $q_k^t = q(V_k, V^t)$ connections with adjacent (on relations) objects O_k;

– value of association $a_j^t = \max_k(0, a_j^{t-1} - c, q_k^t \cdot a_k^t + r_j^t)$, (c<<1), defining significance of the association trace.

Concept forming is based on proportional association values increasing weights of objects belong to the associations' traces. Concepts are built as often occurring substructures consisting of significant objects.

Consequently in this paper, the basics of formal grammar designs in the context of the structural approach with pattern recognition are considered. The system structure of the syntax pattern recognition, which includes the pre-processing module, the description module, and the syntax analysis module, is proposed. The two-dimensional grammar of M. I. Shlezinger for recognizing the simplest binary graphical primitives in static scenes has been studied in detail. It has shown that for the recognition of complex scenes with multi-level motion objects we can apply a three-level grammar including the main vocabulary of objects, temporal events, scene genres, additional vocabulary of structural elements, local features of motion and global features of motion, predicates of regions building, predicates of objects building, and predicates of temporal events. Procedures of object recognition, based on the possible transformations, and recurrent procedure of statistical approximation, depending on the number of possible video object projections on the frontal plane, have been proposed. The association procedure, calculating the nearness of the vector features for scene objects, has been designed for the interpretation of complex scenes with a multi-level motion.

References

1. Favorskaya M. N. To the problem of applying formal grammars in indentifying objects in complex scenes // In the materials of the XIII international science conference "Reshetnev Conference". Part. 2. Krasnoyarsk, 2009. P. 540–541.
2. Shlezinger M. I. The syntactic analysis of two-dimensional visual signals in interference conditions // Cybernetics. № 4. 1976. P. 76–82.
3. Favorskaya M. N. Possible methods of video-flow segmentation as a problem lwith missing data // Vestnik. Scientific Journal of the Siberian State Aerospace University named after M. F. Reshetnev. 2010. Ed. 3 (16). 2009. P. 4–8.
4. Video Event Classification using Bag of Words and String Kernels / L. Ballan [et al.] // ICIAP09. 2009. P. 170–178.
5. Favorskaya M. N. Local time-space signs of events in video sequences // In the materials of the X international science conference "Theoretical and practical problems of modern information technologies", Part. II. Ulan-Ude, 2009. P. 461–466.

E. A. Furmanova, O. G. Boiko, L. G. Shaimardanov
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

THE POSSIBILITIES FOR OPTIMIZING THE FUNCTIONAL SYSTEM STRUCTURE OF CIVIL AVIATION AIRCRAFT

This is an analysis of the traditional approach to systems with individual reserving reliability calculation usage. An alternative calculation method for these systems with an individual reserve system has been developed. Its application is demonstrated.

Keywords: functional systems, analysis of the complicated systems, system structure optimization.

The airplanes functional systems execute many important functions: make the planer's steering surfaces drive by mechanization means, provide the aviation engines fuel supply, provide the cabin air pressure and air conditioning in them, provide all the consumers with electricity, protect the airplane from ice, provide fire extinguishing functions, provide automatic piloting, and air navigation.

The functional systems of the same type on all the route airplanes execute the same functions. At the same time, systems with the same name from different developers, or one developer, which is seen more often, on different types of aircraft, may have a different functional systems structure. With the same development in machine building, which provides a similar level of aggregate systems reliability, the reservation level is different in individual and typical aircraft system types with the same name.

Such a position is connected with the absence of research in the field of the system's optimization structure.

The proposed study shows the possibilities of common and individual reserving in the reliability securing systems.

Let's look through the individual reserving system, which contains n units that are successively linked. Each of the units includes $m = 2$ aggregates connected parallel. The structural scheme of such system is shown in the fig. 1.

Let's think that all the aggregates have the same breakdown stream parameters ω.

Let's consider for a mathematical model of the aggregate's breakdown probabilities the distribution with an even density probability. Then, the aggregate breakdown probability integral function will be:

$$q(t) = \omega \cdot t .$$

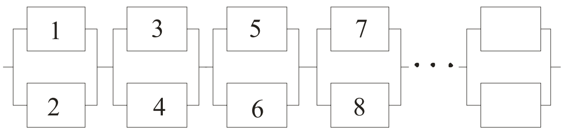


Fig. 1. The individual reserving system structural scheme

If $t = 1$, then we can find the breakdown probability for 1 flight hour:

$$q = \omega.$$

In the traditional approach of the reliability calculation, the first thing we had to do was defining the breakdown probability for parallel switched units at the aggregates:

$$q_\delta = \omega^2 .$$

Their working probability without breakdown [1; 2]:

$$P_\delta = 1 - \omega^2 .$$

So, the initial system, exchanges with the system with successively connected elements, and their without breakdown work probabilities equal P_δ. And then we define the initial system breakdown probability:

$$Q_c^T = 1 - \left(1 - \omega^2\right)^n . \quad (1)$$

Breakdown probability of the system with successively connected units increases with increasing of unit's quantity n.

In reality it is not so. The initial system (fig. 1) will break down only if in one of the units (any) both aggregates will fail. Because the aggregate's breakdown probabilities in 1 hour are identical and equal to ω, various scenarios of the aggregate's breakdown are possible, which will not lead the system to breakdown. For example, aggregates 1, 2, 6, 7 will fail (it doesn't matter in what succession order). The more n is, the less will be the breakdown probability of 2 aggregates in one unit. Any aggregate of the system may fail first with a probability of ω. The probability of the fact that the aggregate, based in one unit with the first, will fail equals to:

$$\frac{1}{2n-1}\omega .$$

Then the studied system breakdown probability during the proposed alternative approach to systems with individual reserving calculations will be:

$$Q_c^a(n) = \frac{\omega^2}{2n-1} . \quad (2)$$

Such an expression defines the most probable breakdown. Other possible scenarios of breakdown development in the system define essentially less breakdown probabilities due to:

$$Q_{ci}^a(n) = \frac{\omega^i}{\prod_{j=2}^{i}(2n - j - 1)} , \quad (3)$$

where i – is the quantity of broken aggregates, which lead the system to breakdown during the observed breakdown development scenario.

It's important to highlight the fact that during the increase of quantity n for successively connected units, breakdown probability according to (2) and (3) decreases, not increases, as the traditional approach shows (1).

The developed technique for systems with reserving reliability calculation allows changing the opinion of the system's reliability possibility increase; changes in the approach allow the usage the combination of common and individual reserving.

Let's view a system, consisting of two parallel working undersystems. The undersystems each contain 16 successively connected aggregates (fig. 2).

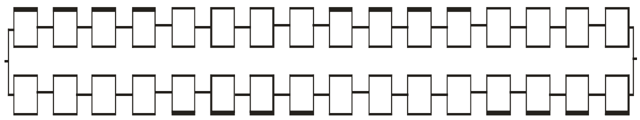


Fig. 2. The common reserving system

The breakdown probability of such a system may be defined as:

$$Q_c^{\text{общ}} = \left[1 - (1 - \omega)^2\right]^{16} , \quad (4)$$

where $n = 16$ – is the number of successively connected aggregates; $m = 2$ – is the number of undersystems, connected in a parallel way; $\omega = 0.01$ is breakdown flow parameter; the same is for all aggregates.

We shall not change the number of the aggregates in the system and its level of reserving. Let's divide the system into z parts, which will be connected successively (fig. 3).

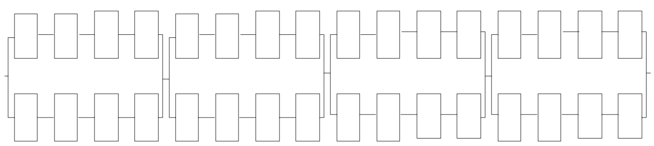


Fig. 3. The initial system divided into $z = 4$ parts of the common reserving

Let's account the reliability level for such a system. The probability of breakdown for each branch for any of the 4 parts will be:

$$Q_в = 1 - (1 - \omega)^4 . \quad (5)$$

Then the system (fig. 3) shall be transformed into.

In this system, the breakdown probability of each element equals to $Q_в$ and it is defined according to (5). The first and the largest system breakdown probability (fig. 4) is realized when 2 elements in one unit fail. It will be defined as:

$$Q^a(z) = \frac{\omega^2}{7}.$$

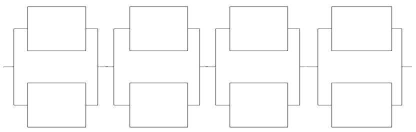


Fig. 4. The essential system, divided into $z = 4$ parts and transformed into the individual reserving scheme

We completed the breakdown probabilities calculations, during common reserving and after division into z parts, in accordance to the traditional and alternative method. The dependences of $\frac{Q_c^{общ}}{Q_{(z)}^a}$ and $\frac{Q_c^{общ}}{Q_{(z)}^T}$ on z are shown in the fig. 5–8.

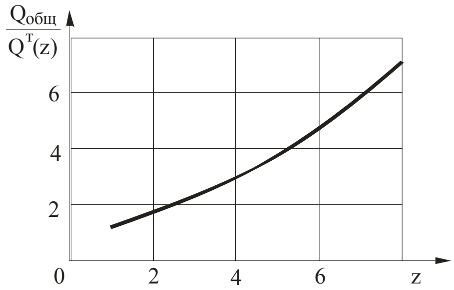


Fig. 5. The dependence of relation for the common reserving initial system breakdown probability and the breakdown probability of the system, divided into z parts from z, applying the standard technique $n = 16$, $m = 2$ and $\omega = 1\times10^{-2}$

From fig. 5 we can conclude that the calculations, made according to the traditional approach, show us the common reserving system reliability increase, during its division into z parts. This will transfer the system into a system containing z units with individual reserving. However, this reliability increase is not much: at $z = 2$ it is 1,8 times, and at $z = 4$ it is 3,5 times greater.

The transference of a system with a common reserving to a system with z units of individual reserving is linked to some difficulties. We shall try to explain them further. A slight increase in reliability didn't stimulate the system developers to overcome the stated difficulties.

Applying the alternative method for the same task of the reliability increasing solution dividing the system into z parts, the reliability is greater (fig. 6). With $z = 4$ achieving 100 and with $z = 6$ – (340–400) times.

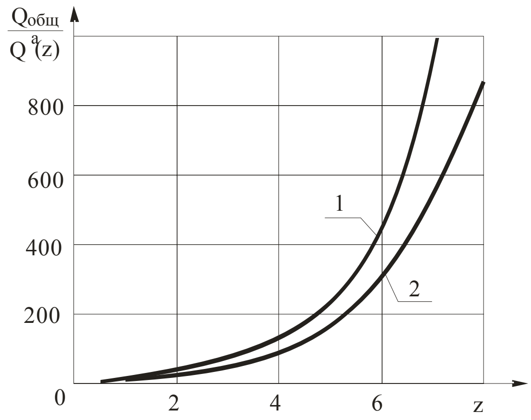


Fig. 6. The dependence of breakdown relation probabilities for the initial system and the breakdown probability of the system, divided into z parts and calculated according to the alternative technique from z, with $n = 16$, $m = 2$;
1 – $\omega = 1\times10^{-5}$ and 2 – $\omega = 10^{-2}$

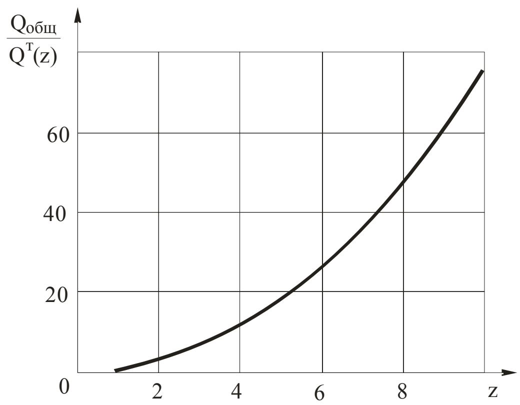


Fig. 7. The dependence of the initial system relation breakdown probabilities from a system divided into z parts and calculated according to the traditional technique breakdown probability from z. With $n = 20$, $m = 3$, and $\omega = 1\times10^{-2}$

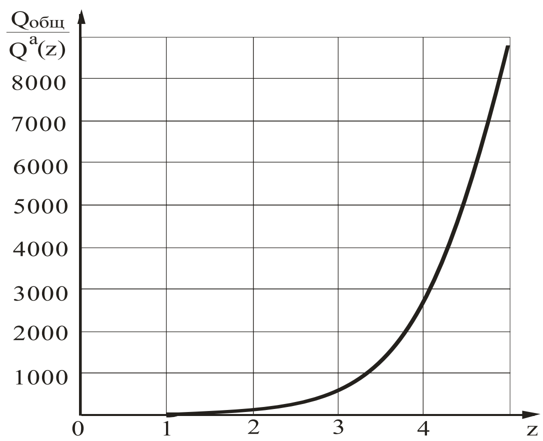


Fig. 8. The dependence of initial system breakdown relation probabilities from the breakdown probability of a system, divided into z parts and calculated according to the alternative technique from z, with $n = 20$, $m = 3$, $\omega = 1\times10^{-2}$

For a system with common reserving, with 3 undersystems working parallel the effect is much greater. In the traditional approach the reliability calculation has a 12 time reliability increase during $z = 4$ (fig. 7), and for the alternative approach (fig. 8) the reliability increases by 2.800 times, for $z = 5$ by 9.000 times.

This may stimulate the developers to overcome the earlier highlighted difficulties, which are connected with the transference of systems with a common reserving to systems with z units of individual reserving.

Let's shortly characterize these difficulties. In the hydraulic system, which consists of 2 identical undersystems, the pipe-line gap or one of the aggregate's core gaps leads to the loss of the entire hydro-liquid undersystem. The second undersystem remains intact and will provide the entire hydro-system's function execution. The changing of a systematical scheme, which would lead it to z units of individual reserving, deprives it of such protection. The undersystems are joined into one and the loss of liquid in one branch of the system's z-part will result in the loss of liquid in the whole system; this is inadmissible. Here it is possible to apply some of the blocking measures. In each branch of the z-part of the system, an expenditure measuring unit should be installed; in the beginning of a system a shutdown valve should be installed, and at the end of the system a reverse valve should be installed. During some flight time there is no liquid loss in the system. Such loss is displayed as pressure loss. The expenditure gauge sends a signal and the shutdown valve closes. The reverse valve eliminates the liquid expenditure. The shutdown valve blocks itself during the functioning of the usual consumers, which have some liquid expenditure during their functioning.

For electric systems the labor saving provisional system is much simpler. There are two kinds of malfunctions in the electric system: it is unnecessary to block the circuit in one of the z-part branches. It stops working and the entire load is transferred to a parallel branch of the z-part. In order to block the influence distribution of short-circuiting on one of the z-part branches on previous system parts, in the beginning of each z-part branch it is necessary to have a network protection device (a fuse).

References

1. Vorobiev V. G., Constantinov V. D. The reliability and effectiveness of aviation equipment. Moscow : Transportation, 1995.

2. Sugak E. V., Vasilienko N. V., Nazarov G. G. The reliability of technical systems. Krasnoyarsk : MPG “Rasko”, 2001. P. 608.

E. N. Garin
Siberian Federal University, Russia, Krasnoyarsk

THE INFLUENCE OF CONDITIONS FOR PASSING GLONASS AND GPS SATELLITE RADIO NAVIGATION SIGNALS ON THE ERRORS OF DEFINING RELATIVE COORDINATES

The influence of the conditions for passing GLONASS and GPS satellite radio navigation signals on the errors of defining relative coordinates is overlooked.

Keywords: signal delay, multipath propagation, error, frequency, navigation spacecraft.

Preliminary analyses show that a significant source of error is the reason for differences in the conditions for the of GLONASS and GPS signals passed on to navigation spacecraft (NSC).

It results in the difference of signal delay that could lead to an additional systematic error in determining relative coordinates.

The differences in signal delays could be caused by several different reasons. The first reason is the influence of the ionosphere and the troposphere. The radio navigation equipment sometimes installed in places where a signal delay impact from the ionosphere and the troposphere can be distinguished. This difference is of random nature but if the distance between the antennas is increased, the error will also increase consequently because of the fact that the ionosphere and the troposphere properties will change. When the distance between radio navigation equipment antennas is more than 100 kilometers the difference in an atmospheric error signal transmitted from a navigation spacecraft could reach several meters.

Another significant factor of measurement error is one caused by interference in the receiving end. This signal phase measurement error is caused by interference at the receiving antenna of the main signal or signals reflected from local items. This error component is often called a multipath error. Multipath errors depend on specific radio navigation equipment operation conditions, and typically can not be predicted. In most cases, this error could be considered as a random low-frequency component.

There is a difficulty in error measurement estimation of signal parameters caused by multipath transmission because of the instability of interfering signals. The presence of objects near the antennas with a large effective reflecting surface such as metallic constructions can generally make it impossible to calculate the results with abnormally large errors of phase measurement. When measuring the distance code delay the error could comprise tens of nanoseconds under difficult conditions.

For radio navigation equipment that performs measurements in a phase of carrier frequency of navigation spacecraft signal the impact of a direct signal and reflected signal interference can be unacceptably high. This can completely destroy the phase information on the signal delay (incremental delay) and lead to the disability of phase algorithms for the determination of navigation parameters.

Thus navigation spacecraft signals could be taken with sufficient signal to noise ratio that makes solving the navigation problem difficult and a necessity to apply a weight signal processing to reduce the negative impact on interference errors. Considering this fact the phase equipment of GLONASS and GPS systems should be provided with conditions for receiving that guarantee the absence of intense return signal. It is obvious that this is a serious limitation, reducing the possible application area of phase radio navigation equipment.

The value of multipath error is affected by the form of directional diagram of radio navigation equipment antenna and the presence of side and back lobe directional diagram. It is important to control the form of the directional diagram not only in the co-polarization (right-hand circular polarization), but also on the return (left) polarization because of the fact that the reflected signals only have return polarization.

Control methods over the given error component are refusal of work with navigation space vehicle signals, having low angles of elevation (reduction of directional diagram), forming of a directional diagram with a small level of side lobes and back lobes, optimal installation of antennas on the object, minimizing the hit of reflected signals in the mouth of directional diagram antennas [1–3]. A perspectives for the given error controlling is an integration of radio navigation equipment with additional independent sources of navigating data, for example inertial sensors.

References

1. Conley R., Lavrakas J. W. The world after Selective Availability // The Satellite Division of the Institute of Navigation 12th International Technical Meeting : Proc. of ION GPS-99. Nashville, Tennessee, September 14–17. 1999. P. 1353–1361.

2. Camargo P. O., Monico J. F. G., Ferreira L. D. D. Application of ionospheric corrections in the equatorial region for L1 GPS users // Earth Planets Space. 2000. Vol. 52. P. 1083–1089.

3. Lipkin I. A. Satellite navigation systems. M. : Vuzovskaya kniga, 2001. P. 288.

A. Sh. Geryukov, V. V. Bogdanov, S. V. Nazarov
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

RESEARCHES OF HYBRID TECHNOLOGY OF THE NONCONTACT ACOUSTIC CONTROL

This work is devoted to the results of experimental research concerning noncontact acoustic control of products. An advanced method of ultrasonic vibrations excitation in materials is offered. Design features of an electromagnetic acoustic converter for radiation – acoustic control are described.

Keywords: noncontact generation, cathode ray, ultrasonic vibrations, signals record.

The provision of high and stable quality of special engineering products, including aerospace equipment, is now one of the basic technological problems. Efforts of many scientific-research and design-technological organizations are put into its solution.

One of the primary factors influencing quality of products are defects of internal structure. The majority of refusals arising during tests and exploitation of products are related to manufacturing defects not found out earlier. Therefore a nondestructive test (NT) plays an important role both in the production process and in the process of technological development of products. The form of the construction, the presence of many inaccessible and out-of-the-way places for traditional devices of NT together with the requirement of maximum sensitivity and high resolution as well as special conditions of manufacturing make heavy demands on used methods and devices of NT, and often limit the application of many traditional methods of control. In this connection the creation of essentially new methods and devices of NT is extremely urgent.

One of the most widespread test method in industry is crack detection by means of acoustic waves. The necessity of acoustic contact creation in overwhelming majority of ultrasonic (US) devices makes it practically impossible to carry out NT in vacuum or controllable environment, at considerable levels of radiation and high temperatures which is typical first of all for beam technologies. All this has caused and stimulated the intensive development of researches of new noncontact methods of acoustic control based on optoacoustic and radiation-acoustic (RA) methods of ultrasonic vibrations generation in materials.

Increase of NT reliability is directly connected with the solution of the problem of acoustic contact stabilization. In its turn it has caused the development of noncontact methods of ultrasonic control. In particular, the use of vibrations of pulse compact electron accelerators for excitation of US vibrations provides the formation of short acoustic signals of nanosecond duration with sharp edges, which is extremely important for measurement accuracy increase. Besides, the radiation method of US vibrations generation provides remote excitation of stable acoustic impulses with parameters poorly depending on the quality of controllable product surfaces treatment that is explained by marked subsurface character of ultrasound formation.

To conduct experimental researches on noncontact registration of acoustic vibrations generated by impulse electron beams in current-carrying non-magnetic materials as well as to define defect detecting opportunities of RA quality monitoring a model set of an installation of RA control has been designed and manufactured.

The basic element of an installation is a small-sized electron accelerator, mounted on the basis of the MIRA-2Д impulse X-ray machine.

The accelerator has the following parameters:

- maximum electron energy in spectrum is 200 kilo electron volt;
- full energy transported by an electron beam for a pulse is about 200 mJ;
- duration of current pulse is 10 nsec;
- amplitude of beam current is 100 A;
- the diameter of beam put into the atmosphere is 6 mm.

The registration of generated acoustic signals was carried out by the electromagnetic-acoustical converter (EMAC) with a magnetic system of one constant magnet of the size 7 × 20 × 30 mm and residual induction 0.6 Tesla. A flat reception coil was made in the form of an ellipse with 12 × 18 mm size made of wire PEV-0.31. Electric pulses from EMAC entered the block of signal amplification, consisting of a preliminary low – noise amplifier and a radio-frequency amplifier UZ-29 with a coefficient of amplification 45 Db, a pass band 20 MHz and a dynamic range 1–200 mV through a parallel diode limiter. The block of amplification by a cable line coordinated on the entry and the exit was connected with the block of visualization made on the basis of the broadband oscilloscope C8-12.

Researches were carried out on samples of AlMg-6 alloy, made in the form of parallelepipeds with 40 × 40 × 60 mm size and plates with the thickness of 18 mm. The sizes of samples were chosen proceeding from the condition of eliminating the effect of lateral faces upon measurement results. Volumetric defects were simulated by flat-bottomed apertures, lateral drilling and rectangular grooves of various sizes. An electron beam was directed on the faultless surface of the sample, the axis of the beam was combined with the axis of defect, as otherwise the characteristics of generated acoustical radiation will depend on the kind and size of an artificial defect. Fig. 1 shows the oscillogram of the audible signals generated by an impulse electron beam in the faultless sample and registered with the help of EMAC.

Identifying signals according to the run time of the acoustical wave in the sample, it is possible to distinguish two types of US waves – longitudinal ones and shift ones.

The amplitude of a longitudinal wave is much bigger than that of a shift wave, which is explained by the design of EMAC used in the experiment and various character of spatial distribution of acoustical radiation of longitudinal and transverse US vibrations. Besides, it is necessary to note, that the second pulse of a shift wave is the result of a longitudinal wave transformation at its reflection from the opposite surface of the sample. The amplitudes of all three registered pulses of a longitudinal wave are practically equal in magnitude which is explained by insignificant losses of acoustical energy in aluminum samples. It is necessary to pay attention to the fact that signals from longitudinal and shift waves are opposite in phase. It confirms the previous conclusion that a longitudinal wave is primary, and a shift wave is the result of its transformation.

Fig. 2. shows the oscillograms illustrating the opportunities of noncontact radiation-acoustical control in the echo – through variant. The prismatic samples contained defects in the form of rectangular grooves of various depths while their width remained constant – 3 mm. From the comparison of the character of registered ultrasonic signals with the oscillogram of EMAC response while checking the faultless sample it is obvious that the presence of a defect in the product results in the appearance of additional intermediate audible signals and their position on a time base depends on the depth of a defect occurrence. The presence of a double signal from a defect is most likely explained by the character of reflection of the acoustical wave from the groove with flat bottom [1; 2].

Fig. 3. shows EMAC response oscillogram while checking the sample as a plate with a defect diameter of 3 mm and the depth of 10 mm. It is seen that a series of the first audible signals is lost in noise. However, the second echo-through pulse and the signal from a defect are clearly seen; the third rereflected pulse also gives a stable signal characterizing the defect.

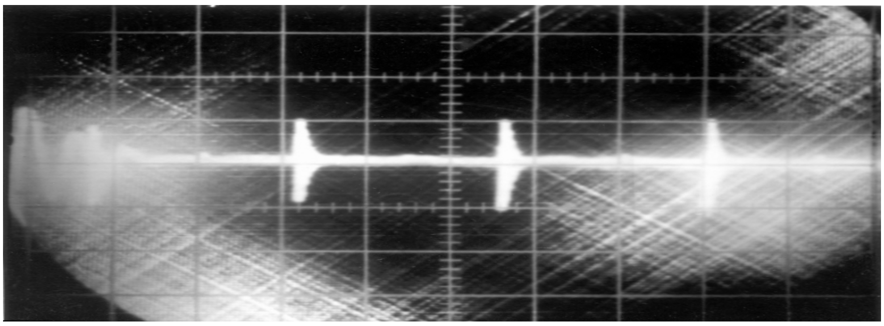
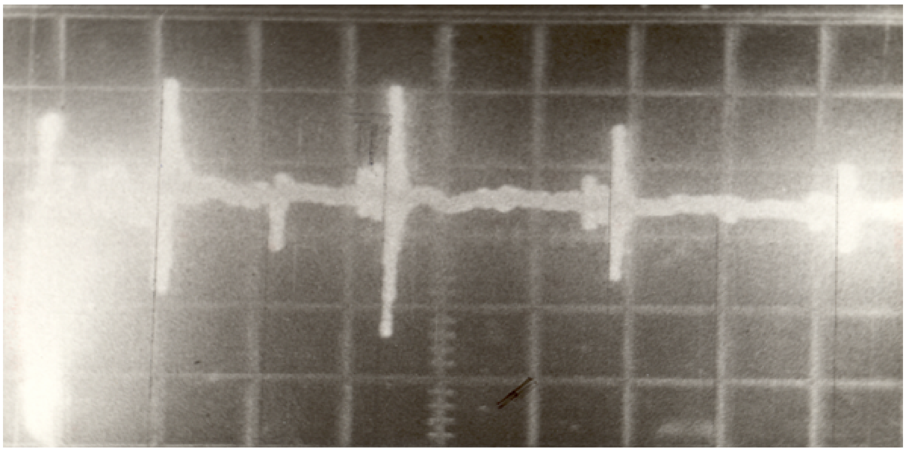
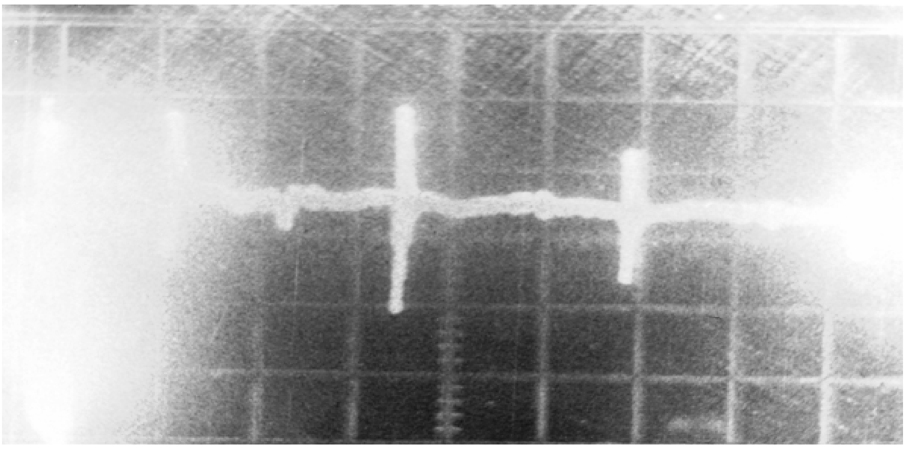


Fig. 1. Oscillogram of electric response of EMAC at registration of the acoustical vibrations generated by an electron beam in the sample made of AlMg-6 alloy



a



b

Fig. 2. The oscillograms received during the research of samples with artificial defects in the form of a rectangular groove with a depth: *a* – 5 mm; *b* – 15 mm

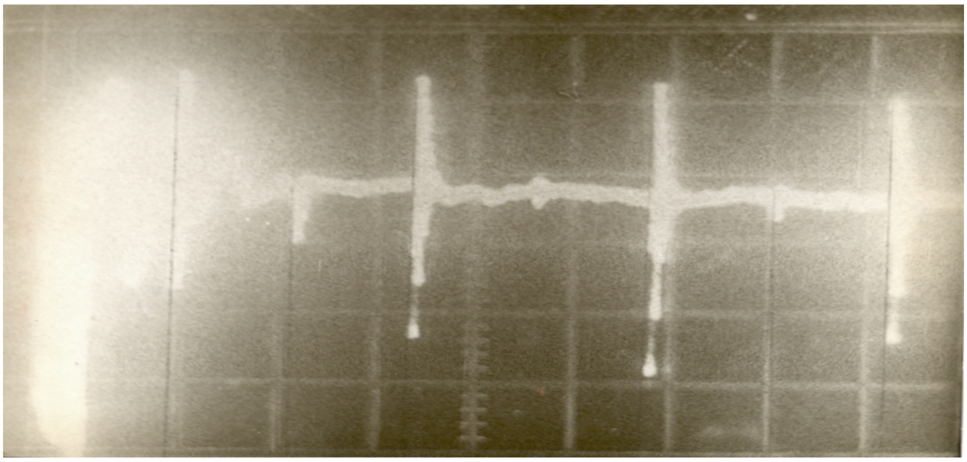


Fig. 3. The oscillogram received during the research of a sample made of alloy AlMg-6 in the form of a plate t = 18 mm with a flat-bottomed aperture

From the point of view of the requirements made to such detectors which were caused by both peculiarities of a radiating source of acoustical radiation and the character of the problem the analysis of technical characteristics of presently used noncontact receivers of US vibrations has shown that the optimal type of noncontact US detectors is EMAC. Such addition to a radiation-acoustical method provides not only the implementation of noncontact technique of US control of current-carrying non-magnetic materials, but also allows to expand the functionalities of EMAC significantly.

Experimental researches aimed at the optimization of design of EMAC working on longitudinal and shift acoustical waves and allowing to register both types of waves were carried out with the purpose of creation of simple and reliable means of noncontact registration of audible signals. Magnetic EMAC system of all types were made of rare-earth materials (REM) on the basis of intermetallic compounds of cobalt possessing high coercive force and magnetic energy ($H_{\text{св}}$ = 0.51–0.7 mA/m; W = 120–200 kJ/m^3) [3], which allowed maximum miniaturization of converters. As intermetallic bonds of rare-earth metals are included in a group of over critic magnetic materials they possess rather attractive properties. This is the independence of a magnetic moment on a magnet form and the ability to completely restore a magnetic stream during magnetic return even in case when the strength of a foreign degaussing field is commensurable with or exceeds the coercive force. The latter property is especially important at EMAC work in conditions of powerful electromagnetic fields the source of which are impulse electron accelerators. Besides magnets of REM are characterized by high thermal stability (a temperature inductance coefficient for samarium alloys is: cobalt – 0.043 %C^o, samarium–praseodymium–cobalt – 0.056 %C^o, samarium–gadolinium–cobalt – 0.0015 %C^o), which is important during the control directly in a technological process when EMAC characteristics should be constant in a wide temperature span.

EMACs developed on the basis of RE magnets are characterized by simplicity of design and reliability.

Detectors being a component of a cylindrical beam guide of an electron beam were developed for the realization of the control in conditions of unilateral access to the product. Thus, EMAC was made in the form of a ring covering the collimator. In one variant two magnets made in the form of concentric rings separated by a dielectric spacer were set in a case made of ferromagnetic material. Concentric coils were placed under magnet poles. By changing a way of their turning on it is possible to form the optimum diagram of reception orientation.

In the second variant (fig. 4), the detector represents a system providing an arrangement of measuring windings 1 EMAC and the face surface of two electron accelerators in one plane, irrespective of changes of the geometrical sizes of used constant magnets 3. It is provided by the mounting of magnetic systems on the steel base 4 with a possibility of coaxial rotation relative to case 5 and the collimator by means of threaded connection. In its turn the case is fixed on the electron accelerator. The design of the detector provides setting up and connecting of a diode limiter and a preliminary amplifier on board 6 as well as the use of several converters with various geometry of magnets and various connection of coils relative to each other. Steel ring 7 screens the magnetic system EMAC made of prismatic magnets.

In practice a signal-to-noise ratio is of major importance in determination of sensitivity limits of EMAC and the efficiency of this or that way of increasing the sensitivity of a detector should be estimated first of all according to this parameter [4].

The way of control with the help of EMAC use was offered, allowing to decrease the influence of an air gap between a converter and a product on the quality of control [5] considerably

$$K_c = \frac{A_0}{A_g}. \quad (1)$$

This method is realized as follows. Beforehand an electromagnetic acoustic converter is placed on the neighboring sites of a sample with a defect of a maximum permissible size. Thus, there is set a minimum gap h between surfaces of a sample and an electromagnetic acoustic converter. Echo-signals of each converter are registered and their difference V_{g0} is defined. Then detectors are placed with a minimum gap h on a surface of a sample without defects. Echo-signals of converters are received and their difference V_0 is defined. Then an electromagnetic acoustic converter is placed on a surface of a controlled product with a defect, echo-signals V_1 and V_1 of corresponding electromagnetic acoustic converters are registered and the revealed defect is qualified according to ΔV value:

$$\Delta V = \left(1 - \frac{V_{g0}}{V_0}\right) V_1 - V_2. \quad (2)$$

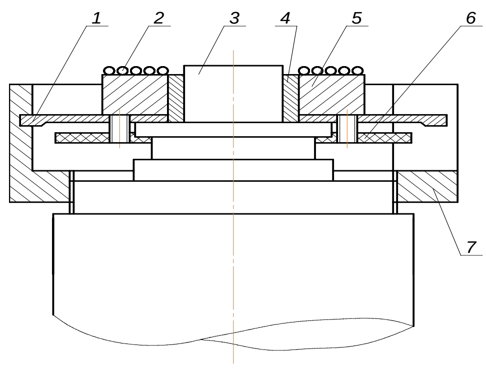


Fig. 4. The design of an acoustic fluctuations detector based on an electromagnetic acoustic converter for realization of noncontact ultrasonic control at unilateral access to a product. *1* – is a steel base; *2* – are inductance coils; *3* – is a collimator; *4* – is a shielding steel ring; *5* – is a magnetic system; *6* – is a board for diode limiter and a preamplifier installation; *7* – is a detector case

When $\Delta V > 0$, a defect is admitted acceptable, when $\Delta V \le 0$ it is admitted unacceptable. Thus h gap change, for example, at the expence of controllable product surface pollution, influences ΔV value a little.

The conducted researches allow to conclude that working out of a noncontact method of acoustic control and hardware means for its realization is possible on the basis of ultrasonic vibrations generation in metals by pulse electron bunches of nanosecond duration.

References

1. Yermolov I. N. Theory and practice of ultrasonic control. M. : Machine-building, 1981.
2. Alyoshin N. P., Belyi V. E., Vopilkin A. H. Methods of metals acoustic control. M. : Machine-building, 1989.
3. Korolyov M. V., Sheveldykin V. G. Small-sized high-sensitive EMA – converter // Theses of X national scientific-technical conf. Nondestructing physical methods and ways of control. Lvov, 1984.
4. Zhukov V. K., Simanchuk V. I., Surkova N. V. Investigation of EMA-converters of shear waves on the basis of magnets of rare-earth metals alloys. // Technical diagnostics and nondestructing control. Kiev, 1992. № 2. P. 58–63.
5. A.c. 16321177. Electromagnetic-acoustic way of control / V. K. Zhukov, V. I. Simanchuk. №29/04 ; appl. 23.11.88 ; published 30.03.90, Bulletin № 11.

A. E. Gorbatenko, A. A. Nazarov
Tomsk State University, Russia, Tomsk

RESEARCH OF THE SEMI-MARKOVIAN PROCESS IN CONDITIONS OF LIMITEDLY RARE CHANGES IN ITS STATE*

In this work the SM-flow in conditions of limitedly rare changes in its states is considered. In the proposed asymptotic condition there is a probable distribution of a number of events coming from the SM-flow in time t. We have shown that this distribution can be multimodal.

Keywords: *SM-process*, *state of flow, limit rare changes of flow states, method of the additional variable, asymptotic analysis method.*

In this work the SM-process has been considered [1]; it is the general flow of flow models with homogeneous events.

We shall give a definition of the Semi-Markovian (SM) process. For this purpose we have considered a two-dimensional homogeneous Markovian stochastic process $\{\xi(n), \tau(n)\}$ with a discrete time. Here $\xi(n)$ the ercodic Markov chain with a discrete time and matrix $P = [p_{\nu k}]$ accept the probabilities of transition for one step [2]; the process $\tau(n)$ accepts non-negative values from the continuous set.

Then we determine the Markovian transition function $F(v, x, k, y)$ for the process $\{\xi(n), \tau(n)\}$:

$$F(k,x;\nu,y)= \\ =P\{\xi(n+1)=k,\tau(n+1)<x|\xi(n)=\nu,\tau(n)=y\}.$$

We shall consider two-dimensional processes such as $\{\xi(n), \tau(n)\}$ for which the following equalities are correct:

$$F(\nu,x;k,y)=F(\nu,x;k),$$

that is $F(v, x, k, y)$ does not depend on the values of the y process $\tau(n)$.

Denote

$$F(k,x;\nu)=A_{\nu k}(x)= \\ =P\{\xi(n+1)=k,\tau(n+1)<x|\xi(n)=\nu\}. \quad (1)$$

Matrix $A(x)$ with elements $A_{\nu k}(x)$ can be called Semi-Markovian.

The stochastic process of homogeneous events:

$$t_1<t_2<...<t_n<t_{n+1}<...$$

is called the Semi-Markovian process or SM, set by matrix $A(x)$; if for moments t_n the approaching of its events is correct, the following equations are performed:

$$t_{n+1}=t_n+\tau(n+1).$$

In view of (1), thee transitive probability matrix of Markov's chain $\xi(n)$ is defined by equation:

$$P=A(\infty).$$

This chain for the Semi-Markovian process is called the embedded Markov chain.

In general case, the elements of the Semi-Markovian matrix have a place in the multiplicative form which can be written as:

$$A_{\nu k}(x)=P\{\xi(n+1)=k,\tau(n+1)<x|\xi(n)=\nu\}= \\ =P\{\tau(n+1)<x,|\xi(n)=\nu,\xi(n+1)=k\}\times \\ \times P\{\xi(n+1)=k,|\xi(n)=\nu\}=G_{\nu k}(x)\,p_{\nu k},$$

where $G_{\nu k}(x)$ – is the conditional distribution function of an interval length of the Semi-Markovian process in condition that at the beginning of this interval the embedded Markov chain has an accepted value v, and at the end of it will accept value k.

*This research has been carried out due to support of the Analytical Departmental Target Program: "The Development of Scientific Potential of the Higher School (2009–2010)" of the Federal Education Agency (Project № 4781).

Note, due to equation:

$$A_{\nu k}(x) = G_{\nu k}(x)\, p_{\nu k}, \qquad (2)$$

matrix $A(x)$ can be written as a Hadamard produce:

$$A(x) = G(x) * P \qquad (3)$$

of two matrixes $G(x)$ and P, and it is possible to suppose, that the Semi-Markovian process is set by two matrixes $G(x)$ and P.

The state of the semi-markovian process at the moment of time $t_n < t \le t_{n+1}$ is called state k of its embedded Markov chain, accepted at the beginning of interval $(t_n, t_{n+1}]$.

The research of the Semi-Markovian process will be will carried out in conditions of limit rare changes of (LRCS) flow conditions; i. e. the transition from one state of the embedded Markov chain to another is realized extremely rarely.

The conditions of limit rare condition changes of the Semi-Markovian process are formalized by the following equation for matrix $P(\delta)$ transition probabilities of its embedded Markov chain:

$$P(\delta) = I + \delta \cdot Q, \qquad (4)$$

where δ – is some small parameter ($\delta \to 0$); I – is an identity diagonal matrix.

Matrix Q with elements $q_{\nu k}$ is similar to a matrix with infinitesimal characteristics and has the same properties: $k \ne \nu$ matrix elements $q_{\nu k} > 0$, and also is correct:

$$\sum_k q_{\nu k} = 0, \quad \sum_{k \ne \nu} q_{\nu k} = -q_{kk}.$$

The Semi-Markovian matrix for the SM process in condition of LRCS:

1. $k = \nu$

$$A_{kk}(x,\delta) = P\{\xi(n+1) = k, \tau(n+1) < x \mid \xi(n) = k\} = G_{kk}(x)\{1 + \delta \cdot q_{kk}\}.$$

2. $k \ne \nu$

$$A_{\nu k}(x,\delta) = P\{\xi(n+1) = k, \tau(n+1) < x \mid \xi(n) = \nu\} = G_{\nu k}(x) \cdot \delta \cdot q_{\nu k}.$$

On the other hand, in the multiplicate notation (3) of the Semi-Markovian matrix let's substitute (4) and get:

$$A(x,\delta) = G(x) * \{I + \delta \cdot Q\}. \qquad (5)$$

Let's state $m(t)$ – as the event number of the Semi-Markovian process, which appeares during t on the interval $[0, t)$.

The process $m(t)$ is non-Markovian, therefore it is necessary to make it Markovian by a method of additional variables. Let's define the process: $z(t)$ – is the length of an interval from time moment t till the moment of approach for the next event in the considered SM-process.

However, the two-dimensional process $\{m(t), z(t)\}$ is not Markovian, therefore let's consider one more stochastic process $s(t)$ with piecewise constant continuous realizations on the left, defined by equation:

$$s(t) = \xi(n+1), \text{ if } t_n < t \le t_{n+1}.$$

This is on the interval $(t_n, t_{n+1}]$ process, $s(t)$ accepts and assures that value of the embedded Markov chain $\xi(n)$ accepts the beginning of the following interval [1].

The three-dimensional stochastic process is thus defined as $\{s(t), m(t), z(t)\}$ with two additional variables $s(t)$ and $z(t)$ – is Markovian with continuous time and with probability distribution:

$$P(s,m,z,t,\delta) = P\{s(t) = s, m(t) = m, z(t) < z\}, \qquad (6)$$

it is simple to create a system of differential Kolmogorov equations:

$$\frac{\partial P(s,m,z,t,\delta)}{\partial t} = \frac{\partial P(s,m,z,t,\delta)}{\partial z} - \frac{\partial P(s,m,0,t,\delta)}{\partial z} + \sum_{\nu=1}^{\infty} \frac{\partial P(\nu,m-1,0,t,\delta)}{\partial z} A_{\nu k}(z), \qquad (7)$$

at the set initial conditions:

$$\begin{cases} P(s,0,z,0,\delta) = R(s,z,\delta), \\ P(s,m,z,0,\delta) = 0, \quad m \ge 1, \end{cases} \qquad (8)$$

where function $R(s, z, \delta)$ – is the stationary distribution of the two-dimensional Markovian process $\{s(t), z(t)\}$.

Let's denote the function:

$$H(s,u,z,t,\delta) = \sum_{m=0}^{\infty} e^{jum} P(s,m,z,t,\delta), \qquad (9)$$

where $j = \sqrt{-1}$ – is an imaginary unit.

For these functions it is possible to write down the following Cauchy problem from system (7) and initial condition (8):

$$\begin{cases} \frac{\partial H(s,u,z,t,\delta)}{\partial t} = \frac{\partial H(s,u,z,t,\delta)}{\partial z} - \frac{\partial H(s,u,0,t,\delta)}{\partial z} + \\ + e^{ju} \sum_{\nu=1}^{\infty} \frac{\partial H(\nu,u,0,t,\delta)}{\partial z} A_{\nu k}(z,\delta), \\ H(s,u,z,0,\delta) = R(s,z,\delta). \end{cases} \qquad (10)$$

Let's denote:

$$H(u,z,t,\delta) = \{H(1,u,z,t,\delta), H(2,u,z,t,\delta), \ldots\},$$

also matrix $A(z, \delta)$ with elements $A_{k\nu}(z, \delta)$, then from (10) we receive the following:

$$\begin{cases} \frac{\partial H(u,z,t,\delta)}{\partial t} = \frac{\partial H(u,z,t,\delta)}{\partial z} + \frac{\partial H(u,0,t,\delta)}{\partial z} \times \\ \times \{e^{ju} A(z,\delta) - I\}, \\ H(u,z,0,\delta) = R(z,\delta), \end{cases} \qquad (11)$$

where I – is an identity matrix, and vector $R(z, \delta)$, is defining the initial condition of problem (11) with components $R(s, z, \delta)$, as shown in [3], given by:

$$R(z,\delta) = \kappa_1(\delta) r \int_0^z \left(P(\delta) - A(x,\delta)\right) dx,$$

where r – is the stationary value probability distribution of the embedded Markov chain $\xi(n)$; magnitude $\kappa_1(\delta)$ – is defined by equation:

$$\kappa_1(\delta) = \frac{1}{rA(\delta)E},$$

where matrix $A(\delta)$ is defined by equation:

$$A(\delta) = \int_0^\infty \left(P(\delta) - A(x,\delta)\right) dx.$$

Asymptotic probability distribution of event numbers, the arrival of which is in the SM-process in time t is found in a condition of limit rare changes for the conditions of process.

Let's denote:

$$\lim_{\delta\to 0} H(u,z,t,\delta) = H(u,z,t). \quad (12)$$

In problem (11), considering (12) we shall execute a limiting transition at $\delta \to 0$, then $H(u, z, t)$ we will receive in the set of an independent Cauchy problems [3; 4]:

$$\begin{cases} \dfrac{\partial H(u,z,t)}{\partial t} = \dfrac{\partial H(u,z,t)}{\partial z} + \dfrac{\partial H(u,0,t)}{\partial z}\left\{e^{ju} A(z) - I\right\}, \\ H(u,z,0) = R(z), \end{cases} \quad (13)$$

where:

$$\lim_{\delta\to 0} P(\delta) = I,\ \lim_{\delta\to 0} A(z,\delta) = A(z),\ \lim_{\delta\to 0} A(\delta) = A,$$
$$\lim_{\delta\to 0} \kappa_1(\delta) = \kappa_1,\ \lim_{\delta\to 0} R(z,\delta) = R(z).$$

Considering the kind of matrix $A(z)$, from (12) are we get a set of independent differential equations:

$$\frac{\partial H(s,u,z,t)}{\partial t} = \frac{\partial H(s,u,z,t)}{\partial z} + \frac{\partial H(s,u,0,t)}{\partial z} \times \times \left\{e^{ju} G_{ss}(z) - 1\right\}, \quad (14)$$

the initial conditions are given in the following way:

$$H(s,u,z,0) = R(s,z). \quad (15)$$

The solution of a problem (14–15) by applying of the transformation of Fourier–Stieltjes:

$$\varphi(s,u,\alpha,t) = \int_0^\infty e^{j\alpha z} d_z H(s,u,z,t). \quad (16)$$

The function $\varphi(s, u, \alpha, t)$ is satisfied by the following equation:

$$\frac{\partial \varphi(s,u,\alpha,t)}{\partial t} = -j\alpha\varphi(s,u,\alpha,t) + \frac{\partial H(s,u,0,t)}{\partial z} \times \times \left\{e^{ju} G^*_{ss}(\alpha) - 1\right\}, \quad (17)$$

and the initial condition:

$$\varphi(s,u,\alpha,0) = \int_0^\infty e^{j\alpha z} d_z H(s,u,z,0) = = \int_0^\infty e^{j\alpha z} d_z R(s,z) = R^*(s,\alpha), \quad (18)$$

where

$$G^*_{ss}(\alpha) = \int_0^\infty e^{j\alpha z} d_z G_{ss}(z).$$

The solution of the differential equation (17) is:

$$\varphi(s,u,\alpha,t) = e^{-j\alpha t}\left\{R^*(s,\alpha) + \int_0^t e^{j\alpha\tau} \frac{\partial H(s,u,0,\tau)}{\partial z} \times \times \left(e^{ju} G^*_{ss}(\alpha) - 1\right) d\tau\right\}. \quad (19)$$

As

$$\lim_{t\to\infty} \varphi(s,u,\alpha,t) = \int_0^\infty e^{j\alpha z} d_z H(s,u,z,\infty) = 0,$$

that, having directed t to infinity in equation (19) we get:

$$0 = R^*(s,\alpha) + \int_0^\infty e^{j\alpha\tau} \frac{\partial H(s,u,0,\tau)}{\partial z}\left(e^{ju} G^*_{ss}(\alpha) - 1\right) d\tau.$$

From this equality we find the Fourier transformation in τ from function $\dfrac{\partial H(s,u,0,\tau)}{\partial z}$:

$$\int_0^\infty e^{j\alpha\tau} \frac{\partial H(s,u,0,\tau)}{\partial z} d\tau = R^*(s,\alpha)\left(1 - e^{ju} G^*_{ss}(\alpha)\right)^{-1}.$$

Having executed the reverse Fourier transformation we identify:

$$\frac{\partial H(s,u,0,\tau)}{\partial z} = \frac{1}{2\pi}\int_{-\infty}^\infty e^{-j\alpha\tau} R^*(s,\alpha)\left(1 - e^{ju} G^*_{ss}(\alpha)\right)^{-1} d\alpha.$$

Equation (19) considering the received transformation will be:

$$\varphi(s,u,\alpha,t) = e^{-j\alpha t}\left\{R^*(s,\alpha) + \int_0^t e^{j\alpha\tau} \frac{1}{2\pi} \times \times \int_{-\infty}^\infty e^{-jy\tau} R^*(s,y)\left(1 - e^{ju} G^*_{ss}(y)\right)^{-1} dy \left(e^{ju} G^*_{ss}(\alpha) - 1\right) d\tau\right\}. \quad (20)$$

Knowing, that $H(s, u, \infty, t) = H(s, u, t) = \varphi(s, u, 0, t)$, $R^*(s, 0) = \kappa_1 r_s A_{ss}$, $G^*_{ss}(0) = 1$, we get an expression for function $H(s, u, t)$:

$$H(s,u,t) = \kappa_1 r_s A_{ss} + \frac{1}{2\pi}\int_{-\infty}^\infty \int_0^t e^{-jy\tau} d\tau R^*(s,y) \times \times \left(1 - e^{ju} G^*_{ss}(y)\right)^{-1} dy \left(e^{ju} - 1\right). \quad (21)$$

Considering:

$$\int_0^t e^{-jy\tau} d\tau = \frac{1}{jy}\left(1 - e^{-jyt}\right),$$

and:

$$R^*(s,y) = \int_0^\infty e^{jyz} d_z R(s,z) =$$

$$= \kappa_1 r_s \int_0^\infty e^{jyz} d \int_0^z \left(1 - G_{ss}(x)\right) dx = \frac{\kappa_1 r_s \left(G_{ss}^*(y) - 1\right)}{jy}.$$

We shall denote the equation (21) as:

$$H(s,u,t) = \kappa_1 r_s A_{ss} + \frac{\kappa_1 r_s}{2\pi} \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \times$$
$$\times \left(1 - G_{ss}^*(y)\right)\left(1 - e^{ju} G_{ss}^*(y)\right)^{-1} dy \left(e^{ju} - 1\right). \quad (22)$$

Let's denote $h(u,t) = \sum_s H(s,u,t)$ as:

$$h(u,t) = \sum_s H(s,u,t) = 1 + \frac{\left(e^{ju} - 1\right)\kappa_1}{2\pi} \times$$
$$\times \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \sum_s r_s \left(1 - G_{ss}^*(y)\right)\left(1 - e^{ju} G_{ss}^*(y)\right)^{-1} dy.$$

Knowing, that $h(u,t) = \sum_s H(s,u,t) \approx \sum_s H(s,u,t,\delta) = = \sum_{n=0}^{\infty} e^{jun} P(n,t)$, and sorting it to the right part of the received equation of the exponent levels e^{ju}, it is possible to write down the following asymptotic equation as:

$$h(u,t) \approx \sum_{m=0}^{\infty} e^{jum} P(m,t) = 1 + \frac{\left(e^{ju} - 1\right)\kappa_1}{2\pi} \times$$
$$\times \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \sum_s r_s \left(1 - G_{ss}^*(y)\right) \sum_{m=0}^{\infty} e^{jum} G_{ss}^{*\,n}(y) dy =$$
$$= 1 - \frac{\kappa_1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \sum_s r_s \left(1 - G_{ss}^*(y)\right) dy +$$
$$+ \frac{\kappa_1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \sum_s r_s \left(1 - G_{ss}^*(y)\right)^2 \sum_{m=1}^{\infty} e^{jum} G_{ss}^{*\,m-1}(y) dy.$$

Then the asymptotic probabilities distribution of possibility $P_1(m,t)$, number of events m which start in the SM time process t, look the following way:

$$\begin{cases} P_1(0,t) = 1 - \frac{\kappa_1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \sum_s r_s \left(1 - G_{ss}^*(y)\right) dy, \\ P_1(m,t) = \frac{\kappa_1}{2\pi} \int_{-\infty}^{\infty} \frac{1}{y^2}\left(1 - e^{-jyt}\right) \sum_s r_s \left(1 - G_{ss}^*(y)\right)^2 \times \\ \times \sum_{m=1}^{\infty} e^{jum} G_{ss}^{*\,m-1}(y) dy, \end{cases} \quad (23)$$

where

$$P_1(m,t) \approx P(m,t).$$

Above were obtained formulas that allow us to find the asymptotic probability distribution of the number of events, which happen in the SM-process in time t. This distribution was also found in prelimit situations [1], by using methods of integrated transformation. It is necessary to find out how close the results received by the asymptotic method of analysis are to the results received in prelimit situations. For this purpose there is a magnitude $\Delta = \max_n \left|\hat{F}(n,t) - F(n,t)\right|$, where $\hat{F}(n,t)$ – is the function of distribution, obtained from the asymptotic analysis; $F(n, t)$ – is the function of distribution for prelimit situations, received in [1].

Let:

$$Q = \begin{Bmatrix} -3 & 1 & 2 \\ 3 & -4 & 1 \\ 4 & 2 & -6 \end{Bmatrix},$$

$$G(x) = \begin{Bmatrix} 1 - e^{-5x} & 1 - e^{-2x} & 1 - e^{-10x} \\ 1 - e^{-4x} & 1 - e^{-x} & 1 - e^{-9x} \\ 1 - e^{-6x} & 1 - e^{-3x} & 1 - e^{-8x} \end{Bmatrix}, \quad t = 6.$$

From table it can be seen that at the given parameters of values the magnitude Δ are:

Deviation results of the asymptotic from prelimited analysis

δ	0,001	0,0005	0,0001	0,00005	0,00001
Δ	0,0427	0,0216	0,0042	0,0021	0,0003

On fig. 1–3 are the probability distributions of the number of events in the SM-process, which occurred in time t = 6, obtained by using the asymptotic analysis and the prelimit situation.

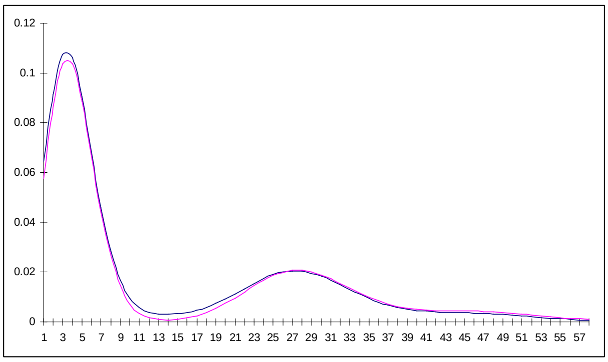


Fig. 1. Distribution probabilities of a number of events for the SM-flow at δ = 0,001

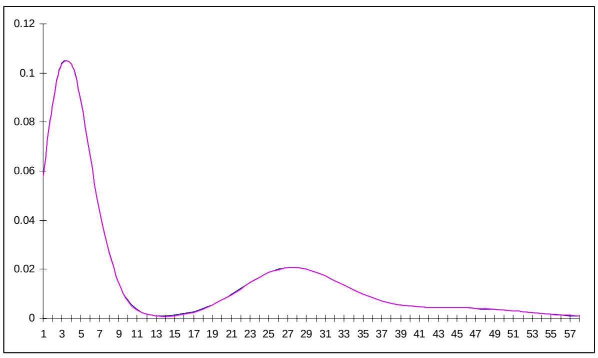


Fig. 2. Distribution probabilities of a number of events for the SM-flow at δ = 0,0001

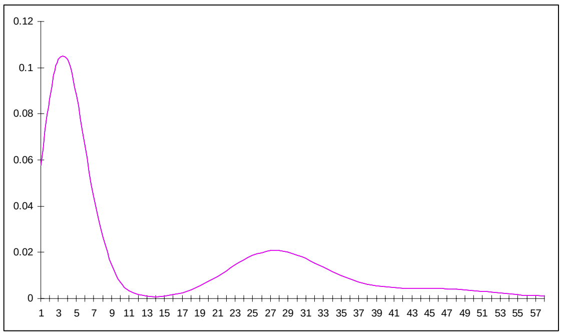


Fig. 3. Distribution probabilities of a number of events for the SM-flow at $\delta = 0{,}00001$

In this research we have found the distribution of the asymptotic probability for a number of events occurring in the SM-flow in time *t*. By reducing the parameters δ, the deviation results of the asymptotic analysis for the prelimited varies: at $\delta \leq 0{,}0001$ it is equal to less than 1 %. Also, it is necessary to notice, that the obtained distribution is multimodal.

References

1. Lopukhova S. V. The asymptotic and numerical methods of studying special currents of homogenous events : candidate paper in physical-mathematical sciences. Tomsk, 2008.

2. Nazarov A. A., Terpugov A. F. The theory of probabilities and accidental processes : textbook. Tomsk : NTL publishers, 2006.

3. Gorbatenko A. E. Studying quasi-decomposable semi-Markov flow потока // The probability theory, mathematical statistics and appendix : journal of scientific article to an international science conference. Minsk, 2010. P. 53–59.

4. Nazarov A. A., Moiseeva S. P. Methods of asymptotic analysis in the theory of mass services. Tomsk : NTL publishers, 2006.

A. A. Hvalko, S. B. Suntsov
JSC “Academician M. F. Reshetnev” Information Satellite Systems”, Russia, Zheleznogorsk

V. G. Butov, A. P. Zhukov, A. A. Yaschuk
Tomsk State University, Russia, Tomsk

MECHANICAL ANALYSIS OF ONBOARD EQUIPMENT AND THE ISSUE OF FINITE-ELEMENT MODEL ADEQUACY

In this article we have considered issues related to the applied software package, designed to conduct onboard mechanical analysis of spacecraft equipment using FEM, together with an algorithm that simplifies the aforementioned models and estimates their correctness.

Keywords: hardware and software complex, mechanical analysis, satellite-borne equipment, simplified model, finite-element model.

The design of a complicated high-tech product, which includes space onboard equipment, is based on the currently used analysis methods. This allows building complicated structures and analyzing them thoroughly. Today these methods are successfully used is computers and modern CAE systems (Computer-Aided Engineering – engineering analysis systems). By applying this method, difficulties during the design and the strength analysis could be resolved, by bringing design models much closer to the existing structures.

The Finite Element Method (FEM) is one of the most universal methods of mechanical analysis, which is performed for different structures. It is, in fact, a variation method used to calculate detailed irregular models. Its capabilities of constructing overall systems, which can calculate all structural parts without separating the elements of the structure, give this method more advantages. The design of onboard equipment is characterized by the tendency of its design model improvements; this can be explained in two ways:

– in some cases it is possible to significantly reduce the mass of the onboard equipment, owing to more a detailed analyses of voltage and the distorted construction;

– this analysis is often the only method to estimate structure behavior in space conditions and on the way to space after onboard equipment simulation; ground operations in most cases are extremely limited.

Nevertheless, together with a number of beneficial characteristics FEM also has some disadvantages. The continuous progress in computational technologies stipulates the building of more detailed models, in the size of which there could be approximately 10^5 equation members. It is easier to perform a dynamic analysis of a reduced model while calculating the detailed models; this is still restricted in terms of the obtained results validity.

To eliminate such difficulties, the created applied programs packages are used to perform mechanical analysis, including analysis of the reduced models. ASONIKA-TM is the most popular product used for these purposes. It is used for analyzing mechanical characteristics of the upper level structures (cabinet, rack, and unit). Printed circuit boards with lumped EEE parts and associated material data package are not taken in account for this analysis. More over, it does not foresee a possibility to generate a design model automatically.

Due to the aforementioned, a dedicated program package had to be developed. In result of this necessity, the hardware & software system (HSS) was developed to perform onboard equipment mechanical analysis. This software has been generated to determine eigenfrequencies, effective masses, stressed-deformed states of models being exposed to linear acceleration, sine and wideband random vibrations, shocks. The objects of the HSS analysis are reduced Finite Element Models from onboard equipment accounting PCBs with EEE parts simulators.

CAE-system ANSYS and program "onboard equipment mechanical analysis" are the main parts of this HSS, since they are generated to analyze the mechanical characteristics of unspecified and unified onboard equipment structures. The dynamic analysis of onboard unit linear models is performed to verify if they can be reduced and implemented as a numerical mathematical model. Based on the analysis results, the main ways of reducing (simplifying) onboard structural elements have been defined. Also considered was the fact that implemented modifications should not cause any unacceptable errors of numeric analysis results.

For the EEE parts to be included into the onboard equipment models a special algorithm was developed; it can generate the APDL file (ANSYS Parametric Design Language) and allow automating some time-consuming modeling procedures. This algorithm substitutes EEE parts to use lumped mass elements, but not the associated increased density of materials used for frame parts; as it is done in similar calculation systems. HSS adaptation for the final user is also ensured by a friendly interface and database of onboard equipment unified structures.

Based on the requirements specified for HSS, a relay type database has selected with access of being performed via ODBC (Open Data Base Connectivity) under Microsoft Access control. Database comprises service tables, providing a description of the contents and links, existing within the data tables; and the data tables as well. The database is designed accounting its possible enlargement.

Service tables contain data referred to other tables, types of information in the database (unit, frame, fixation, material, data source, etc.), and data fields of all tables (size, variables range, description of meaning, etc.).

The following informational blocks are included in the data tables:

– devices;

– independent units;

– materials;

– unit structural elements;

– location and contact of units in each device;

– mechanical loads;

– performed analyses (calculations);

– EEE parts (automatically imported from PCB design systems).

Onboard equipment database of the structural design can be seen as an example, since it has a "forest" structure, comprising different "tree" elements. Thus, the top of the tree is the onboard equipment – a structural element that has a reference to its description in a dedicated line of the table. Each line in the onboard equipment table shall refer to record of layout schematics.

Onboard equipment layout schematics may have the following affiliated (daughter) elements: units comprised, additional structural elements (bracket, plank, etc.), and the calculation results. It should be noted, that each affiliated element shall have a line of its own description and its own affiliated elements. Thus, a unit with affiliated elements may have EEE parts that can be finally automatically accommodated into random groups, to be further mounted on different unit functional parts (PCBs, modules, sides of the basic plate (aluminum plate under the PCB), etc). Affiliated element calculation may result in onboard equipment reaction to particular types of mechanical loads.

Generating the database content, performing database operations, constructing onboard equipment models, and mechanical analysis procedures are performed via the user interface, which is the program shell of the HSS "onboard equipment mechanical analysis" demonstrated in fig. 1.

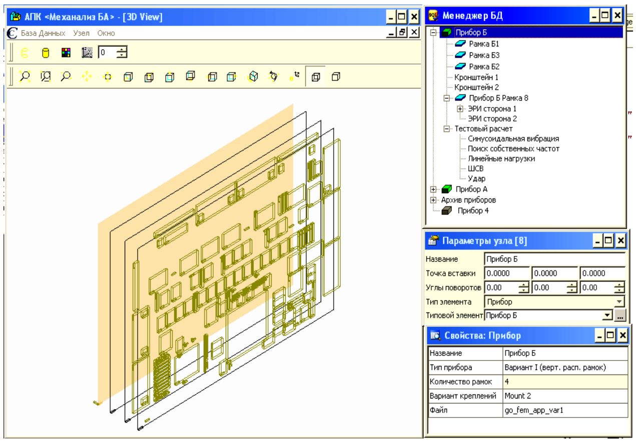


Fig. 1. Graphic user interface

The program shell has 4 main windows:

– the Main window of the HSS "onboard equipment mechanical analysis" comprises the Manu with the main functions, allowing to manage database information; it is responsible for the visualization of the built models and the coordination of the other windows operations;

– the window "Data base Manager" ensures operations with the onboard equipment layout tree and automated building of a model with the accommodation of EEE part simulators;

– the window "Nod parameters" specifies the location of any structural element in relation to a parental, and displays these parameters;

– the window "Properties" allows the user to review and specify the parameters of the structural elements, referring to a layout element specified in the "Database Manager" window.

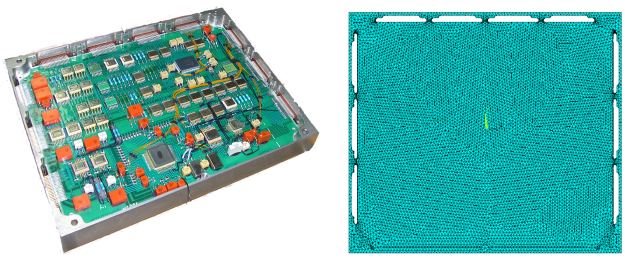
Reduction of onboard equipment structural elements 3D-models. Based on the experienced mechanical analyses of different structure FEMs built with CAD, it can be stated that it is not reasonable to use detailed CAD 3D-models to perform numerical analysis, since the recourses required for that overcome the computer's capabilities. HSS foresees the following algorithms to be used for the unit models' reduction:

– reduction of onboard equipment structural elements' 3D-models;

– automatic generation of reduced onboard equipment FEMs with EEE parts' mass simulators, accommodated in it.

The reduction of a complete onboard equipment model is performed by reducing its structural elements (frames, brackets, planks, etc), and their further connection. For the onboard equipment unified structures, recorded into the HSS database, it is predictable to automatically generate a point or allocated contacts by using additional elements which produce a mechanical join between structural parts, which are compliant with the actual equipment connected on a sufficient extent.

The operation of the reduction algorithm is demonstrated with an aluminum frame of the onboard equipment unit given in fig. 2, *a*; its complete FEM is demonstrated in fig. 2, *b*. This model is based on 92 074 solid-state elements (2 order tetrahedron SOLID92) and comprises 172 615 nods.



a *b*

Fig. 2. Appearance of the onboard unit (*a*); detailed FEM of its frame (*b*)

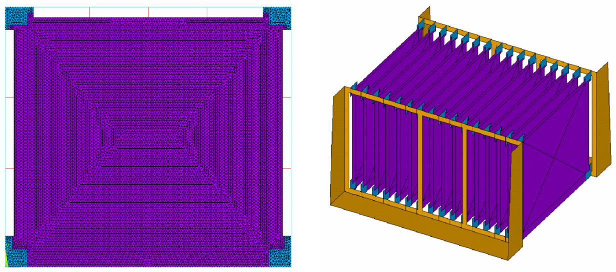
The reduction of structural elements is performed in two ways: the first one is by decreasing the task size; the second one is by excluding insignificant details. The result of this structure's modification was beam elements BEAM188 (frame) and BEAM188 (supplementary elements), shell elements SHELL181 (plate). It should be noted that model of this plate was built up as a multi-layer structure with a number of layers referring to the number of structural elements on it – such as glue, insulating spacers, PCB, coating, etc. The result of the 3D-model reduction is demonstrated in fig. 3, *a*. The reduced FEM of the frame structure comprises 21 278 elements and 11 684 nods.

The general mathematical description of element types used for complete and reduced models may be demonstrated in the form of a matrix equation:

$$[M]\left\{\ddot{u}\right\}+[C]\left\{\dot{u}\right\}+[K]\{u\}=\{F\}, \qquad (1)$$

where {u} is a vector of nodal displacement for the entire unit body; $\left\{\dot{u}\right\}$, $\left\{\ddot{u}\right\}$ are velocity vectors and vectors of each body nod velocity [*K*], [*C*], [*M*] are "global" matrixes of stiffness, damping, and masses of the entire body, {F} is a vector of equivalent nodal forces for the entire body. As it has been shown in equation (1), the mathematical representation of the reduced model is more beneficial for numerical modeling.

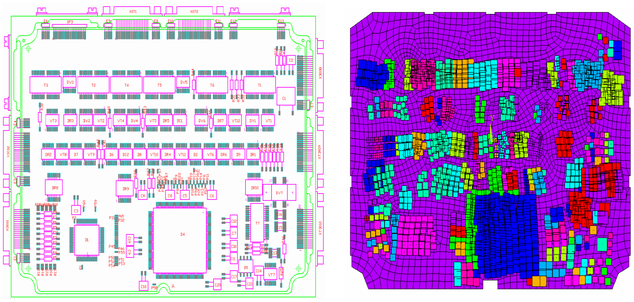
Further, the entire onboard equipment model is automatically built by comprising 16 units as shown in fig. 3, *b*. After that, the EEE parts' mass simulators are automatically accommodated onto units of the reduced onboard equipment model. This accommodation can be demonstrated with one of the units, used as an example.



a *b*

Fig. 3. Unit reduced model (*a*); onboard equipment reduced model (*b*)

The general view of the unit with PCBs and EEE parts accommodated on PCBs is shown in fig. 4, *a*. The generation of the data file for the parts' automatic accommodation on PCB is based on their attributes, generated accounted on the topology of the developed PCB. For the mass simulators to be included in the unit FEM, a special algorithm and input file APDL for ANSYS have been developed. The presence of the elements on PCB has been considered by including additional masses in the dedicated nods of the FEM. For this purpose, their each size and location onboard were used to define the network nods that were covered by the element plane. Then mass of each element was evenly distributed between two selected nods. In these nods, elements as MASS21 were generated; they had a distributed mass. The results of this algorithm operation are demonstrated in fig. 4, *b*.



a *b*

Fig. 4. General view of the unit with PCB and EEE parts (*a*); unit plate with automatically generated EEE parts' mass simulators (*b*)

Estimated conformity of the developed models. Eigenfrequencies of one of the units were calculated; the obtained results were compared to the experimental in order to estimate the conformity of the onboard equipment reduced model. One of the main parameters of the reduced model is a number of finite elements (FE), comprised into the model. Consequently, when changing the number of FE, it is possible to vary deviations of the eigenfrequencies of the design model relatively to the empirical one. The following equation was used to calculate the relative deviation:

$$\varepsilon = \frac{f_e - f_P}{f_e} \cdot 100\ \%, \tag{2}$$

where f_e stands for frequencies obtained by applying empirical methods; f_P stands for frequencies, obtained while calculating the reduced model.

Regarding fig. 3, *a*, the main quantity of the nods and of the reduced model is accommodated on a plate; this means that it is reasonable to use the plate to perform the experiment with a changing number of FE. Fig. 5 demonstrates relative deviations' dependability on the number of elements, comprised in the model, when determining eigenfrequencies, used in modeling the plate.

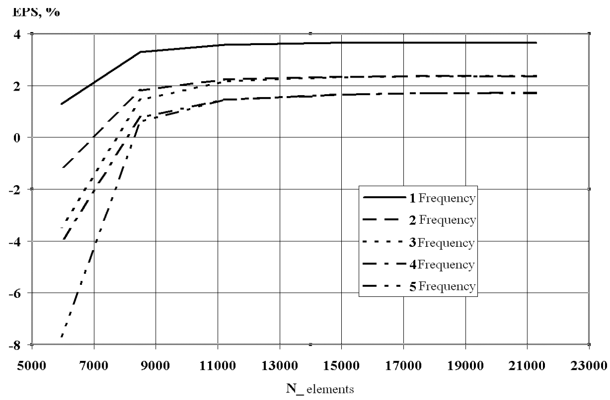


Fig. 5. Relative deviations' dependability upon the number of the model FE

As it is shown in the figure above: with the growth of FE quantity, the deviations at first sustain their actual rate, but as soon as they are equal to 9 000, their quantity falls down abruptly, and goes the same way up to 20 000. There is no sense in the further increase in the quantity of FE in the model, since the mathematical model becomes significantly complicated, while its stability does not improve much. Consequently, it is acceptable to decrease the number of FE to 9 000. Fig. 6 demonstrates the results of the dynamic analysis, performed with the unit model under sine vibration, and with 20 000 of FE at the plate of the unit. The acceleration was calculated at the point of sensor installation which is perpendicular to the unit plane. The damping factor was equal to 2 %. The figure also provides results obtained during the tests of this unit.

The calculation results performed for the entire unit were compared with the results of the actual experiment in order to confirm the application of a correct philosophy, used for design model reduction and adjustment. The vibration test and mechanical analysis were performed using real onboard equipment, comprising 16 units, and using its numeric model, provided in fig. 3, *b*. It should be noted that the mechanical analysis was performed by using two numerical models: I – less-weight model with less than 20 000 FE on the unit plate; II – light model with at least 9 000 FE on the unit plate. Only clearly marked peaks were considered. The frequencies, defined by experience were placed into compliance with the design frequencies, the values of which were closest to the experienced ones.

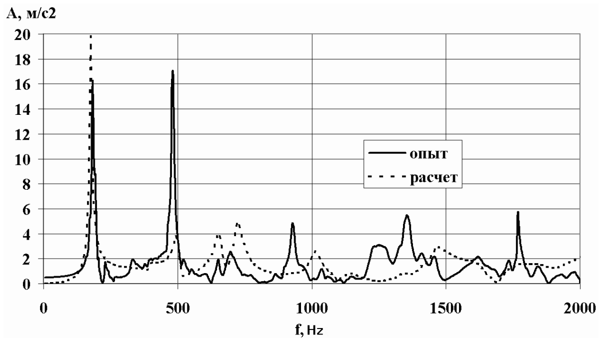


Fig. 6. Results of the design and experimental models' dynamic analysis

Table demonstrates the design and experimental frequencies for type I and II models. Besides that, the effective masses are also provided. If the effective masses are analyzed it can be seen, that some frequencies induced during the experiment have an effective mass close to 0 during calculations. The probable reason can be that the boundaries specified for the calculations do not fully comply with the experimental conditions.

The results compared for type I and II models have demonstrated that the type II light model can provide positive results as well. More evident changes could be observed for some effective masses. For instance, is type II model: the frequency effective mass with number 6 increased approximately by two times, while the frequency effective mass with number 7 decreased.

Besides the implementation of functions and tasks assigned to the developed HSS, it can also smooth interaction between the designer and CAE-system (for example, ANSYS). The developed structure of the database, interface, and algorithms of design models' reduction allow the designer to be rapid with the necessary data, and perform the mechanical analysis of onboard equipment when using unified structures. The schematics of model conformity estimations provided above, permitted the provision of qualitative content for the HSS database, comprising unified structures of onboard equipment and performing valid calculations. There is also a possibility to perform a mechanical analysis for the models created with other CAD programs. The activities performed in terms of this analysis, have demonstrated the effectiveness of the applied methods in resolving difficulties of Designer/CAE-system operation; thus, activities related to the content of the database and HSS improvement are in progress.

Design and experimental frequencies for the models type I and II

Frequency №	Object				
	Real onboard equipment	Model I		Model II	
	Frequency, experimental value, Hz	Frequency, design value, Hz	Effective mass, kg	Frequency, design value, Hz	Effective mass, kg
1	183.0	171.1	0.291695	176.0	0.288 266
2	–	320.2	3.90×10^{-04}	324.6	6.96×10^{-04}
3	–	388.8	7.63×10^{-06}	398.0	1.31×10^{-03}
4	482.7	490.9	2.13×10^{-02}	498.2	1.83×10^{-02}
5	–	545.8	3.54×10^{-03}	554.7	2.77×10^{-03}
6	653.0	660.0	1.30×10^{-02}	655.5	2.53×10^{-02}
7	700.0	729.9	2.12×10^{-02}	722.9	1.80×10^{-02}
8	–	740.2	3.98×10^{-03}	745.5	2.00×10^{-03}
9	–	854.6	1.42×10^{-04}	857.2	1.06×10^{-04}
10	–	885.6	1.36×10^{-04}	890.0	7.68×10^{-05}
11	930.9	962.7	9.71×10^{-05}	943.5	2.19×10^{-04}
12	1 043.0	1 018.2	6.74×10^{-03}	1 013.5	6.64×10^{-03}
13	–	1 071.9	2.01×10^{-03}	1 083.9	1.23×10^{-03}
14	–	1 151.1	9.62×10^{-05}	1 121.4	9.41×10^{-04}
15	–	1 152.9	1.06×10^{-03}	1 148.3	1.02×10^{-03}
16	1 237.0	1 258.4	7.21×10^{-05}	1 259.3	6.51×10^{-07}
17	1 276.0	1 276.1	7.74×10^{-05}	1 288.8	2.98×10^{-04}
18	–	1 336.3	3.11×10^{-03}	1 341.6	4.63×10^{-06}
19	1 359.0	1 362.4	1.58×10^{-03}	1 343.1	3.83×10^{-03}
20	1 418.0	1 406.9	7.05×10^{-04}	1 410.4	1.15×10^{-03}
21	1 462.0	1 456.8	1.21×10^{-02}	1 460.1	1.13×10^{-02}

I. V. Ignatyev, A. E. Kovrov
Bratsk State University, Russia, Bratsk

ALGORITHM OF CHOICE FOR AUTOMATIC EXCITATION REGULATOR SETTINGS IN MULTIMACHINE ELECTRIC POWER SYSTEMS

A program-realized coordinated algorithm of choice for automatic excitation regulator settings has been developed. This algorithm is based on the resultant theory and applies a mathematical model which is synthesized by experimental frequency characteristics of an electric power system.

Keywords: automatic excitation regulator, resultant, stability.

The ensuring of stability in electric power systems (EPS) and the damping of fluctuation are realized by automatic excitation regulators (AER) which are placed in power station generators [1].

Today it is matter of current coordinated choices in AER stability coefficients, in conditions of ensuring multimachine EPS with the required intermediary processes quality.

To solve this problem, algorithms have been developed based on the D-dividing method [2–4]. These algorithms assume serial choices for AER settings and for each separate station by a calculated field of stability. In this case, transition from one station to another is realized by increasing the system stability extent. However, in connection with the complexity of using special-purpose functions, these algorithms do not allow us to ensure acceptable EPS property damp with a greater AER number.

Another way to solve this problem is to use algorithms based on the principal matrix calculation values of the Gorev-Park linear differential equation [5–7]. However, these algorithms have a complex operative control in multimachine EPS faults:

1) they are characterized by a high order of differential equation demanding significant calculations;

2) they admit average data value for elements of large EPS units and subsystem during large intervals. The last results in non-conformity mathematic model forming during a current mode situation.

In [8; 9] it has been shown that EPS mathematic models may be obtained through the result of experimental frequency characteristic stability parameters as a characteristic polynom. It allows avoiding of many admission and simulation errors, which are typical for calculated methods; the describing the upper and lower dimension of the AER choice problem settings to coordinate them betimes during the current EPS working conditions.

In this article new coordinated choice algorithms of AER settings for multimachine EPS were obtained. This allowed the consideration of the decision merits offered in works [8; 9] and excluded the mistakes in previously developed algorithms. The realization of the new algorithm program was developed by MATLAB software.

The coordinated choice algorithm of AER setting description. The new algorithm is based on the resultant theory [10]. Resultant is a two polynomial's coefficient function, turning into null; it is essential and satisfies the condition for exciting the common root of these polynomials.

Let's examine the first polynomial characteristic of the studied EPS polynom

$$D(p) = a_0 p^n + a_1 p^{n-1} + a_2 p^{n-2} + \ldots + a_{n-1} p + a_n, \quad (1)$$

in the coefficient where the AES stabilization setting of channels is included. In the final form we can write down:

$$a_i = a_{i0} + \sum_{j=1}^{r} a_{ij} k_j + \sum_{j=1}^{r}\sum_{k=1}^{r} a_{ijk} k_j k_k + \sum_{j=1}^{r}\sum_{k=1}^{r}\sum_{l=1}^{r} a_{ijkl} k_j k_k k_l + \ldots,$$

where r is the amount of generators equipped by AER; k_j, k_k, k_l, ... are stabilizing coefficients.

Polynom's roots with almost no real part modulus determine the static stability degree of analyzing EPS and are called dominate.

The second polynomial is an auxiliary function

$$Q(\lambda) = b_0 \lambda^m + b_1 \lambda^{m-1} + b_2 \lambda^{m-2} + \ldots + b_{m-1}\lambda + b_m, \quad (2)$$

the roots of which are chosen equally for the required meanings of dominate roots.

The issue is to ensure polynom equality of dominate roots (1) with roots of polynom (2) to account the stabilization coefficient's variation and therefore a_i. The acquired AER coefficients should meet this realization in practice settings.

Let us examine the stages of algorithm work in the EPS, for the improvement of damp property; it is necessary to coordinate the AER settings of three equivalent generators. Let us suppose that all these generators are equipped by AER microprocessors.

At the first stage, based on the work procedure offered [8], we shall obtain a mathematic model for studying the EPS as a polynom (1). In the beginning, we shall pick control loops out of the system; these include stabilization channels on a chosen voltage frequency Δf, and its derivative f'. Such an approach is stipulated by placing low-frequency fluctuating damp on the selected AER channels. For the mathematical description of these channels we will use transmission function $F_1(p)$, $F_2(p)$ and $F_3(p)$, the indexes of which correspond to ordinal numbers of the generators.

The remaining part of the EPS, including other AER channels, symmetries about the chosen control loops (fig. 1) through the principal $W_{11}(p)$, $W_{22}(p)$, $W_{33}(p)$ and crosses $W_{12}(p)$, $W_{13}(p)$, $W_{21}(p)$, $W_{23}(p)$, $W_{31}(p)$, $W_{32}(p)$ transmission functions of mode stabilization parameters.

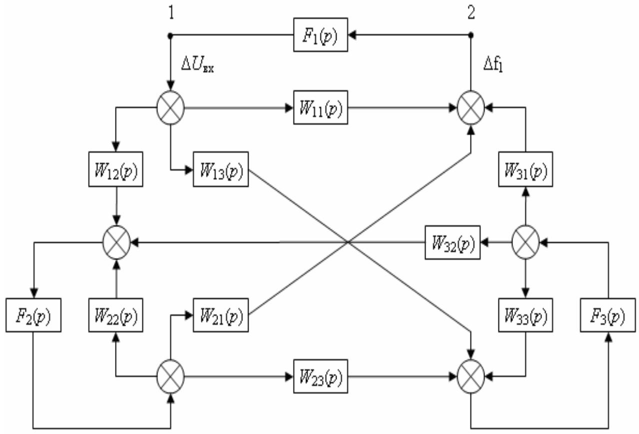


Fig. 1. Three-loop structural representation of EPS

Further, using Mayson's formula we write down the transmission function of regulated EPS from input 1 to output 2 in accordance to fig. 1:

$$W_p = \frac{W_{11}\overline{W}_{12} + \overline{W}_{12}F_2 + \overline{W}_{13}F_3 + \overline{W}_{123}F_2F_3}{1-(W_{11}F_1 + W_{22}F_2 + W_{33}F_3 + \overline{W}_{12}F_1F_2 + \overline{W}_{13}F_1F_3 + \overline{W}_{23}F_2F_3 + \overline{W}_{123}F_1F_2F_3)}, \quad (3)$$

where $\overline{W}_{12}$, $\overline{W}_{13}$, $\overline{W}_{23}$, $\overline{W}_{123}$ are symmetrical cross transmission functions which are accordingly formed to minors and matrix determinants:

$$\overline{W} = \begin{pmatrix} W_{11} & W_{12} & W_{13} \\ W_{21} & W_{22} & W_{23} \\ W_{31} & W_{32} & W_{33} \end{pmatrix}. \quad (4)$$

Note that in expressions (3) and (4) (as it shall be in following complex formulas), we do not use operator "p" for simplification.

The authors of this article have analyzed the dynamic properties of the modern microprocessor AER before. The results of the analysis have produced a transmission function, obtained for choosing stabilizing channels in EPS:

$$F(p) = \frac{2p}{(1+2p)} \cdot \frac{k_{0f}}{(1+0.2p)} + 0.5p\,\frac{k_{1f}}{1+0.5p}, \quad (5)$$

where k_{0f} is the stabilization coefficient of frequency deviation, k_{0f} = 0–15 puU_f/Hz; k_{1f} is the stabilization coefficient of frequency derivative, k_{1f} = 0–5 puU_f/Hz/s.

Reducing the formula (5) to the same denominator, we can write the transmission function of stabilizing channel for AER i-th generator:

$$F_i(p) = \frac{(p^2+2p)\,k_{0fi} + (0.2p^3 + 1.1p^2 + 0.5p)\,k_{1fi}}{0.2p^3 + 1.5p^2 + 2.7p + 1} = \frac{f_i'}{f_i}. \quad (6)$$

Substituting expression (6) to denominator (3) we obtain the characteristic EPS polynom in the general form:

$$D(p) = \Delta_N f_1 f_2 f_3 - W'_{11} f'_1 f_2 f_3 - W'_{22} f_1 f'_2 f_3 - W'_{33} f_1 f_2 f'_3 - \\ - \overline{W}'_{12} f'_1 f'_2 f_3 - \overline{W}'_{13} f'_1 f_2 f'_3 - \overline{W}'_{23} f_1 f'_2 f'_3 - \overline{W}'_{123} f'_1 f'_2 f'_3, \quad (7)$$

where W'_{ij}, $\overline{W}'_{ij}$, $\overline{W}'_{ijk}$ are own numerators and the cross transmission function of mode stabilization parameters; Δ_N is the same denominator for these functions.

For calculating numerical values of zeros and poles for transmitted functions, which are included in formula (7), the authors, according to the method of EPS parametric identification [9] had developed the following procedure:

1) determine the complex sample of principal and cross frequency characteristics for unlocked systems by measuring the EPS time characteristics locked in points 1 and 2 (fig. 1) with help of the fast Fourier transformation;

2) configure the analysis of obtaining frequency characteristics with the purpose of exposing dominating poles and zeros; determine EPS basic dynamic properties in the essential frequency rate;

3) approximate the remainder frequency characteristic by a smooth fractional-rational function with help of the least-square method.

By substituting the (7) zeros and poles obtained previously, we represent a multi-parametric model of EPS analysis as a polynom (1). In result each coefficient of the polynom will include a non-linear combination of AER required settings.

The second stage of the algorithm includes the realization of two procedures: the calculation of characteristic polynom roots and the analysis of their location on the complex plane.

The first procedure making after substituting in expression (1) is the stability value coefficient set in the AER of each generator.

The second procedure can conduct a quantity estimate of the analyzed system's stability degree for the current mode situation. This parameter is determined by the true part modules of the polynom conjugate complex root (1) pair, nearest to the imaginary axis.

The third stage is related to the procedure of auxiliary function calculation. This procedure includes the following steps:

1) from all roots combination of the characteristic polynom, corresponding to electromechanical system traffic component, we select m dominating roots (m = 6). The number of m is determined by the amount of coordinated stability coefficients;

2) proceeding onward from the condition of required quality for the transmitted processes in the EPS, we set desirable locations of the chosen dominant roots on the complex plane; meaning their real and imaginary parts;

3) coefficients b_i of function (2) are calculated by preset values of dominant roots.

Thus, in the examined case it is necessary to obtain such stability coefficients, which will ensure the equality of six polynom dominant roots (1) from their total amount n, to roots of auxiliary functions.

On the fourth stage of the solution, we shall compound the resultant matrix by the Silvester rule – the order of which is equal to $m + n$

$$M = \begin{pmatrix} 0 & 0 & \dots & 0 & a_0 & a_1 & \dots & a_{n-2} & a_{n-1} & a_n \\ 0 & 0 & \dots & a_0 & a_1 & a_2 & \dots & a_{n-1} & a_n & 0 \\ \vdots & \iddots & & \vdots & \vdots & & \iddots & & & \\ a_0 & a_1 & \dots & a_{n-2} & a_{n-1} & a_n & \dots & & 0 & 0 \\ b_0 & b_1 & \dots & b_{m-1} & b_m & & \dots & & 0 & 0 \\ 0 & b_0 & \dots & \dots & b_{m-1} & b_m & \dots & & 0 & 0 \\ \vdots & & \ddots & & & \vdots & \ddots & & & \vdots \\ 0 & \dots & & \dots & 0 & b_0 & \dots & \dots & b_m & 0 \\ 0 & & \dots & & \dots & 0 & b_0 & \dots & \dots & b_m \end{pmatrix}, \quad (8)$$

where elements, above a_0, b_m and below b_0 and a_n, are equal to null.

Concordantly, in the resultant theorem [10], for the existence of m common polynom roots (1) and (2), it is quite necessary to perform the following:

$$R(D,Q) = R_1(D,Q) = R_2(D,Q) = \ldots = R_{m-1}(D,Q) = 0, \quad (9)$$

where $R(D, Q)$ is the matrix determinant, which is formed from matrix M by the disposal its m–1 first and last rows, and $2m$–2 last columns; $R_1(D, Q)$, $R_2(D, Q)$, ..., $R_{m\text{-}1}(D, Q)$ are matrix determinants, which are formed from the previous matrix by serial change in its last column for each of the following columns of matrix M.

Note that the pointed determinant in the theorem is of nonlinear function, containing the AER stability coefficient as a variable.

In conformity with the theorem let us calculate determinants $R(D, Q)$, $R_1(D, Q)$, ..., $R_5(D, Q)$ and then write down a system of nonlinear equations:

$$\begin{cases} R(D,Q) = 0, \\ R_1(D,Q) = 0, \\ R_2(D,Q) = 0, \\ R_3(D,Q) = 0, \\ R_4(D,Q) = 0, \\ R_5(D,Q) = 0. \end{cases} \quad (10)$$

Numerical solutions of the equation system (10), computed by procedures executed at Matlab, allow us to obtain the stability coefficients, provided by polynom dominant roots (1) and required values with a set accuracy.

If in result of calculations the acquired AER settings do not correspond to practice values, then the auxiliary function roots should be changed; the calculation procedure will provide once more, beginning from the third stage. In our case the modification of the function roots (2) is based on the motion of two extreme pair dominant roots directed towards each other.

Thus, the solution of the set problem can have an iterative character; however we can reach the acceptable values of AER settings for some iteration as a rule.

Calculation examples. For the validation of algorithm functionability, a computer experiment was conducted by providing calculations of optimal AER settings for Irkutsk EPS generators.

The realization of given the purpose was reached by solving the following problems:

– data collecting for Irkutsk EPS's elements (parameters of transmit lines, transformers, generators, reactor and control unites);

– forming of EPS model in Simulink;

– checkout of the developed model's adequacy in the real work of the EPS;

– calculating a typical mode which is in the Irkutsk EPS;

– submission of revolt influence in choosing the EPS's points (fig. 1); the recording of obtained time characteristics, reflecting dynamic properties of the system;

– procedure execution of the coordinated choice of AER setting algorithm.

The analysis was provided by SimPowerSystems and Simulink packages of Matlab expansions. SimPowerSystems was used for different EPS device model forming: power systems, systems of automatic regulating, control and measurement systems. The Simulink package was used for AER microprocessor block development and the realization of imitative modeling.

The Irkutsk EPS is a model equivalent to the generators of the Bratsk, Ust-Ilimsk and Krasnoyarsk Hydroelectric Power Plants (HPP). The generating of the Irkutsk HPP is taken into account in load. The fragment of the model realized in Matlab is shown in fig. 2.

During the typical mode data formation for the supervisor lists, the transmit lines (500 kV) of the Irkutsk EPS were used.

To begin the simulation from the set mode, the initialization of synchronic generators and AER for a set load level had been made. Initialization means a setting mode of bidirectional load for a generator equivalent to the Bratsk HPP. For the Ust-Ilimsk and Krasnoyarsk HPP a mode of set active power production supported by a constant stator voltage was chosen.

After the supply revolt influences and executes the procedures of the first two algorithm stages, we get a mathematic description of the EPS as polynom (1); the order of which is $n = 28$. Results of the dominant root calculation for the polynom of the established settings for generators AER are listed in the tab. 1.

From tab. 1 it can be seen that the analyzed system has low damp properties because the pair of dominant and imaginary roots $\omega = 7.22$ rad/s has low a damping $\alpha = 0.23$. For such a mode it is necessary to change AER settings to increase the EPS stability. Conforming this are the roots of auxiliary function with set values' real part are $\alpha < -0.6$.

In result of the following calculation procedures, the next stages of algorithm stability coefficients provided polynom dominant roots (1) with set damping. The results of these calculations are listed in tab. 2.

The calculations obtained by us, with the use of such software programs as "Regim", "Sborka", and "Poisk" (developed by the Saint Petersburg State Polytechnic University); completely confirm the results (see tab. 2). This is the correct work of the algorithm and the possibility of its exploitation for the coordinated choice of AER settings in conditions of ensuring the EPS's required level damping transients.

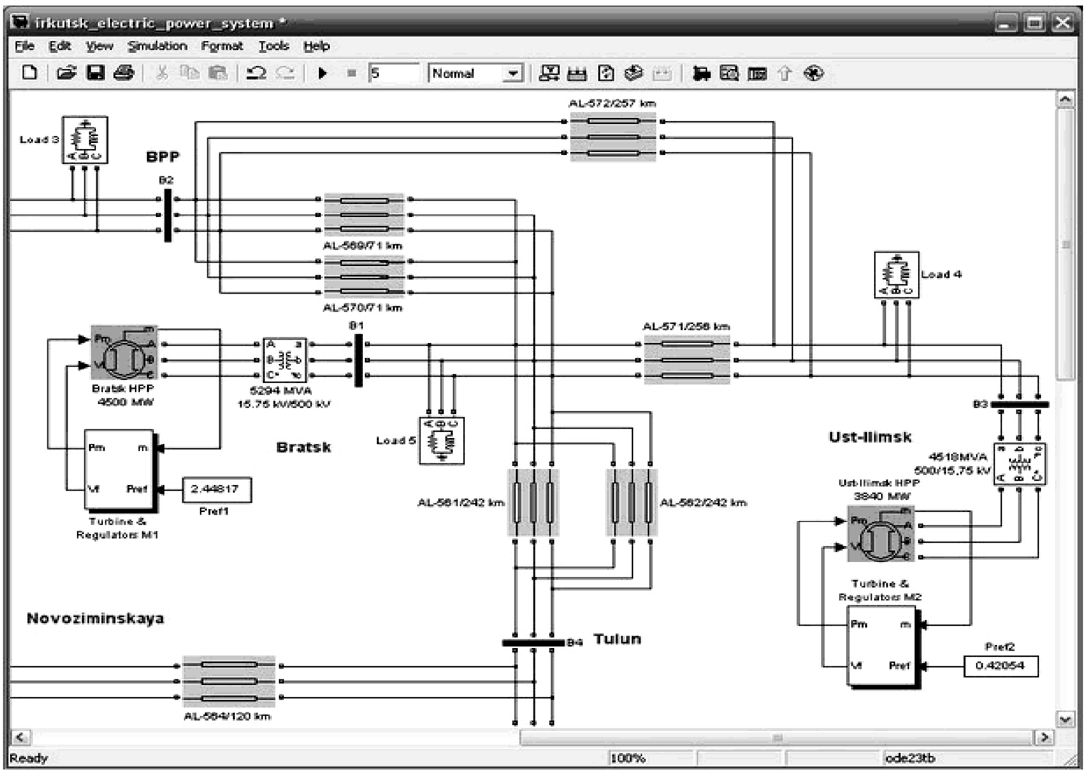


Fig. 2. Fragment of Irkutsk EPS in Simulink

Table 1

Results of dominant root calculations

Dominant roots			AER settings					
Number of root	Real part	Imaginary part	Bratsk HPP		Krasnoyarsk HPP		Ust-Ilimsk HPP	
			k_{0f}	k_{1f}	k_{0f}	k_{1f}	k_{0f}	k_{1f}
1	–0.23	7.22	10	2	15	0	7	5
2	–0.23	–7.22						
3	–0.77	5.35						
4	–0.77	–5.35						
5	–0.97	6.34						
6	–0.97	–6.34						

Table 2

Results of AER setting calculations

Roots of auxiliary function			AER settings					
Number of root	Real part	Imaginary part	Bratsk HPP		Krasnoyarsk HPP		Ust-Ilimsk HPP	
			k_{0f}	k_{1f}	k_{0f}	k_{1f}	k_{0f}	k_{1f}
1	–0.65	7.22	7	2	10	5	5	2
2	–0.65	–7.22						
3	–0.77	5.35						
4	–0.77	–5.35						
5	–0.8	6.34						
6	–0.8	–6.34						

These are the final results:

– a concept resultant was proposed and a new algorithm of AER settings coordinated choice in multimachine EPSs had been developed;

– the new algorithms include mathematic models, synthesized by experimental data;

– a computer experiment has shown that the algorithm allows efficient solving of the ensured quality problem in the EPS transition process;

– the obtained algorithm allows dimension reduction of the choice AER settings' problem essentially; it can be used in real operating condition for the EPS, increasing its stability.

References

1. Ovcharenko N. I. Automation of Electric Station and Electric Power system : Tutorial for High Schools / Under ed. of A. F. Dyakov. M. : Publishing house SC ENAS, 2003.

2. Litkens I. V., Filinskaya N. G. Choice of AER settings in multimachine EPS // Electricity. 1986. P. 15–19.

3. Zekkel A. S. Estimation of regulating quality and method of stability adjustment of AER generators // Electricity. 1988. Vol. 5. P. 15–21.

4. Abdul-Zade V. M, Aliev D. G., Guseynov A. M. Choice of AER generators settings by results of statistical stability analyze // Electricity. 1990. Vol. 3. P. 54–58.

5. Simeonova K. Zh., Stroev V. A. Problems of AER parameters choice in complex electric power system // News of AS SSSR. Power engineering and transport. 1987. Vol. 5. P. 6–71.

6. Maslennikov V. A. Software for calculation of oscillating static stability of electric power system // News of high schools : Power engineering. 1995. Vol. 3–4. P. 33–38.

7. Maslennikov V. A., Ustinov S. M. Software "POISK" – Advanced Information Technology for Power Systems Stability Control // Proc. of International Conference on Informatics and Control (ICI&C'97). St.-Petersburg, 1997. Vol. 2. P. 696–703.

8. Doynikov A. N. Construction of mathematic model for centralize regulating of generators excitation in complex EPS by samples of real signals / Electrification of Siberian metallurgical enterprises / Compose and common ed. B. I. Kudrin. Iss.8. Tomsk : Publishing house Tomsk Univers., 1999. P. 198–202.

9. Doynikov A. N., Grigorieva T. A. Analyze of dynamic properties and model synthesis of electric power system by mode frequency characteristics // Information technology and problems of mathematic simulation of complex systems. Irkutsk : ISURE Institute "Modern Technologies. System Analysis. Modeling", 2005. P. 91–102.

10. Kalinina E. A., Uteshev A. U., Theory of exception: The manual. St. Petersburg. : Publishing house SPb. SU SRI of chemistry, 2002.

A. B. Ivanovsky
RosNeft – Inform Ltd. Krasnoyarsk Subsidiary, Russia, Krasnoyarsk

V. V. Shishov
Krasnoyarsk State Institute of Trade and Economy, Russia, Krasnoyarsk

A PARAMETERIZATION ALGORITHM FOR THE VAGANOV–SHASHKIN MODEL OF SEASONAL GROWTH AND TREE-RING FORMATION*

In order to simulate the Vaganov–Shashkin model for seasonal growth and tree-ring formation, a solution algorithm for the parameterization problem of the model is being proposed in cases, when a modulation is possible. The algorithm is realized as dll-library (or as a text file), tested on extensive data. A concept of difference in criterion between the actual tree-ring chronology and its model is introduced. Two new difference criteria are developed.

Keywords: the Vaganov–Shashkin model, tree-ring chronology, parameterization algorithm, difference criterion, optimal model parameters.

The Vaganov- Shashkin model of seasonal growth and tree-ring formation [1–3] (hereinafter referred to as VS-model) describes the influence of climatic conditions on the cellular structure of annual rings. The main destination of the model consists in using it as a tool permitting any individual or generalized tree-ring chronology (TRC) during a set of years with available meteorological station data (see below description of the VS-model input): 1) extracting the climate-driven component of the TRC under consideration; 2) pointing, if modeling quality of the involved TRC by the VS-model is satisfactory, for arbitrary day of the given years set, what of two factors – air temperature or precipitation – had limited the growth of woody plant which corresponds to the TRC. TRC is a time series of values of some numerical characteristic of tree/trees of this TRC [4].

The main destination of the VS-model, in particular, is that it makes possible to use it as a quality tester for individual and generalized TRCs which will be used as mediators, carrying information about climatic data [2]. A TRC is considered to be suitable for using for the purpose of a mediator for carrying climatic information if and only if its climate-driven component had been extracted by the VS-model conforms with the TRC satisfactorily enough.

The VS-model is a deterministic dynamic simulation model. Its input data consist of two blocks. The first block is the climatic data of daily resolution at air temperature, precipitation, and doses of solar radiation coming down to the Earth's surface. The data of this block can refer to an arbitrary set of years, which does not have to be continuous. However, climatic data per year must not contain missing values.

The second block is a set of the VS-model's parameter values. The model has 42 parameters. All parameters, except two, are real single-valued variables and three parameters are integer single-valued variables. Two parameters are vectors of equal dimension. One of the vectors contains real single-valued variables, the second – boolean single-valued variables. The dimension of the vectors is a parameter of the VS-model.

The value of each parameter provides the VS-model with information either on the actual tree or on its site. The VS-model has 10 options. A set of option values defines the VS-model variant which will be used for modeling. The options of the VS-model define a collection of its modules being run under modeling, calculating accuracy under modeling, etc.

The output of the VS-model consists of two blocks. The data of the first block consists of numerical characteristics reflecting the dynamics in time, the time step of whose tracking does not exceed one day – for some processes taking place in the modeled tree and its modeled site (soil moisture, value of tree transpiration, number of cambial cells, etc). The contents of this block depend on the used set of the VS-model option values.

The second output block of the VS-model is a model of TRC. The VS-model supposes the actual individual TRC for being modeled to be a time series of values either of width of a formed annual ring or of number of wood cells in such a ring. The VS-model supposes all individual TRC are used for the construction of an actual generalized TRC to be such a one described above. For each year of the VS-model, the climatic input of the second block contains the value of modeled TRC.

This value does not have any units if a generalized TRC has been modeled. If an individual TRC has been modeled, this value either does not have any units or is number of wood cells in an annual ring (variants, taking place are determined by the used set of the VS-model options values).

A parameterization problem and approach to its solution for the VS-model. In the field of mathematical modeling the term "parameterization" is usually understood as either an activity on describing some process/phenomenon by means of a finite number of parameters, i. e. the creation of a parametric mathematical model of this process/phenomenon, or choosing certain values of parameters of a parametric mathematical model that were already created. In the article the term is being understood according to the second interpretation.

*This work has been supported by Russian Foundation of Basic Research (Project № 09-05-00900-a).

To process the VS-model it is necessary to have certain numerical values of its parameters. Sets of these values must have the necessary internal structures that are specified by interrelations between parameters of the VS-model and by ranges of parameter values. Meaning, many parameters, the VS-model structure, semantics of its output components, and specific context in which modeling is performed determine the mentioned interrelations, and value ranges. As a rule, the checking a set of values of the VS-model parameters for the presence of the necessary internal structure in it requires an analysis of the VS-model output that corresponds to this set.

In an attempt to set certain values to the model's parameters to realize the modeling, it often happens that the available information is not enough for the identification of exact values for some parameters. For such parameters it is possible to provide only a qualitative assessment of their values, for example, by means of determining the boundaries in which the values of these parameters lie. In addition to a number of properties of some the VS-model's parameters, it complicates the empirical measurement of the values and makes the exact value conception for these parameters meaningless. It has as a consequence in a situation when for each parameter of the VS-model, the range of its values was known, and for some parameters their specific numerical values are known.

The parameterization problem for the VS-model is a problem of choosing values of those parameters, for which only the ranges of their values are known. The choice has to be realized within these ranges in such a way that the derived set of values of all the VS-model parameters has the necessary internal structure. How do we realize this choice? What should one be guided by to realize the choice? These two questions are the essence of the parameterization problem. The parameterization problem of an existing mathematical model is a problem of choosing the model input data in conditions of insufficient information.

There are different approaches used in the practice of solving the parameterization problem of a parametric mathematical model [5]. A solution to the parameterization problem for the VS-model, presented below, uses the existing arbitrariness in choosing values of parameters, for which information is necessary for determining their exact values is absent, purposively. Values of the parameters are selected, which conform to the researcher's objective formalized in a form of reaching the optimum of an objective function in the optimal, over the domain of definition of the objective function, way (conform in the optimal way). The proposed solution to the parameterization problem for the VS-model can be applied only in cases when entity for being modeled is available, in particular, under using the VS-model according to its main destination.

A concept of difference criterion between actual TRC and its model. Introducing the concept "difference criterion between actual TRC and its model", being significant out of the context of the VS-model parameterization problem, to state the solution being proposed in this article to the latter problem. The actual TRC is a TRC for being modeled. The difference criterion is a real single-valued nonnegative function of two arguments, denote it by $DC(trCr, mCr)$, that is defined over a set of pairs of real vectors of equal dimension with nonnegative components. The dimension of the vectors is the power of set *Yrs*. The set *Yrs* is the intersection of two sets – set of years, which the modeling has been performed for, and set of those years, which values of the actual TRC are available for.

The first argument is that *trCr* must be a vector of values, relating to the years *Yrs*, of the actual TRC, which are situated in the vector in increasing order of the years. The second argument must be *mCr* a vector of respective values of the TRC modeled by the VS-model.

A value of the difference criterion characterizes the remoteness of the TRC model, obtained by means of the VS-model, from this actual TRC. Given a fixed actual TRC, a smaller value of the difference criterion corresponds to a better model of this TRC. A difference criterion introduces the relation of equivalence on a set of models of the certain fixed TRC and orders the equivalence classes linearly.

A difference criterion between actual TRC and its model must have two additional properties: symmetry property on its domain of definition and reflexive property ($trCr = mCr$ implies $DC(trCr, mCr) = 0$). These two properties are not imposed on the difference criterion between actual individual TRC and its model since semantics of the actual TRC values (including units of measurement) coincides with semantics of values of the TRC model, obtained by means of the VS-model, guaranteed for generalized TRCs only.

The necessary properties of a difference criterion do not guarantee its continuity. Multivalued difference criteria have not been considered in this article.

The testing and approbation of the parameterization algorithm for the VS-model, which is presented below, have been done with the following two difference criteria which are being offered for using as default difference criteria. Difference criterion DC_{ITRC} between actual individual TRC and its model, and difference criterion DC_{GTRC} between actual generalized TRC and its model are defined with formulas:

$$DC_{ITRC}(trCr, mCr) = 1 - \mathrm{crln}(trCr, mCr) + \\ + 1{,}25\left(1 - \mathrm{sync}(trCr, mCr)\right)^2,$$

$$DC_{GTRC}(trCr, mCr) = DC_{ITRC}(trCr, mCr) + \\ + 0{,}4\left(\max_{i \in Yrs}\left|trCr_i - mCr_i\right|\right)^2,$$

where $\mathrm{crln}(trCr, mCr)$ and $\mathrm{sync}(trCr, mCr)$ are coefficients of Pearson's correlation and of synchronism (coincidence) between vectors *trCr* and *mCr*, $trCr_i$ (mCr_i) is a corresponding to the year *i* component of the vector *trCr* (*mCr*).

The selection of such difference criteria is conditioned by a desire: 1) to get calculated TRCs the most possible positive correlated with their respective actual TRCs;

2) of not having small values of synchronism coefficient between actual TRC and its model; 3) of absence of serious visual differences between broken lines that represent an actual generalized TRC and its model. The difference criteria DC_{ITRC} and DC_{GTRC} have been tested on extensive data, are compatible with the parameterization algorithm for the VS-model (see below), and reflect the view of most researchers on concept of proximity of two TRCs.

The parameterization algorithm. Let us introduce definitions and denotations. Let $p = (p_1, \ldots, p_n)$ denote vector of the VS-model parameters used under modeling (PUuM). Vector p is uniquely defined by a set of the VS-model options values used under modeling. We denote that the numbers of these n parameters of the VS-model, which certain numerical values are known for, by $i_1, \ldots, i_k$; let $c_{i_1}, \ldots, c_{i_k}$ denote the certain numerical values known for these parameters. The range of values of the VS-model parameter i is denoted by $[a_i ; b_i]$. For the ease of exposition, it is supposed that $a_{i_j} = b_{i_j} = c_{i_j}$ for $j = 1, \ldots, k$.

Let us call the following subset of $(n - k)$ a dimensional parallelepiped:

$$P = \left\{ p \in \mathbf{R}^n : \quad a_i \le p_i \le b_i \quad \text{for} \quad 1 \le i \le n \right\}$$

lying in $\mathbf{R}^n$ the optimization space S. S is defined by the condition: an element of P, considered as a set of values of the PUuM, belongs to S if and only if it has the necessary internal structure (see above). Call a set, belonging to S, of values of the PUuM a feasible. We consider $S \neq \varnothing$. The structure of S is conditioned by the interrelations, used under modeling, between the PUuM. S resembles a rectangular piece of cheese. It is a typical situation: among 100 000 values of the continuous n-dimensional random variate uniformly distributed on P only one belongs to S.

Define the predicate function $pFail(p)$: $P \to$ {“TRUE”, “FALSE”} taking the logical value “TRUE” only if $p \notin S$. As a rule, for calculation of $pFail(p)$ the VS-model output data obtained for the set p are necessary.

Let us call the quantity:

$$\sup\left\{ x \in \square : \forall InpOpt_1 \forall InpOpt_2, \left(\left| p_i^1 - p_i^2 \right| \le x \quad \to \quad Q(Output_1, Output_2) \right) \right\}$$

the accuracy along the i-th axis, $1 \le i \le n$. Here $InpOpt_1$ and $InpOpt_2$ – are two collections of the VS-model input data and of values of its options, which differ only in values p_i^1 and p_i^2 of the i-th PUuM, $Output_1$ and $Output_2$ – the VS-model output data corresponding to $InpOpt_1$ and $InpOpt_2$ respectively. The predicate Q is true if and only if $Output_1$ and $Output_2$ differ from each other so little that this difference may be ignored and is considered to be negligible. Denote by h_i an estimate, used under modeling, of the accuracy along the i-th axis.

We introduce S – a metric d: $S \times S \to [0; +\infty)$ defined as:

$$d(x, y) = \max_{1 \le i \le n} \left| \frac{x_i - y_i}{h_i} \right|.$$

Selecting a metric on S is not a trivial problem. Such a metric is introduced in S since: 1) the optimization space has, generally speaking, different physical units of measurement along with different axes; 2) the constants h_i corresponding to axes with equal physical units of measurement can be different; 3) it is a natural analog of the metric ρ_∞ on $\mathbf{R}^n$ [6].

Let us state the parameterization algorithm for the VS-model. The proposed algorithm selects a set of values of the PUuM from the optimization space S; the selected set provides the global minimum over S of the objective function $F(p)$. The algorithm requires that $F(p)$ is a real single-valued nonnegative function defined everywhere on S. The algorithm does not require any additional properties of $F(p)$. The objective function is defined by a following equality:

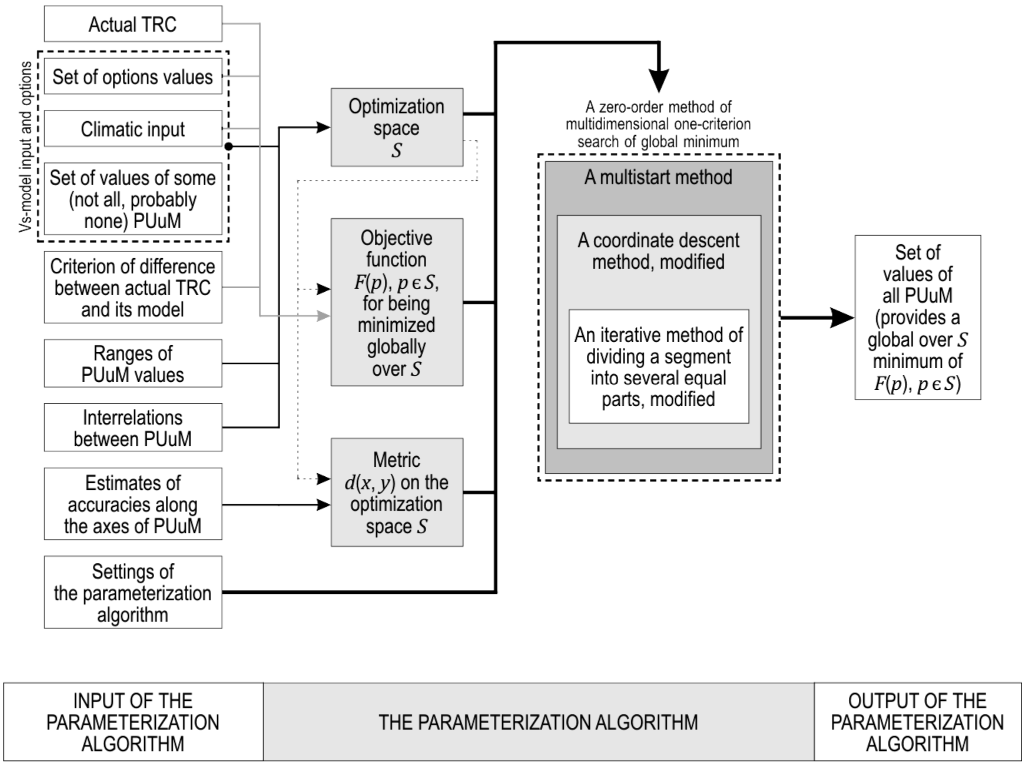
$$F(p) = DC(trCr, mCr(p)), p \in S,$$

where $mCr(p)$ is the vector of values, relating to the years Yrs, of the TRC model (see above) derived by using the VS-model with the set p of the PUuM values. The proposed parameterization algorithm for the VS-model is applicable with any difference criterion. It considers only those properties of the difference criterion being minimized by the one in which any difference criterion has necessarily.

The objective function $F(p)$ and its domain of definition S are individual for every situation which the parameterization algorithm being stated is applied in. Investigating the problem of searching the global over S minimum of $F(p)$ have shown that $F(p)$ has more than one local minimum as a rule. Input and output algorithm data and its structure are presented in figure.

A zero-order method of multidimensional one-criterion search of global minimum, used by the parameterization algorithm, is a multistart method of a coordinate descent method [7]. The question of existence of a more effective method for solving a family of problems, searching the global over S minimum of $F(p)$ is not considered in this article. The used coordinate descent method differs from the classical one in the following ways: 1) order of tracing coordinate axes of S is defined by a user; 2) a metric different from the Euclidean one is used on S; 3) the used one-dimensional optimization method performs a search of a global optimum; 4) absence of requirement of $F(p)$ being continuous, and complicated structure of S require a modification of the classical coordinate descent method.

Settings of the parameterization algorithm are a collection of six constants m, $level$, $maxIter$, ε, q, $maxStart$, and two vectors τ and $wght$. The algorithm uses the same values of its settings over a period of all its operations being performed. The following settings of the parameterization algorithm are being offered for using as its default settings: $m = 3$, $level < \inf_{p \in S} F(p)$, $maxIter = 500$, $\varepsilon \le (n - k)^{-1}$, $q = 1$, $maxStart = 1000$, $\tau_i \le h_i \cdot \min(1, \varepsilon)$ $(1 \le i \le n)$, $wght_j = 0$ for $j = i_1, \ldots, i_k$ (default values of remaining $n - k$ positive components of the vector $wght$ are not noticed in the article). The oddness of m, and truth of inequality $\tau_i \le h_i \cdot \varepsilon$, $1 \le i \le n$, are necessary for correct functioning of the algorithm.



Structural scheme of the parameterization algorithm for the VS-model

The multistart method launches the coordinate descent method from *maxStart* starting points in S with the same settings. The least of *maxStart* local minima found is an estimate of a global minimum and is delivered as a result of the multistart method run. Starting points are generated in a random manner and are different values of the continuous n-dimensional random variate uniformly distributed on P.

The order of tracing the coordinate axes of S with the coordinate descent method is defined by the weights *wght*. Each iteration of the latter consists in performing one-dimensional minimizations along the coordinate axes of S that have positive weights; halting condition is checked at the end of the iteration. One-dimensional minimizations along the axes that have maximum weight are carried out in the first instance. Axes with equal positive weights are traced in the order induced by their numbers – the smaller the number, the earlier the minimization is made.

The minimization along axis i, $1 \le i \le n$, means searching a global minimum of function $f(x) = F(\dots, p_{i-1}, x, p_{i+1}, \dots)$ of a real variable x on a set that lies inside the segment $[a_i; b_i]$ and which is specified by predicate function $xFail(x) = pFail(\dots, p_{i-1}, x, p_{i+1}, \dots)$. After having performed the search, a value x^* at which global minimum of $f(x)$ occurs is being written into the i-th coordinate of the current point p in S. In entering the first iteration, the current point p coincides with the starting one which has been selected by the multistart method. Coordinates of p with zero weights are not altered over a period of all operations being performed by the coordinate descent method.

The halting condition holds in case of truth of at least one of conditions: 1) number of iterations have reached a threshold *maxIter*; 2) $F(p) < level$, where p is a current point at the end of iteration (after executing all one-dimensional minimizations of the iteration); 3) $d(p^N, p^{N-i}) < \varepsilon$ for all $i = 1, \dots, q$ (N – the current iteration number, p^j – current point in S at the end of iteration j, q and ε – positive constants); 4) $d(p^N, p^{N-1}) = 0$.

The one-dimensional minimizations along the coordinate axes of S are carried out by the method that is a modification of an iterative method of dividing segments into several equal parts [7]. During each iteration the method divides current segment into m equal parts and either makes one of them a current segment or completes the operation informing the user about the necessity to change the value m. The iterations stop when the value, on the current segment, of the function being minimized becomes less than the value *level*, or when the length of the current segment becomes less than the value δ defined before launching the method.

All one-dimensional minimizations along axis i, is carried out in the course of running the coordinate descent method, are performed with $\delta = \tau_i$ ($1 \le i \le n$).

The realization and approbation of the parameterization algorithm. The parameterization algorithm for the VS-model is realized in programming language C++, using the means of the Standard C++ Library. Its textual code satisfies the C++03 standard (ISO/IEC 14882:2003). The executable code of the realization for a certain operating system can be derived without changing the textual code of the realization and is arranged as a dll-library (if the objective platform is an

OS of the Windows family) or a text-file (if the objective platform is an OS of the Unix family).

The main factors influencing the duration of parameterization for a certain TRC by means of the realization, being considered, of the parameterization algorithm are: 1) the set of values of the VS-model options used under modeling; 2) the power of the set of the years which modeling is performed on; 3) the interrelations between PUuM used under modeling; 4) the value of the quantity $n - k$ and the lengths of the values ranges of those PUuM which specific numerical values are not known for; 5) the difference criterion used under parameterization; 6) the used settings of the parameterization algorithm; 7) hardware and software of a device carrying out the code of the parameterization algorithm realization. Varying any of these factors, one can get a significant change (of many times) of the duration of parameterization.

Using laptop Asus F3JP with dual core processor T5300 (maximum clock frequency of each core is 1.73 GHz), the parameterization of a TRC by using the realization of the parameterization algorithm usually lasts from several hours to several days.

The proposed parameterization algorithm for the VS-model was used at more than 700 different pairs (S, $F(p)$). Having performed these parameterizations as a consequence: 1) a number of statements about the VS-model, both the new and the previously formulated, are first confirmed by acomputing experiment; 2) a hypothesis about the relationship between the models of individual TRCs and the model of the generalized TRC (all the models are derived by using the VS-model with the same set of values of its options) is for the first time formulated and confirmed by a computing experiment; 3) the VS-model has been first used in full as the quality tester of individual and generalized TRCs which are going to be used in the role of mediators carrying climatic information; 4) several properties of the considered parameterization algorithm have been established. The successful approbation of the parameterization algorithm and its realization is due to these results.

The results obtained in this article are demanded mainly in dendroclimatology and dendrochronology, and in mathematical modeling of processes taking place in woody plants. The created parameterization algorithm and its realization: 1) significantly extend the range of situations in which using the VS-model is already realizable in practice; 2) allow the testing and analyzing of the model at a qualitatively new level; 3) allowing to solve in practice inverse problems of restoring certain growing conditions for woody plants via its existing TRC.

References

1. Hughes M. K., Swetnam T. W., Diaz H. F. Dendroclimatology: Progress and Prospects (Developments in Paleoecological Research) // Berlin : Springer-Verlag, 2009.

2. Vaganov E. A., Hughes M. K., Shashkin A. V. Growth Dynamics of Conifer Tree Rings: Images of Past and Future Environments. Berlin : Springer-Verlag, 2006.

3. Vaganov E. A., Shashkin A. V. Growth and tree-ring structure of conifers. Novosibirsk : Nauka, 2000.

4. Cook E. R., Kairiukstis L. Methods of Dendrochronology: applications in the environmental sciences. Dordrecht : Kluwer Acad. Publ., 1990.

5. Papalambros P. Y., Wilde D. J. Principles of Optimal Design: Modeling and Computation. Cambridge : Cambridge University Press, 2000.

6. Kolmogorov A. N., Fomin S. V. Elements of the Theory of Functions and Functional Analysis. N. Y. : Dover Publications, 1999.

7. Horst R., Pardalos P. M. Handbook of global optimization. Dordrecht : Kluwer Academic Publishers, 1995. Vol. 1.

D. V. Kapulin, I. S. Korosteleva, D. A. Pupaeva
Siberian Federal University, Russia, Krasnoyarsk

CALCULATION METHOD FOR AUTOMATIC SECONDARY POWER SOURCES*

The article describes methods for calculating the parameters of secondary power sources (SPS): rectifying devices and voltage impulse stabilizers. Algorithms for calculating parameters of SPS power circuits have been developed.

Keywords: secondary power sources, algorithms for calculating the parameters of secondary power sources.

Secondary power sources (SPS) are an integral part of any radio set. Modern SPSs of electronic equipment have developed far beyond the class of simple radio-electronic devices, containing a small number of elements as it was 25–30 years ago. Today the secondary power sources are fairly complex devices which contain a large variety of functional units, performing certain functional transformations of electrical energy improving its quality. All this demands constant increase of performance from the SPS, while the time requirement for the design of the SPSs is being reduced. Electronic circuits of SPSs are characterized by a presence of components with nonlinear characteristics.

*The article has been written due to the Young Scientists Support Program of the Siberian Federal University.

The calculation of these circuits represents a significant challenge not only for manual exploit, but for computer operation. Nevertheless, there are methods of SPS approximate calculations that allow with sufficient accuracy to determine the main parameters of the projected scheme.

Analytical calculation of rectifying device parameters and voltage stabilizers represents a significant difficulty when applying a manual approach. Existing computer-aided calculation schemes enable SPSs with sufficient accuracy for engineering practice to determine the main parameters of the projected scheme. However such tools are calibrated to a particular class of circuits, such as specific types of rectifiers, stabilizers, converters, and other devices [1].

Therefore, the purpose of this article is to research methods of calculating the basic schemes for rectifying devices and voltage impulse stabilizers, as well as the development of algorithms for future program implementation, which would allow the calculation of parameters for basic SPS types (rectifying devices and voltage stabilizers), electronic equipment without the use of other software funds.

Methods and algorithms for calculating the parameters of secondary power sources.

The function of a rectifier device is the transformation of AC to DC (which is used for various electronic power devices). The design of the rectifier is reduced according to the selected scheme and type of valves, the valve calculation mode, the effective values of currents and voltages of the transformer windings, and the features of the smoothing filter. Fig. 1, 2 shows the basic circuits of rectifiers used to power electronic equipment.

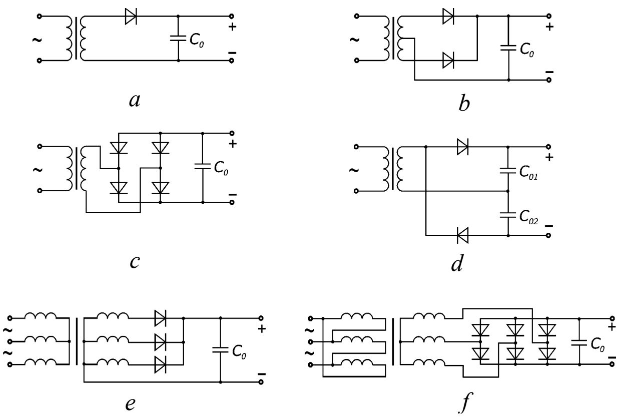


Fig. 1. Rectifiers operating at capacity:
a – half-wave circuit; *b* – full-wave circuit; *c* – full-wave bridge circuit; *d* – double-voltage circuit; *e* – three-phase circuit; *f* – six-phase (Larionov) scheme

Generally all the elements of the rectifier unit should be computed in a complex, because each successive element significantly affects the behavior of the previous as well as of the subsequent scheme (element) block. For example, a smoothing filter can dramatically change the mode of operation and the estimated ratio of currents and voltages in a circuit of valves and transformers. Nevertheless, in many cases it is preferable to calculate each block of the device separately, specifying, and considering the extent to which it affects the mutual influence of the modes' other circuit blocks, using data from the calculation of one block as an input for subsequent calculations for the following and previous blocks. For example, in result of the rectifier calculation (meant for obtaining source data for the transformer calculation) and the smoothing filter, the smoothing filter calculations can be applied to obtain the necessary data for calculating the rectifier and voltage stabilizer.

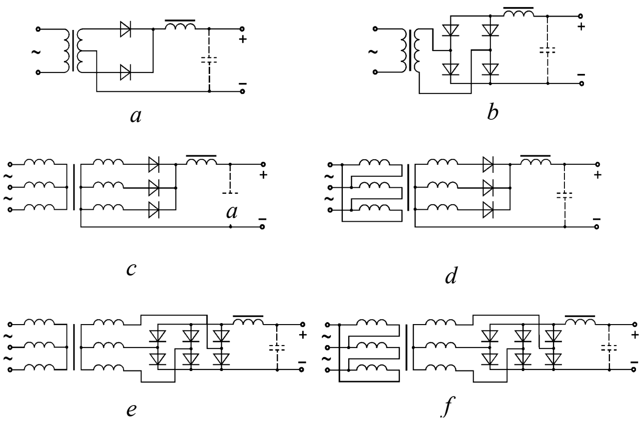


Fig. 2. Rectifiers operating at inductance;
a – full-wave circuit; *b* – full-wave bridge circuit; *c* – three-phase circuit; *d* – three-phase circuit; *e* – six-phase (Larionov) scheme; *f* – six-phase (Larionov) scheme

The initial data is required to count the value of the output voltage and the current for given admissible values of the pulsation coefficient. It also considers the conditions for rectifier power access, given the additional operational requirements and properties of the calculated rectifier device. In the process of calculating individual blocks it is essential to compare the obtained data with that known or obtained by calculating the blocks in a scheme of pre-computed clusters.

The method of calculating a rectifier device depends on the operating modes of the device and its type. Similar methods of calculating various types of rectifiers are given in [1–4]. Currently, electronics equipment widely uses rectifying devices with an inductive and capacitive load response for power. To calculate the rectifier a with capacitive load response a method based on approximate graph analytical calculations is used. The following data is implied:

– rectified voltage E_0;
– rectified current I_0;
– coefficient of pulsation a_{p0};
– supply voltage U_c;
– voltage frequency f.

The same method is used when calculating any similar scheme. The method given in [1] is for automated use. The only difference in these schemes is in the varying coefficients K_i in the formulas. The coefficients introduced in the calculation method (for the convenience of automation) and do not play a significant role in understanding the algorithm. For each scheme, these factors have an impact; they are given in [1].

After choosing a desired scheme, we proceed to the evaluation of the valves. The average direct current of the diode:

$$I_{0v} = K_1 I_0. \quad (1)$$

There are approximate formulas for reverse voltage and pulse current of the diode, but their exact values can be obtained after calculations of the transformer. The calculation begins with the definition of the transformer's induction: B_T. An approximating formula is used for software implementation:

$$B_T = 1.2 - 0.4\sin(0.003\, E_0 I_0). \quad (2)$$

Then the resistance phase of the secondary winding are calculated:

$$r_T = K_2 \frac{E_0}{I_0 f B_T} \cdot \sqrt[4]{\frac{f B_T}{E_0 I_0}}. \quad (3)$$

The rectifier phase resistance:

$$r = K_8 R_i + r_T, \quad (4)$$

where $R_i \approx 0{,}2/I_{0v}$ – is the internal resistance of valves. Next the auxiliary parameter is defined:

$$A_0 = \mathrm{tg}\Theta - \Theta = \frac{\pi r I_0}{K_3 E_0}. \quad (5)$$

Solving the transcendental equation: $\mathrm{tg}\Theta - \Theta = A_0$, it is possible to find the current cutoff angle Θ. In order to find Θ in software implementation, a numerical method dividing the segment in half is used. The EMF of the secondary winding of the transformer is:

$$U_{2x} = \frac{K_4 E_0}{\sqrt{2}\cos\Theta}. \quad (6)$$

Then we define the auxiliary coefficient:

$$D_0 = \frac{\sqrt{\pi[\Theta\,(1 + 0{,}5\cos 2\Theta) - 0{,}75\sin 2\Theta]}}{\sin\Theta - \Theta\cos\Theta}, \quad (7)$$

the effective current of the secondary winding:

$$I_2 = K_5 D_0 I_0, \quad (8)$$

the effective current primary winding:

$$I_1 = \frac{K_{10} I_0 U_{2x}}{U_c}. \quad (9)$$

The overall capacity of the transformer is:

$$P_{ovr} \approx K_7 P_0 = K_7 E_0 I_0. \quad (10)$$

Then we define the diode parameters. Reverse voltage of the diode:

$$U_{rev} = K_9 U_{2x}. \quad (11)$$

Efficient diode current:

$$I_d = K_6 I_2. \quad (12)$$

The support coefficient F_0:

$$F_0 = \frac{\pi(1 - \cos\Theta)}{\sin\Theta - \Theta\cos\Theta}, \quad (13)$$

and calculate the surge current of the diode:

$$I_m = \frac{F_0 I_0}{K_3}. \quad (14)$$

Power allocated on a valve when the current flows in a forward direction:

$$P_{all} \approx \frac{I_d^2 \cdot 0{,}2}{I_{0v}}. \quad (15)$$

The required value of filter capacitor is calculated:

$$C_o = \frac{H_f}{r \cdot a_{p0} \cdot f}, \quad (16)$$

where H_f – is the auxiliary factor depending on the angle Θ and the number of rectification phases (coefficient K_3) [1]. It is counted by:

$$H_f = 25330 \cdot (2\Theta - \sin 2\Theta) \cdot \cos\Theta, \quad (17)$$

if $K_3 = 1$, or:

$$H_f = 101000 \cdot \frac{[\sin(K_3\Theta)\cdot\cos\Theta - K_3 \cdot \cos(K_3\Theta)\cdot\sin\Theta]}{K_3 \cdot (K_3^2 - 1)\cdot\cos\Theta}, \quad (18)$$

if $K_3 > 1$. For calculation the internal resistance of the rectifier is necessary to define the intermediate point of the load characteristics:

$$E = \frac{U_{2x}\sqrt{2}\cos\Theta}{K_4}, \quad (19)$$

$$I = 0.45 \cdot \frac{K_3 U_{2x} \cdot \left(\sin\frac{\Theta}{2} - \frac{\Theta}{2}\cdot\cos\frac{\Theta}{2}\right)}{K_4 \cdot r}. \quad (20)$$

Then internal resistance of the rectifier is:

$$R_r = \frac{E - E_0}{I_0 - I}. \quad (21)$$

Calculation of rectifiers with inductive loads is similar to the aforementioned method. The following inputs are required for the calculation:

– rectified voltage E_0;
– rectified current I_0;
– pulse rate at the filter output a_{p1};
– supply voltage U_l;
– voltage frequency f.

Differences in the calculation scheme are only in the various coefficients K_i in the formulas. For each scheme, these factors have certain significance; they are given in [1].

The calculation begins determining the parameters of the transformer. First, we find the induction B_T and the resistance r_T of transformer. Leakage inductance of the transformer:

$$L_S = K_{13} \cdot \frac{E_0}{I_0 f B_T} \cdot \frac{1}{\sqrt[4]{\frac{f B_T}{E_0 I_0}}}. \quad (22)$$

Then the rectified voltage at idling is:

$$E_{0X} = E_0 + \Delta E_r + \Delta E_X + \Delta E_{B\,CX} + \Delta E_{thr}, \quad (23)$$

where voltage drop across the active resistance of the transformer $\Delta E_r = K_2 I_0 r_T$, at reactive – $\Delta E_X = K_3 I_0 f L_S$, at valves in circuit – $\Delta E_{B\,CX} = K_4 \cdot \Delta E_B \approx K_4 \cdot 0{,}6$; voltage drop during throttle $\Delta E_{thr} \approx 0{,}005 \cdot E_0$. The EMF of the secondary winding of the transformer is:

$$U_{2x} = K_6 E_{0x}. \quad (24)$$

The effective secondary current:

$$I_2 = K_7 I_0. \quad (25)$$

The effective primary current:

$$I_1 = K_{14} I_0 n, \quad (26)$$

where $n = U_{2x}/U_1$ – is the transformation ratio. The overall capacity of the transformer is:

$$P_{ovr} = K_9 E_{0x} I_0. \quad (27)$$

The parameters of the diodes. The reverse voltage of the diode is equal to:

$$U_{rev} = K_5 U_{2x}. \quad (28)$$

The average forward current of diode:

$$I_{0v} = K_1 I_0. \quad (29)$$

The forward pulse current $I_m = I_0$. Diode power dissipation:

$$P_d = \Delta E_B I_{0v} \approx 0{,}6\, I_{0v}. \quad (30)$$

The minimal throttle inductance:

$$L_{thr\,\min} = \frac{2E_0}{(K_3^2 - 1)\, K_3 \pi f I_0}, \quad (31)$$

then the filter capacity:

$$C = \frac{q_p \cdot 10^6}{K_3^2 \cdot 4\pi^2 f^2 L_{thr\,\min}}, \quad (32)$$

where $q_p = a_{p0}/a_{p1} = K_{10}/a_{p1}$, a_{p0} – rate fluctuations at the filter input (constant for this scheme). The capacitors' operating voltage must be calculated during idling rectifier, i. e. $U_w \geq E_{0x\,C} > 0 = K_{11} U_{2x}$, where $E_{0x\,C} > 0$ – is the rectified circuit voltage at $C > 0$.

The internal resistance of the rectifier is:

$$R_r = \frac{E_{0x} - E_0}{I_0}. \quad (33)$$

The critical point of the load characteristic is determined by:

$$I_{0cr} = \frac{E_0}{(K_3^2 - 1)\, K_3 \pi f L_{thr\,\min}}, \quad (34)$$

$$U_{0\,cr} = E_0 + (I_0 - I_{0cr}) \cdot R_r. \quad (35)$$

To calculate the rectifier, supplied by rectangular voltage, the initial data for the calculation technique [1] is:

– rectified voltage E_0;
– rectified current I_0;
– pulse rate a_p;
– supply voltage U_l;
– voltage frequency f;
– duty cycle β.

Below is a generalized method of calculation for a case of inductive and capacitive load reaction. With a few exceptions, the difference in calculation for the two types of reaction load (inductive and capacitive) differs in coefficients K_i in the formulas. Some of these factors depend on the value of the main calculations of the duty cycleβ; it is desirable to calculate them in advance. The values and formulas for calculating the coefficients are presented in [1]. The remaining differences between the calculations for the capacitive and inductive load reactions are described as the methodology.

First, we determine the resistance of the transformer:

$$r_T = K_1 \cdot \sqrt{\frac{1.2E_0}{f I_0^3}}. \quad (36)$$

Then we find the EMF of the secondary winding:

$$U_{2x} = K_2 E_0 + K_3 r_T I_0 + K_4, \quad (37)$$

and the overall transformer power:

$$P_{ovr} = 1.1 \cdot U_{2x} I_0 K_5. \quad (38)$$

The effective secondary current:

$$I_2 = I_0 \sqrt{K_6 \beta}. \quad (39)$$

The effective primary current:

$$I_1 = 1.1 \cdot K_7 I_0 n, \quad (40)$$

where $n = U_{2x}/U_l$ – is the ratio of transformation.

We define the parameters of the diodes. The reverse voltage of the diode is:

$$U_{rev} = 2 K_8 U_{2x}. \quad (41)$$

The average forward current when calculating diode rectifier devices with inductive load response is calculated using:

$$I_{0v} = \frac{I_2}{2}, \quad (42)$$

in the case of a capacitive load reaction:

$$I_{0v} = \frac{I_0 K_9}{2}. \quad (43)$$

The forward pulse current of the diode:

$$I_m = I_0 \beta K_9. \quad (44)$$

The diode's power dissipation:

$$P_r = K_{10} I_0. \quad (45)$$

We then compute the capacitance filter. With inductive load is equal to the reaction:

$$C = \frac{10^7 \cdot I_0}{4\pi^3 a_p E_0 f}. \quad (46)$$

In a case of capacitive reaction load, the capacitance is calculated by:

$$C = \frac{K_9 I_0 \cdot K_9 \beta \cdot \sin\left(\dfrac{\pi}{K_9 \beta}\right) \cdot 10^5}{2\, a_p E_0 f}. \quad (47)$$

When calculating the response of the rectifier with an inductive load, we still must calculate the inductance of the throttle and the internal resistance of the rectifier. The throttle inductance:

$$L_{thr} = 0{,}1 \cdot \frac{E_0 \beta \sin \frac{\pi}{\beta}}{I_0 f}. \quad (48)$$

The internal resistance of the rectifier:

$$R_r = \frac{U_{2x} - E_0}{I_0} \cdot \frac{2}{K_4}. \quad (49)$$

Voltage impulse stabilizers (VIS) along with the voltage rectifiers also are widely used in power radio-electronic equipment. The advantages of VISs, compared to continuous stabilizers, are high efficiency, small size and weight, and high power density; hence there is a widespread use of VISs in the design of autonomous power facility systems. The VIS power circuits are shown in fig. 3–5.

There are different methods for calculating the parameters of the VIS power circuits [1; 4]. The initial data for calculations is:

– input voltage E_{in};
– output voltage E_0;
– operating frequency f;
– voltage fluctuations U_p;
– pulse rate a_p;
– minimum $I_{\min}$ and rated I_0 load currents.

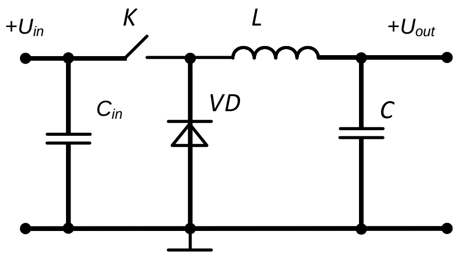


Fig. 3. Decrease-type VIS

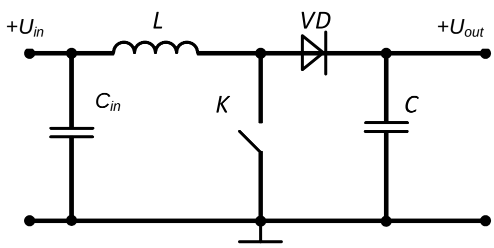


Fig. 4. Increase-type VIS

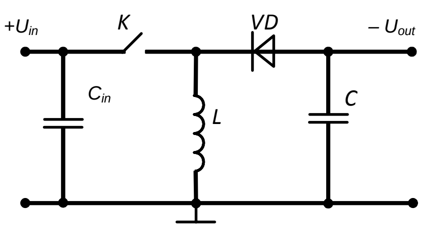


Fig. 5. Invert-type VIS

A method for computing VIS power circuits is presented in table. In addition to these parameters it is necessary to determine the capacity of the filter. It is calculated differently for each type of stabilizer. Calculation can be started after the determination of the critical inductance throttle L. The next method is for decrease-type VIS calculations. First we determine the coefficient of the smoothing filter Q:

$$Q = \frac{E_0 \cdot \left[2.2 - 2.5 \cdot \left(\frac{E_0}{E_{in}} - 0.1\right)\right]}{U_p}. \quad (50)$$

The inductance produce on the capacity of the filter is:

$$X(N) = \frac{25 \cdot (Q^{\frac{1}{N}} + 2)}{f^2}, \quad (51)$$

where N – is the number of filter links. Then the capacity of the filter is:

$$C(N) = \frac{X(N)}{L}. \quad (52)$$

The amplitude of voltage fluctuations can be calculated by:

$$U_{p\max}(N) = U_p \cdot Q^{\left(1 - \frac{1}{N}\right)}. \quad (53)$$

For increase-type VIS the capacity of the filter is determined by:

$$C = \frac{10^6 \cdot I_0 (E_0 - E_{in})}{f\, E_0\, U_p}. \quad (54)$$

For invert-type VIS the capacity of the filter is:

$$C = \frac{10^6 \cdot I_0\, E_0}{f (E_{in} + E_0) U_p}, \quad (55)$$

and capacity of the input filter is:

$$C_{in} = \frac{10^6 \cdot I_0\, E_0}{f a_p E_{in} (E_{in} + E_0)}. \quad (56)$$

Using the aforementioned computational procedures (1–56), we have developed algorithms for calculating the parameters of pulse rectifiers and voltage stabilizers, which are graphically presented in fig. 6–9.

Algorithms for the calculation of rectifying devices include a number of restrictions at the input data under which the obtained result has a sufficient degree of accuracy. If certain terms are violated the algorithm will provide output data which will reduce accuracy. The diagnosis of the entered data for correctness is carried out after the entry procedure.

There are following limitations to the input data:

– inappropriate negative figures;
– output power ($E_0 I_0$) should not exceed 500 W;
– output voltage E_0 not less 3 V;
– maximum frequency voltage f is 5 kHz;
– for fullwave three-phase circuits the rate fluctuations a_{p0} must be less than 0,07.

Methods for calculating parameters of voltage impulse regulators

Measured parameter	VIS type		
	Decrease	Invert	Increase
Critical throttle inductance L	$\frac{500E_0(E_{in}-E_0)}{fI_0E_{in}}$	$\frac{500E_0E_{in}^2}{fI_{\min}(E_{in}+E_0)^2}$	$\frac{500E_{in}^2(E_0-E_{in})}{fI_{\min}E_0^2}$
Average throttle current I_{tr}	I_0	$I_0(E_{in}+E_0)/E_{in}$	I_0E_0/E_{in}
The variable component of the throttle current $I_{tr\sim}$	$\frac{E_0(E_{in}-E_0)}{2fE_{in}L}$	$\frac{E_{in}E_0}{2f(E_{in}+E_0)L}$	$\frac{E_{in}(E_0-E_{in})}{2fE_0L}$
Amplitude of collector current $I_{k\max}$	$I_{thr}+I_{thr\sim}$		
Effective collector current I_k	$I_0\sqrt{\frac{E_0}{E_{in}}}$	$\frac{I_0}{E_{in}}\sqrt{E_0(E_0+E_{in})}$	$\frac{I_0}{E_{in}}\sqrt{E_0(E_0-E_{in})}$
Collector-emitter voltage U_{ke}	E_{in}	$E_{in}+E_0$	E_0
Power dissipated transistor P_{tr}	$I_k\left(0.07\cdot E_{in}+\frac{2E_0}{E_{in}}\right)$	$I_k\left(0.07(E_{in}+E_0)+\frac{2E_0}{E_{in}+E_0}\right)$	$\frac{I_kE_0\,(0.07E_{in}+2)}{E_{in}}$
Average diode current I_d	$I_0\frac{E_{in}-E_0}{E_{in}}$	I_0	I_0
Reverse diode voltage U_{rev}	E_{in}	$E_{\text{in}}+E_0$	E_0
Power dissipated diode P_d	$0.8\cdot I_d\frac{E_{in}-E_0}{E_{in}}$	$0.8\cdot I_0\frac{E_{in}}{U_{rev}}$	
Input stabilizer current I_{in}	$\frac{E_0I_0+P_{tr}+P_d}{E_{in}}$		
Efficiency η	$\frac{E_0I_0}{E_{in}I_{in}}$		

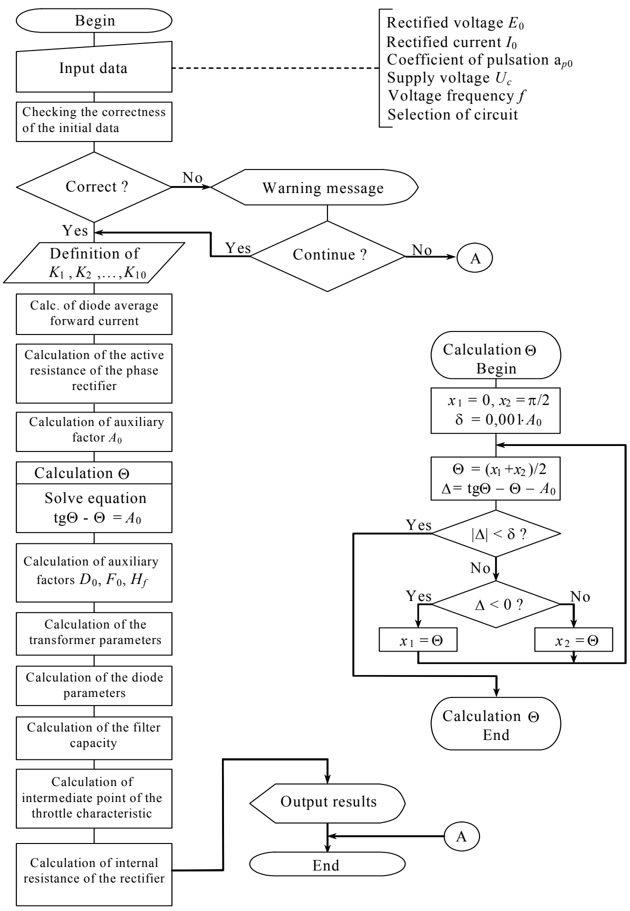


Fig. 6. Algorithm for calculating the parameters of the rectifier with capacitive load reaction

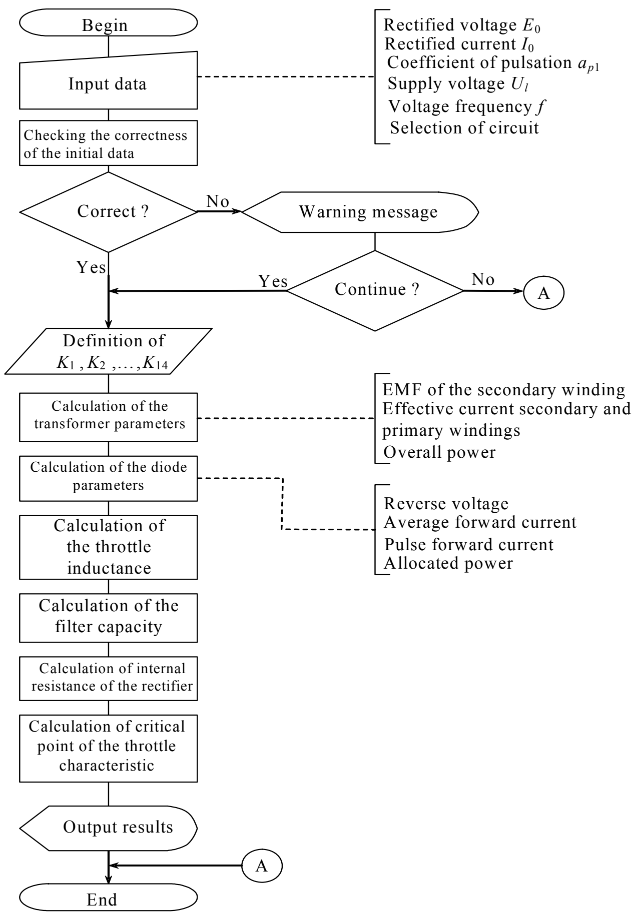


Fig. 7. Algorithm for calculating the parameters of the rectifier with inductive load reaction

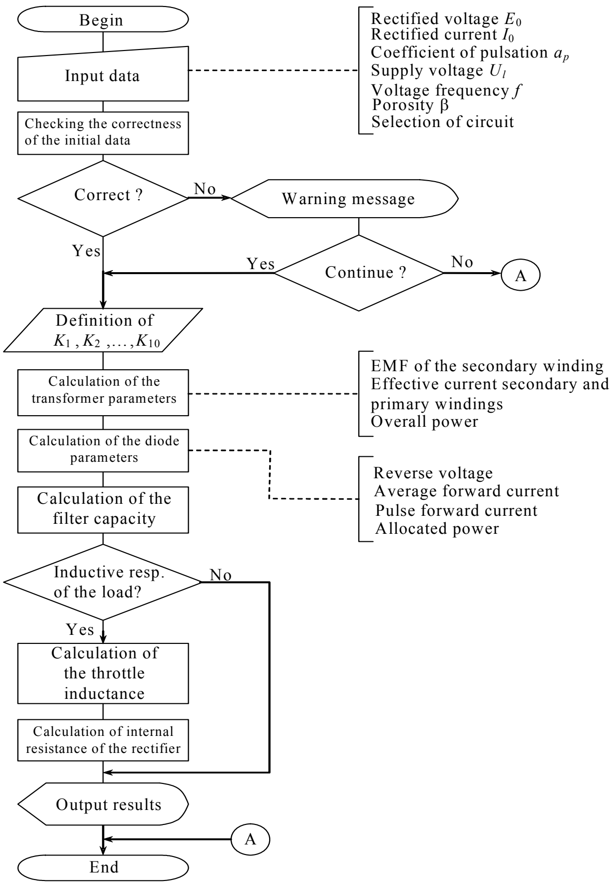


Fig. 8. Algorithm for calculating the rectifier account settings, fed by rectangular voltage

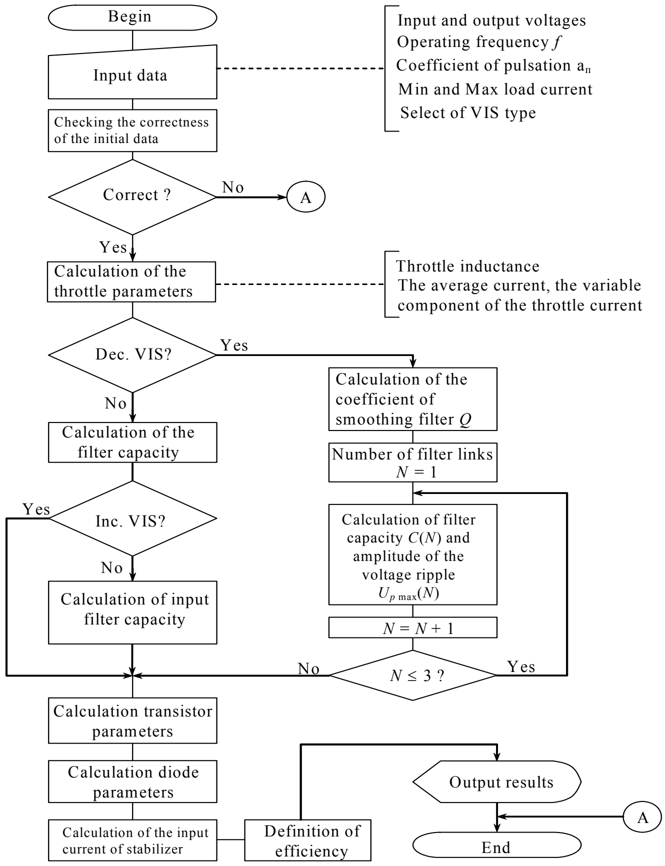


Fig. 9. Algorithm for calculating the parameters of the voltage impulse stabilizer

If the there are negative figures in the initial data, there will be errors that shall make further calculation impossible. If other conditions are violated, there will be a warning of reduction in accuracy for the result and the necessity to precede further calculations.

An algorithm for calculating the parameters of the voltage impulse stabilizer after initial data input is carried out to check the preciseness. The following conditions are checked:

– negative figures are not allowed in the original data;

– input voltage for decrease-type VISs E_{in} must be greater than the rectified voltage E_0;

– on the contrary, input voltage for increase-type VISs E_{in} must be less than the rectified voltage E_0.

If any of these conditions are not observed, the result will occur in an error message, making further calculations impossible.

In the article, each of the main secondary power sources have been presented with a summarized methodology of power circuits parameter calculations. For rectifying devices the key parameters are those of the transformer, diodes, the filter, internal resistance of the rectifier, voltage impulse stabilizer, parameters of the throttle, diodes, transistors, and filters.

The developed methods of calculation had been used to form algorithms for accounting rectifying devices and pulse voltage stabilizers. The developed algorithms have a sufficient accuracy for engineering practice (deviation of the calculation results, using software applications does not exceed 5–7 % compared to manual calculations for any of the reviewed devices); this has been confirmed by MathCAD simulation. The developed algorithms will dramatically reduce the required time for designing electronic devices, and reduce the total number of design errors by automating routine operations for accounting pulse rectifiers parameters and voltage regulators of various types.

References

1. Kozharsky G. V., Orechov V. I. Methods of computer-aided design of secondary power sources. M. : Radio & Communications, 1985. P. 184.

2. Power sources of electronic equipment: Directory / ed. G. S. Nayvelt. M. : Radio & Communications, 1985. P. 576.

3. Secondary power sources / ed. Yu. I. Konev. Second edition, revised and enlarged. M. : Radio & Communications, 1990. P. 280.

4. Moin V. S. Stabilized transistor converters. M. : Energoatomizdat, 1986. P. 376.

V. S. Komorovsky, G. A. Dorrer
Siberian State Technological University, Russia, Krasnoyarsk

CALCULATION METHOD OF PARAMETERS OF FOREST FIRES AS DYNAMIC PROCESSES ON THE EARTH'S SURFACE ON THE BASIS OF SPACE MONITORING DATA

A simple method for calculating the parameters of a forest fire is offered. The method is based on the representation of a fire as a mobile set on the surface of the Earth. The information base for the proposed methodology is the data of space monitoring of forest fires.

Keywords: forest fires, forecasting, space monitoring, dynamic processes.

One of the areas where it is required to use methods of mathematical modeling and control theory is the protection and rational use of natural resources. As a rule objects that belong to this area are characterized by their spatial distribution and dependence on a large number of hard-to-control factors.

A special group includes objects representing mobile sets or plane waves on the surface of the Earth. Forest and steppe fires, pollution zones of land and water surfaces, areas affected by plant pests, areas of soil erosion and desertification can serve as examples. The importance of studying such objects is increasing in connection with the increasing anthropogenic load on the biosphere and processes of climatic changes.

Enumerated objects have different physical nature, different spatial and temporal scales, but under certain assumptions their dynamics including the processes of their control can be described in a unified manner. A forest fire can serve as an example to illustrate a set of problems connected with the study of these processes and the urgency, goals and objectives of such researches most fully.

The problem of forest fires is very acute in the world today. In the Russian Federation, according to the data of an information system for remote monitoring of Federal Forestry Agency (ISDM-Forestry) 35,337 fires took place during the fire-hazardous seasons in 2008 and 2009, fire covered 15,565,510 hectares, including 9,772,782 hectares of forested areas [1]. Since forest is one of the most important strategic resources, the task of protecting forests from fires, including modeling and forecasting the spread of fires is an extremely important task. By now there has been developed a sufficient number of mathematical models of the forest fires spread and forecasting methods based on them. These models and methods contain different approaches to considering the process of burning in natural fires, however their common feature is the high demands for information support of modeling, that is, the exact knowledge of the set of terrain characteristics, where fire occurs, the state of plant combustible materials, weather conditions are the necessary condition for an adequate model building and, accordingly, forecasting. At the moment it is impossible to create an information base for provision of forecasting on the basis of such models in Russia.

The most complete information about forest fires is currently available in ISDM-Forestry. This information system has existed and is developed since 1995. The sources of the information provided for ISDM-Forestry is the data of space, aviation and ground monitoring of forest fires. Because of vast areas of observation, methods of space monitoring play a leading role. Satellites of such series as NOAA, TERRA, AQUA, SPOT, LANDSAT ETM +, МСУ-Э [1] take part in data collection.

The presence of a reliable forecast of the spread and development of a forest fire allows to evaluate the threat to environment, economic resources and population aggregates, to take necessary measures to prevent damage, to plan the work of fire fighting forces. One of the most important factors in predicting of forest fires is the construction of their contours. In the present paper the method of calculating of large (more than 200 hectares) forest fires contours on the basis of limited information available in ISDM-Forestry [2; 3] is presented.

The most complete information available in ISDM-Forestry and data bases of territorial airbases is connected with the dynamics of changesof the forest fires area. That's why it is reasonable to begin the modeling of the fire spreading process with this figure. When simulating the configuration of fire it is convenient to use Huygens' principle, which describes the propagation of waves in an anisotropic plane environment. This technique can be used for modeling of other dynamic processes mentioned above.

Original assumptions. The following assumptions were taken:

1. The dynamics of change of the forest fires area is determined by

$$S(t) = k_0(t - t_0)^{\alpha}, \quad (1)$$

where t is current time, day; t_0 is the time of fire appearance, day; α is the indicator of fire dynamics; k_0 is the constant coefficient having the dimension $n/d^{\alpha.}$ As it is clear from geometrical considerations and will be shown below, the change in rate of the fire front is also connected with the indicator α : when $\alpha = 2$ the fire front rate is constant, when $\alpha < 2$, this rate decreases with time, and if $\alpha > 2$ the rate increases.

2. The rate of the fire front in accordance with Huygens' principle is presented as

$$v(\varphi, t) = v_0(t)\xi(\varphi),$$

where v_0 (t) is the maximum rate of the fire front propagation to be determined (for example downwind); $\xi(\varphi)(|\xi| \le 1)$ is an indicatrix of the front full rate, which determines the configuration of the fire; φ is the propagation direction ($0 \le \varphi \le 2\pi$). Thus, in general

$$\nu_0 = \nu_0(w,t),$$
$$\xi = \xi(w,\varphi),$$

where w is wind speed rate, m / s. We omit the dependence on w where it is not required.

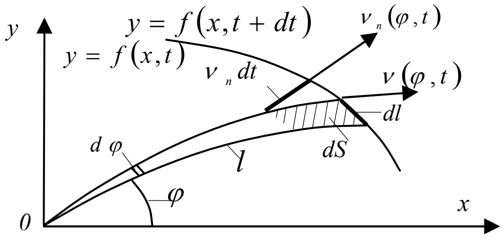
3. The indicatrix is determined by speed and wind direction, it has been adopted unchanged in the settlement period of time. The angle φ in the formula below is measured from the wind direction clockwise.

Formulas for calculating of the fire front rate. To simplify the formulas in this section let $t_0 = 0$. Let's consider an elementary area increment dS. Figure shows the edge of the fire in two close points of time t and $(t + dt)$. From the figure it follows that $dS = \nu_n dt dl$, where ν_n is normal front rate; dl is the increment of the contour length, which is calculated by the formula [4]:

$$dl = ld\varphi = d\varphi\sqrt{\xi^2(\varphi)+\xi^2(\varphi)}\int_0^t v_0(\tau)d\tau . \quad (2)$$

In its turn, the normal rate is associated with the full rate by the ratio:

$$v_n(\varphi,t) = \frac{v(\varphi,t)}{\sqrt{1+(\frac{v'(\varphi,t)}{v(\varphi,t)})^2}} = \frac{v_0(t)\xi^2(\varphi)}{\sqrt{\xi^2(\varphi)+\xi'^2(\varphi)}} .$$



Derivation of the formula of the fire front rate

Hence

$$dS = d\varphi dt v_0(t)\xi^2(\varphi)\int_0^t v_0(\tau)d\tau .$$

The area increment in all directions of propagation:

$$dS_{2\pi} = \int_0^{2\pi} dS d\varphi = 2dt v_0(t)\int_0^{\pi} d\phi\xi^2(\varphi)\int_0^t v_0(\tau)d\tau .$$

The rate of growth throughout the area of fire:

$$\frac{dS}{dt} = 2v_0(t)\int_0^{\pi}\xi^2(\varphi)d\varphi\int_0^t v_0(\tau)d\tau , \quad (3)$$

and the dynamics of the total area becomes

$$S(t) = 2\int_0^{\pi}\xi^2(\varphi)d\varphi\int_0^t v_0(\tau)\int_0^{\pi} v_0(\xi)d\xi d\tau .$$

Now let us consider the law of variation of the fire front rate. We search for the function $\nu_0(t)$ in the form $\nu_{0S}(t^{\beta})$, where the quantities β and ν_{0S} are to be determined. Then the time integral over in equation (3) will be:

$$\int_0^t v_0(\tau)d\tau = \frac{1}{\beta}v_{0S}t^{\beta+1} ,$$

And the rate of area growth:

$$\frac{dS}{dt} = 2\frac{v_{0S}^2}{\beta+1}t^{2\beta+1}\int_0^{\pi}\xi^2(\varphi)d\varphi .$$

On the other hand, from the initial assumption $S(t) = k_0 t^{\alpha}$ follows:

$$\frac{dS}{dt} = k_0\alpha t^{\alpha-1} ,$$

whence we get the equation

$$2\frac{v_{0S}^2}{\beta+1}t^{2\beta+1}\int_0^{\pi}\xi^2(\varphi)d\varphi = k_0\alpha t^{\alpha-1} . \quad (4)$$

Accepting in equation (4) exponents of t, we obtain $2\beta + 1 = \alpha - 1$, whence

$$\beta = \frac{\alpha}{2} - 1 .$$

Next, equating the factors of t, we define the value of ν_{0S}:

$$v_{0S} = \left(\frac{\alpha k_0(\beta+1)}{2\int_0^{\pi}\xi^2(\varphi)d\varphi}\right)^{1/2} = \frac{\alpha\sqrt{k_0}}{2\int_0^{\pi}\xi^2(\varphi)d\varphi} = \frac{\alpha\sqrt{k_0}}{2P(w)} , \quad (5)$$

where

$$P(w) = \int_0^{\pi}\xi^2(w,\varphi)d\varphi .$$

Equation (5) is the desired formula for calculating of the fire rate.

On the basis of the formula (2) we can also calculate the perimeter of the fire burning edge. In a particular case, when a fire develops from a seat of small radius (point seat):

$$L(t) = 2\int_0^{\pi} ld\varphi = 2\int_0^t v_0(\tau)d\tau\int_0^{\pi}\sqrt{\xi^2(w,\varphi)+\xi'^2(w,\varphi)}d\varphi .$$

The formula for the rate of growth of the fire perimeter the becomes

$$\frac{dL(t)}{dt} = 2v_0(t)Q(w), \quad (6)$$

where

$$Q(w) = \int_0^{\pi}\sqrt{\xi^2(w,\varphi)+\xi'^2(w,\varphi)}d\varphi .$$

Indicatrices of front velocity. Let's consider the specific expressions for the indicatrices.

1. The exponential indicatrix [4]

$$\xi(\varphi) = \exp(\alpha(w)(\cos(\varphi) - 1)),$$

where $\alpha\,(w)$ is the coefficient dependent on wind speed:

$$\alpha(w) = 0{,}785w - 0{,}06w^2.$$

The formula is valid under the conditions $0 \le w \le 3$ m/s. Then the integrals appearing in expressions (5) and (6) take the form

$$P(w) = \int_0^{\pi} \xi^2(w,\varphi)d\varphi =$$

$$= \exp(-2a(w))\int_0^{\pi} \exp(2a(w)\cos(\varphi))d\varphi,$$

$$Q(w) = \int_0^{\pi} \sqrt{\xi^2(w,\varphi) + \xi'^2(w,\varphi)}d\varphi =$$

$$= \int_0^{\pi} \xi(w,\varphi)\sqrt{1 + \frac{a^4(w)}{4}\sin^2(2\varphi)}d\varphi.$$

The calculated values of these integrals for certain values of wind speed are shown in tab. 1.

2. The elliptical indicatrix

$$\xi(\varphi) = \frac{1 - e(w)}{1 - e(w)\cos(\varphi)},$$

where $e(w)$ is the eccentricity of the ellipse, which depends on wind speed.

To assess the dependence of the elliptic indicatrix eccentricity on the wind velocity the diagrams of indicatrices shown in the work of F. Albini [5] were approximated by ellipses; the corresponding eccentricities were calculated for a larger range of wind speeds. The following approximation of the dependence of eccentricity on the wind speed was obtained:

$$e(w) = 1 - \exp(-0{,}4w).$$

Then the values of the integral $P(w)$ and $Q(w)$ were calculated. The results are given in tab. 2.

Comparing tab. 1 and 2, we can see that the values of the functions $P(w)$ and $Q(w)$ for both indicatrices can differ markedly at the same wind speeds. It should be born in mind that in the first case the wind speed was calculated at a height of two meters from the ground and in the second – at a height of ten meters, and this also explains a more elongated form of an elliptical indicatrix.

The error of the proposed technique. It is clear that the errors of all inputs result in an error in the final result. In this paper we confine ourselves to the simplest calculation of errors associated with the estimation of the fire area and its daily increment. A more accurate calculation would require involvement of the laws of distribution of variables.

Let the error in determining the fire area is δS ha. We'll consider the model parameters errors caused by this fact (1). The error in the estimate of the coefficient k

$$\delta S = \frac{\partial S}{\partial t}\delta k = t^{\alpha}\delta k,$$

whence

$$\delta k = \frac{\delta S}{t^{\alpha}}. \tag{7}$$

The error in the estimation of the fire propagation time:

$$\delta S = \frac{\partial S}{\partial t}\delta t = k\alpha t^{\alpha-1}\delta t,$$

whence

$$\delta t = \frac{\delta S}{k\alpha t^{\alpha-1}}. \tag{8}$$

The error in estimating α:

$$\delta S = \frac{\partial S}{\partial \alpha}\delta\alpha = kt^{\alpha}\ln t\,\delta\alpha,$$

whence

$$\delta\alpha = \frac{\delta S}{kt^{\alpha}\ln t}. \tag{9}$$

Formulas (7)–(9) show that the error of all variables that affect the model (1) are time-dependent and decrease with increasing *t*.

A numerical example of calculating the rate of fire propagation. The fire K-1002 registered in the system ISDM-Forestry was burning in Idrinskoye forestry from 10.05.2008 to 16.05.2008. The fire was registered on the area of 393 ha, the area of fire suppression was 1.163 hectares. The air temperature was 22.7 degrees, the wind speed – 1 m/sec, wind direction – 180 degrees.

Table 1

The values of integrals of the square exponential indicatrix, depending on the wind speed

w, m /s	0	0,1	0,2	0,5	1,0	2,0	3,0
$\alpha(w)$	0	0,078	0,155	0,378	0,725	1,33	1,815
$P(w)$	3,142	2,705	2,361	1,1695	1,178	0,818	0,686
$Q(w)$	3,142	2,911	2,708	2,234	1,758	1,455	1,535

Table 2

The values of the integrals of the elliptic indicatrix square depending on the wind speed

w, m /s	0	1,25	2,5	5	10	15
$e(w)$	0	0,393	0,632	0,865	0,982	0,998
$P(w)$	3,142	1,469	0,782	0,588	0,539	0,193
$Q(w)$	3,142	2,151	1,60	1,424	1,379	1,194

Let's assume that the error in determining the area δS = 100 ha, Δt = 16–10 = 6 days. We accept the indicator of the fire area growth rate α = 2, i. e. assume that the speed of the fire front is constant.

We calculate the coefficient k and its error in the model:

$$K = \Delta S/\Delta t^2 = 770/36 = 21.4 \text{ ha/per 24 hours},$$

$$\delta k = 100/36 = 2.7 \text{ ha/per 24 hours}.$$

An error estimate of the indicator α according to formula (9) yields:

$$\delta\alpha = \frac{100}{21.4 \cdot 6^2 \cdot \ln 6} = 0.07 .$$

As the wind speed is not great, we use the exponential indicatrix. For the wind speed w = 1 m/sec, from tab. 1 we determine the factor $P(w)$ = 1,178 and $Q(w)$ = 1,758.

In accordance with the formula (5), converting acres into square meters, we get:

$$v_{0S} = \frac{\sqrt{k}}{P(\omega)} = \sqrt{\frac{21.4 \cdot 10^4}{1.178}} = 426 \text{ m/day}.$$

In view of a possible error δk the value of the front velocity will be in the following ranges: $275 \le v_{0S} \le 577$ m/day.

With the help of formula (6) we can estimate the rate of the fire perimeter growth:

$$\frac{dL(t)}{dt} = 2v_0(t)Q(w) =$$
$$= 2 \cdot 426 \cdot 1.758 = 1498 \pm 531 \text{ m/day}.$$

The resulting estimate of the fire front rate together with the selected indicatrix of propagation allows us to form the predictive estimates of the fire contour over time.

A simple method of estimating the parameters of the distribution process presented in the article is based on the use of geometric models and allows to predict the forest fires propagation on the basis of information stored in ISDM-Forestry. The methodology uses the indicatrices of two kinds, depending on the wind speed in the area of fire: the exponential one and the elliptic one. The received formulas were used for the numerical calculation of a real fire parameters, their errors were estimated. The described method of calculation was used to construct the contours of fires on the basis of neural network forecasting in a software package "Taiga-2" and in the development of an information system of forest fires forecasting for ISDM-Forestry.

References

1. Information system of remote monitoring of Forestry Federal Agency [Electronic resource]. Pushkino.: FGU Avialesoohrana, 2010. URL: http://www.pushkino.aviales.ru/rus/main.sht.

2. Dorrer G. A. The problems of large forest fires predicting // Materials of XI All-Russian theoretical and practical conf. : collected articles. Krasnoyarsk, 2009. P. 5–8.

3. Komorovsky V. S. The estimation of the possibility to forecast the forest fires propagation according to the data of «ISDM-Forestry» [Electronic resource] // I International scientific conf. «Modern problems of informatization in the systems of modeling, programming and telecommunactions» : collected articles. URL: http://econf.rae.ru/article/4679.

4. Dorrer G. A. Dynamics of forest fires. Novosibirsk : Publishing House SO RAN, 2008.

5. Albini F. A. Estimating wildfire behavior end effects. USDA Forest Service. Gen. Tec. Rep. INT-30, Ogden, 1976. (Intermountain Forest and Range Exp. Stn.).

S. A. Krat, V. V. Hristich
JSC "Academician M. F. Reshetnev" Information Satellite Systems", Russia, Zheleznogorsk

A. A. Filatov
JSC Scientific Industrial Enterprise of Fiber Optical and Laser Equipment, Ltd, Russia, Saint Petersburg

A THERMAL VACUUM SPACECRAFT TEST: THE EXPERIENCE OF CREATING SOLAR SIMULATORS USING HIGH PRESSURE GAS-DISCHARGE LAMPS

In this article we have considered a possibility to use the commercial XBO xenon lamps to create a source of radiation, integrated in solar simulators. We have conducted experimental studies of photonic lamp performance.

Keywords: thermal vacuum test, solar simulator, radiation source, light flux, XBO-lamp.

One of the main stages of ground spacecraft testing is considered to be the thermal vacuum test. The tested object is a thermophysical spacecraft model of its elements. Instead of flight devices and blocks, the platform of the thermophysical model has its own thermal simulators – heater sections, simulating thermal dissipation during flight operations. The resistance sensors are installed in the model's elements to read the temperature information during the thermal vacuum test (TVT).

The purpose of the TVT is an experimental test of the thermal mode and thermal schemes of newly developed external spacecraft elements, devices, and equipment during the injection simulation during orbital operations.

There are some tasks that should be fulfilled during the test: verification of a mathematical conducted heat exchange model between the platform structure elements; verification that the temperature of the structural elements and devices meets the requirements for thermal control subsystem, as well as requirements for external elements when simulating extreme external and internal heat loads; determination of temperature fields by external structural elements of the articled parts; verification of sufficient electrical power for heat from external devices and articled equipment; development of recommendations, concerning the updating of technical documentation on heat schemes for the articled external elements.

During the preparation of test results, the temperature parameters registered and controlled during the test shall be assessed if they are in compliance with the permitted values. If required, some recommendations will be given in a report concerning the thermal scheme update. Thermal vacuum tests are performed on dedicated test benches: vacuum units, equipped with cryogenic screens and simulators of external heat and light fluxes. The test conditions are similar to those at spacecraft in orbit. The thermal vacuum test benches comprise: a thermal vacuum chamber, a vacuum pumping system, a cryogenic system, panels of the heat fluxes simulator, and a solar simulator. The vacuum pumping system provides pressure inside the chamber not exceeding $5 \cdot 10^{-5}$ mm of mercury. The cooled cryogenic screens simulate the environment of black space with a temperature not exceeding –180 ºC.

The solar simulator is the main and most complicated element of the thermal vacuum test. The requirements for the solar simulator are so: the radiation energy spectral distribution must be within (0.2–2.5) micrometer wavelength range, close to solar radiation distribution; the radiation flux density must be 1340–1440 W/m^2 with a simulation error not exceeding 10 % from the nominal values; the light spot sizes must be corresponding to the working field size; the light flux nonuniformity must not exceed 10 %.

The solar simulator is a source of radiation and an optical shaping system, directing flux into the working area.

In the recent studies considered is the possibility to apply gas-discharge XBO-lamps with a high-pressure as a source of radiation for solar simulators based on the thermal vacuum chamber TVC-120 (a unique test bench for spacecraft ground tests).

Quite recently (in 2005), during a thermal vacuum test, the solar simulator setup in the TVC-120 was arranged on ДКсРМ-55000 lamps of Russian production. ДКсРМ-55000-УХЛ-4 is a xenon arc lamp with ultrahigh pressure, with power of 55.000 W. The lamp has a metallic water cooling body with internal reflective optics and a domical output quartz window, which is cooled by distilled water. The moveable cathode and copper anode has a slit water cooling system. The lamp spectrum (fig. 1) is close to solar [1]; however, today it is not produced due to its unprofitability. Since the lamp's resource is limited (200 hours) and the available lamps have already been used, there is a necessity to change the source of radiation.

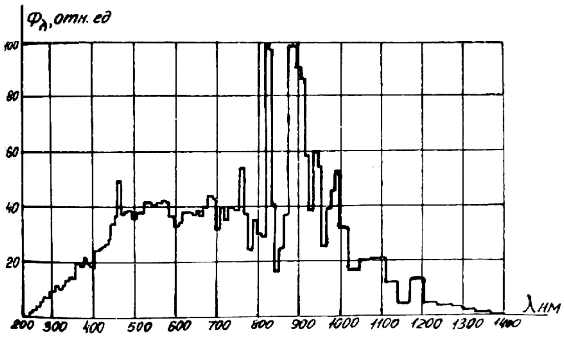


Fig. 1. Radiative energy spectral distribution of the ДКсРМ-55000-УХЛ-4 lamp

A market analysis has shown that XBO lamps can be competitive. These lamps have a reasonable price (since there is a serial production); they are widely applied in film production.

The XBO-lamps belong to the gas-discharge lamp type. In these lamps the light is generated by a discharge arc, which freely ignites in pure xenon between two electrodes. The arc's length is the same as the distance between two electrodes composing several millimeters [2]. The brightness distribution of the lamp arc is shown in fig. 2.

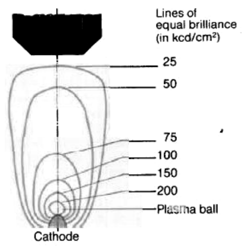


Fig. 2. Brightness distribution in the XBO-lamp arc

As it is shown in fig. 2, the main energy is concentrated inside a plasma ball which is 2–3 mm in size.

This means that the lamps are very close to be an ideal spotlight, which is particularly important for application in solar simulators – the lamp accommodation in optical element focus allows to create and direct the light flux to a desired field with greater accuracy.

In fig. 3 is the XBO-lamp's spectral radiation rate in comparison to black body radiation in given color temperature.

The XBO-lamp has a continuous spectrum of visible radiation range, the intensity of which coincides with solar radiation [2].

The confirmation of the XBO-lamps performance and the possibility of their application in solar simulation in thermal vacuum test were proved during an experimental test bench (fig. 4).

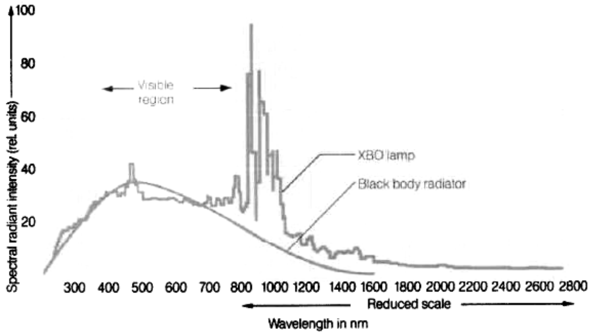


Fig. 3. Spectral radiation rate of the XBO-lamp

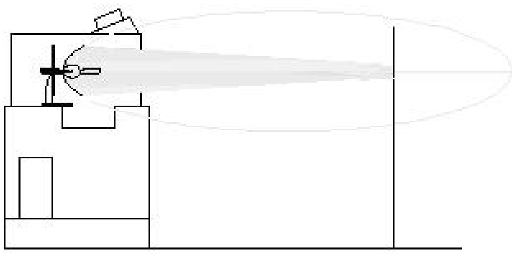


Fig. 4. The scheme of the experimental test bench
for verification of XBO 10000 W/HS OFR lamp performance

The test bench comprises of: film projector Kinoton with a XBO 10000 W/HS OFR lamp (fig. 5) [3]; light flux intensity sensor КФЛП (silicon photocell of radiant flux), attached to a moveable device (fig. 6); a screen for light spot shaping.

Fig. 5. Kinoton film projector with a XBO lamp

The КФЛП sensor consists of a set of silicon plates that convert incoming solar energy into a current. According to the electric signal level it has been defined as light flux power hitting a plate. An ellipsoidal reflector is used for gathering the beams radiated by the lamp in all directions. The XBO lamp installed in the first focus of the reflector radiates the light flux, which then reflects it from the reflector mirror walls to be gathered in the second focus. This occurs in conditions with an ideal spot source. Since the real source had its final sizes, the objective of the experiment was to clarify whether the XBO-lamp is applicable as a source of radiation and what quantity of lamps is required for the desired illumination in the thermal vacuum chamber working area.

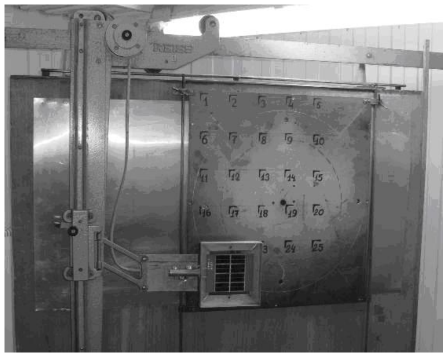


Fig. 6. The screen with marks for flux density scanning

The screen is a flat metallic surface with markings corresponding to dimensions of the photocells' working area and is divided into 25 areas. The screen is placed vertically in the second focus of the ellipsoidal reflector. The central screen area, arc of the lamp, and reflector are located on one optical axis.

In order to verify the compliance with the allowable range of КФЛП values, a rough estimation of power going through this sensor had been performed:

$$P = P_л \cdot K_{кпд} \cdot K_{охв} \cdot K_{отр} \cdot K_{пл},$$

where $P_л$ – power set in the lamp; $K_{кпд}$ – lamp efficiency; $K_{охв}$ – coefficient of light coverage from the lamp by the reflector surface; $K_{отр}$ – reflection ratio of reflector; $K_{пл}$ = 125 considering dimensional equation of КФЛП output value (W/m^2) and its own area (0,008 m^2).

Assuming that $P_л$ = 2560 W , $K_{кпд}$ = 0,6, $K_{охв}$ = 0,6, $K_{отр}$ = 0,85, we get:

$$P = 97920 \text{ (W/m}^2\text{)}.$$

Considering the КФЛП sensor measurement range to be (0–3000 W/m^2), a decision had been made to manufacture mesh from brass with attenuation of approximately about 97920/3000 = 32.6 times.

During the experiment the lamp was switched on at 2 560 W, the light spot on the screen was shaped, and by a КФЛП sensor moving around the screen area, field scanning of 400 × 500 mm had been performed. The average data (illumination values in W/m^2) for 6 sizes are shown in table.

Screen illumination in the areas (W/m^2)

984	1 599	1 968	1 722	1 107
1 722	3 321	5 658	4 428	1 845
26 44.5	11 931	24 600	12 423	3 444
2 263.2	6 519	11 070	4 674	2 337
1 107	1 845	2 214	1 845	1 230
			total	114 500

Based on the obtained data, power distribution inside the shaped spot had been created (fig. 7).

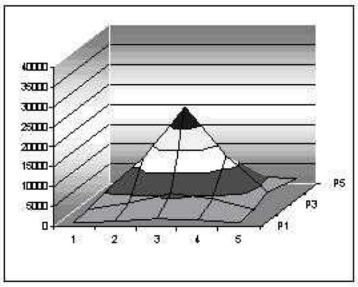


Fig. 7. Power distribution inside the shaped spot

The measurement analysis has shown that in the central area (240 × 300 mm – 9 average elements of the table) 74 % of light energy falling on the screen is concentrated; on the periphery there is only 26 % of energy.

Taking into account the diameter of the thermal vacuum chamber input block (the block through which the light flux passes inside the chamber – 350 mm); for the following calculation it is accepted that energy falling on the input block is on 9 average elements. The total light power coming to the input block is calculated by the formula:

$$P_{сумм} = (\sum_{i=1}^{9} P_{кфлп})/K_{пл},$$

Thus:

$$P_{сумм} = 677 \text{ (W)}.$$

If we consider the way of the light flux through the input block and farther inside the thermal vacuum chamber TVC-120; it also has certain losses.

The input block comprises a mixer and light port, providing hermiticity in the vacuum chamber during the test. Here, there are 4 quartz surfaces. By defining losses on one quartz surface as 4 % [4], we will have losses on the input block of 16 %.

Inside the vacuum chamber there is a collimating mirror. After the mixer, the light flux passes to this mirror, and then it is reflected and goes inside the thermal vacuum chamber, shaping the light flux with section 2 × 1 m^2.

Assuming that the reflection ration of the collimating mirror is 0.85, we get losses inside the chamber of 31 %.

So, according to the experimental data, in the 2 m^2 area there is 467.13 W of light power from one lamp (under the set lamp power of 2.560 W); in the working area of the thermal vacuum chamber there is a light flux density of:

$$467.13 / 2 = 233.6 \text{ (m}^2\text{)}.$$

Based on the requirements for the solar simulator (flux density of 1.340–1.440 W/m^2), it is calculated that the quantity of lamps, needed for generating desired flux density is:

$$N = 1\,440 / 233.6 = 6.2 \text{ (pieces)}.$$

In order to generate the light flux of a desired density during the thermal vacuum test it is sufficient to sum up the fluxes to 7 lamps.

Considering that the experiment was carried out for the lamp with power of 2.560 W, and the maximum lamp power is 10,000 W, we must speak about additional lamp sources. This allows using the lamps in a partial load mode that can provide longer operational lifetime.

This way, the experimental results confirm that it is possible and desirable to use the OSRAM XBO 10 000 OFR lamps in solar simulators. At the same time it is necessary to carry out calculations and make a light optical scheme for summing up the fluxes to 7 XBO-lamps.

If we compare the operational performance of lamps previously used in solar simulators (ДКсРМ-55,000 lamps and OSRAM XBO 10,000 OFR lamps), it is necessary to mention the following: the ДКсРМ-55,000 lamp has an electric power of 55 kW, the optical power is approximately 15 kW, the lamp requires water cooling (too expensive), the average lamp resource is 200 hours. The OSRAM XBO 10,000 lamp requires electric power of 10 kW, provides optical power of approximately 6 kW, and requires only air cooling; the average lamp resource is 500 hours.

Considering other lamp performance (stable light arc; operation under direct current; the possibility of a repeated ignition in a hot state; the full light flux immediately after ignition), it is possible to state that these lamps are quite suitable for application.

A broad application of the XBO-lamps in medicine and film production, pointing to mass production will make them financially available.

References

1. ДКсРМ-55000-УХЛ-4 Lamp. Technical specification. ИКБЖ. 675637. 004 ТО.

2. XBO Theatre Lamps // Technology and Applications. Copyright OSRAM SYLVANIA Inc, 2000. P. 10–12

3. Kinoton. Operating Manual. Gigalight Special Lamphouse. Copyright by KINOTON Filmtheater- und Studiotechnik GmbH, 2002

4. Applied optics: Study guide for Ins. Departments of Instrument Design / L. G. Bebchuk [et al.] ; under the editorship of N. P. Zakaznov. M. : Mechanical Engineering, 1988. P. 312.

A. V. Lapko
Institute of Computational Modelling, Russian Academy of Sciences, Siberian Branch, Russia, Krasnoyarsk

V. A. Lapko
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

THE ANALYSIS OF NONPARAMETRIC MIXTURE PROPERTIES WITH A PROBABLITY DENSITY OF A MULTIDIMENSIONAL RANDOM VARIABLE

The asymptotic properties of a mixture with nonparametric estimations of probability density with a multidimensional random variable are researched in this article. They are compared with the properties of the traditional Rosenblatt–Parzen type nonparametric probability density estimation, depending on the quantity of the composed mixture and dimension of the random variable.

Keywords: mixture of probability densities, nonparametric estimation, large samples, asymptotic properties.

The application of nonparametric statistics methods based on the estimations of Rosenblatt–Parzen type probability density [1; 2] is a rapidly developing modelling method of priori uncertainty systems. However, when the research conditions of the system are complicated, there appear methodical and computing difficulties in traditional nonparametric algorithms and models; this can be clearly observed during the processing of statistical data in great amounts.

The perspective “detour” direction of the arisen problems consists in the application of decomposition principles of training samples according to their size, and the application of the parallel calculation technology.

The purpose of this work is to prove the effective usage of decomposition principles when processing large-scale arrays of statistical data, on the basis of the asymptotic properties’ analysis for a nonparametric estimation of probability density mixture.

Let sample $V = \left(x^i, i = \overline{1, n}\right)$ from n independent observations of $k-$ dimensional random variable $x = \left(x_v, v = \overline{1, k}\right)$ be with a probability density $p(x)$. The type $p(x)$ is a priori unknown.

Let’s divide sample V into T observation groups $V_j = \left(x^i, i \in I_j\right)$, $j = \overline{1, T}$. Multiple observation numbers x in the group with number j shall be identified as I_j. While: $\bigcup_{j=1}^{T} I_j = I = \left(i = \overline{1, n}\right)$. The quantity $n_j = \left|I_j\right|$ of units in samples $V_j = \left(x^i, i \in I_j\right)$ is equal and equals $\overline{n} = \frac{n}{T}$.

At each sample V_j let us construct a nonparametric estimation of probability density with a multidimensional random variable x [1]:

$$\overline{p}_j(x) = \frac{1}{\overline{n} \prod_{v=1}^{k} c_v} \sum_{i \in I_j} \prod_{v=1}^{k} \Phi\left(\frac{x_v - x_v^i}{c_v}\right), \; j = \overline{1, T}. \quad (1)$$

In statistics (1), the nuclear function $\Phi(u_v)$ is satisfied to conditions of normalization, positivity, and symmetry. The parameters of nuclear $c_v = c_v(\overline{n})$ functions decrease with the increase of $\overline{n}$.

Let the intervals of component x_v value change for vector x be identical. In these conditions it is reasonable to assume that the values of coefficients c_v in nonparametric estimations of probability densities $\overline{p}_j(x)$, $j = \overline{1, T}$ are identical and equal to c. Then estimation (1) will look as:

$$\overline{p}_j(x) = \frac{1}{\overline{n}\, c^k} \sum_{i \in I_j} \prod_{v=1}^{k} \Phi\left(\frac{x_v - x_v^i}{c}\right), \; j = \overline{1, T}. \quad (2)$$

As for magnifying $p(x)$ with statistical sample V we shall use a mixture of nonparametric estimations of a probability density type:

$$\overline{\overline{p}}(x) = \frac{1}{T} \sum_{j=1}^{T} \overline{p}_j(x). \quad (3)$$

Statistics (3) allows the usage of parallel calculation technology while estimating the probability density in conditions of large samples.

The asymptotic properties $\overline{\overline{p}}(x)$ are defined by the following statement.

The theorem. Let $p(x)$ and its first two derivatives from each component x_v, $v = \overline{1, k}$ be limited and continuous; the nuclear functions satisfy $\Phi(u_v)$ conditions:

$$\Phi(u_v) = \Phi(-u_v), \quad 0 \le \Phi(u_v) < \infty,$$
$$\int \Phi(u_v)\, du_v = 1, \quad \int u_v^2\, \Phi(u_v)\, du_v = 1,$$
$$\int u_v^m\, \Phi(u_v)\, du_v < \infty, \quad 0 \le m < \infty; \; v = \overline{1, k},$$

of sequence $c = c(\overline{n})$ for blur coefficient in nuclear functions are such, that at $\overline{n} \to \infty$ values, $c \to 0$ and $\overline{n}\, c^k \to \infty$.

Then at finite values T the nonparametric estimation (3) of the probability density $p(x)$ has a property of asymptotic unbiasedness and competence.

Hereinafter infinite limits of integration are omitted.

The proof:

1. By definition:

$$M\left(\bar{\bar{p}}(x)\right)=\frac{1}{T}\sum_{j=1}^{T}M\left(\bar{p}_j(x)\right)=\frac{1}{T}\sum_{j=1}^{T}\frac{1}{\bar{n}c^k}\sum_{i\in I_j}\int\ldots$$

$$\ldots\int\prod_{v=1}^{k}\Phi\left(\frac{x_v-x_v^i}{c}\right)p(x_1^i,\ldots,x_k^i)\,dx_1^i\ldots dx_k^i=$$

$$=\frac{1}{c^k}\int\ldots\int\prod_{v=1}^{k}\Phi\left(\frac{x_v-t_v}{c}\right)p(t_1,\ldots,t_k)\,dt_1\ldots dt_k=$$

$$=\int\ldots\int\prod_{v=1}^{k}\Phi(u_v)\,p(x_1-cu_1,\ldots,x_k-cu_k)\,du_1\ldots du_k,$$

where M – is a mathematical expectations sign. When performing the conversion, it is considered that statistical sample units V_j, $j=\overline{1,T}$ are values of the same random variable t with a density probability of $p(t_1,\ldots,t_k)$.

Let's spread out $p(x_1-cu_1,\ldots,x_k-cu_k)$ in the Taylor row at point $x=x_1,\ldots,x_k$ and being limited by the first two terms of the series, we get:

$$W_1=M\left(\bar{\bar{p}}(x)-p(x)\right)\sim\frac{c^2}{2}\sum_{v=1}^{k}p_v^{(2)}(x), \qquad (4)$$

where $p_v^{(2)}(x)$ – is the second derivative of the probability density $p(x)$ at component x_v.

From here, in condition that $c\to 0$ at $\bar{n}\to\infty$, appears the property of the asymptotic unbiasedness for a mixture of nonparametric probability density estimations (3).

2. For convergence proof of $\bar{\bar{p}}(x)$ in square mean we shall consider the following expression:

$$M\int\ldots\int\left(p(x)-\bar{\bar{p}}(x)\right)^2dx_1\ldots dx_k=$$

$$=M\int\ldots\int\left[\frac{1}{T}\sum_{j=1}^{T}\left(p(x)-\bar{p}_j(x)\right)\right]^2dx_1\ldots dx_k=$$

$$=\frac{1}{T^2}M\left[\sum_{j=1}^{T}\int\ldots\int\left(p(x)-\bar{p}_j(x)\right)^2dx_1\ldots dx_k+\right.$$

$$\left.+\sum_{j=1}^{T}\sum_{\substack{t=1\\t\neq j}}^{T}\int\ldots\int\left(p(x)-\bar{p}_j(x)\right)\left(p(x)-\bar{p}_t(x)\right)dx_1\ldots dx_k\right]. \qquad (5)$$

Let's find the asymptotic component expression for the second part of expression (5):

$$M\int\ldots\int\left(p(x)-\bar{p}_j(x)\right)\left(p(x)-\bar{p}_t(x)\right)dx_1\ldots dx_k=$$
$$=\int\ldots\int p^2(x)\,dx_1\ldots dx_k-M\int\ldots\int\bar{p}_t(x)p(x)\,dx_1\ldots dx_k-$$
$$-M\int\ldots\int\bar{p}_j(x)p(x)\,dx_1\ldots dx_k+$$
$$+M\int\ldots\int\bar{p}_j(x)\bar{p}_t(x)\,dx_1\ldots dx_k. \qquad (6)$$

Let's transform its last part:

$$M\int\ldots\int\bar{p}_j(x)\bar{p}_t(x)\,dx_1\ldots dx_k=$$
$$=\int\ldots\int M\left(\bar{p}_j(x)\right)M\left(\bar{p}_t(x)\right)dx_1\ldots dx_k,$$

which, with great enough volumes of statistical data considering expression (4) is presented as:

$$\int\ldots\int\left(p(x)+\frac{c^2}{2}\sum_{v=1}^{k}p_v^{(2)}(x)\right)^2dx_1\ldots dx_k. \qquad (7)$$

Notice that the asymptotic statistics expression of type:

$$M\int\ldots\int\bar{p}_t(x)p(x)\,dx_1\ldots dx_k$$

corresponds to:

$$\int\ldots\int\left(p(x)+\frac{c^2}{2}\sum_{v=1}^{k}p_v^{(2)}(x)\right)p(x)\,dx_1\ldots dx_k. \qquad (8)$$

Substituting expression (7), (8) in (6), after a series of simple conversions will give:

$$M\int\ldots\int\left(p(x)-\bar{p}_j(x)\right)\left(p(x)-\bar{p}_t(x)\right)dx_1\ldots dx_k\sim$$

$$\sim\frac{c^4}{4}\int\ldots\int\left(\sum_{v=1}^{k}p_v^{(2)}(x)\right)^2dx_1\ldots dx_k=\frac{c^4}{4}B. \qquad (9)$$

In V. A. Epanechnikov's research [2] – an asymptotic expression for the purpose of square deviation in nonparametric probability density estimation $p(x)$, composing the first part of expression (5), is received:

$$M\int\ldots\int\left(p(x)-\bar{p}_j(x)\right)^2dx_1\ldots dx_k\sim$$
$$\sim\frac{\prod_{v=1}^{k}\int\Phi^2(u_v)\,du_v}{\bar{n}\,c^k}+\frac{c^4}{4}B. \qquad (10)$$

Accounting (9) and (10), expression (5) with enough $\bar{n}$ values is represented as:

$$M\int\ldots\int\left(p(x)-\bar{\bar{p}}(x)\right)^2dx_1\ldots dx_k\sim$$
$$\sim\frac{\prod_{v=1}^{k}\int\Phi^2(u_v)\,du_v}{T\,\bar{n}\,c^k}+\frac{c^4}{4}B. \qquad (11)$$

It is not difficult to notice that in conditions $c\to 0$ at $\bar{n}c^k\to\infty$ the estimation $\bar{n}\to\infty$ of probability density mixture (3) converges in square mean to $p(x)$; considering the property of its asymptotic unbiasedness is well-founded.

At $T=1$ the received result (11) coincides with Epanechnikov's theorem [2], which confirms the correctness of the fulfilled conversions.

The analysis of approximating properties of statistics $\bar{\bar{p}}(x)$. For the efficiency analysis of a nonparametric estimation of probability densities mixture (3) and the Rosenblatt–Parzen estimations of a probability density:

$$\tilde{p}(x)=\frac{1}{n\,c^k}\sum_{i=1}^{n}\prod_{v=1}^{k}\Phi\left(\frac{x_v-x_v^i}{c}\right)$$

let's consider the ratio of asymptotic expressions, corresponding to deviation squares for the best coefficients of blur values in nuclear functions.

Let's define the minimum value W_2 of expression (11) with optimal coefficient c^* values of blur nonparametric estimations $\bar{p}_j(x)$ composing the probability densities mixture. In the accepted assumption value:

$$c^* = \left(\frac{k \prod_{v=1}^{k} \int \Phi^2(u_v)\, du_v}{\bar{n} B} \right)^{\frac{1}{(k+4)}} .$$

Then:

$$W_2 = \left[\left(\frac{\prod_{v=1}^{k} \int \Phi^2(u_v)\, du_v}{\bar{n}} \right)^4 B^k \right]^{\frac{1}{(k+4)}} \frac{4 + T\,k}{4T\,k^{\frac{k}{(k+4)}}} . \quad (12)$$

If $k = 1$, then W_2 – is coincides with the minimal asymptotic expression of square deviation for the mixture of nonparametric probability densities estimations, obtained in study [3].

At $T = 1$ and $\bar{n} = n$ expression (12) corresponds to the minimal asymptotic expression W_2' for a deviation square of the probability density Rosenblatt–Parzen type estimation [2].

After simple conversions we get:

$$R_2 = \frac{W_2}{W_2'} = \frac{4 + T\,k}{(4+k)\,T^{\frac{k}{(k+4)}}} .$$

By analogy we shall calculate the ratio for the minimal values of the main dispersing composing statistics $\bar{\bar{p}}(x)$ and $\tilde{p}(x)$:

$$W_3 = \frac{1}{T\,k^{\frac{k}{(k+4)}}} \left[\left(\frac{\prod_{v=1}^{k} \int \Phi^2(u_v)\, du_v}{\bar{n}} \right)^4 B^k \right]^{\frac{1}{(k+4)}} ,$$

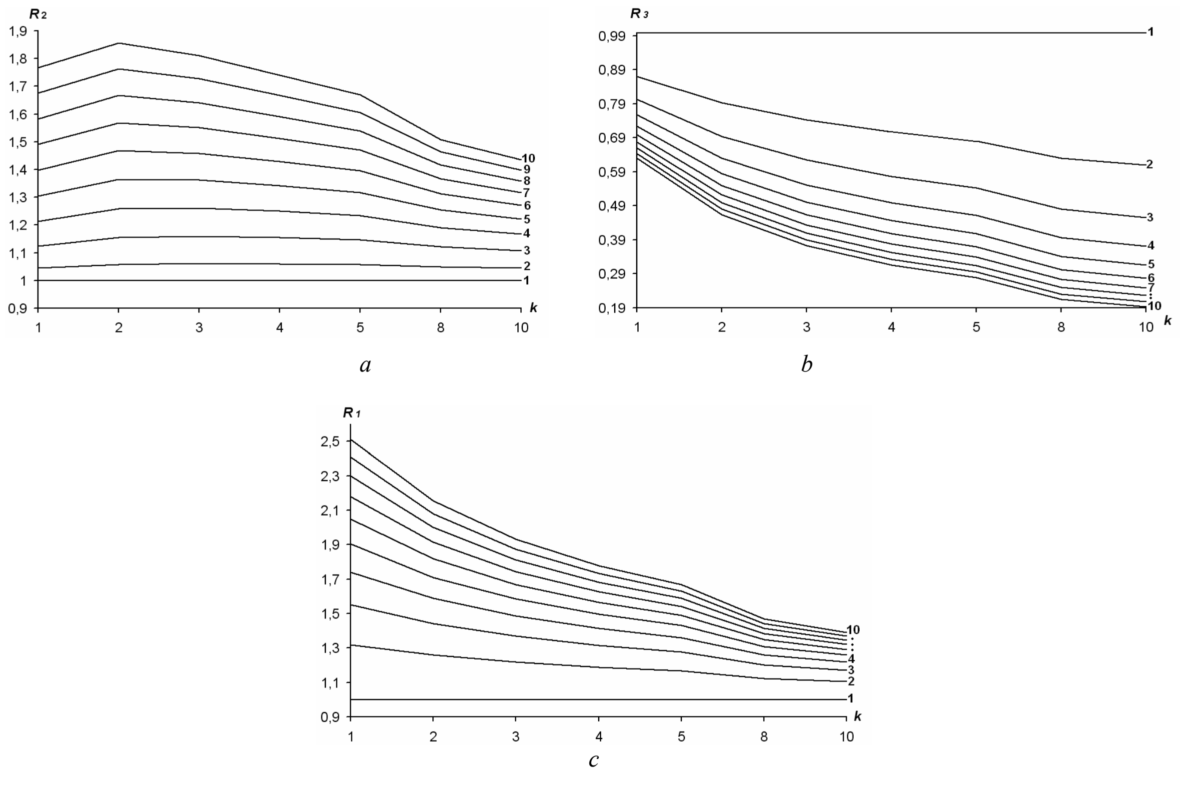
$$W_3' = \frac{1}{k^{\frac{k}{(k+4)}}} \left[\left(\frac{\prod_{v=1}^{k} \int \Phi^2(u_v)\, du_v}{n} \right)^4 B^k \right]^{\frac{k}{(k+4)}} .$$

Their ratio looks as:

$$R_3 = \frac{W_3}{W_3'} = \frac{1}{T^{\frac{k}{(k+4)}}} .$$

It is easy to be convinced, that the ratio of asymptotic expressions offset: W_1, W_1' for the estimated probability density $\bar{\bar{p}}(x)$ and $\tilde{p}(x)$ at optimal blur coefficients for nuclear functions, is equal to:

$$R_1 = \frac{W_1}{W_1'} = T^{\frac{2}{(k+4)}} .$$



Dependences of ratios R_2 (*a*), R_3 (*b*), R_1 (*c*) from the dimension of random variable k and $x = \left(x_v,\, v = \overline{1, k}\right)$ quantity T = 1–10 (curves 1, …, 10), composing the nonparametric estimations mixture of probability density $\bar{\bar{p}}(x)$ (3)

With growth of component quantity T of the nonparametric estimations mixture of probability density, there is an increase in ratio values $R_2 > 1$ (figure, *a*), $R_1 > 1$ (figure, *c*). The noticed deterioration of approximating mixture properties $\bar{\bar{p}}(x)$ in comparison to traditional nonparametric estimation of density probability $\tilde{p}(x)$ (12), points to the decrease in sample sizes used during the estimation of compositions $\bar{\bar{p}}(x)$. This is a special feature of minor dimensions k of random variables. When complicating the estimating probability density with efficiency k, the growth of nonparametric estimations $\bar{\bar{p}}(x)$ also decreases $\tilde{p}(x)$. Criteria corresponding to them W_2, W_2' and W_1, W_1' become commensurable; this is evident in the decreasing of ratio R_2 and R_1 values.

The offered mixture $\bar{\bar{p}}(x)$ of probability density estimations has a lesser dispersion in comparison to the nonparametric estimation $\tilde{p}(x)$, which is identified by its structure, since statistics synthesis $\bar{\bar{p}}(x)$ is carried out on the basis of an averaging operator (figure, *b*). With a quantity increase in T composing the mixture of nonparametric estimations $\bar{\bar{p}}(x)$, the density probability and dimension k of random dimensions increases.

On the basis of the asymptotic properties analysis for nonparametric estimations mixtures of probability density with a multidimensional random variable, the decomposition possibility for initial statistical data under a synthesis of nonparametric statistics in large samples conditions is justified. The researched statistics, in comparison to the traditional Rosenblatt – Parzen nonparametric evaluation, has a considerably smaller dispersion and allows using parallel calculating technologies.

References

1. Parzen E. On estimation of a probability density function and mode // Ann. Math. Statistic. 1962. Vol. 33. P. 1065–1076.

2. Epanechnikov V. A. Nonparametric estimation of a many-dimensional probability density // Teoriya veroyatnosti i ee primeneniya, 1969. Vol. 14. № 1. P. 156–161.

3. Lapko V. A., Varochkin S. S., Egorochkin I. A. Development and research of a nonparametric estimation of the probability density grounded on a principle of decomposition of learning sample on its size // Vestnik SibSAU. 2009. Vol. 1 (22). P. 45–49.

D. V. Lichargin
Siberian Federal University, Russia, Krasnoyarsk

GENERATION OF THE STATE TREE BASED ON GENERATIVE GRAMMAR OVER TREES OF STRINGS

In the article the principle of state trees generation is considered based on the generative grammars over trees of strings in such objects as the sentences of natural languages, as well as two and tree dimensional images. The image of the object as a forest is considered; including the trees of object different layouts for the purpose of complex system modeling.

Keywords: natural language generation, generative grammars, semantics.

The problem of natural language sentences generation is one of the key issues in the field of computer science and formal grammar theories. The issue of meaningful speech generation applies to the area of semantics and computer science [1–7]. The states tree generation issue is studied well enough in computer science and in system analysis. In respect to the question of meaningful phrases tree generation the problem is first of all connected to the method of sentence generation by means of Chomsky's generative grammars. Generative grammars are successfully applied in software such as electronic translation systems, expert systems, systems of orthography checking, etc.

The flash point of the article is the analysis prospects for using generative grammars not over strings, but over trees of strings. In this respect it is possible to solve the task of generating grammatically and semantically meaningful speech more effectively and increasing the efficiency of different images analysis and synthesis aspects.

The importance of the issue on effective generating language meaningful constructions and two or three dimensional images is generally understood and is connected with the demands of linguistic and other software.

The purpose of this research is to apply generative grammars on the necessity basis over trees as means of meaningful speech generation connected with greater heterogeneous context.

The novelty of the work is in the application of generative grammars not over strings but over trees of strings.

It is well-known that standard generative grammar over strings have the form of the four: G<S, T, N, R>, where S is an initial symbol of the generative grammar, T is a set of terminal symbols, N is a set of non-terminal symbols, and R is a set of rules for transforming one string into another.

For generative grammars over trees, the strings of symbols *t* and *n* are substituted by trees (or forest – trees with equal nodes). $t = t <t', t'', \ldots, t^n>$, where $t' = t' <t^1, t^2, \ldots, t^m>$ etc, $n = n <n', n'', \ldots, n^n>$, where $n' = n' <n^1, n^2, \ldots, n^m>$ etc.

One of the main particularities of any system is the presence of hierarchy elements in the system. Meanwhile the hierarchy relations can sometimes be presented as a set of hierarchies in different layouts of the system consideration. For example, the sum of three systems: a sentence as a distributed narration, a sentence with the purpose to order the tea, and a sentence with the purpose to support polite dialogue can result in a meaningful sentence of the natural language. At the same time, for the generation of such complex systems with several purposes and layouts of consideration it is necessary to use more complex means of generative grammar over the trees of strings, for the purpose of generating the tree of possible natural language sentences.

The generative grammar over trees of strings is composed in the following way. Let A<…B<…C1→C2…>,…, B'<…C1'→C2'>…> is a rule of generative grammar over trees, from a set of such rules with the trees of strings for terminal symbols T and non-terminal symbols N, «→» is a symbol of transferring one string to another. S<> is an initial symbol of generative grammar over trees.

During each stage of deepening the tree of states into another generated tree, or a forest of strings reduced to the multiplication of the obtained generated tree by the generative grammar rule.

It is also possible to consider the trees of heterogeneous information A<B{B1, B2}, C{C1, C2}> = {A<B1, C1>, A<B1, C2>, A<B2, C1>, A<B2, C2>} = {A<B1, C1>, A<B1, C2>, A<B2, C{C1, C2}>}. In this respect, the tree of states can be included into the elements' tree and vice versa.

As a result, the sentence can be considered as the union (addition) of trees from different consideration layouts over the whole space (tree) of natural language points [4–6].

Let's have a tree A <B <B'<…>, B''<…>,…, B'''<…>>, C <C'<…>, C''<…>, C'''<…>>,…, D <D'<…>, D''<…>, …,D'''<…>>> or briefly A <…B <…B''…>…>, then the forest of trees can be considered as a set of trees with equal nodes over a set of the trees' nodes: F<A<…B<…B''(=L1)…>…>, X<…Y<…Y''(=L1)…>…>,…>, where L1 is an equal node of the first two trees from the example above.

Let's consider an example; the tree maneuvers in a chess game: Board <Column [1] <Cell [1], Cell [2], …>, …>, such a tree is formed by position multiplication over the chess board by a set of rules for possible half-moves.

The half-move of a knight can be such: Board <…Column [X] <…Cell [Y] <Knight→Empty>>, …, Column [(X + 1) or (X – 1)] <Cell[(Y + 2) or (Y – 2)] <Empty→Knight>…>…> .

The generation of a chair image for example, can presuppose a potential image of a person on the chair. Chair <Seat, Legs, Back, Person(= L1) <Arms(= L2), Legs(= L3), Trunk(= L4), Head(= L5)>> + Gentleman(= L1) <Body <Arms(=L2), Legs(=L3), Trunk(=L4), Head(= L5)>, Clothes <Jacket <Trunk<= L4>>, Boots, Top Hat <Head(=L5)>>> = Image<Chair<…>, Gentleman<…>,…>.

The principle of reducing or adding the images is the following: semantically analogue elements – tree nodes are declared to be identical, in the case for several reduction variants an additional sub-tree of possible system states is built as a result of adding system element trees or generating system state trees.

The sentence of the natural language can be presented in the form of a tree as well. For example, the tree of grammatical sentence analysis the can be simplified as: Clause <Introductory Word, Modifier, Subject <Determiner, Attribute <Adverb of Degree, Adjective Group>, Nominal Group>, Predicate <Modality, Modifier, Verbal Part>, Object <Determiner, Attribute <Adverb of Degree, Adjective Group>, Nominal Group>, Modifier>.

The tree can be added to (reduced by) the tree of semantic analysis, for example, the Topic "Building" <Relation-Creature-Building {enter, build}, Properties-Building {marble, multistoried}, Здание {house, library}, Modifier 1 <with/without {with, without}, Essence of – Building/Rooms {corridor, hall}>, Modifier 2 <with/without {with, without}, Property-Thing (Essence of – Building/Architectural Element {large, beautiful}), Essence of -Building/Architectural Element {wall, corner}>>.

A tree of the following type can be used for the generation of natural language sentences:

1. Subject – Essence (the … / person / man / woman).
2. Modality – Action over Relation (want / wish / love / adore).
3. Predicate – Action with Clothes (buy / get / try on / wear).
4. Object – Clothes (the … / jeans / sweater / footballer).

The given tree can be multiplied by the following rule of generative grammar.

1. 0 → the.
2. 0 → Attribute – Property of Clothes (stylish / fashionable / checked.
3. Object – Clothes (The … → 0 / jeans / sweater / footballer).

In result, a sentence like: "the person wants to get the fashionable sweater" or "the woman wishes to buy the checked footballer" is obtained.

It can be assumed that the analyzing of the image recognition problem, natural language analysis, and a number of other problems can be effectively solved only based on their synthetic joint consideration. For example, for the translation of the word-combination "up-link communication" not into the English language as "communication with a satellite" it is necessary to use a visual image of the facts discussed in the text. This way, in a system of translations, while the text translating is a semantically visual image of narration that should be grown, a translation without a latter close to the human one is impossible.

For the realization of the principles aforementioned, it is necessary to start the elaboration of the dictionary for semantic trees of heterogeneous data: images, patterns of sentences composition, algorithms, and so on. It will be necessary to use the already existing dictionaries of sentences generation in the "Electronic Dictionary" software for the system basis.

In conclusion it is necessary to mention that generative grammar over the trees of strings is an effective means of generating the state trees for such systems, like natural language sentence and semantically loaded images. It is thought to apply the generative grammars over the trees of strings on the basis of the Semantic Trees' Dictionary, which is a classification of heterogeneous semantic data.

References

1. Agamdjanova V. I. Contextual Redundancy of the Lexical Meaning of a Word. M. : Higher School, 1977.
2. Apresyan Yu. D. Ideas and Methods of Modern Structural Linguistics. M. : Science, 1966.
3. Verdieva Z. N. Semantic Fields in the Modern English Language. M. : Higher School, 1986.
4. Lichargin D. V. Operations over the Natural Language Words Semes in Machine Translation // Works of the Conf. of Young Scientists. Krasnoyarsk : ICM SB RAS, 2003. P. 23–31.
5. Lichargin D. V. Elimination of Semantic Noise as the Means of Adequate Translation // Questions of the Theory and Practice of Translation. Works of All-Russian Conference. Penza : Privolzhye Region Knowledge House, 2003. P. 90–92.
6. Lichargin D. V. Generation of the Natural Language Phrases within the Task of Creating Natural Language Interface with Software // Problems of the Territory Information Development : Materials of the Eighth All-Russian Conf. PTID. 2003. Vol. 2. Krasnoyarsk : IP Centre of Krasnoyarsk State Technical University, 2003. P. 152–156.
7. Nikitin M. V. Lexical Meaning of a Word. M. : Higher School, 1983.

P. K. Lopatin

Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

AN ALGORITHM FOR AN OBJECT GRASPING BY A MANIPULATOR IN AN UNKNOWN STATIC ENVIRONMENT

An algorithm for a n-link manipulating robot (MR) control in an environment with unknown static obstacles is considered. A theorem is proved which states that following the algorithm a MR in a finite number of steps will either grasp an object or will give a proved conclusion that an object cannot be grasped in any configuration.

Keywords: robot, unknown environment, obstacles, reachability.

In MR control the following typical problem arises: a MR should move from a start configuration $\boldsymbol{q}^0$ and grasp an object *Obj* by its gripper. Herewith sometimes the *Obj* may be grasped not in one but in several and sometimes in an infinite number of target configurations $\boldsymbol{q}_i^T$. The target configurations are united into a target set B_T. The set B_T has an arbitrary shape.

Let us consider that the B_T does not grow during the whole movement of MR. Consider also that the coordinates of every point from B_T are known and defined reliably.

A MR is represented in the configuration space (generalized coordinate space) as a point. MR functioning should take place in the bounded region X of the configuration space. Let's consider that X is such that for any $\boldsymbol{q} \in X$ the following inequalities are fulfilled:

$$\boldsymbol{a}^1 \le \boldsymbol{q} \le \boldsymbol{a}^2, \quad (1)$$

where $\boldsymbol{a}^1 = (a_1^1, a_2^1, ..., a_n^1)$ is a vector of lower bounds on the generalized coordinates values, $\boldsymbol{a}^2 = (a_1^2, a_2^2, ..., a_n^2)$ is a vector of upper bounds on the generalized coordinates values of a MR, $\boldsymbol{q} = (q_1, q_2,..., q_n)$ is a vector of the generalized coordinates of a MR. So X is a hyper parallelepiped. We will consider all points not satisfying (1) as forbidden.

Moreover, it is necessary to take into account that there also may be forbidden states inside X. Firstly these are the states (configurations) conditioned by constructive limitations of a MR, for example, those in which inadmissible intersection of MR links takes place. It is possible to calculate such forbidden configurations in advance. Secondly we will consider a configuration as forbidden in case when it intersects obstacles. It is impossible to calculate all such configurations in advance in the conditions of an unknown environment. So we will consider a configuration as forbidden if a MR cannot be present in it because of constructive limitations or because of intersection with an obstacle. Before the MR movement beginning we do not have information about forbidden states in X or it is incomplete. If we do not have exact information that a point $\boldsymbol{q}^* \in X$ is forbidden, we will consider such point as allowed.

Now let us consider points from B_T. We will consider a point $\boldsymbol{q}^T \in B_T$ as allowed if it satisfies both criteria: 1) it is not forbidden in senses described in previous paragraphs, 2) it may be reached from $\boldsymbol{q}^0$ in a finite number of steps moving in X by allowed states. We will consider points from B_T as forbidden if they do not satisfy at least one criterion. As far as we need to find out whether the set B_T is reachable at least in one point we will consider that before the MR movement we do not have information about any point from B_T whether it is forbidden or allowed. Now let us formulate the following Problem of a MR control in an unknown static environment: a start configuration $\boldsymbol{q}^0$ and a target set B_T are given. It is necessary to propose an algorithm which in a finite number of steps will either move the MR from $\boldsymbol{q}^0$ to a point from B_T or will give a proved conclusion that there is no allowed state in B_T.

Review of related works. Currently there are many works dedicated to algorithms for dynamic systems (DS) control and, in particular, for robotic systems (RS) in known and unknown environments. There are good reviews of such algorithms [1; 2]. There are algorithms, for example, the forward search algorithm and the *A** algorithm [3] which in a finite number of steps will either find a path from $\boldsymbol{q}^0$ to a point from B_T or inform that there is no point in B_T reachable from $\boldsymbol{q}^0$.

Some algorithms for planning in a known environment in principle may be used for movement in an unknown environment. If we discretize the configuration space then we may use graph algorithms for a DS path searching [2; 3]. But these algorithms have one common feature which makes their application for the DS control in an unknown environment very difficult. The feature is that they demand to carry out the breadth-first search in a certain volume otherwise reaching of a target point $\boldsymbol{q}_i^T$ is not guaranteed [4]. But during the breadth-first search the following situation often arises: suppose we have just finished considering the vertices adjacent to a vertex $\boldsymbol{q}$ and we have to consider vertices adjacent to a vertex $\boldsymbol{q}'$ and the $\boldsymbol{q}$ and $\boldsymbol{q}'$ are not adjacent. In order to consider vertices adjacent to the $\boldsymbol{q}'$ the manipulator has to come to the $\boldsymbol{q}'$ at first. So we get a problem of the manipulator movement from $\boldsymbol{q}$ to $\boldsymbol{q}'$. The necessity of searching and following paths for multiple different $\boldsymbol{q}$ and $\boldsymbol{q}'$ makes the total sum of the manipulator movements very big [4]. In case when we plan a path in a known environment a computer simply "switches its attention" from $\boldsymbol{q}$ to $\boldsymbol{q}'$, which are stored in its memory.

According to classification [2] it is possible to outline the following representatives of the breadth-first approach: proper breadth-first searching algorithm, A* algorithm, best-first heuristic search, lazy PRM, dynamic programming. The methods based on a randomized potential field, Ariadne's Clew algorithm, rapidly-exploring random trees [2] have such feature that new vertices are generated randomly and therefore using these methods for the unknown environment leads to the same difficulties. The approaches based on cell decomposition, visibility (bitangent) graphs, Voronoi diagrams [2] are reduced to alternate graph building and searching a path on it and have the above mentioned disadvantage connected with multiple mechanical movements.

In the algorithm presented in this article the vertices $\boldsymbol{q}$ and $\boldsymbol{q}'$ are always neighbor vertices and it reduces the number of movements.

It is also known that the "depth-first" algorithms do not guarantee reaching the goal [4].

There is a common difficulty for the methods of path planning in the presence of known obstacles: it is very difficult to borrow full information about workspace of a robot in advance and to represent this information in a form suitable for trajectory planning. Considering our algorithm one can see that there is no need for a control system to have full information about workspace in advance, a manipulator will borrow necessary information by itself in limited quantities and in terms of generalized coordinates which is suitable for path planning.

The attempts of creating algorithms for the robot control in presence of unknown obstacles were made. Most of them cover various two-dimensional cases [5].

In [6–9] different approaches for a robot control in two-dimensional unknown environment are considered. In [6; 9] the approaches are based on Voronoi diagrams, in [8] a tabu search approach is presented. As these approaches demand alternate graph building and searching a path on it they lead to multiple robot movements. In [7] obstacles should have polygonal form. The application of methods proposed in [6–9] to a *n*-link manipulator control in an unknown environment is not presented.

In [5] an algorithm for the control of manipulators in the presence of unknown obstacles in three-dimensional space is given. A MR must have not more than three elements and the last kinematic pair should be sliding. In the given conditions the algorithm in a finite number of steps either transfers a MR into a target configuration or informs that it is unreachable.

In [10] the n-dimensional case is considered. The algorithm is based on the solution of the system of nonlinear equations using Newton method and therefore it cannot guarantee the reaching of a target position.

In [2] algorithms for moving a robot in the presence of uncertainty (including cases of an unknown environment) are considered. The algorithms are based on the sequential solution theory. In general the algorithms do not guarantee reaching the goal. In cases when the algorithms use searching on a graph the above mentioned difficulty arises connected with multiple mechanical movements.

In [11] an algorithm for controlling dynamic systems in an *n*-dimensional state space in presence of unknown forbidden static states is proposed. The algorithm in a finite number of steps either moves the DS from a start point $\boldsymbol{q}^0$ to a target point $\boldsymbol{q}^T$ or gives a proved conclusion that the $\boldsymbol{q}^T$ is unreachable. The algorithm's disadvantage is that it assumes that in the set X any state may be reached from any state. In a number of cases such demand may contradict the purpose of the algorithm to define in a finite number of steps whether a target state is reachable from a start state.

In [4] an approach to robot's (including manipulating robots) control in a n-dimensional state space is proposed. The essence of this approach is that the robot generates a path, connecting a start point $\boldsymbol{q}^0$ and a target point $\boldsymbol{q}^T$, not intersecting known forbidden states and tries to follow the trajectory either till reaching the $\boldsymbol{q}^T$ or till meeting an earlier unknown forbidden state in a point $\boldsymbol{q}^n$. In the last case a new path $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$ is generated, connecting $\boldsymbol{q}^n$ and $\boldsymbol{q}^T$ and not intersecting any known forbidden state. It is shown [4] that the problem of the robot movement from $\boldsymbol{q}^0$ to $\boldsymbol{q}^T$ in an unknown environment is reduced to a solution of a finite number of problems of generating and following the path $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$ (in other words to a solution of a finite number of problems PI (planning in a known environment)). Herewith it is supposed that we have a priori information that the $\boldsymbol{q}^T$ is reachable.

In this article we used the approach [4] for a solution of the Problem. The Problem is reduced to investigation of reachability of a finite number N_{BT} of points $\boldsymbol{q}_i^T$, $I = 1, 2, \ldots, N_{BT}$. The demand of the a priori knowledge of $\boldsymbol{q}_i^T$, $I = 1, 2, \ldots, N_{BT}$ reachability is omitted. New demands to the PI procedure necessary for the Problem solution were formulated.

In [12–15] algorithms were given which in a finite number of steps move a DS from $\boldsymbol{q}^0$ to $\boldsymbol{q}^T$ in an environment with unknown static states. Herewith it was supposed that there is a priori information that $\boldsymbol{q}^T$ is reachable.

Preliminary conditions

1. Let us extract from the set B_T a finite number N_{BT} of points $\boldsymbol{q}_i^T$, $I = 1, 2, \ldots, N_{BT}$. These are the configurations whose reachability will be explored. Further we will consider B_T as a list of configurations $\boldsymbol{q}_i^T$, $I = 1, 2, \ldots, N_{BT}$. Let's consider that B_T is not replenished and therefore N_{BT} is not increased, consider that coordinates of every point from B_T are defined reliably.

2. Consider that we have a procedure PI which in a finite number of steps either generates a path from an arbitrary allowed point $\boldsymbol{q}^n \in X$ to an arbitrary point $\boldsymbol{q}^T \in X$ in the presence of known forbidden states or informs that $\boldsymbol{q}^T$ is unreachable. Such procedures already exist, for example, the forward search algorithm or the *A** algorithm [3], which for any start point $\boldsymbol{q}^n$ and any target point $\boldsymbol{q}^T$ under given known forbidden states in a finite number of states either find a path from $\boldsymbol{q}^n$ to $\boldsymbol{q}^T$ or inform that a path from $\boldsymbol{q}^n$ to $\boldsymbol{q}^T$ cannot be found.

3. The obstacles' disposition inside the MR working area does not change during the whole time of the MR movement.

4. The obstacles' number inside the MR working area does not change during the whole time of the MR movement.

5. The MR movement including the resultant path should take place inside the hyperparallelepiped (1).

6. The MR has a sensor system (SS) which supplies information about a r-neighborhood of a current MR point $\boldsymbol{q}^n \in X$. The current point of the MR is the point where the MR is situated right now. The *r*-neighborhood of the $\boldsymbol{q}$ is a hyperball in X with a center in $\boldsymbol{q}$ and radius $r > 0$. We denote the set of all points comprising the *r*-neighborhood of the $\boldsymbol{q}$ as $Y(\boldsymbol{q})$. The words "supplies information about the *r*-neighborhood of the point $\boldsymbol{q}$" mean that the SS defines whether every point from $Y(\boldsymbol{q})$ is allowed or forbidden. Herewith all forbidden points from $Y(\boldsymbol{q})$ are stored in a set $Q(\boldsymbol{q})$ and all allowed points from $Y(\boldsymbol{q})$ are stored in a set $Z(\boldsymbol{q})$. There may be different ways of the sets $Y(\boldsymbol{q})$, $Q(\boldsymbol{q})$, $Z(\boldsymbol{q})$ representation – in a form of formulas, lists, tables etc., but we consider that we have such representation. We will not consider the SS structure.

7. Consider that we have a program *Procedure1*(B_T, N_{BT}, $Q(\boldsymbol{q}^n)$). *Procedure1*($\cdot$) in the moment of call gets the set B_T, the number N_{BT} of points in the set B_T, the set $Q(\boldsymbol{q}^n)$ which was formed during the last call of the SS. *Procedure1*($\cdot$) throws out of the B_T those points which coincide with points from $Q(\boldsymbol{q}^n)$. After the throw the points left in the B_T are renumerated by the continuous numeration beginning from 1 and the number of points left in B_T after execution of *Procedure1*($\cdot$) is inscribed in N_{BT} .

8. Consider that we have a program *Procedure2*(B_T, N_{BT}, $\boldsymbol{q}^T$). *Procedure2*($\cdot$) in the moment of call gets the set B_T, the number N_{BT} of points in the set B_T and the point $\boldsymbol{q}^T$. *Procedure2*($\cdot$) throws out of the B_T the point $\boldsymbol{q}^T$. After that the points are renumerated by the continuous numeration beginning from 1 and the number of points left in B_T after execution of *Procedure2*($\cdot$) is inscribed in N_{BT}.

Below the Algorithm for the solution of the Problem is given. Before a movement beginning the current configuration $\boldsymbol{q}^c$ of the MR is $\boldsymbol{q}^0$, during the movement the Algorithm 1 may be called from other current configurations of the MR.

Algorithm

If $N_{BT} = 0$ then the Algorithm terminates its work with a message that the *Obj* cannot be grasped. Otherwise we consider that the first point from B_T is $\boldsymbol{q}^T$.

STEP 1. The MR is in a configuration $\boldsymbol{q}^c$. $n := 0$, $\boldsymbol{q}^n := \boldsymbol{q}^c$. Execute target_point_is_forbidden*:* = Algorithm 1 ($\boldsymbol{q}^c$, $\boldsymbol{q}^T$, B_T, N_{BT}). If target_point_is_forbidden*:* = *NO* then the Algorithm successfully terminates its work with a message that the object is grasped in the point $\boldsymbol{q}^T$.

If target_point_is_forbidden = *YES* go to STEP 2.

STEP 2. If $N_{BT} = 0$ the Algorithm terminates its work with a message that an object cannot be grasped in any target configuration. If $N_{BT} \neq 0$ consider the first point from B_T as $\boldsymbol{q}^T$ and go to STEP 1. The End of the Algorithm.

Algorithm 1 gets values in the format: Algorithm 1($\boldsymbol{q}^n$, $\boldsymbol{q}^T$, B_T, N_{BT}) and defines whether point $\boldsymbol{q}^T$ is reachable from $\boldsymbol{q}^n$ in an unknown environment or not.

Algorithm 1

STEP 1. MR is in $\boldsymbol{q}^n$ (let us call it “a path changing point”). SS supplies information about $Y(\boldsymbol{q}^n)$, $Z(\boldsymbol{q}^n)$, $Q(\boldsymbol{q}^n)$.

STEP 2. Execute

$$N_{BT} := Procedure1(B_T, N_{BT}, Q(\boldsymbol{q}^n)).$$

If $N_{BT} = 0$ then target_point_is_forbidden*: = YES* and return to the Algorithm. If $N_{BT} \neq 0$ check whether the point $\boldsymbol{q}^T$ was thrown out of B_T. If yes then target_point_is_forbidden*: = YES* and return to the Algorithm, if no (that is $\boldsymbol{q}^T$ was left in B_T) go to STEP 3.

STEP 3. Call the PI procedure in order to generate a path $L(\boldsymbol{q}^n,\boldsymbol{q}^T)$ satisfying the following conditions:

– $L(\mathbf{q}^n,\boldsymbol{q}^T)$ connects $\boldsymbol{q}^n$ and $\boldsymbol{q}^T$;

– $L(\boldsymbol{q}^n,\boldsymbol{q}^T)$ does not intersect the set $\bigcup_{i=0}^{n} Q(\boldsymbol{q}^n)$ that is, it does not intersect any forbidden point;

– $L(\boldsymbol{q}^n,\boldsymbol{q}^T)$ satisfies the limitations (1).

There may be two results of the PI procedure execution:

1. PI returns the generated path and Algorithm 1 goes to STEP 4;

2. PI informs that the $L(\boldsymbol{q}^n,\boldsymbol{q}^T)$ cannot be generated, that is, the $\boldsymbol{q}^T$ is unreachable. In this case:

$$N_{BT} := Procedure2(B_T, N_{BT}, \boldsymbol{q}^T),$$

make assignment target_configuration_is_forbidden: = =*YES* and return to the Algorithm.

STEP 4. MR begins to follow the $L(\boldsymbol{q}^n,\boldsymbol{q}^T)$. There may be two results:

1. MR comes to a point $\boldsymbol{q}_i^T \in B_T$. In this case make assignments $\boldsymbol{q}^T := \boldsymbol{q}_i^T$ and target_point_is_forbidden*: = NO* and return to the Algorithm;

2. MR will come to such point $\boldsymbol{q}^*$ that the next point after that is forbidden. In this case execute $n := n + 1$, $\boldsymbol{q}^n := \boldsymbol{q}^*$ and Algorithm1 goes to STEP 1. The End of Algorithm 1.

Theorem. Executing the Algorithm the MR will solve the Problem in a finite number of steps.

Proof. The Algorithm defines the reachability of a finite number of points from B_T. In order to define whether a point $\boldsymbol{q}_i^T \in B_T$ is reachable it is necessary to call Algorithm 1 one time. Therefore the Algorithm execution is reduced to a finite number of calls of Algorithm 1. Therefore in order to demonstrate that the Algorithm will be executed in a finite number of steps it is necessary to show that Algorithm 1 will be executed in a finite number of steps for arbitrary $\boldsymbol{q}^n$ and $\boldsymbol{q}^T$.

Algorithm 1 defines whether a point $\boldsymbol{q}^T$ is reachable in an unknown environment from a path changing point $\boldsymbol{q}^n$ or not. In Algorithm 1 when the MR is in point $\boldsymbol{q}^n$, $n = 0, 1, 2, \ldots$ the SS and the PI procedure are called. If after executing of these actions the point $\boldsymbol{q}^T$ will be defined as forbidden (because of intersection with obstacles or unreachability) the return to the Algorithm will take place and another point $\boldsymbol{q}_i^T \in B_T$ will be considered. If after execution of these actions the point $\boldsymbol{q}^T$ will not be defined as forbidden, a path $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$ will be generated and the MR will begin to follow this path. There may be two results of following this path: either MR will not meet forbidden points and therefore will reach the $\boldsymbol{q}^T$ (and therefore successful termination of the Algorithm work will occur) or the MR will come to such a point $\boldsymbol{q}^n$, $n = 1, 2, \ldots$ that the next point will be forbidden. Let us show that all path changing points $\boldsymbol{q}^n$, $n = 0, 1, 2, \ldots$will be different and their number will be finite.

Let us prove that all points where the MR changes its trajectory will be different. Suppose that the manipulator changed its trajectory being in point $\boldsymbol{q}^s$, and later it again changed its trajectory, being in point $\boldsymbol{q}^p$, that is $s<p$. Let us show that $\boldsymbol{q}^s \neq \boldsymbol{q}^p$. Suppose, at first, that, on the contrary $\boldsymbol{q}^s = \boldsymbol{q}^p$. Then $Q(\boldsymbol{q}^s) = Q(\boldsymbol{q}^p)$. As the manipulator changed its trajectory being in point $\boldsymbol{q}^s$ it generated the trajectory which did not intersect the sets $Q(\boldsymbol{q}_i)$, $I = 0, 1, \ldots, s$. But as it changed the trajectory in point $\boldsymbol{q}^p$ it means that its trajectory intersected $Q(\boldsymbol{q}^p) = Q(\boldsymbol{q}^s)$ (besides, $\boldsymbol{q}^s = \boldsymbol{q}^p$ is the center of r-neighborhood of the point $\boldsymbol{q}^s = \boldsymbol{q}^p$ and the following point is forbidden). That is, $Q(\boldsymbol{q}^p) = Q(\boldsymbol{q}^s)$ was unknown. Here is a contradiction. It means that all points where the manipulator changes its trajectory are different.

Now let us show that the number of such points is finite. Suppose that it is infinite. All points of a trajectory changing must satisfy the inequalities (1). It means, that the sequence of these points is limited. According to the Boltsano-Weierstrass theorem it is possible to extract a convergent subsequence $\boldsymbol{q}_i$, $I = 1, 2, \ldots$ from this sequence According to Cauchy property of the convergent sequences it is possible for any ε to find such a number s that all points $\boldsymbol{q}_i$, $i > s$ will lie in an ε-neighborhood of $\boldsymbol{q}_s$. Let us take $\varepsilon < r$. Consider an arbitrary point $\boldsymbol{q}_i$ of the trajectory changing lying in the ε-neighborhood of $\boldsymbol{q}_s$. As the manipulator had to change the trajectory in the $\boldsymbol{q}_i$, it means that that trajectory intersected $Q(\boldsymbol{q}_s)$ (because $\boldsymbol{q}_i$ and its neighbor points belong to $Q(\boldsymbol{q}_s)$). From this fact it is possible to draw the conclusion that the set $Q(\boldsymbol{q}_s)$ was not taken into account when that trajectory was generated. But such situation is impossible if we strictly follow the conditions of the Algorithm 1. The situation when a trajectory changing point belongs to the ε-neighborhood of another trajectory changing point will necessarily appear if the number of the points where trajectory is changing is infinite. But we showed that such situation is impossible and it means that a number of the points where it is necessary to change trajectory will be finite.

So, the number of path changing points $\boldsymbol{q}^n$, $n = 0, 1, 2, \ldots$ is finite and they are different. In every point $\boldsymbol{q}^n$ the SS and the procedure PI are called and PI generates a $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$. As a result we either get information that $\boldsymbol{q}^T$ is forbidden or do not get it. If we get the information that $\boldsymbol{q}^T$ is forbidden then we consider the $\boldsymbol{q}^T$ as unreachable. Otherwise an attempt of the path $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$ following occurs. If in the last path changing point $\boldsymbol{q}^n$ the point $\boldsymbol{q}^T$ was not qualified as forbidden, a path $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$ will be generated, this path will be followed and the $\boldsymbol{q}^T$ will be reached.

So it was shown that executing Algorithm 1 the MR will either reach the $\boldsymbol{q}^T$ or will make the conclusion that the $\boldsymbol{q}^T$ is unreachable. The Algorithm is reduced to a finite number of the Algorithm 1 calls. Therefore one may see that executing the Algorithm the MR will solve the Problem in a finite number of steps. The Theorem is proved.

*Note 1.*We have already mentioned that Algorithm 1 is reduced to a finite number of paths $L(\boldsymbol{q}^n, \boldsymbol{q}^T)$ generation and following. The Algorithm is reduced to a finite number of the Algorithm 1 calls. Therefore one may see that the Problem is reduced to the solution of a finite number of PI problems of a path generating and following in an environment with known forbidden states.

Note 2. On the first call of the Algorithm 1 $\boldsymbol{q}^c = \boldsymbol{q}^0$, on the next calls, generally speaking, $\boldsymbol{q}^c \neq \boldsymbol{q}^0$. Upon the Algorithm 1 execution a conclusion about a $\boldsymbol{q}^T$ reachability/unreachability from a $\boldsymbol{q}^c$ is drawn. But, as the MR has come to $\boldsymbol{q}^c$ from $\boldsymbol{q}^0$ by continuously following each other allowed points, the conclusion about reachability/unreachability of $\boldsymbol{q}^T$ from $\boldsymbol{q}^c$ will be simultaneously considered as the conclusion about reachability/unreachability of $\boldsymbol{q}^T$ from $\boldsymbol{q}^0$.

An algorithm for a n-link manipulating robot control in an environment with unknown static obstacles was considered. A theorem was proved which states that following the algorithm the MR will either grasp an object or will give a proved conclusion that an object cannot be grasped in any configuration in a finite number of steps .

References

1. Ahrikhencheikh C., Seireg A. Optimized-Motion Planning: Theory And Implementation. John Wiley & Sons. Inc, 1994.
2. LaValle S. M. Planning Algorithms. 1999–2006. URL: http://msl.cs.uiuc.edu/planning.
3. Nilson N. Problem-Solving Methods in Artificial Intelligence. N. Y. : McGraw-Hill Book Company, 1971.
4. Ilyin V. A. Intellectual robots: theory and algorythms. Krasnoyarsk : Publishing house SAA, 1995.
5. Lumelsky V. J. Sensing, intelligence, motion. How robots and humans move in an unstructured world. John Wiley & Sons, 2006.
6. Chen C., Li H.-X., Dong D. Hybrid Control for Robot Navigation. A hierarchical Q-learning algorithm // IEEE Robotics & Automation Magazine. 2008. Vol. 15, № 2. P. 37–47.
7. Online Algorithms with Discrete Visibility. Exploring Unknown Polygonal Environments / S. K. Ghosh, J. W. Burdick, A. Bhattacharya, S. Sarkar // IEEE Robotics & Automatiom Magazine. 2008. Vol. 15. № 2. P. 67–76.
8. Masehian E., Amin-Nasari M. R. Sensor-Based Robot Motion Planning. A Tabu Search Approach // IEEE Robotics & Automation Magazine. 2008. Vol. 15. № 2. P. 48–57.
9. Rawlinson D., Jarvis R. Ways to Tell Robots Where to Go. Directing autonomous robots using topological instructions // IEEE Robotics & Automation Magazine. 2008. Vol. 15. № 2. P. 27–36.
10. Yegenoglu F., Erkmen A. M., Stephanou H. E. On-line Path Planning Under Uncertainty // Proc. 27th IEEE Conf. Decis. and Contr. (Austin, Tex. Dec.7–9, N. Y.). 1988. Vol. 2. P. 1075–1079.
11. Algorythmic provision and software of intellectual manipulation robots in conditions of uncertainty. Theory and modelling of control tasks, trajectories planning and actions of manipulation intellectual robots in conditions of uncertainty : report on state budget scientific-research work B6-12 for 1996 / Thesis supervisor V. A. Ilyin. Krasnoyarsk, 1997. № state registration 01.9.80.001231. Inv. № 02.9.8.0000526.
12. Lopatin P. K. Computer simulation of manipulation robots control in unknown environmen on the basis of exact and simplified algorythms // Mechatronics, automatization, control . Appendix. 2006. № 8. P. 7–14.
13. Lopatin P. K. Algorythm of dynamic systems control in unknown static environment // Vestnik. Scientific Journal of Siberian aerospace university named after academician M. F. Reshetnev. Issue. 4(11). 2006. P. 28–32.
14. Lopatin P. K. Algorythm of dynamic systems control in unknown static environment // Mechatronics, automatization, control. № 2. 2007. P. 9–13.
15. Lopatin P. K. Algorithm of a manipulator movement amidst unknown obstacles // Proc. of the 10th Intern. Conf. on Advanced Robotics (August 22–25, Budapest). Budapest : Hotel Mercure Buda, 2001. P. 327–331.

E. V. Malinkin
Siberian State University of Telecommunications and Computer Science, Russia, Novosibirsk

INVARIANT SIGNALING WITH PROCESSING IN THE FREQUENCY AREA

A method of sending signals to the subsequent processing in the frequency domain is offered. To implement such a treatment, a structure is proposed based on the direct and inverse Fourier transformation with an element of division, making the division of information signals and training signals possible.

Keywords: immunity, invariant, invariant relative amplitude modulation, the probability of pairing transition, the signal/noise.

The author [1] suggested a relative amplitude modulation (RAM). Its essence lies in the fact that the previous assumption is core to the next, and their attitudes at the reception, despite the impact of multiplicative noise with the channel transmission coefficient K, equal to the ratio of these parcels for transfer. The RAM can significantly reduce the influence of multiplicative noise. However, the influence of additive noise has not been eliminated.

The author of this work used the idea of the RAM, but bearing messaging of equal size allocated in the same sequence of training signals as their subsequent averaging to reduce the impact of additive noise. Information signals of various amplitudes are identified in the information sequence.

The relation of the information sequence signals to value of the average training signals also reduces the influence of a multiplicate hindrance.

This method of signal formation on transfer and their demodulation on the reception party allows improving the noise stability of the offered structure at least by two orders in comparison with the classical RAM systems.

The given work is devoted to the synthesis of the processing of the information algorithm by the invariant system of information transfer in frequency areas; and to the analysis of the received technical characteristics.

We have an analogue communication channel with a pass-band from $f_{\min}$ to $f_{\max}$ on which there is transferred peak-modulated rectangular bending around the signals. The sequence of information part signals with those of the training part is united into blocks.

Transferred signals are treated to actions of hindrances: a multiplicate, described by the change of a channel broadcast of communication factor $k(t)$ at certain frequencies; and additive, representing white noise with no correlated readout and a dispersion σ^2 (while modelling the dispersion accepts a value equal to 1). Besides this, the accepted signal is influenced by peak-frequency and fazo-frequency distortions of the communication channel.

Let's consider that the communication channel is subject to a smooth general fading and consequently, it is possible to allocate the interval's stationary channel properties. Particularly, to allocate intervals at which the factor of a channel broadcast of communication is constant, and the amplitude of the readout AFC is constant (does not vary).

Let's suppose that the duration of the transferred block does not exceed the duration of an interval stationarity.

To resist the additive hindrance accumulation the averaging of training signals is entered. For this purpose, arithmetically there are readouts of training signals with the same name and then this sum is divided by the sum of the composed.

It is necessary to synthesise processing algorithms of the signal steadily against the influence of the hindrance complex (multiplicate and additive) and also to synthesize the system's structure of the information transfer, allowing demodulate transfer signals according to the offered algorithm with the subsequent analysis of quantitative characteristics of the information transfer system.

Synthesis of the signal processing algorithm. To solve the task it is required to first, generate a signal transfer (opposite side) and, secondly – to synthesize the algorithm of processing on reception.

We shall generate the transfer signal in blocks. Each block will consist of a training signal (the pilot of the signal) and an information signal. The training signal will be identical on all subsequent blocks. For the simplicity of training signal formation $S_{tr}(nT)$ we shall present it in the form of readout of an equal amplitude. According to the laws of digital signals to training signal processing $S_{tr}(kT)$ there corresponds a power spectrum $S_{tr}(jk)$. A power spectrum of an information signal shall be presented in:

$$S_{inf}(jk) = S_{tr}(jk) S_{mod}(jk), \quad (1)$$

where $S_{tr}(jk)$ – is the power spectrum of the training signal, $S_{mod}(jk)$ – is the power spectrum of the transfer signal.

In fig. 1. the structure of the transferring part of the invariant system is presented.

Thus, the power spectrum of a transfer signal at input of block OBPF shall look like:

$$\left[S_{tr}(jk), S_{inf}(jk)\right]_1;\ \left[S_{tr}(jk), S_{inf}(jk)\right]_2;\ \ldots;$$
$$\left[S_{tr}(jk), S_{inf}(jk)\right]_i;\ \ldots$$

In block the OBPF transformation of the power spectra readout to the readout of the transfer signal $S_{inf}(nT)$ is accomplished. By means of the modulator the modulation of the information signal by means of AM modulations is done.

The first, during T_{tr}, transfer the radio impulses of identical amplitude representing are signals of training sequence. Values of readout amplitudes of the training signals on a communication channel exit are remembered subsequently in a memory element and serve for the demodulation together with values of readout of information signals amplitudes. Besides, the values of amplitudes of signals of training sequence readout on a communication channel exit are used for the definition of channel broadcast of communication factors. During T_{inf}, the sequence of radio impulses for various amplitudes, representing the sequence of information signals is transferred. Their quantity is equal to the quantity of training signals. The transfer time for information sequence is equal to the time of training sequence transfer and consequently $T_{inf} = T_{tr}$. Then the transfer of training sequence follows its information and so on, until the end of the transfer of the whole block.

After the reception of training signals, the value of the readout of training signals amplitudes in a memory element arithmetically develops and is averaged by the quantity composed for the purpose of resisting to additive hindrance, which raises noise stability in the whole system of information transfer.

On fig. 2 the readout of information and the training sequence of signals (fig. 2, *a*) with their power spectra (fig. 2, *b*) are presented. It is necessary to notice, that the amplitudes of readout for a training signal on a communication channel input are equal among themselves in time space; the values of amplitudes of power spectra readout for training signals are equal to the communication channel input. Amplitudes of signals of information sequence readout in a communication channel input differ in size and in time, and have various values of power spectra readout of signal amplitudes.

The set N of readout for an information part in time space corresponds to N readout of its power spectrum.

The similarly set N of readout for a training signal in time area corresponds to N readout of its power spectrum.

For a signal of information sequence of i block, its power spectrum will be equal to:

$$S_{i\,out}(jk) = S_{i\,in}(jk) \cdot k_i(jk) \cdot H_i(jk) + N_{1i}(jk)\,, \qquad (2)$$

where k – is a harmonic number; $S_{i\,out}(jk)$ – is a power spectrum of a signal on a communication channel exit on i block; $S_{i\,in}(jk)$ – is a power spectrum of a signal on a communication channel input on i block; $k_i(jk)$ – is a power spectrum of a multiplicate hindrance on i block; $H_i(jk)$ – is the transfer characteristic of the channel on i block of processing; $N_{1i}(jk)$ – is the spectral density of capacity of additive noise of a signal of information sequence on i block of processing.

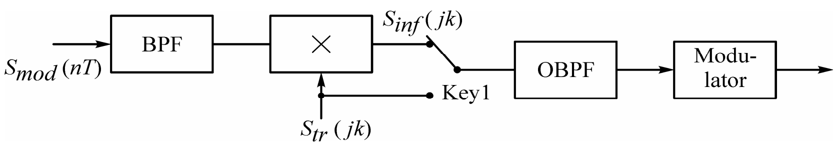


Fig. 1. Sending device structure

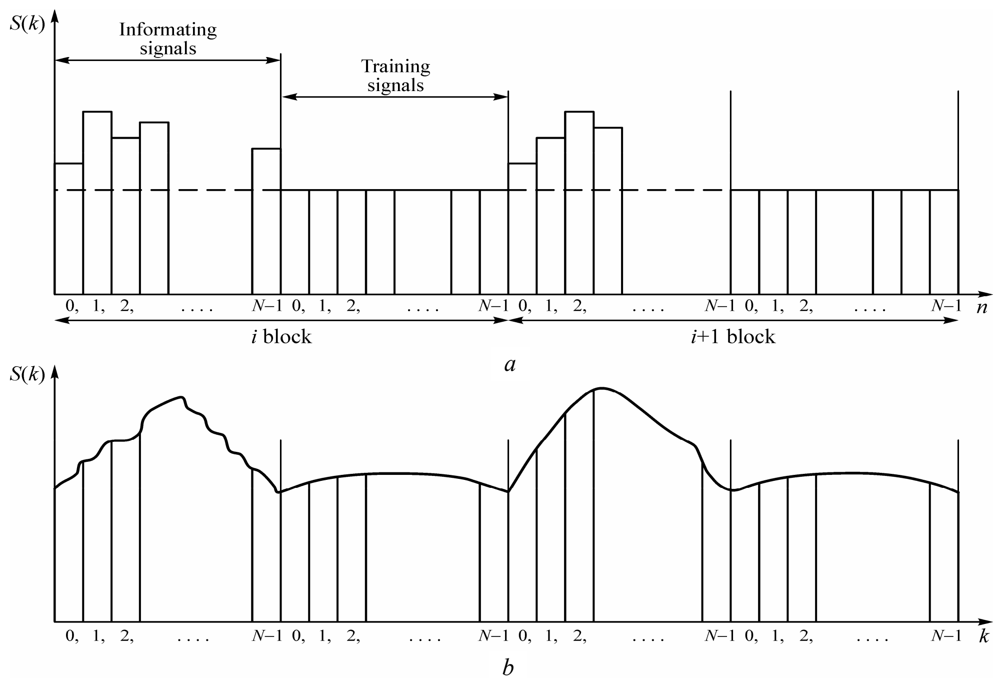


Fig. 2. Information and training sequences and their power spectra

For a signal of training sequence of *i* block its power spectrum will be equal to:

$$S_{i\,tr.out}(jk) = S_{i\,tr.in}(jk)\cdot k_i(jk)\cdot H_i(jk) + N_{2i}(jk)\,, \qquad (3)$$

where $S_{i\,tr.out}(jk)$ – is a power spectrum of a training signal on a communication channel exit; $S_{i\,tr.in}(jk)$ – is a power spectrum of a training signal on a communication channel input; $N_{2i}(jk)$ – is the spectral density of capacity of additive noise of a signal of training sequence on *i* block of the processing.

It is necessary to notice, that after the averaging of training signals, the value of the amplitudes of readout $N_{2i}(jk)$ it is much less, than the values of amplitudes of readout $N_{1i}(jk)$ at the expense of the accumulation effect with averaging. Values of the readout amplitudes $N_{1i}(jk)$ are more than values of the amplitudes of readout $N_2(jk)$ by 1 000 times, i.e. $N_{1i}(jk) = 1\,000 N_{2i}(jk)$ as the quantity of averaging equal to 1 000 [2].

As it was said earlier, the modulating parameter on transfer modulates the relation of power spectra of information and training parts. For the demodulation of reception signals it is necessary to divide a power spectrum of an information part into a power spectrum of a training part. In result we will receive:

$$S_{i\,mod.out}(jk) = S_{i\,out}(jk)\big/S_{i\,tr.out}(jk) = \\ = \frac{S_{i\,in}(jk)\cdot k_i(jk)\cdot H_i(jk) + N_{1i}(jk)}{S_{tr.in}(jk)\cdot k_i(jk)\cdot H_i(jk) + N_{2i}(jk)}\,. \qquad (4)$$

Due to the fact that the readout values of amplitudes of a power spectrum of additive noise in training sequence is essentially less than the values of amplitudes of the power spectrum readout of additive noise of information sequence, the size $N_{2i}(jk)$ in expression (4) can be neglected [3].

Then expression (4) after elementary transformations is reduced to:

$$S_{i\,mod.out}(jk) = S_{i\,mod.in}(jk) + \frac{N_{1i}(jk)}{S_{i\,tr.in}(jk)\cdot k_i(jk)\cdot H_i(jk)}\,, \qquad (5)$$

where $S_{i\,mod.out}(jk)$ – is the power spectrum of modulating sequence on a communication channel exit on *i* block; $S_{i\,mod.in}(jk)$ – is the power spectrum of modulating sequence on a communication channel input on *i* block.

According to Parseval's theorem the energy of the signal calculated in time space is equal to the energy of a signal calculated in frequency space. Therefore the relation of the information signal energy to the energy of a training signal remains both in time, and in frequency space. So, the results of processing in time and frequency space yield identical result and, consequently noise stability estimation will be made in time space.

For this purpose, look at fig. 3, which shows the principle of signal processing.

At point "a" information and training signals are given in the form of a readout in time area. At exit BPF ("b") this signal is presented by the readout of a power spectrum. At point "c" the signal is presented by readout of the power spectrum of an information part, and in point "d" – is the readout of the power spectrum of a training part. In point "*e*" – is the readout of the power spectrum of modulating sequence together with the readout of a power spectrum of the additive noise, the size of which is estimated by formula (5).

At point "f" is the signal before modulating the sequence of readings presented in the time domain.

To calculate the probability of erroneous reception of the time domain, we introduce a notion of the invariant. Given the use of the relative value of the invariant amplitude modulation; this can be found as [2]:

$$\text{INV}_l = \frac{\sum_{i=1}^{n} \text{INV}_l\cdot S_{tr}}{\sum_{j=1}^{N} S_{tr}} = \frac{N\cdot \text{INV}_l\cdot S_{tr}}{N\cdot S_{tr}}\,.$$

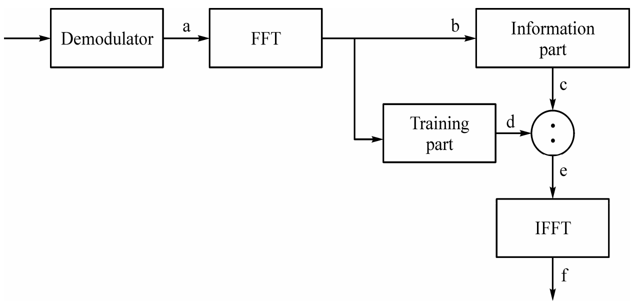


Fig. 3. Block diagram of the ISPR:
FFT – fast Fourier transform, *IFFT* – inverse Fourier transform

We assess the probability of the enumeration invariants by selecting the first pair as the smallest value of the invariant – $\text{INV}_1 = 1$ and the second invariant is taken from 2–6. When passing an invariant by the other there is an error in the reception. Instead of the first invariant of the pair it takes a second, compared to certain probability, and vice versa, instead of the second invariant of the pair being compared invariants is taken first with the same probability.

We perform the calculation of the probability of the erroneous reception. For this we use the well-known approach [2]:

$$P = P_1\int_{-\infty}^{z_p} W_2(z)\,dz + P_2\int_{z_p}^{\infty} W_1(z)\,dz\,, \qquad (6)$$

where P – is the transition probability of the first invariant of the second and vice versa; P_1 – is the probability of occurrence of the first invariant, P_2 – is the probability of a second constant, the first integral – is the probability of occurrence of the second invariant, when sent to the first and the second integral – is a probability of occurrence of the first invariant when sent second invariant; z_p – is the threshold needed to calculate P, if known P_1 and P_2.

The value of z_p is determined by using the best Bayesian estimates by minimizing P to z_p. We believe in the equally probable appearance of the first and second invariants, so we choose $P_1 = P_2 = 0{,}5$.

As seen from expression (6), it is necessary to know the analytical expression of $W_1(z)$ and $W_2(z)$.

For coherent reception with a sinusoidal subcarrier calculation, the functions $W_1(z)$ and $W_2(z)$ are known and are given in [2].

In our case, a rectangular envelope of the signal is allocated using the synchronous detector, and hence, the interference has a normal distribution [4].

Therefore, in our case, you can use the same approach for finding the analytical expression of probability density estimates of the invariant [2].

Thus, the value assessment of the invariant in our system is calculated as:

$$\mathrm{INV}_l^* = \frac{\sum_{i=1}^{N}\left(k\cdot \mathrm{INV}_l + \xi(i)\right)}{\frac{1}{L}\sum_{m=1}^{L}\sum_{j=1}^{N}\left(kS_{tr} + \eta(m,j)\right)} \cdot S_{tr}, \quad (7)$$

where INV_l – l is the transmitted invariant; $\xi(i)$ – i is the value of the Gaussian noise; in the denominator: S_{tr} – is the meaning training signal; $\eta(m, j)$ – j is the value of the Gaussian noise in the m realization of the signal S_{tr}; k – is the coefficient of transmission of the communication channel; N – is the number of samples taken from the envelope INV_l or S_{tr}; L – is the number of training signals.

To calculate P we need to know the expectation and variance of the numerator and the denominator of expression (7).

For their calculation we use the following approach.

The expectation of the numerator (7) will be equal to:

$$m_1 = k{\cdot}N{\cdot}\mathrm{INV}_l. \quad (8)$$

The expectation of the numerator (7) will be equal to:

$$D_1 = N{\cdot}\sigma^2, \quad (9)$$

where σ^2 – is the variance of Gaussian noise.

The expectation of the denominator (7) after the transformation will be equal to:

$$m_2 = k{\cdot}N. \quad (10)$$

The dispersion of the denominator after transformation will be:

$$D_2 = \frac{N}{LS_{tr}^2}\cdot\sigma^2. \quad (11)$$

Then the expression of the probability density estimates of the invariant is equal to [3] taking into account expressions (8–11):

$$W(z) = \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma_1\sigma_2} e^{-\frac{(zx - kN\,\mathrm{INV}_l)^2}{2N\sigma_1^2}} e^{-\frac{LS_{tr}^2(x-kN)^2}{2N\sigma_2^2}} |x|\,dx, \quad (12)$$

where $\sigma_1 = \sqrt{D_1}$; $\sigma_2 = \sqrt{D_2}$; L – is the number of signal processing.

Calculation P performed numerical approximation formula (12).

The proposed system was compared with a classical coherent system.

The probability of pairing transition was calculated in both cases for the same values of h signal/noise ratio, which is calculated by formula:

$$h^2 = \frac{\sum_{i=1}^{N} k^2\,\mathrm{INV}_{li}^2\,\Delta t}{N\Delta t\sigma^2} = \frac{k^2\,\mathrm{INV}_l^2}{\sigma^2}.$$

Thresholds were calculated by minimizing Z_p P in formula (6). For $k = 1$ and $\mathrm{INV}_1 = 1$, $\mathrm{INV}_2 = 2, 3, 4, 5, 6$ calculations give the result $Z_p = 1{,}5; 2; 2{,}5; 3; 3{,}5$.

For $k = 0{,}7$ and INV*1* = 1, INV*2* = 2, 3, 4, 5, 6 calculations give the result $Z_p = 1{,}5; 2; 2{,}5; 3; 3{,}5$.

The simulation results are shown in fig. 4 and 5.

Invariant feature of any system based on the principle of relativity invariant amplitude modulation is that the channel transmitted amplitude-modulated signals, producing the INV– and S_{tr}.

The transfer of these signals is provided on the basis of classical processing algorithms, which are generally with low immunity [4].

And, only after processing these signals in accordance to the algorithm of private expression (7), we obtain the invariant essential to the number, rather than signal.

As seen in fig. 4 and 5, the probability of pairing transition of one invariant in the other at high signal-noise ratio is defined by (10^{-30}–10^{-80}). The same values of the SNR probability of erroneous reception for a single character in the classical systems lie in the range (10^{-6}–10^{-10}).

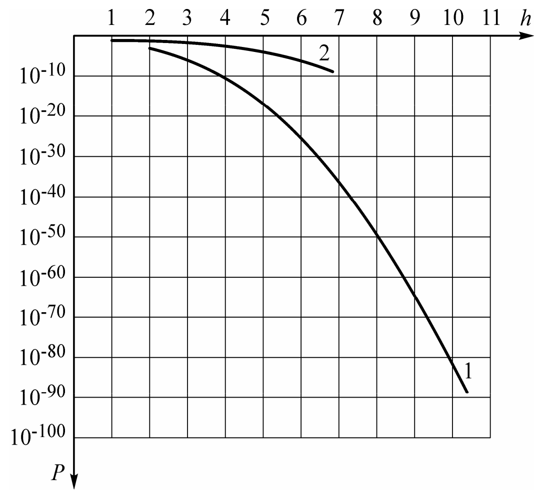


Fig. 4. The probability of pairing transitions for one invariant inside the other inthe following prescribed conditions: $k = 1$; $\mathrm{INV}_1 = 1$; $\mathrm{INV}_2 = 2, 3, \ldots, 10$ curve *1* – is the coherent reception, invariant relative amplitude modulation, rectangular envelope, curve *2* – is the classical relative amplitude modulation

According to the author, the immunity investigated invariant system should be compared to the noise immunity similar invariant systems, which will be done in subsequent papers.

We have proposed an invariant coherent information transfer system and defined its qualitative characteristics.

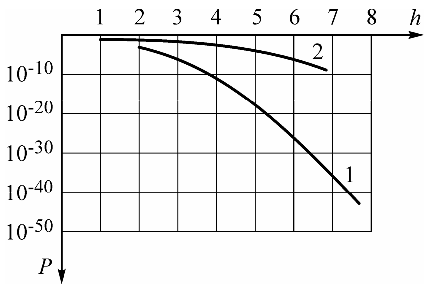


Fig. 5. The probability of pairing transition of one invariant inside the other inthe following prescribed conditions: $k = 0{,}7$; $INV_1 = 1$; $INV_2 = 2, 3, \ldots, 10$ curve *1* – is the coherent reception, invariant relative amplitude modulation, rectangular envelope, curve *2* – is the classical relative amplitude modulation.

The analysis of this system has shown that it has a high level of noise immunity. The chance of error in the classical algorithm with a relative amplitude modulation of at least two orders of magnitude is greater than the probability of pairing transition in the invariant system. Therefore, this system can be applied in telecommunication systems, remote control systems, and other systems, which demand high indicators for noise immunity.

References

1. Petrovich N. T. Digital data transmission in channels with phase-shift keying. M. : Sov. radio, 1965.

2. Invariant method for the analysis of telecommunication systems of information transmission : monograph / V. B. Malinkin, E. I. Algazin, D. N. Levin, V. N. Popantonopulo. Krasnoyarsk, 2006.

3. Levin B. R. Theoretical Foundations of Statistical Radio Engineering. 3rd ed. M. : Radio and Communications, 1989.

4. Teplov N. L. Noise immunity of digital data transmission systems. M. : Communications, 1964.

V. B. Malinkin, H. F. Kurash, O. V. Soboleva, E. V. Malinkin
Siberian State University of Telecommunications and Computer Science, Russia, Novosibirsk

INVARIANT METHOD OF INFORMATION TRANSMISSION IN FIBER-OPTIC TRANSMISSION SYSTEMS

We have synthesized a method of controlling the distortions introduced by fiber-optic communication lines. This method is based on the use of invariant equality. Its main technical characteristics have been determined.

Keywords: invariant fiber-optic transmission system.

Classical AM Modulation in fiber-optic transmission systems (FOTS) is used to transfer information signaling in most cases.

The probability of erroneous reception in regenerators is 10^{-10}, according to the recommendations of the International Telecommunication Union (ITU) [1]. More recent ITU [2] recommendations have proposed to apply device error protection (RCD) the performance of which is based on a special transmission signal coding by means of cyclic codes. It is difficult to create a residual operating in real time at a speed of 10 Gb / s and more.

Meanwhile, the reduction of error probability can be achieved in other ways. One is suggested below.

We have the FOTS (fig. 1). A laser is used as a transmitter. A photo detector is used as a receiver. The second window of transparency is used to send the information signal.

The information signal transmission algorithm must be synthesized based on the invariant method of information processing.

Fig. 1 shows the structure of FOTS, which includes the transmitting and receiving devices and the fiber-optic transmission.

It should be noted that the cross-cutting path of the FOTS at the second window of transparency [3] is linear, provided the power output of the transmitter to not exceed the permissible value of 1 mW. The *Z*-transform signal reception *Y* (*Z*) on the output of the FPU will be equal to (for i-volume processing unit):

$$Y_i(Z) = \left[G(Z) \cdot H_0(Z) \cdot H_1(Z) \cdot H_2(Z)\right]_i, \qquad (1)$$

where $G_i(Z)$ – is the *Z*-transform of the signal transmission at the *i*-volume processing unit; $H_{0i}(Z)$ – the transfer characteristic of the signal shaper at the *i*-volume unit; $H_{1i}(Z)$ – is the transfer characteristic of fiber-optic transmission at the *i*-volume unit; $H_{2i}(Z)$ – the transfer characteristic of the unit (FPU) on the *i*-unit.

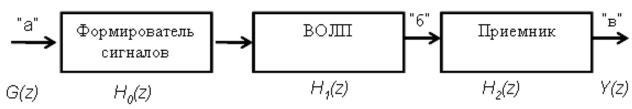


Fig. 1. The FOTS structure

In [4] an invariant equation is given, fair for any linear quadrupole:

$$\frac{G_i(Z)}{G_{i-1}(Z)} = \frac{Y_i(Z)}{Y_{i-1}(Z)}. \quad (2)$$

Equality (2) is valid for physically realizable systems, when the denominators are not equal to zero.

FOTS are conservative systems, the characteristics of which are divided into intervals of stationarity. Thus:

$$H_{\Sigma(i-1)}(Z) \approx H_{\Sigma i}(Z) \approx H_{\Sigma(i+1)}(Z), \quad (3)$$

where $H_{\Sigma i}(Z) = H_{0i}(Z) \cdot H_{2i}(Z) \cdot H_{2i}(Z) -$ is the transfer characteristic of through-tract FOTS on the i-unit.

Substituting expression (3) for (2), considering (1), equality (2) transforms into an identity.

In the transition from the Z-image to the amplitude-phase spectrum, we have:

$$\frac{G_i(jk\omega_1)}{G_{i-1}(jk\omega_1)} = \frac{Y_i(jk\omega_1)}{Y_{i-1}(jk\omega_1)}. \quad (4)$$

Equation (4), is in turn, divided into equal relations of the amplitude spectra and the equality of the difference in digital spectra:

$$\left.\begin{array}{l} \dfrac{G_i(k\omega_1)}{G_{i-1}(k\omega_1)} = \dfrac{Y_i(k\omega_1)}{Y_{i-1}(k\omega_1)} \\ \varphi_i(k\omega_1) - \varphi_{i-1}(k\omega_1) = \psi_i(k\omega_1) - \\ -\psi_{i-1}(k\omega_1) \end{array}\right\}, \quad (5)$$

where $G_i(k\omega_1)$ and $G_{i-1}(k\omega_1)$ – are the amplitude spectra at the input of the signals on the i and (i–1) blocks; $Y_i(k\omega_1)$ and $Y_{i-1}(k\omega_1)$ – are the amplitude spectra at the output of the FPU on the i and (i–1) blocks; $\varphi_i(k\omega_1)$ and $\varphi_{i-1}(k\omega_1)$ – are the phase spectra of signals at the input of the signals on the i and (i–1) blocks; $\psi_i(k\omega_1)$ and $\psi_{i-1}(k\omega_1)$ – are the phase spectra at the output of FPU signals on the i and (i–1) blocks. The first equation (5) reiterates the principle of relative amplitude modulation (OAM), and the second – the principle of relative phase modulation (RPM). Thus, to achieve a minimum probability of error in the FOTS it is necessary to "invest" the modulation parameter into the relation of the Z-image signal transmission at neighboring processing units; and at the receiving side of the modulating parameter it should be extracted by comparing neighboring blocks.

The formation of information signals in such a system is carried out at the input of the signals. The demodulation – at the output of the FPU. We shall call this system – the "invariant fiber-optic transmission system" (IFOTS).

The formation of the signals will be like so:

$$\frac{G_1(Z)}{G_0(Z)} = S_{\text{mod}1}(Z) \rightarrow G_1(Z) = G_0(Z) \cdot S_{\text{mod}1}(Z),$$

$$\frac{G_2(Z)}{G_1(Z)} = S_{\text{mod}2}(Z) \rightarrow G_2(Z) = G_1(Z) \cdot S_{\text{mod}2}(Z) =$$

$$= G_0(Z) \cdot S_{\text{mod}1}(Z) \cdot S_{\text{mod}2}(Z),$$

$$G_N(Z) = G_0(Z) \cdot \prod_{i=1}^{N} S_{\text{mod}i}(Z), \quad (6)$$

where G_0 (Z) – is the Z-image data signal in the initial block (signal learning).

However, it is impossible to realize the algorithm of modulation according to expression (6), because at long communication sessions $N \rightarrow \infty$ and the non-recursive filter can not be physically realized. Fig. 2 shows the realized structure of the signal shaper for N = 4. It contains four blocks of delay, a key, and multipliers of FFT and IFFT. The number of taps can be different.

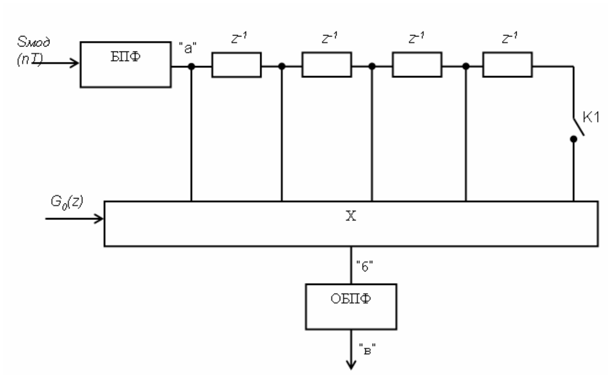


Fig. 2. The structure of the signal shaper for the N = 4 IFOTS

Modulating parameter S_{mod} (nT) in the FFT block is converted into S_{mod} (Z).

The process of forming the transfer of signals in each block consists of 2 stages. During the first stage the key K1 is locked. The signal at point "b" shall be so:

$$G_i(Z) = G_0(Z) \cdot \prod_{k=0}^{5} S_{\text{mod}(i-k)}(Z). \quad (7)$$

In the second stage the key K1 is open. The signal at point "b" will be:

$$G'_i(Z) = G_0(Z) \cdot \prod_{k=0}^{4} S_{\text{mod}(i-k)}(Z). \quad (8)$$

The signal transmission is shown is in fig. 3.

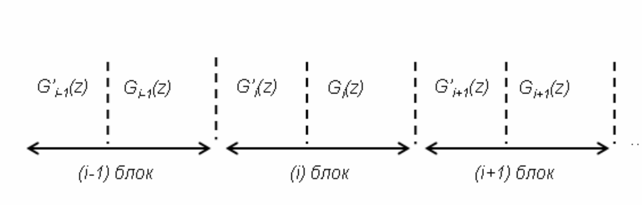


Fig. 3. Signal transmission

In accordance to the laws of digital filtering, each block at the receiving side is multiplied by the transfer characteristic of the through path. Imagine the Z-imaging signals in the form of:

$$\left.\begin{array}{l} Y_{i-1}(Z) = G_{i-1}(Z) \cdot H_{\sum i-1}(Z) \\ Y'_{i-1}(Z) = G'_{i-1}(Z) \cdot H_{\sum i-1}(Z) \\ Y_i(Z) = G_i(Z) \cdot H_{\sum i}(Z) \\ Y'_i(Z) = G'_i(Z) \cdot H_{\sum i}(Z) \end{array}\right\}. \quad (9)$$

The process of demodulation is in the division of the first part of $G_i(Z)$ by $G_i'(Z)$. Then:

$$S'_{\text{mod}i}(Z) = \frac{G_0(Z)\cdot\prod_{k=0}^{5} S_{\text{mod}(i-k)}(Z)\cdot H_{\sum i}(Z)}{G_0(Z)\cdot\prod_{k=0}^{4} S_{\text{mod}(i-k)}(Z)\cdot H_{\sum i}(Z)}. \quad (10)$$

The validity of expression (10) is based on the properties relative to the medium through-tract IFOTS spread and the validity of equation (3).

Fig. 4 shows the structure of the IFOTS receiving part.

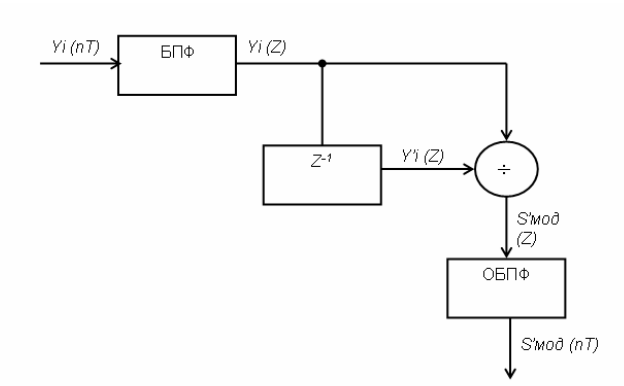


Fig. 4. Structure of the IFOTS receiving part

Note that in this algorithm there is a compensation of amplitude frequency distortions of the FCHI in the IFOTS pass-through. This in turn, leads to the compensation of the dispersion properties of fiber-optic transmission, an increase in the signal-to-noise ratio and reduction of error possibilities.

This section should indicate the advantages and disadvantages of the presented method. The undeniable benefits may include the compensation for amplitude frequency distortions and medium phase-frequency distortions of propagation. This makes it possible to increase the length of the regeneration area, while maintaining the probability of false acceptance; or substantially reduce the probability of error for fixed length area regeneration.

Disadvantages include speed increase of the basic signal transmission. Basically, redundancy was introduced into the signal transmission, making it possible to improve the quality indicators.

However, along with the compensation of amplitude frequency and phase-frequency distortions, an increase in additive noise is observed.

Let us estimate the intrinsic noise using the known relation [5]:

$$\sigma^2 = \frac{\Delta_0^2}{12}\sum_{n=0}^{\infty} h^2(nT) + \frac{\Delta^2}{12}\sum_{j=1}^{N}\sum_{n=0}^{\infty} h_j^2(nT), \quad (11)$$

where Δ_0 – is the quantization step in input words; $h(nT)$ – is the impulse response digital filter; $h_j(nT)$ – is the a truncated quantization impulse response of the digital filter of the j-order noise source; Δ – is the step of signal processing in the digital filter (DF); N – is the number of DF taps.

Usually the calculations are done as so: $\Delta_0 = \Delta$, and $h(nT) = h_j(nT)$. Then expression (11) is simplified.

The value of additional self-noise in gear will be equal to (for $N = 4$):

$$\sigma^2_{\text{собств.ПРД}} = \frac{\Delta^2}{12}\cdot 5\cdot\sum_{n=0}^{\infty} h^2(nT) = \frac{5\Delta^2}{3}. \quad (12)$$

The extra noise value at reception will be:

$$\sigma^2_{\text{собств.ПРМ}} = \frac{4\Delta^2}{12} = \frac{\Delta^2}{3}. \quad (13)$$

The total value of additional noise will be equal to:

$$\sigma^2_{\sum\text{собств}} = \sigma^2_{\sum\text{собств.ПРД}} + \sigma^2_{\sum\text{собств.ПРМ}} = 2\Delta^2. \quad (14)$$

If you receive a communication channel with noise (photon noise), the value of its IFFT block output will be equal to [4]:

$$\sigma^2_{\text{доп.КС}} = \sigma^2_{\text{КС}}\sum_{n=1}^{\infty} h^2(nT) = 2\sigma^2_{\text{КС}}, \quad (15)$$

where $\sigma^2_{\text{КС}}$ – communication channel noise power.

The structure of the IFOTS for the compensation of amplitude frequency distortions and medium phase-frequency distortions of propagation has been developed. The specifications technical had been defined. The developed method can be widely applied in fiber-optic transmission systems.

References

1. ITU-T G.707. Network node interface for the Synchronous Digital Hierarchy. 2004

2. ITU-T G.975. Forward error correction for submarine systems. 1996.

3. Zaslavsky K. E. Optical fibers for communication systems : textbook / Siberian State University of Telecommunications and Informatics. Novosibirsk, 2008.

4. Malinkin V. B. Improved noise immunity Modified Kalman filter in relative compensation methods / Omsk State Technical University. Omsk, 2003.

5. Goldenberg L. M., Matyushkin B. V., Polak M. N. Digital Signal Processing. M. : Radio and Communications, 1990.

A. M. Mikheenko, S. S. Abramov, I. I. Pavlov
Siberian State University of Telecommunication and Information Sciences, Russia, Novosibirsk

ABOUT THE ANALYSIS OF THE PULSE-WIDTH SYSTEM WITH FEEDBACK

In the article research results of the pulse-width system (PWS) captured by a negative feedback circuit have been depicted. Based on an asymptotic method of order decrease in the linear part of system, we have offered a technique for decreasing the PWS to an equivalent nonlinear pulse-amplitude system for which known methods of research are applied.

Keywords: stability, width-pulse systems, complex linear chain, nonlinear distortion.

Pulse-width systems (PWS) find application in automatic control devices, as well as in amplifiers of capacity of “D” class with an intermediate pulse-width modulation, captured by a negative feedback for the reduction of nonlinear distortions. Feedback presence in PWSs is inevitably connected with the problem of stability maintenance. The solution of this problem, in a comprehensible with practical application way for the PWSs generally does not exist. The possibilities of private decisions have been considered in [1; 2].

The aim of the present work consists in the search for new methods of the PWS stability analysis, applicable in practical applications.

Methods of pulse-width systems analysis. An analysis of the pulse systems is made, as a rule, basing on the theory of trellised functions and discrete transformation of Laplasa [1]. The theory of pulse-amplitude systems has now developed in enough detail: for a linear path and for systems containing nonlinear inertialess elements [1; 2].

Unlike the APS, the PWS is much more difficult to analyze, therefore it is investigated using data equivalent to the nonlinear APS.

A nonlinear APS block diagram is presented in fig. 1.

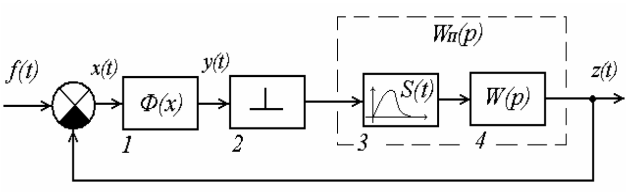


Fig. 1. The block diagram of a nonlinear APS: $f(t)$, $z(t)$ – are the input/output signals; $w(t)$ – is the pulse characteristic of the PWS continuous (linear) part; γ – is the duration of an impulse; $x(t)$ – is the error signal; $y(t)$ – is the nonlinear transformation $x(t)$; *1* – nonlinear inertialess tract converters; *2* – the generator clock δ – are the functions modulated on amplitude by signal $y(t)$; *3* – the generator of form impulses $s(t)$; *4* – a linear part of a path

According to [2], the formula of the control systems (fig. 1) looks as:

$$x(n,0) = f(n,0) - BT \times \\ \times \sum_{m=0}^{n=1} \Phi(x) \int_0^{\gamma} S(\bar{\tau})\omega(n-m-1,1-\bar{\tau})d\bar{\tau}, \quad (1)$$

where $\bar{\tau} = \frac{\tau}{T}$ – is the discrete time; B – is the coefficient of linear strengthening $x(t)$ in the APS tract.

The integral in (1) represents the convolution of the function for a forming element and the pulse characteristic $\omega(t)$. In the research [2] this has received the name of the resulted pulse characteristic $\omega(t)$.

If equation PWS manages to be reduced to (1) to it is possible to apply all known methods of the analysis and calculation of nonlinear APS.

A simplified block PWS diagram is presented in fig. 2.

For PWS the function of a forming element $s(t)$ looks like fig. 3, where T – is the period of clock frequency.

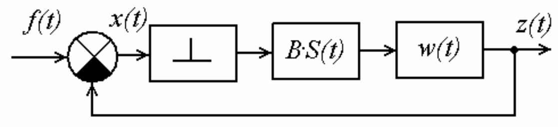


Fig. 1.2. The block diagram of the PWS

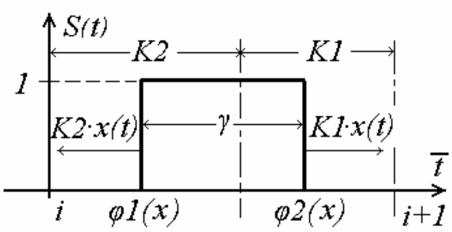


Fig. 1.3. Forming the PWS element

According to fig. 3, $s(\bar{\tau})$ corresponds to the bilateral PWS.

In $K1 = 0$, or $K2 = 0$ the unilateral takes the place of PWS.

By analogy with (1), for PWS it is possible to write down following expression:

$$x(n,0) = f(n,0) - BT \times \\ \times \sum_{m=0}^{n-1} \int_{\varphi_1(\chi)}^{\varphi_2(\chi)} S(\bar{\tau})\omega(n-m-1,1-\bar{\tau})d\bar{\tau}. \quad (2)$$

Unlike (1), the integral in (2) is functioning and $x(t)$, defines the features of the PWS analysis.

Accorging to fig. 3

$$\varphi_1(x) = K_2(1-x)\ ; \quad \varphi_2(x) = K_2 + K_1 x\ ;$$
$$K_1 + K_2 = 1\ . \tag{3}$$

Let's include the concept of factor of PWS symmetry,

$$A = \frac{K_1 - K_2}{K_1 + K_2}\ ,$$
$$K_1 = \frac{1+A}{2}\ ;\ K_2 = \frac{1-A}{2}\ . \tag{4}$$

On an integration interval in (2) $s(\tau)$ = 1, and ω (*n*–*m*–1; 1) it is possible to present a number of elementary pulse characteristics:

$$\omega(n-m-1, 1-\tau) =$$
$$= \sum_{v=0}^{s} \sum_{\mu=0}^{r_v - 1} C'_{v\mu} \frac{(n-m-\tau)^{\mu}}{\mu!} \exp\left[q_v (n-m-\tau) \right], \tag{5}$$

where s – is a number of different poles; r_v – is the frequency rate *v* of the poles:

$$C'_{v\mu} = \frac{1}{(r_v - \mu - 1)!} \cdot \frac{d^{(r_v-\mu-1)}}{dq^{(r_v-\mu-1)}} \left[\frac{P_H(q)}{TQ_H(q)} (q-q_v)^{r_v} \right]_{q=q_v},$$

$P_H(q)$; $Q_H(q)$ – are the polynoms in the numerator and a denominator of the transfer function of continuous part of the PWS $W(q)$.

Let's assume that all poles are simple and are not equal to zero (the presence of zero or multiple poles does not complicate and does not simplify a problem of the PWS and APS data.

For a case with simple poles:

$$C'_{v\mu} = C'_v = \frac{1}{T} \left| W(q)(q-q_v) \right|_{q=q_v}$$
$$\omega(n-m-1, 1-\tau) = \sum_{v=1}^{s} C'_v \exp\left[q_v (n-m-\tau) \right]. \tag{6}$$

Substituting (6) in (2), we will receive:

$$x(n,0) = f(n,0) - BT \sum_{m=0}^{n-1} \sum_{v=1}^{s} \left[\int_{\varphi_1(x)}^{\varphi_2(x)} \exp\left[-q_v \tau \right] d\tau \right] \times$$
$$\times C'_v \exp\left[q_v (n-m) \right] = f(n,0) - BT \times \tag{7}$$
$$\times \sum_{m=0}^{n-1} \left\{ \sum_{v=1}^{s} \Phi_v \left[x(m,0) \right] w_{\Pi v}(n-m) \right\} \ldots,$$

where:

$$\Phi_v \left[x(m,0) \right] = \int_{\varphi_1[x(m,0)]}^{\varphi_2[x(m,0)]} \exp\left[-q_v \tau \right] d\tau \ldots, \tag{8}$$
$$w_{\Pi v} = C'_v \exp\left[q_v (n-m) \right]. \tag{9}$$

Comparing (7) and (1) it is possible to draw a conclusion that the PWS it is reduced to the equivalent multidimensional nonlinear APS.

The number of parallel branches equivalent to the APS is equal to the number of roots in the typical equation.

The analysis of the multidimensional nonlinear APS allows us to define only the absolute stability of the system in a general view; the results comprehensible to practical management are to be received only for two-dimensional systems, or for systems reduced to two-dimensional during the stability analysis [3].

In this case the analysis of PWS stability resorts to the approached methods, allowing the decrease in the APS order equivalent. In [4], for example, it is offered to present $\Phi_v(x)$ as a sedate polynom of such a kind:

$$\Phi_v(x) = \sum_{k=1}^{N} r_k^{(v)} x^k\ . \tag{10}$$

In this case (7) becomes:

$$x(n,0) = f(n,0) - BT \sum_{m=0}^{n-1} \sum_{v=1}^{s} \sum_{k=1}^{N} C'_v \exp\left[q_v (n-m) \right] \times$$
$$\times x^k(m,0) = f(n,0) - BT \sum_{m=0}^{n-1} \sum_{k=1}^{N} x^k(m,0) \varpi'_{\Pi k}(n-m), \tag{11}$$

where:

$$\varpi'_{\Pi v}(n-m) = \sum_{v=1}^{N} r_k^{(v)} C'_v \exp\left[q_v (n-m) \right]. \tag{12}$$

When comparing (7) and (11), it is easy to notice that in (11) the number of branches is linear and is defined by the number of polynom members (10), from which it occurs.

If the equivalent nonlinearity $\Phi_v(x)$ is rather small in a polynom (10) it is possible to leave only two or three composed. Thus, the equivalent APS will be two – three measurements, even in case of constant usages of linear PWS parts.

If: $\quad \Phi_v(x) = r_1^{(v)} \cdot x + r_2^{(v)} \cdot x^2$.

As shown in [4], the PWS block diagram can be shown as a one-dimensional nonlinear APS.

Such a method of approached analysis of stability is especially effective, when the nonlinearity equivalent APS is expressed poorly.

The analysis of PWS stability; the method of artificial order drop in its linear part. The approached calculation of transients is based on an offered method of the analysis in difficult linear chains. This way is offered and in details investigated by J. S. Itshoki [5]. Its issues are following:

The order of the initial differential equation of a linear part of the system artificially goes down (the differential equation "is shortened"), and equivalent delay (in certain cases probably and not a late decision) is placed into the system description.

The parameters of the truncated equation steal up so (fig. 4), that a steepness of transitive function increase $h(t)$ in in-between decline space (DS) and the amplitude of the oscillatory process (if possible) corresponds to the initial $h(t)$.

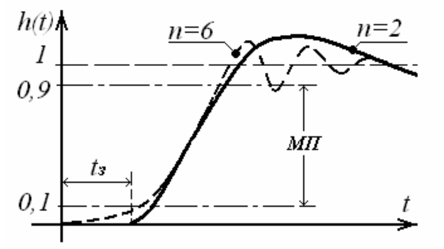


Fig. 4. Approximation of the transitive feature

Such a method can be rather effective even during the fall of difficult linear chains of a high order to the first and the second order.

For the solution of the approached description of the PWS linear part, its transfer function is necessary to lead the normalized equation to:

$$W(p)=\frac{1+g_1p+g_2p^2+\cdots+g_np^n}{1+a_1p+a_2p^2+\cdots+a_kp^k}.$$

Where at least $n < k - 1$ (at $n > k - 1$ the approximation by late function does not exist).

The required approximation $W(p)$ in order m is found in:

$$W_m(p)=\frac{\exp[-pt_{зm}]}{1+b_1'p+b_2'p^2+\cdots+b_m'p^m}, \quad (13)$$

where $t_{зm}$ – it is defined by the solution of the following equation:

$$\frac{t_{зm}^{m+1}}{(m+1)!}-\frac{t_{зm}^{m}}{m!}+\cdots+(-1)^{m+1}\cdot\Delta_{m+1}=0,$$

where:

$$\Delta_1=a_1-g_1,$$

$$\Delta_2=a_2-g_2-a_1\cdot g_1,$$

$$\Delta_3=a_3-g_3-a_2\cdot g_2-a_2\cdot g_1.$$

Let's notice that necessary value of delay tз corresponds to the least material (always positive) root $t_з = t_{зm}$. Finding this value it is uneasy even at $m = 3$.

Other parameters of approximation in (13) are found as follows:

$$b_1'=\Delta_1-t_{зm}$$

$$b_2'=\Delta_2-\Delta_1\cdot t_{зm}+\frac{t_{зm}^2}{2!}$$

…………………………………………

$$b_m'=\Delta m-\frac{\Delta_{m-1}}{1!}+\frac{\Delta_{m-2}\cdot t_{зm}^2}{2!}-\cdots\cdots+(-1)^m\frac{t_{зm}^m}{m!}$$

For the approximation of this or that order there are certain living conditions.

Order approach $m = 0$ exists always.

Approach $m = 1$ is limited by a condition $\frac{\Delta_2}{\Delta_1^2}<\frac{1}{2}$.

In case when $m = 2$, two variants are possible:

1. Approach $m = 1$ does not exist, i. e. $\Delta2 \geq 0,5\Delta12$. The condition of the existing approach $m = 2$ looks like: $\frac{\Delta_3}{\Delta_1^3}<\frac{\Delta_2}{\Delta_1^2}-\frac{1}{3}$.

2. Approach $m = 1$ exists, but the accuracy of approach is insufficient. Then the condition of approach existence $m = 2$ becomes:

$$\frac{\Delta_3}{\Delta_1^3}<-\frac{1}{3}+\frac{1}{3}\sqrt{\left(1-\frac{2\Delta_2}{\Delta_1^2}\right)^3}\approx$$

$$\approx\frac{1}{2}\left(\frac{\Delta_2}{\Delta_1^2}\right)^2\left[1+\frac{\Delta_2}{3\Delta_1}+\frac{1}{4}\left(\frac{\Delta_2}{\Delta_1^2}\right)^2+\cdots\right].$$

Because the increase of an approximation order sharply complicates a problem of search for the approached decision and the analysis of PWS stability, it is not necessary to apparently use, $m > 2$.

Let's assume that in result of the considered method of approximation PWS linear function in the form of the first order approach ($m = 1$) is defined:

$$W_1(p)=\frac{\exp[-pt_{з1}]}{1+b_1'p}=\frac{1}{b_1'}\cdot\frac{\exp[-pt_{з1}]}{p+\frac{1}{b_1'}}.$$

The corresponding trellised (discrete) pulse characteristic on the basis of (6) will become:

$$w(n-m-k,k-\overline{\tau}-\overline{t_1})=$$

$$=\frac{1}{b_1}\exp\left[-\frac{1}{b_1}(n-m-k)\right]\exp\left[-\frac{1}{b_1}(k-\overline{\tau}-\overline{t_1})\right],$$

where $\overline{t_1}=\frac{t_{з1}}{T}$; $b_1=\frac{b_1'}{T}$; k – comes out such, that it makes time discrete time.

The resulted pulse characteristic according to (9) can be written down as:

$$w_{n1}=\frac{1}{b_1}\exp\left[-\frac{1}{b_1}(n-m-1)\right]\cdot\exp\left[-\frac{1}{b_1}(1-\overline{t_1})\right].$$

Nonlinearity equivalent APS we will be according to (8):

$$\Phi_1\left[X(m,0)\right]=\int\limits_{K_2(1-X)}^{K_2+K_1X}\exp\left[\frac{1}{b_1}\tau\right]d\tau=$$

$$=b_1\left[\exp\left[\frac{1}{b_1}(K_2+K_1X)\right]-\exp\left[\frac{K_2}{b_1}(1-X)\right]\right]\cdots \quad (14)$$

where X – has normalized a signal changing within $0 < X < 1$. Taking into account, expression can be copied as so:

$$\Phi_1(X)=2b_1\exp\left[\frac{1}{2b_1}(1-A)\right]\cdot\exp\left[\frac{AX}{2b_1}\right]sh\frac{X}{2b_1} \quad (15)$$

Schedules of this dependence for each special case $b_1 = 0.5$ are presented in fig. 5. As one would expect, at asymmetrical kinds PWS ($A = \pm 1$) nonlinearity $\Phi_1(x)$ – expresses not much more strongly, than in $A = 0$.

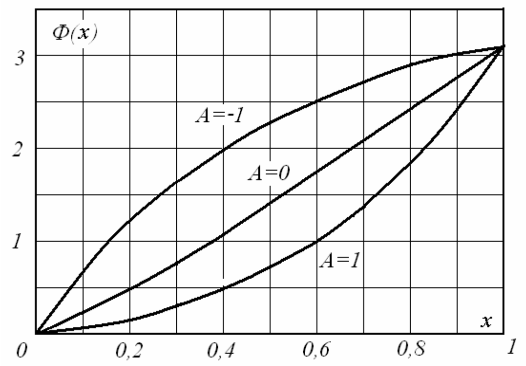


Fig. 5. Nonlinear PWS characteristics

The transfer function of the resulted linear part will be found in the form of discrete Laplasa transformation from (14):

$$W_n^*(q,0) = \frac{1}{b_1} \cdot \frac{\exp\left[-\frac{1}{b_1}(1-\bar{t}_1)\right]}{\exp[q] - \exp\left[-\frac{1}{b_1}\right]}.$$

Having replaced in this expression q by j (= ωT), we shall receive the peak-phase characteristic of the resulted linear PWS part. Using this detail, it is possible to estimate the PWS stability applying the criteria For example, according to criterion of the absolute position stability balance [2]:

$$\frac{1}{\sigma} + \mathrm{Re}\left[\frac{B}{b_1} \cdot \frac{\exp\left[-\frac{1}{b_1}(1-\bar{t}_1)\right]}{\exp[j\bar{w}] - \exp\left[-\frac{1}{b_1}\right]}\right] > 0, .$$

where $\sigma = \left|\frac{\partial \Phi_1(x)}{\partial x}\right|_{\max}$ – is the maximum differential value of factor of equivalent nonlinear element transfer.

In – is the factor of linear strengthening in the PWS pulse tract.

Introducing the connecting processes, offered by L. S. Iuhoni, the method drawn near calculation, has greatly allowed simplifying the problem of the analysis to a stable width-pulsed system in the event of a high order even by its linearities.

References

1. Tsipkin J. Z. The theory of linear pulse systems. M. : Fizmatgiz, 1963.
2. Tsipkin J. Z., Popkov J. S. The theory of nonlinear pulse systems. From-in Science. M., 1973.
3. Polov K. P. Amplifier Stability conditions in a mode “D” with a feedback // Radio engineering. 1971. № 6.
4. Polov K. P. To research of stability of the amplifier in a mode “D” with a feedback // Radio engineering. 1974. № 1.
5. Itshoki J. S. The approached method of the analysis of transients in difficult linear chains // The Soviet radio. 1969.

E. E. Miloserdov
Siberian Federal University, Russia, Krasnoyarsk

DYNAMIC MODELING OF A BUCKET-WHEEL EXCAVATOR PROPELLING MOTOR

The article reviews algorithmization of a dynamic model of a bucket-wheel excavator propelling motor during its motion. There are essential schemes and formulas used in dynamic modeling algorithmization.

Keywords: algorithmization, propelling motor, contour, moment, reaction.

Multi-support contours as analytical models are considered in order to determine bearing reaction, bearing load and moments which appear in the process of a bucket-wheel excavator motion at the face. This, in its turn, will help to evaluate a technical state of a bucket-wheel excavator [1].

When analyzing the impact of a traveling gear on a bucket-wheel excavator dynamics, we don't consider a set of an excavator motion equations, but focus on specific issues of traveling gears algorithmization and their connections with the whole machine in various operating modes.

The reactions to G force action are determined depending on the position of this force projection on the support contour taking into consideration a projection point shift relative to the support contour center. If an axis of a superstructure does not coincide with a force G projection point and a geometric center, both shifts are taken into account. In this case we should consider two centers of mass, where G_1 is a gravity of a fixed part of a machine, with running-out Q and G_2 is a gravity of a swivel part with a resultant running-out S_0 relative to the steering axis.

Thus, the resultant reaction in the support point of a contour:

$$R_{Tj} = R'_{Tj} + R''_{Tj},$$

where R'_{Tj} is a permanent part of G_1 force reaction; R''_{Tj} is a variable part of G_2 force reaction.

Calculation is carried out by a rigid lever method by solving an algebraic equation set with a Kutta–Joukowski theorem. Support reactions to the action of horizontal inertia forces are determined on the basis of data got in the process of solution of an equation set of an excavator dynamic model motion. In this case we use an algebraic equation system, that turns a moment action of horizontal resultant inertia force applied to a center of mass at a height of H into support reactions [2].

Then the complete support reaction is determined by a sum of three summands:

$$R_{Tj}= R'_{Tj1} + R'_{Tj2}+ R''_{Tj},$$

where R'_{Tj1}, R'_{Tj2} are reactions, respectively, to G_1 и G_2 forces; R''_{Tj} are reactions of a moment action from horizontal inertia force action.

The evaluation of a part of R''_{Tj} support reaction relative to a complete R_{Tj} value shows that if $R''_{Tj} / R_{Tj} \le 0.1$, then R_{Tj} value can be neglected. The general analytical models to determine support reaction for multi-support contours are shown in fig. 1, where γ_0 is a constant angular displacement of a fixed part center of mass relative to a geometric support contour center; r_0 is a displacement radius of a swivel part center of mass relative to a turning center and its α_0 angle of displacement from a central point. Calculated dependences to determine support reactions in fig. 1, *a* result from equations of forces moments relative to a reference triangle sides:

$$\left.\begin{aligned} &\Sigma M_{AB}(R) = 0 \to R'_{Cj}\\ &\Sigma M_{AC}(R) = 0 \to R'_{Bj}\\ &\Sigma M_{CB}(R) = 0 \to R'_{Aj}\end{aligned}\right|.$$

To determine support reactions of a support contour in fig. 1, *b* we use the following set of equation:

$$\left.\begin{aligned} &M_{AB}(R_D) + M_{AB}(R_C) = M_{AB}(G) = (R_D + R_C)\, l_{AD}\\ &M_{CB}(R_A) + M_{CB}(R_D) = M_{CB}(G) = (R_A + R_D)\, l_{AB}\\ &M_{DC}(R_A) + M_{DC}(R_B) = M_{DC}(G) = (R_A + R_B)\, l_{AD}\\ &M_{AD}(R_B) + M_{AD}(R_C) = M_{AD}(G) = (R_C + R_B)\, l_{AB}\end{aligned}\right|.$$

Therefore:

$$\left.\begin{aligned} &R'_D + R'_C = M_{AB}(G)/l_{CB}\\ &R'_A + R'_D = M_{CB}(G)/l_{AB}\\ &R'_A + R'_B = M_{DC}(G)/l_{CB}\\ &R'_C + R'_B = M_{AD}(G)/l_{AB}\end{aligned}\right|. \qquad (1)$$

The equation (1) determines R'_A, R'_B, R'_C, R'_D. The support reactions to the action of horizontal component inertia forces are found according to the scheme in fig. 2. for a three-support contour (fig. 2, *a*). For a three-point contour we take into account the moments from horizontal forces projections in a direction orthogonal to corresponding triangle sides, for instance:

$$M_{AC}(F) = F_{AC1}H_1 + F_{AC2}H_2$$
$$M_{BC}(F) = F_{BC1}H_1 + F_{BC2}H_2 ,$$

where F_{AC1} , F_{AC2} are vertical projections of inertia forces on AC side of the first and second masses at a height of H_1 and H_2 respectively.

Then the reactions:

$$\left.\begin{aligned} &R''_B = M_{AC}(F)/h_B(AC)\\ &R''_C = M_{AB}(F)/h_C(AB)\\ &R''_A = M_{BC}(F)/h_A(CB)\end{aligned}\right|,$$

where $h_B(AC), h_C(AB), h_A(CB)$ are the arms up to support points of corresponding sides.

For a four-support contour (fig. 2, *b*) we take into account the moments from horizontal forces projections, such as $M_{AB}(F) = F_{AB1}H_1 + F_{AB2}H_2$. Then, the reactions are determined by the following set of equations:

$$\left.\begin{aligned} &R_D + R_C = M_{AB}(F)/l_{CD}\\ &R_A + R_B = -M_{AB}(F)/l_{CD}\\ &R_C + R_B = M_{AB}(F)/l_{AB}\\ &R_D + R_A = -M_{AD}(F)/l_{AB}\end{aligned}\right|.$$

This equation system determines R''_A, R''_B, R''_C, R''_D.

The complete support reactions are found by formulas:

$$R_A = R'_A+ R''_A;\ R_B = R'_B + R''_B;$$
$$R_C = R'_C + R''_C;\ R_D = R'_D + R''_D.$$

The values of complete support reactions are used to evaluate the resistance to a machine motion.

An excavator turning resistance during its motion is determined by three factors: tractive force for F_i driving bogies, linear motion resistance for every W_i driving bogie and turning resistance moments relative to general dynamic turning center M_i^W; M_i^S; $F_i^W \cdot l_i$. The machine turning force balance is considered through the following equation system.

A sum of moments relative to a general dynamic turning center:

$$\sum F_i\rho_i - \sum M_i^W - \sum M_i^S - \sum F_i^W \cdot l_i - \sum W_i\ \rho_i = 0 .$$

an overall projections sum on X axis:

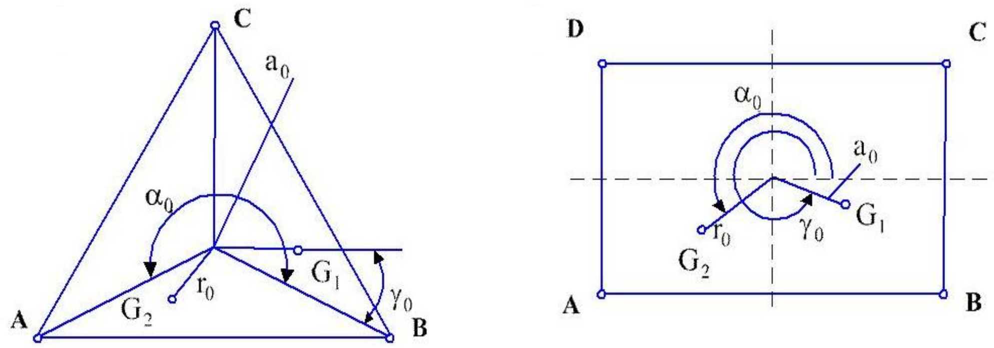
$$\sum F_i \cos\alpha_i - \sum F_i^W \sin\alpha_i - \sum W_i \cos\alpha_i = 0;$$

an overall projections sum on Y axis:

$$\Sigma F_i \sin\alpha_i - \Sigma F_i^W \cos\alpha_i - \Sigma W_i \sin\alpha_i = 0 ,$$

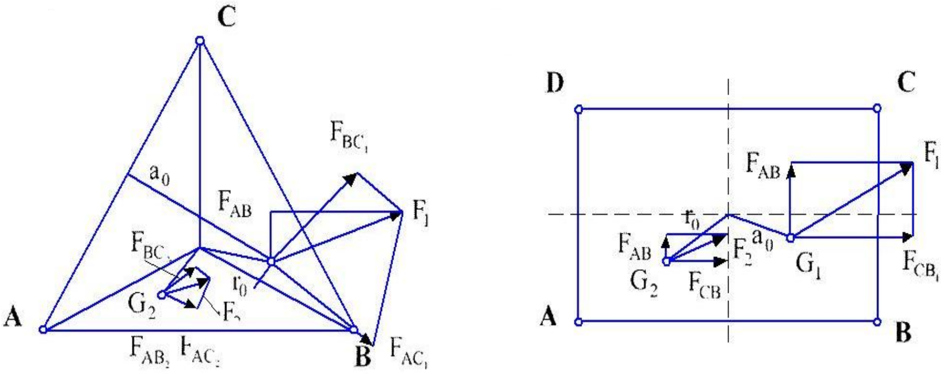
$$F_i^W = 2\mu \cdot R_{Ti} l_i / L' ,$$

where F_i^W is resistance force for F_i driving bogies lateral deviation; L' is the length of a track contacting area; l_i is a displacement of a force application point; F''^W_i is relative to a dynamic turning center (arm of force is F^W_i); ρ_i is the radius from a dynamic turning center to the main bogie direct axis; $M_i^{(S)}$ is a moment applied to a tandem of bogies under different motion resistances.



a b

Fig. 1. The general analytical models to determine support reaction for multi-support contours:
A, *B*, *C*, *D* are contour determining points; a_0 is an angle of displacement from a central point; r_0 is a displacement radius of a center of mass; G_1 и G_2 are gravities; γ_0 is a constant angular displacement of a fixed part center of mass relative to a geometric support contour center



a b

Fig. 2. The general analytical models to determine support reaction for four-support contours:
A, *B*, *C*, *D* are points determining the contour; a_0 is the angle of displacement from a central point; r_0 is a displacement radius of a center of mass; G_1 и G_2 are gravities; γ_0 is a constant angular displacement of a fixed part center of mass relative to a geometric support contour center; F_{AC1} , $F_{AC2} - F_{AC1}$, F_{AC2} are projections on a perpendicular to inertia forces on *AC* side of the first and second masses at a height of H_1 and H_2 respectively

For instance, if a tandem of bogies $i = 1, 2$:

$$M_i^S = \frac{1}{2}(W_2 - W_1)B/2,$$

where *B* is wheel track (in meters).

Taking into account the adopted dependences we propose the following solution algorithm:

1. One should determine ρ_{Hi} radii up to bogies relative to a dynamic turning center, when $l_i = 0$, which are the initial conditions:

$$\rho_{\text{н}i} = f(\alpha_i).$$

2. We assign tractive force:

$$F_i = A_i K_{1i} K_{2i} - B_i K_{2i} \cdot (\omega_m \rho_i).$$

A machine turn rate is:

$$\omega_m = \frac{\Sigma A_i K_{1i} K_{2i} - \Sigma M_i^{(W)} - \Sigma F_i^{(W)} l_i - \Sigma M_i^{(S)} - \Sigma W_i \rho_i}{\Sigma B_i K_{2i} \rho_i}.$$

3. Excess tractive force of bogies is:

$$S_i = F_i - W_i.$$

4. Running-out of $F_i^{(W)}$ force is determined as a function of α turn angels; bogies relative to tandem of bogies centers (coherent support points), R_{Ti} support reactions and some other main design parameters of a support contour:

$$l_i = \Sigma S_i Д_i,$$

where $Д_i = f(a_i, R_{Ti})$; $Д_i$ is a function, depending on a turn angel and R_{Ti} support reactions.

5. One finds the radii from the dynamic machine turning center to the bogies :

$$\rho_i = f(a_i, l_i).$$

6. A substitution of r_i, l_i new values and comparison with acceptable ε deviation is to be done before the inequity is stated:

$$\Delta\Sigma M(\rho_i, l_i) \le \varepsilon \rangle 0.$$

When the iterative calculations are completed, the derived parameters ρ_i, l_i, S_i, ω_i are taken and forces and velocity values are determined.

To sum up, a modeling algorithm of a propelling motor drag resistance should take into account the whole set of drag resistances as well as traction performance of electric motors of driving track bogies.

References

1. Belyakov Yu. I., Vladimirov V. M. Operating devices of bucket-wheel excavators. M. : Machine-building, 1967.

2. Dombrovsky N. G., Mayevsky A. G. Theory and calculation of a caterpillar mover of digging machines. Kiev : Tehnika, 1970.

A. B. Mishurenko, S. N. Kozlovsky
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

MATHEMATICAL MODEL FOR CONTACT RESISTANCE OF COLD CONTACT AT SPOT WELDING

The paper is devoted to the problem of optimizing spot welding modes and contains the description of a mathematical model for the calculation of cold contact initial resistance.

Keywords: spot welding, resistance of cold contact.

The practice of spot welding and the numerous results of published researches unambiguously confirm that one of main conditions of forming quality-welded joints is the optimality of the initial resistance of part -to- part $r_{\text{КТ}}$ contact. Its value and stability essentially influences the sizes of a kernel, the stability of the process against creating splashes and poor penetration. At the same time, until now, the development of spot welding technologies of $r_{\text{КТ}}$ value had been achieved experimentally for each particular welding condition; this is rather labor-intensive.

According to the conducted research it has been concluded that the rod model is the most acceptable engineering technique among all known methods applied in rough surface models for welded contacts; mainly because it describes the mechanism of contact interaction between two rough surfaces with more simplicity and precision (fig. 1).

According to the accepted model of two rough surface contacts, the conductivity in the contact layer is carried out according to n_r number of individually parallel micro conductors of d ($d \to 0$) diameter, and of a length, formed by deformable rods (micro lugs). One of the constituents of the total electric resistance in such a micro contact $r_{\text{КТ}}$ which is caused by the resistance of micro conductors in the contact layer, which has properties different to those of parent metal, is called the internal contact resistance $r_{\text{КВ}}$. The other part, which is formed by current line curvatures j in near contact areas where properties of the parent metal are assumed to not change, is called the micro geometrical contact resistance $r_{\text{МГ}}$. Then the total electric resistance contact will be equal to the sum of these two components:

$$r_{\text{КТ}} = r_{\text{КТ}} + r_{\text{КТ}}.$$

The total internal electric contact resistance $r_{\text{ВН}}$ can be defined in the following manner:

$$r_{\text{ВН}} = \frac{r_{\text{ВН}}}{n_r} = \rho_\Delta \frac{a}{n_r \Delta S} = \rho_\Delta \frac{2R_{\max}(1-\varepsilon)}{A_r}.$$

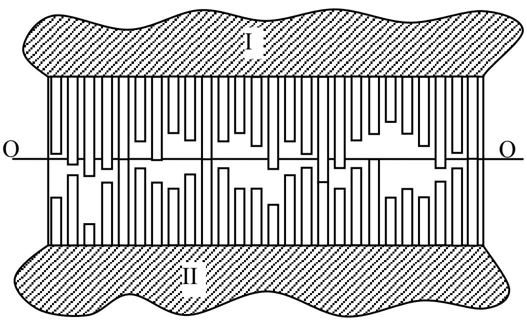


Fig. 1. Contact of two rough surfaces. The rod model

For the accepted model, the micro geometrical resistance for individual contact $r^*_{\text{МГ}}$ can be defined according to the familiar dependence, considering its presence in two parts (fig. 2):

$$r_{\text{МГ}} = 2\rho\left(\frac{1}{d} - \frac{1}{D}\right).$$

The total micro geometrical contact resistivity $r_{\text{МГ}}$ can be defined accepting the following dependence:

$$r_{\text{МГ}} = \frac{r_{\text{МГ}}}{n_r} = \frac{2\rho}{n_r}\left(\frac{1}{d} - \frac{1}{D}\right).$$

To calculate $r_{\text{ВН}}$ and $r_{\text{МГ}}$ it is necessary to define the mechanical parameters of the welded contact and first of all, define the deformation the micro lugs undergo:

whether they are elastic or plastic. This may be done by comparing the pressure σ_{CP} operating the in contact outline area, and the critical pressure σ_K at which the contact transforms to a plastic state.

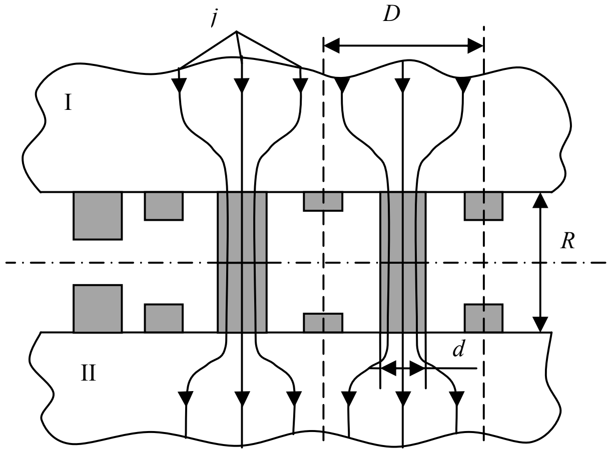


Fig. 2. Contact conductivity in the rod model

The comparison of σ_{CP} and σ_K values at static electrode force defined by the practice of spot welding modes for parts with a thickness of 0.5–4 mm for various steels and alloys shows that the micro roughness in the welded contacts always deforms plastically. For example, at a change R_{max} within 2.4–37 µm the values σ_K / σ_{02} and σ_{CP} / σ_{02} (in brackets) at spot welding of parts with a thickness of 0.5–4 mm made of 08 and 12X18P10E type steels, and a AMг6 alloy change within: $2\cdot10^{-6}$–0,03 (1.6–1.9), $1,2\cdot10^{-5}$–0.02 (1.8–2.7) и $2.5\cdot10^{-5}$–0.4 (1.8–2.9) accordingly. For comparing plastic and elastic contacts the problem of its mechanical parameter definition is considerably simplified; the actual area of the contact depends only on surface's micro geometry, and the rapprochement depends only on the material distribution within the rough layer (fig. 3). For these conditions, value ε can be defined applying the dependence:

$$\varepsilon = \left(\frac{\sigma_{CP}}{C_\varepsilon \cdot \sigma_{02} \cdot b} \right)^{\frac{1}{v}},$$

where C_ε is the factor considering the metal hardening in micro roughness at its plastic deformation; b, v are the parameters of a basic curve b_1, v_1 that are defined by the contacting of two equally rough surfaces according to the formula:

$$b = k_{v_{1,2}} \cdot 2^{v_1} \cdot b_1^2,\ v = 2 \cdot v_1 .$$

The number of micro contacts n_r and their average diameter d can be defined from the following ratio:

$$n_r = \frac{A_r}{\Delta \bar{S}} = \frac{A_C * \eta}{2\pi\ \Theta\varepsilon},$$

$$d = \sqrt{\frac{4\Delta A_C}{\pi}} = 2\sqrt{2\Theta\ \varepsilon}.$$

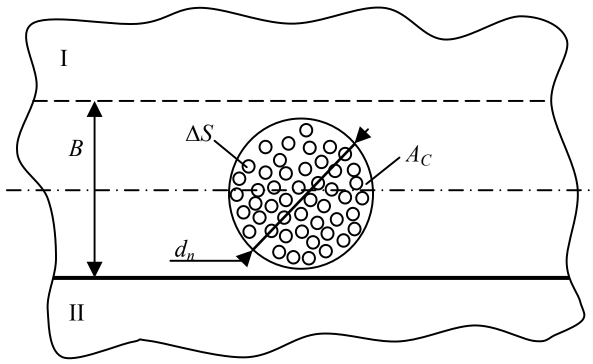


Fig. 3. Model of sieve contact conductivity

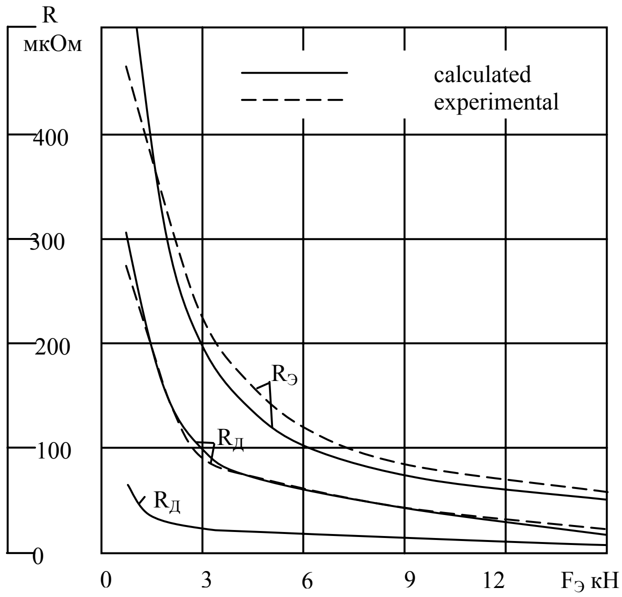


Fig. 4. Results of experimental and calculated resistance values

The average value of the current line spread area from an individual micro conductor can be defined as the area of contacting micro roughness to a number of micro contacts:

$$\Delta A_C = \frac{A_C}{n_r}$$

and its diameter d according to the following dependence:

$$D = \sqrt{\frac{4\Delta A_C}{\pi}} = 2\sqrt{\frac{2\Theta\ \varepsilon}{\eta}}\ .$$

Using the results of the researches, studying contact resistance and carrying out reasonable calculations, we have obtained a dependence to calculate the resistance of the welded contact (fig. 4):

$$r_{KT} = \rho_\Delta \frac{2\sigma_{CP} R_{max}(1-\varepsilon)}{F_Э\ \eta} + \rho \frac{\pi\sigma_{CP}\sqrt{2\Theta\ \varepsilon}}{2F_Э\eta}\left(1-\sqrt{\eta}\right).$$

To define it we need to experimentally measure the value r_{KT} at any $F_Э$ (it is best to measure r_{KT} at $F_Э$ as closely to the recommended spot welding area as possible), and then to calculate value ρ_Δ according to the following dependence:

$$\rho_\Delta = \frac{F_Э \eta}{2\sigma_{CP} R_{max}(1-\varepsilon)} \times$$
$$\times \left(r_{КТ} - \rho \frac{\pi \sigma_{CP} \sqrt{2\Theta\,\varepsilon}}{2F_Э \eta} \left(1-\sqrt{\eta}\right) \right).$$

The developed mathematical model describes, with sufficient accuracy, the action of electric resistance in part-to-part contact, which is testified by experimental and calculated values of this resistivity [1–4].

References

1. Kozlovsky C. N. Modeling of part interaction in the contact area at spot welding // News of high schools. Mashinostroenie. 1990. № 9. P. 89–94.
2. Kochergin K. A. Resistance welding. L. : Mashinostroenie, 1987.
3. Moravsky E. V. Features of initial data definition for cold transition contact resistance calculation // Machine welding. 1988. № 11. P. 32–35.
4. Kragelsky I. V., Dobychin M. I, Kombalov V. C. Calculation bases for friction and wear. M. : Mashinostroenie. 1977.

S. N. Nazarov
Ulyanovsk Higher Civil Aviation School, Russia, Ulyanovsk

A. A. Shagarova
Ulyanovsk State Technical University, Russia, Ulyanovsk

THE REALIZATION OF A MECHANISM FOR RECEPTION DIVERSITY IN THE HYBRID FIBER OF WIRELESS INFORMATION TRANSFER

In this article we have considered the possibilities of realizing diversity reception algorithms in hybrid fibers of wireless information transfer. At the heart of diverse reception algorithms lies the fact that when solving problems of optimal field processing, their correlation curves are definite for the description of the Gauss and any kind of stochastic fields.

Keywords: a hybrid fiber of wireless information transfer, algorithm of diversity reception, correlation factor, optimal reception.

The further development of information transfer in wireless networks (ITWN) consists in the provision of reaching subscribers with various telecommunication services in the principle: "anywhere, everything, when necessary". It is possible to solve the problem of construction networks of such kind basing on the principle of standards convergence, providing compatibility in the management objectives. International standards, reports, and recommendations, which specify the physical level of access management (MAC): IEEE 802.15, 11, 16, 20, 21 cellular and decameter radio communications, are developed in order to provide effective performance in wireless networks. The application of the given standards allows building hybrid networks for wireless information transfer (HNWIT). For this purpose it is necessary to solve a set of problems; one such problem is the application of diversity reception in HNWIT.

The application of diversity signal reception in radio channels is an effective way of increasing communication reliability in conditions of signal fading and the presence of additive hindrances [1–8]. The greatest interest is in the space diversity of signal reception, which consists in reception on different antennas. For HNWIT it is possible to consider the following ways of space diversity: reception on different antennas at one base station in a cellular; reception on antennas with different base stations in a cellular operator area; reception on antennas of interconnected radio centers – repeaters of the radio communication network of a decameter range [9–14]. The realization of these is shown in fig. 1.

The realization of carried reception algorithms demands developing a multiband switch, which depending on the user's inquiries on the transferred information service quality in a network, will perform the connection of the subscriber to corresponding networks. The complexity of this problem's solution consists in the mobile subscriber's equipment. Therefore, an ideal way of realizing the HNWIT substantive provisions is by wireless aeronautical telecommunication (WATC). The primary target of WATC is providing information interchange maintenance between air traffic control points and aircraft situated in their responsibility zones. The solution of given WATC problems are realized by DCM networks, radio communication VHF, and satellite communication network. In order to organize this network structure, the air traffic control points and the aircraft is equipped with necessary means of radio communication.

The aeronautics air communication (AAC) for air traffic control around the airdrome is regulated and organized according to the scheme of air traffic control accepted for the given airdrome.

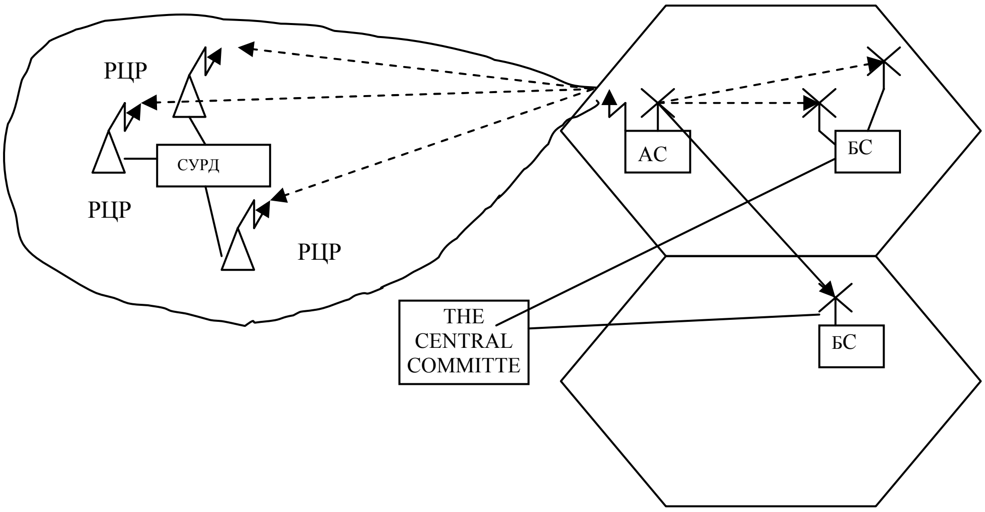


Fig. 1. Ways of spacious separation for signal reception in hybrid networks of wireless information transfer

There are radio networks of air traffic control points around the airdrome: of approach (by quantity of sectors); of a traffic control circle; of takeoff and landing systems; of a start traffic control center; of a command and flight control center; of emergency situation – this is typical for all ATC centers. The analysis of the existing VHF network in radio communication functioning efficiency was carried out based on the accounting technique for expected possible radio communication [15].

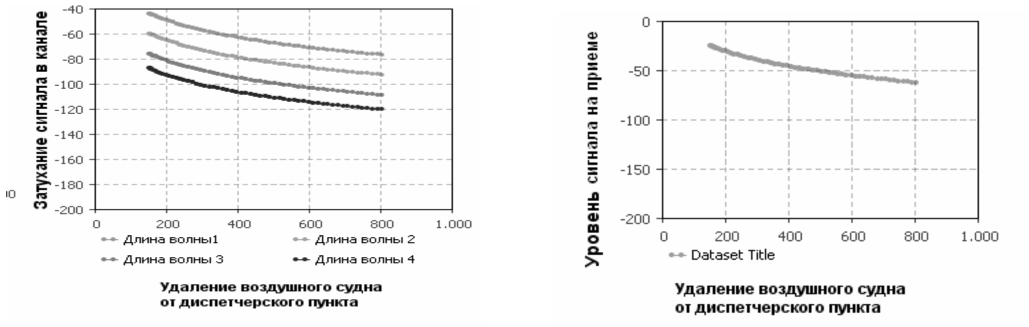
A model, imitating the functioning of the VHF network in radio has been developed using the Any Logic environment program version 6.04. As it can be seen from the graphs in fig. 2, the value of the signal level during reception decreases with as the aircraft distances from the control center. However the radio communication VHF provides the demanded value for the signal at the set distances. Thus, existing VHF networks allow necessary information transfer for necessary aircraft management. However, today the system of air transport consists out of a number of airlines which compete on the market of air transportations. An important factor, making the functioning of such a company successful is providing its passengers with modern information and communication services in real-time. These services can be realized by introducing elements of carried reception for mobile communication networks with broadband access, based on the family of IEEE 802.11, 802.16 aviation telecommunication standards (fig. 3).

AAC on airlines and local air-lines is organized according to the used scheme of the ATC for each airline; the same is for local airlines. The basic means of providing ATC for airlines are those of radio communication, the range of which provides management during all aircraft flight ranges. The following radio networks are organized for this purpose: for management in RC zones (according to the number of sectors) and in auxiliary centers – within VHF range; zone RC aeronautics communication within DCM range; long-distance communication within DCM range; emergency communication within VHF range; radio network of the dispatching office of the air company within VHF range. Long-distance communication radio networks of DCM range are organized to communicate with the aircraft crews, carrying out distant and international flights.

The efficiency analysis of DCM network radio communications functioning has conducted according to a technique considered in [15], based on a model developed in the environment imitating modelling program Any Logic version 6.04. As it can be seen from the graphs in fig. 4, the toughening on requirements to the reliability of information transfer leads to a drop in the realization of radio communication at the given range. In order to increase the SRS DCM in wireless aeronautical telecommunication we have offered to use a network of interconnected radio repeater-centers taken out of the communication zone on a distance of more than 2.000 km at least 250 km apart from each other. This shall provide a low level of spatial correlation in relation values for signal-hindrance between the repeaters (fig. 5). As it can be seen from fig. 6, *a*, *b* the application of the taken out interconnected spatially independent reception points allows raising the values for the probabilities of radio communication in comparison with the use of direct channels.

The performed functioning efficiency analysis of modern wireless communication networks and standards for their design shows that further development of wireless communication is the creating of HSWTI. At the current the level of electronics development the best application of HSWTI is in WATC. The modelling of WATC has shown that the introduction of HSWTI will be carried out on the basis of the taken out radio center networks – repeaters, elements of broadband access, and satellite communication networks.

In order to decreases the communication expenses of local air companies, services to carry out wireless communication construction according to the block diagramme shown in fig. 7 are offered. The aircraft board is considered as a uniform local network where there are passengers and pilots' functions. From a local aircraft network the user can be provided with access to global and corporate network resources in a real-time. According to an inquiry of the user for the type of service and demanded quality, a communicator connects him/her to the corresponding telecommunication network.



a *b*

Fig. 2. Value graphs:
a – attenuation of a signal in channel; *b* – signal level on at receiver input

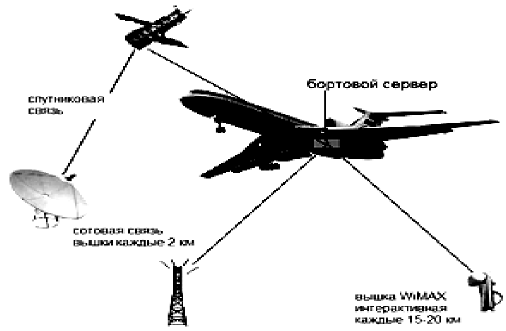


Fig. 3. Application of broadband network elements access in aeronautics telecommunication

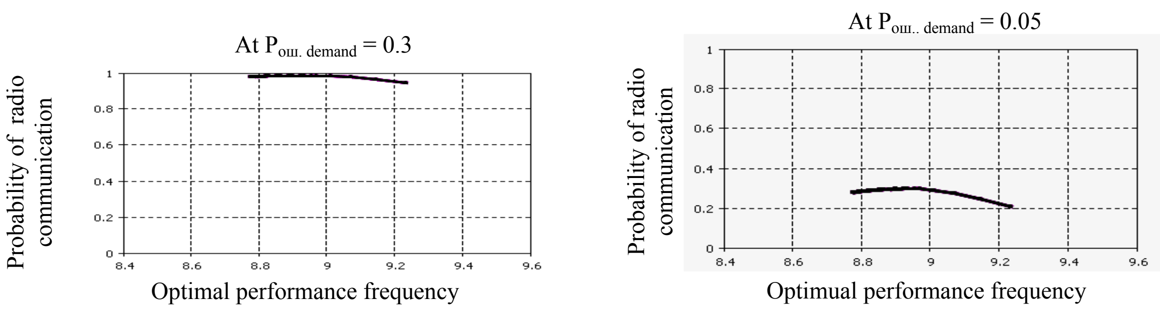


Fig. 4. Value graphs of radio communication probability dependence from the values of optimal performance frequency in existing DCM wireless aeronautics communication

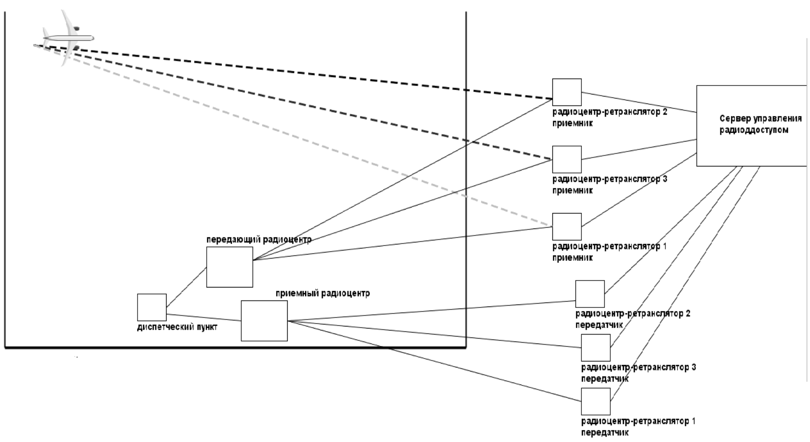
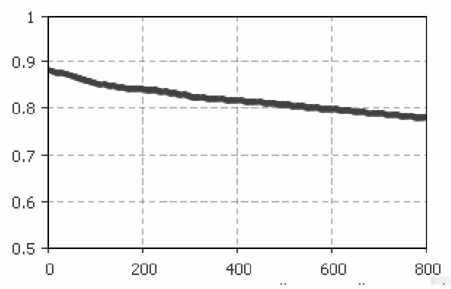
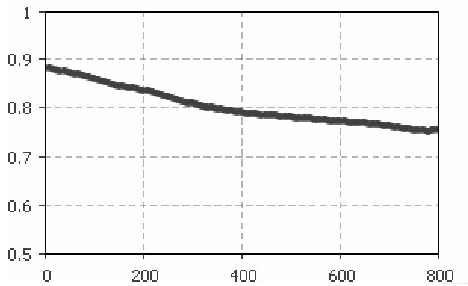


Fig. 5. The application of the “taken out” interconnected radio repeater centers
in a hybrid network of wireless aeronautics telecommunication and information transfer



a



b

Fig. 6. Radio communications of an aircraft with a flight control center:
a – at demanded error probability = 0.01; *b* – demanded error probability = 0.03

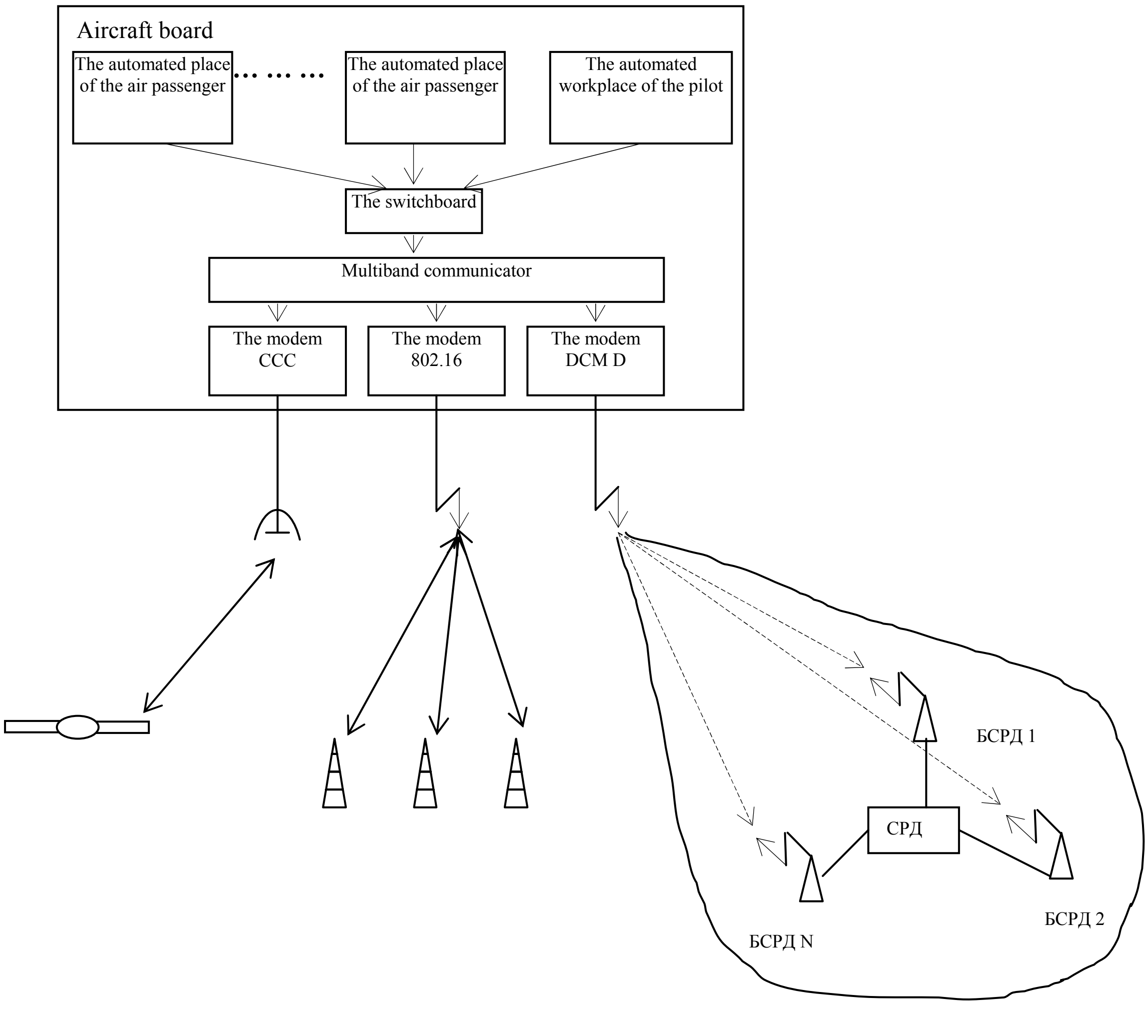


Fig. 7. The offered HSWTI structure for aeronautics telecommunication

References

1. Попов В. И. Основы сотовой связи стандарта GSM. М. : Эко-Трендз, 2005. С. 296.
2. Системы мобильной связи / В. П. Ипатов [и др.]. М. : Горячая линия – Телеком, 2003. С. 272.
3. Комашинский В. И., Максимов А. В. Системы подвижной связи с пакетной передачей информации. Основы моделирования. М. : Горячая линия – Телеком, 2007. С. 176.
4. Ратынский М. В. Основы сотовой связи / под ред. Д. Б. Зимина. М. : Радио и связь, 1998. С. 248.
5. Шахнович И. В. Современные технологии беспроводной связи. М. : Техносфера, 2006. С. 288.
6. Назаров С. Н. Применение динамического программирования при распределении пространственного ресурса радиосвязи декаметрового диапазона // ИКТ. 2007. Т. 5, № 2. С. 70–74.

7. Головин О. В., Простов С. П. Системы и устройства коротковолновой радиосвязи / под ред. О. В. Головина. М. : Горячая линия – Телеком, 2006. С. 598.

8. Назаров С. Н. Общий подход к построению современных гибридных сетей беспроводной связи // Тр. Рос. науч.-техн. общества радиотехники, электроники и связи имени А. С. Попо-ва. М., 2009. Вып. LXIV, С. 22–24. (Сер.: Научная сессия, посвященная Дню радио).

9. Назаров С. Н. Использование стохастических моделей для оценки характеристик современной беспроводной сети передачи информации // Современ-ные проблемы создания и эксплуатации радиотехни-ческих систем : Тр. VI Всерос. науч.-практ. конф. (с участием стран СНГ). Ульяновск : УлГТУ, 2009. С. 170–174.

10. Назаров С. Н., Назаров А. С. Анализ методов моделирования беспроводной сети передачи информации // Современные проблемы создания и эксплуатации радиотехнических систем : тр. VI Всерос. науч.-практ. конф. (с участием стран СНГ). Ульяновск : УлГТУ, 2009. С. 174–177.

11. Назаров С. Н. Применение элементов декаметровой радиосвязи в современных беспроводных сетях // Тр. Рос. науч.-техн. общества радиотехники, электроники и связи имени А. С. Попо-ва. М., 2009. Вып. XI-1. С. 228–230. (Сер: Цифровая обработка сигналов и ее применение).

12. Назаров С. Н., Назаров А. С. Обобщенная модель беспроводной сети передачи информации авиационного предприятия // Современные научно-технические проблемы транспорта : сб. науч. тр. V Междунар. науч.-техн. конф. Ульяновск : УлГТУ, 2009. С. 108–111.

13. Назаров С. Н. Применение гибридной беспроводной сети передачи информации в автоматизированной системе управления воздушным движением // Современные научно-технические проблемы транспорта : сб. науч. тр. V Междунар. науч.-техн. конф. Ульяновск : УлГТУ, 2009. С. 112–116.

14. Назаров С. Н. Основные положения методики определения места расположения сети удаленных взаимосвязанных радиоцентров-ретрансляторов // ИКТ. 2009. Т. 7, № 2. С. 79–82.

15. Прохоров В. К., Шаров А. Н. Методы расчета показателей эффективности радиосвязи Л. : ВАС, 1990. С. 132.

A. I. Pakhirka
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

THE APPLICATION OF IMAGE ENHANCEMENT METHOD FOR FACE RECOGNITION SYSTEMS

Three-step face recognition algorithm which includes non-linear enhancement (dynamic range compression) and faces localization on the basis of skin color segmentation with subsequent extraction of anthropometric face points is proposed. The process of face recognition on the basis of principal component analysis is also considered.

Keywords: dynamic range compression, face localization, face recognition.

Face recognition has always caused great interest in computer vision, especially in connection with increasing practical needs such as biometrics, search engines, video compression, video conferencing systems, computer vision in robotics, intelligent security and access control systems.

Face recognition algorithms can be divided into two categories: methods based on extracting features of images and methods based on representation of a facial image. The first group of methods uses properties and geometric relationships such as areas, distances and angles between feature points of a facial image. The second group of methods considers global features of a facial image.

Usually these methods try to represent facial data more efficiently, for example, as a set of main vectors. Typically, a face recognition algorithm includes three steps: image preprocessing, face localization, face recognition. In this paper we present an algorithm which includes nonlinear image enhancement (dynamic range compression), face localization on the basis of skin color segmentation and face recognition on the basis of principal components analysis [1].

In practice images captured by digital devices often differ from what an observer remembers. It happens due to the fact that a camera captures the physical values of light data, while an observer's nervous system processes these data. For example, an observer can easily see details both in deep shadows and in illuminated areas while a capture device will get the given scene with too dark areas or light-struck areas. A human observer easily perceives scenes with a high range of light intensities while the ratio between the highest and the lowest luminance exceeds the capabilities of a capture or output device.

The human observer deals with high dynamic range scenes by adapting locally to each part of the scene and thus is able to retrieve details in low luminance as well as high luminance areas. Using a digital device is more problematic. The dynamic range of the scene has to be compressed, which often causes the captured image to lack details in areas of low and high illumination. Some recent developments made it possible the capture high dynamic range scenes.

The principle is to capture multiple pictures of the same scene with different exposure times. A so-called radiance map is built from the acquired pictures. This technique allows obtaining an accurate estimation of the scene despite a capture device limitation. Nevertheless, the problem of mapping the high dynamic range values expressed in floating point to the low dynamic range of the output device remains [2].

These problems can be solved by an algorithm which simulates a human visual system. For this purpose we can use Multi-Scale Retinex (from retina and cortex) algorithms. These algorithms compress dynamic range of images with saving (increasing) local contrast in areas of low and high illumination [3].

A classical multidimensional MSR-algorithm is a weighed sum of one-dimensional SSR (Single-Scale Retinex) algorithms for different scales. A univariate output function of i-th color channel $R_i(x, y, \sigma)$ is calculated like this:

$$R_i(x, y, \sigma) = \log\{I_i(x, y)\} - \log\{F(x, y, \sigma)*I_i(x, y)\},$$

where $I_i(x, y)$ is an output function for color channel i, σ is a scale factor, "*" denotes convolution, $F(x, y, \sigma)$ is Gaussian function given by:

$$F(x, y, \sigma) = Ke^{-(x2 + y2)/\sigma 2}.$$

where parameter K is chosen so as to meet the requirement:

$$\iint_{\Omega_{x,y}} F(x, y, \sigma)\, dx\, dy = 1,$$

where $\Omega_{x,y}$ is a number of pixels of the whole input image.

Then a multidimensional output function of i-th color channel $R_{mi}(x; y; \mathbf{w}; \sigma)$ is calculated like this::

$$R_{M_i}(x, y, \mathbf{w}, \sigma) = \sum_{n=1}^{N} w_n R_i(x, y, \sigma_n),$$

where $\mathbf{w} = (w_1, w_2, \ldots, w_m)$, $m = 1, 2, \ldots, M$ is a weight vector of univariate output functions of i-th color channel $R_i(x, y, \sigma)$; $\boldsymbol{\sigma} = (\sigma_1, \sigma_2, \ldots, \sigma_n)$, $n = 1, 2, \ldots, N$ is a vector of scales of univariate output functions, and $\sum_{n=1}^{N} w_n = 1$. The length of scales vector is usually not less than 3. In different sources we can find different recommended scale values, in our experiments they were 15, 90, 180. A weight vector $\boldsymbol{w}$ has, as a rule, elements with equal values.

A block diagram of an image enhancement module is shown in fig. 1. Conversion from RGB to YCbCr is conditioned by the fact that in it color space luminance is presented separately. Therefore, the algorithm is applied only to Y component, without affecting Cb and Cr, which improves the performance of the algorithm. For Gaussian blur recursive implementation of Gaussian filter is used which approximates gaussian, with calculation of filter coefficients for a desired value of sigma (σ). Such representation of the filter is faster than standard filtering using convolution kernel [4].

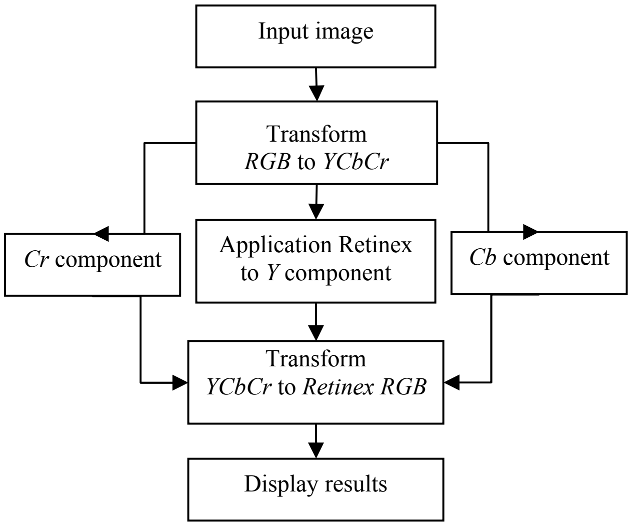


Fig 1. A block diagram of an image enhancement module

Fig. 2 shows the results of application of Retinex algorithm to low illumination image.

a *b*

Fig 2. Results of *SSR*:
a – input image; *b* – output image

The next step is faces localization on the basis of skin color segmentation. The process of face localization can be divided into two stages:

– extraction of image areas with the color similar to human skin color (skin color segmentation);

– analysis of extracted regions after segmentation.

Determination of skin color can significantly reduce a search area and is the first step in many methods of face localization.

Human skin has a characteristic color which allows to successfully segment skin in color images. The independence of hue color component on the face orientation as well as its small dependence on brightness make color a stable characteristic of skin. The advantages of skin color segmentation are:

– low computational complexity;

– stability to change in scale and face rotation;

– stability to change in lighting;

– stability to change in facial expressions and face overlapping.

Skin color segmentation requires building of some rules which distinguish between facial color pixels and pixels not related to skin color. For this purpose a metrics is introduced which allows to measure the distance between pixel color and skin hue. This metrics is a model of skin color distribution in a selected color space.

We use the metrics for NCC RGB color space. Skin color distribution for NCC RGB is shown in fig. 3:

$$Skin(r,g)=\begin{cases}1 & if \ \ (g<g_u)\cdot(g>g_d)\cdot(W>0.0004)\\ & 0 \ \ otherwise\end{cases},$$

where u is an upper boundary, d is a bottom boundary. Values g_u, g_d, W are defined as:

$$g_u = J_u r^2 + K_u r + L_u\,,\ g_d = J_d r^2 + K_d r + L_d\,,$$

$$W = (r-0.33)^2 + (g-0.33)^2\,,$$

coefficients take the following values:

$$J_u = -1.377,\ K_u = 1.074,\ L_u = 0.145,$$

$$J_d = -0{,}776\ \ K_d = 0.560,\ L_d = 0.177.$$

Results of skin segmentation are shown in fig. 4. The segmented image is processed morphologically (compression with subsequent expansion) which allows to separate poorly connected regions and delete regions of small sizes (noise). Later on the connected areas are marked (fig. 5) and anthropometric points are detected in each area (eyes, lips, nose).

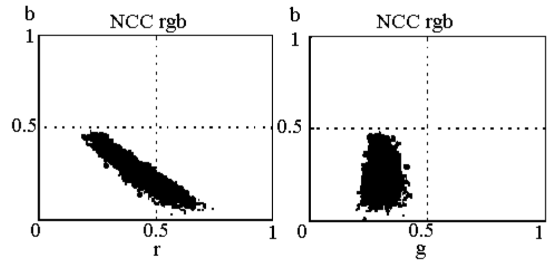


Fig. 3. Skin color distribution for NCC rgb

We use face recognition based on principal component analysis (PCA). Principal component analysis is a standard technique used to approximate the original data with lower dimensional feature vector. PCA is probably the most widely used subspace projection technique for face recognition. This method projects image space into the space of smaller signs. The main idea of PCA is to represent human faces images as a set of image principal components called eigenfaces. The calculation of principal components is reduced to the calculation of eigenvectors and eigenvalues of a covariance matrix which is calculated form image [5].

Any image may be considered as a vector of pixels each value of which is presented by a value in gray scale gradation. For example, an 8×8 image may be unwrapped and treated as a vector of length of 64 pixels. Such vector representation describes the image input space. To present and recognize faces we use a subspace created by eigenvectors of a covariance matrix of investigated images. Eigenvectors corresponding to nonzero eigenvalues of a covariance matrix form an orthogonal basis which rotates and/or reflects the images in the N–dimensional space. Specifically, each image is stored in a vector of size N.

$$X^i = \left[x_1^i \cdots x_N^i\right]^{\mathrm{T}}, \quad (1)$$

where x^i are master images, X is a matrix of master images. The images are centered by subtracting the mean image from each image vector

$$\overline{x}^i = x^i - m, \quad (2)$$

where $m = \frac{1}{P}\sum_{i=1}^{P} x^i$

a *b*

Fig. 4. Skin segmentation:
a – input image; *b* – output image after the use of metrics

a *b* *c*

Fig. 5. Face localization:
a – input image; *b* – morphological processing with marking of connected components; *c* – anthropometric facial points detection

These vectors are combined to create a data matrix of size $N \times P$ (where P is the number of images):

$$\bar{\mathbf{X}} = \left[\bar{x}^1 \;\; \bar{x}^2 \;\; \cdots \;\; \bar{x}^P\right]. \quad (3)$$

The data matrix $\bar{\mathbf{X}}$ is multiplied by data transposed matrix to calculate the covariance matrix:

$$\Omega = \bar{X}\bar{X}^{\mathrm{T}}. \quad (4)$$

This covariance matrix has up to P eigenvectors associated with non-zero eigenvalues, assuming $P < N$. The eigenvectors are sorted from higher to lower value according to their associated eigenvalues. The eigenvector with the largest eigenvalue represents the greatest dispersion in images.

Images recognition through eigenspace projection has three basic steps.

1. Eigenspace must be created using master images (training stage).
2. Master images are projected into the eigenspace (training stage).
3. The projected input image is compared with the projected test images (recognition stage).

Let's consider the first stage – creation of eigenspace consisting of the following steps:

– data centering: each image is centered by subtracting an averaged image from each master image. An averaged image is a column vector including mean pixel values of all pixels of master images (2);

– creation of a data matrix: once the input images are centered, they are combined into a data matrix of size $N \times P$ (equation (3));

– creation of a covariance matrix: the data matrix is multiplied by its transposed representation (4);

– calculation of eigenvalues and eigenvectors: eigenvalues and corresponding eigenvectors are calculated from the covariance matrix:

$$\Omega\mathbf{V} = \Lambda\mathbf{V}$$

where $\mathbf{V}$ is a set of eigenvectors associated with the eigenvalues Λ;

– ordering eigenvectors: the eigenvectors $v_i \in \mathbf{V}$ are ordered according to their corresponding eigenvalues $\lambda_i \in \Lambda$ from higher to lower value. The eigenvectors associated with non-zero eigenvalues are kept. This matrix of eigenvectors is the eigenspace V , where each column of V is an eigenvector:

$$\mathbf{V} = \left[v_1 \;\; v_2 \;\; \cdots \;\; v_P\right].$$

Projecting of master images into eigenspace takes place at the second stage. Each centered input image ($\bar{x}^i$) is projected into an eigenspace:

$$\tilde{x}^i = \mathbf{V}^T \bar{x}^i.$$

Identifying input images takes place at the third stage. Each input image is first centered by subtracting the averaged image, and is then projected into the same eigenspace defined by $\mathbf{V}$:

$$\bar{y}^i = y^i - m,$$

where

$$m = \frac{1}{P}\sum_{i=1}^{P} x^i$$

and

$$\tilde{y}^i = \mathbf{V}^T \bar{y}.$$

The projected input image is compared with every projected master image. Images can be compared using any simple metrics, for example, Euclidean.

Currently a nonlinear image enhancement system in different color spaces is being developed. We plan to use MultiScale Retinex algorithm with color restoration for capture and processing video streams having a big range of brightness. A system of face capturing from video sequence with subsequent processing and "averaging" of images, reducing the influence of illumination, correction of face position, choosing the best face image from video data is being developed.

In this paper we propose an improved approach to face recognition in images, using a nonlinear image enhancement algorithm that allows compensating shadows and highlights. Also, the analysis of color spaces can improve the quality of skin color segmentation recognition and anthropometric face points.

References

1. Jain K., Flynn P., Ross A. Handbook of Biometrics. Springer, 2008.

2. Meylan L., Susstrunk S. Bio-inspired color image enhancement // SPIE Electronic Imaging. San Jose, 2004, P. 46–56
3. Tao L., Asari K. V. Nonlinear enhancement of color images // SPIE Journal of Electronic Imaging. 2005. Vol. 14.
4. Young T., Van Vliet L. J. Recursive Implementation of the Gaussian filter : Signal Processing 44. Elsevier, 1995.
5. Yambor W. Analysis of PCA-based and Fisher discriminant-based image recognition algorithms : Technical Report CS-00-103. 2000.

A. N. Pakhomov, M. V. Krivenkov, V. I. Ivanchura
Siberian Federal University, Russia, Krasnoyarsk

MODAL REGULATORS OF A DIRECT CURRENT ELECTRIC DRIVE WITH A PULSE-WIDTH CONVERTER

The technique of modal regulators synthesis of instantaneous coordinates values of direct current digital electric drives with a pulse-width converter by a state space method taking into account the influence of variable pure delay in the control channel is presented.

Keywords: a modal regulator, a direct current electric drive, a pulse-width converter.

The theory of digital repeated systems of the subordinated regulation of the electric drive [1–3] is developed well enough. Less attention has been paid to electric drive systems constructed on the basis of summation of feedbacks on a state vector that allows to expect the decrease of sensitivity to variations of control object parameters. In domestic literature such regulators are usually called modal because factors of a feedback vector directly influence eigenvalues (modes) of the closed system matrix. In the present paper the task of designing of such regulators is set. The task includes:

– reception of discrete equations of a control object condition according to its differential equations;

– definition of feedback factors by state variables according to the set spectrum of a matrix of the closed digital control system dynamics;

– introducing the corrective amendments connected with some features of the real pulse-width converter (PWC). It is supposed that the influence of variable pure delay in the control channel of a double digital system which is rather typical for an electric drive system with microprocessor control is taken into account.

The reasons of occurrence of two periods of discreteness as well as the nature of pure delay in microprocessor systems are considered in works [4; 5]. It is supposed that the interruption period (IP) of work of a control microcomputer contains integer N of the switching periods (SP) PWC: $T = NT_k$, where T and T_k are values IP and SP accordingly. In view of two periods of discreteness it is expedient to introduce two types of relative time – global τ and local one inside IP θ, whereby

$$\tau = \frac{t}{T}; \quad \theta = \frac{t}{T_k} = \frac{N}{T}t = N\tau; \quad \theta \in [0, N]. \qquad (1)$$

Below only thus determined relative values of time are used. Let the calculation of a control signal $u[n]$ on IP with number n is completed after computing delay τ_d on k of the SP (fig. 1). Local computing delay θ_d is counted off from the beginning of SP with number k. If the local time delay θ_{PWC1} necessary for realization of $u[n]$ is more than θ_d (fig. 1) then this realization can be carried out already on k of the SP. The previous value of a control signal $u[n-1]$ is realized on the first k of the SP.

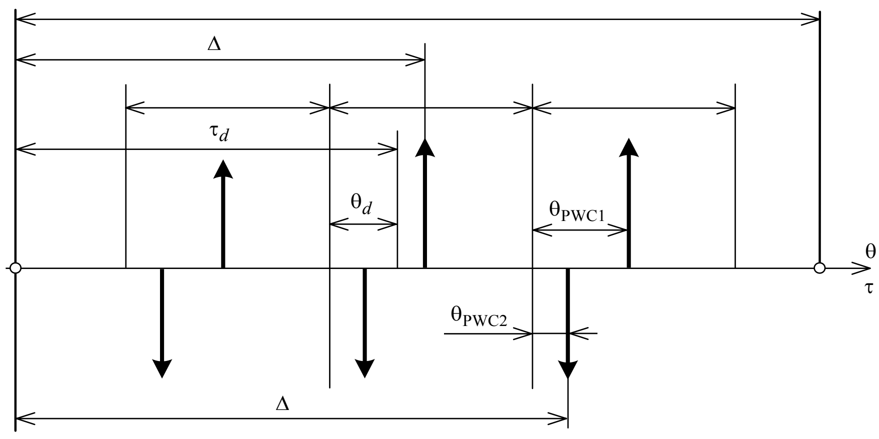


Fig. 1

If $\theta_{PWC2} < \theta_d$ (fig. 1) then the realization of $u[n]$ is postponed up to SP with number $k + 1$. In that specific case when $k = N - 1$ it will begin only from the SP with number zero on $n + 1$ of the IP.

The delay of θ_{PWC} is caused by the nature of semiconductor PWC the outpot voltage of which is adjusted due to the change of its size.

In a local time scale the size of pure delay belongs to the range: $0 \le \theta_\Delta \le N + 1$ [5]. Therefore it is convenient to break it into whole K and fractional δ parts: $\theta_\Delta = K + \delta$.

Digital regulators designing includes calculation of discrete state equations parameters of a control object in view of the influence of pure delay having a general view according to [4; 5]:

$$\tilde{X}[n+1] = \tilde{\Phi}_{IP}\tilde{X}[n] + \tilde{W}_{IP}u[n], \qquad (2)$$

where the vector of state variables, transitive and weight matrices for IP will be written down as:

$$\tilde{X}[n] = \begin{bmatrix} X[n] \\ y_1[n] \end{bmatrix}; \ \tilde{\Phi}_{IP} = \begin{bmatrix} \Phi_{IP}(1) & F \\ 0 & 0 \end{bmatrix};$$

$$\tilde{W}_{IP} = \begin{bmatrix} H \\ 1 \end{bmatrix}; \quad y_1[n+1] = u[n]. \qquad (3)$$

In matrices (3) an additional variable of a condition y_1 is introduced for the case when the whole part of pure delay on a global scale to time is equal to one IP.

If regulation is carried out according to instant values of coordinates then the calculation of matrices F and H is carried out according to formulas [4]:

$$F = \Phi_{SP}^{N-K}(1)\underbrace{\{\Phi_{SP}(1)...\{\Phi_{SP}(1)}_{K-1} + \underbrace{E\} + ... + E\}}_{K-1} W_{SP}(1-\delta); \qquad (4)$$

$$H = \underbrace{\{\Phi_{SP}(1)...\{\Phi_{SP}(1)}_{N-K-1} + \underbrace{E\} + ... + E\}}_{N-K-1} W_{SP}(1-\delta),. \qquad (5)$$

where Φ_{SP} and W_{SP} are transitive and weight matrices for SP, E is an individual matrix.

The synthesis of a digital modal state regulator by the set spectrum of a matrix of dynamics is possible in the only case when [6] the pair of matrices $[\tilde{\Phi}_{IP}, \tilde{W}_{IP}]$ is completely controlled. Let P be a vector of feedback factors of dimension l, where l is the amount of state variables (the order of a characteristic polynomial of a system). Value P can be found from [6]:

$$P = [(GS^T)^{-1}(\gamma - g)]^T, \qquad (6)$$

where

$$S = \begin{bmatrix} \tilde{W}_{IP} & \tilde{\Phi}_{IP}\tilde{W}_{IP} & \tilde{\Phi}_{IP}^2\tilde{W}_{IP} & \cdots & \tilde{\Phi}_{IP}^{n-1}\tilde{W}_{IP} \end{bmatrix};$$

$$\gamma = \begin{bmatrix} \gamma_{l-1} & \gamma_{l-2} & \cdots & \gamma_0 \end{bmatrix}^T; \ g = \begin{bmatrix} g_{l-1} & g_{l-2} & \cdots & g_0 \end{bmatrix}^T;$$

$$G = \begin{bmatrix} g_l & 0 & 0 & \cdots & 0 \\ g_{l-1} & g_l & 0 & \cdots & 0 \\ g_{l-2} & g_{l-1} & g_l & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ g_1 & g_2 & g_3 & \cdots & g_l \end{bmatrix},$$

where γ и g are factors of a characteristic polynomial of a closed system and a control object.

If factor $g_l \ne 0$ then triangular matrix G is not degenerate [6]. Therefore, for the existence of vector P the matrix of controllability S should have a rank l, or, which is the same, the pair of matrices $[\tilde{\Phi}_{IP}, \tilde{W}_{IP}]$ should be controlled.

In a real system with the PWC the size of pure delay is variable and it is necessary to take into account this fact while synthesizing the regulators of digital systems. Let us consider the synthesis of a modal regulator and the research of a closed digital system of a direct current electric drive with the PWC. In view of variable character of pure delay Δ some values of feedback factors corresponding to the range of changes Δ of the latter are counted. Intermediate values are derived by interpolation.

The system of continuous differential equations for the SP does not depend on the model of a power converter and looks like [3]:

$$\left.\begin{aligned} \frac{di}{d\theta} &= -\frac{1}{\theta_a}i - \frac{1}{\theta_a}\omega + \frac{1}{\theta_a}u, \\ \frac{d\omega}{d\theta} &= \frac{1}{\theta_m}i. \end{aligned}\right\}, \qquad (7)$$

where i, ω are relative values of armature current and frequency of an engine shaft rotation accordingly; θ_a и θ_m are relative electromagnetic and electromechanical constants of time. Let, for example, $\theta_a = 8$, $\theta_m = 32$ и $N = 4$ then matrices of dynamics and an input of continuous equations of condition (7) are equal:

$$A = \begin{bmatrix} -0.125 & -0.125 \\ 0.03125 & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0.125 \\ 0 \end{bmatrix}.$$

Factors $P(\Delta)$ in the model are set as the table of conformity between size Δ and eigenvalues of feedback factors which should be calculated prior to the beginning of modelling. The values of vector P are calculated for a linear model with an ideal pulse element in the assumption of the steady-state mode. A part which realizes variable factor $P(\Delta)$ carries out linear interpolation between two values nearest to it if size Δ does not coincide with the calculated values. It allows to reduce the number of elements of the table of conformity because calculations show that the curve of dependence of factors values $P(\Delta)$ is monotonous and smooth enough .

In table the values of feedback factors in a direct current electric drive system with PWC calculated according to the technique described above for a number of values Δ are shown. Here P_I is a vector of feedback factors on an instant armature current, P_ω is a vector of feedback factors on instant frequency of rotation of an engine shaft, P_U is a vector of feedback factors on the value of a control signal on previous IP which is an additional state variable.

A binomial spectrum of a matrix of a closed system dynamics with an equivalent constant of time τ_ω describing its speed and equal to 1,5 was chosen as an example for calculation of values of factors $P(\Delta)$.

Factors of feedback

Δ	0	0,2	$0{,}25_-$	$0{,}25_+$	0,45	0,65	0,85	1,05	$1{,}25_-$
P_I	1,1103	1,0963	1,0926	0,5029	0,4912	0,4795	0,4677	0,456	0,4442
P_ω	3,9081	3,7977	3,7704	1,5461	1,4964	1,4478	1,4004	1,3543	1,3093
P_U	0	0	0	–0,223	–0,16	–0,099	–0,04	–0,017	–0,017

One can see in the table that if the size of pure delay is greater than the SP of the PWC then there is an additional feedback [4] – nonzero factor P_U in the model. At the same time the presence of a feedback on one more coordinate of a system can appear to be useful from the point of view of increase of roughness of an electric drive system.

Let's consider the calculation of transients in the described system by means of a package of imitating modelling *Matlab* of the firm *The MathWorks*, *Inc*. The corresponding structure of a closed system for modelling in a tool *Simulink* of package *Matlab* is shown in fig. 2.

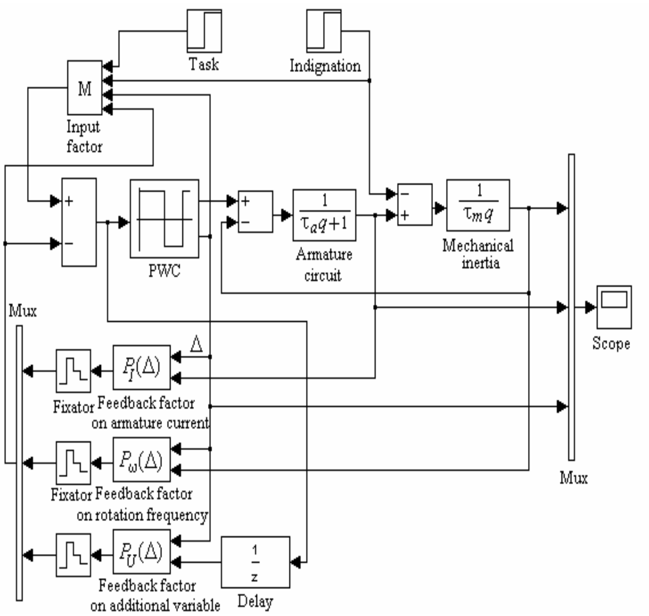


Fig. 2

The block diagram of a model will consist of connected in series model PWC and parts of a control object of a direct current electric drive system, where q is a dimensionless Laplace operator. The variable factors of feedback dependent on value Δ given out by the PWC which are switched N times for IP are included in models. Relative constants of time τ_a и τ_m are N times less than θ_a and θ_m [4].

To correct the static mistake caused by the application of a load moment an adjustable input factor M was introduced into the structure. Its size depends, apart from everything else, on the value of a load moment. As it is known, the measurement of the latter is inconvenient, therefore below a variant of absence of such dependence is considered. An input factor is calculated according to the formula:

$$M = 1 + \sum_l \frac{x_l \cdot p_l}{x^*},$$

where l is the amount of state variables, x_l is an instant value of l-variable at the end of IP, p_l is the value of a feedback factor on l-variable, x^* is the sum of values of master control and a reference signal.

It should be noted that an input factor for a reference signal which was introduced to provide a zero static mistake depends on the size of pure delay as the factors of a characteristic polynomial of a closed system transfer function will change together with the change of vector P.

The size of a delay θ_{PWC} represents the factor of PWC filling β which is calculated according to the formula [3]:

$$\beta = \frac{\omega^* + i_f + 1}{2},$$

where ω^* is a preset value of an engine speed, i_f is the static current accepted as a disturbing signal for a system.

In fig. 3 one can see the curves of transients of the described system which have a typical for a binomial spectrum aperiodic character and a preset speed. Along the horizontal axis there are values of global relative time. The presence of pulsations can be explained by a pulse character of armature current. Disturbance compensation provides a zero static mistake. The size of computing delay introduced by a microcomputer is equal to 0.1. The maximal control signal and EMF of a voltage source of the PWC are both equal to 1. The signal of the task is a jump from a level of 0.1 up to 0.2 with displacement on three IP (for clearness). A static current step from 0.05 up to 0.1 is displaced on 25 IP to provide the established mode in a system after performing the task.

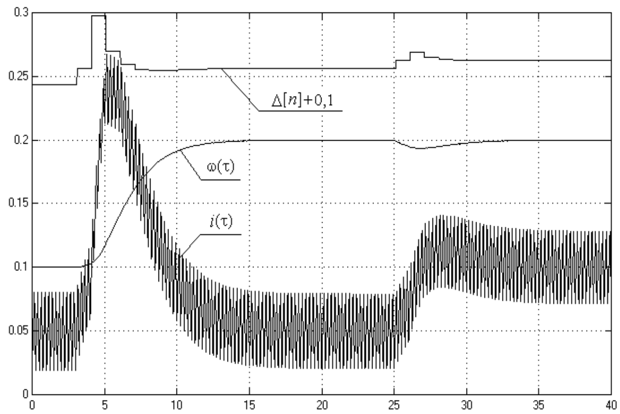
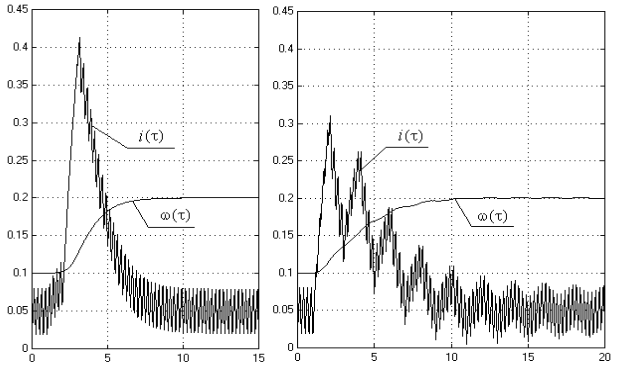


Fig. 3

As it has already been mentioned the actual size of pure delay can vary rather widely and under its significant deviation from the calculated value transients can be completely unacceptable. The developed technique of modal regulators synthesis allows to exclude subharmonic oscillations typical for high-speed systems with a part of pure delay in the control channel. Subharmonic oscillations are oscillations the period of which exceeds the IP by an integer of times [3]. Fig. 4 shows schedules of processes in a digital electric drive system for two cases:

– feedbacks with variable factors in the function of pure delay (fig. 4, *a*);

– feedbacks with constant factors calculated in the assumption of an average value of pure delay equal to one IP [3] (fig. 4, *b*). It is obvious, that in the second case in the process of performing the task there appear converging subharmonic oscillations as a result of which the speed of a system decreases approximately in one and a half time. If the speed of a system increases considerably subharmonic oscillations can even increase. In any case their occurrence is extremely undesirable, as they are badly filtered by flyweights because of their low frequency and it leads to additional losses of energy.



a *b*

Fig. 4

Conclusions:

– the analysis of time delays introduced by a microcomputer together with the PWC is made and the principle of their account is established in the process of calculation of the discrete state equations of a control object;

– the technique of numerical calculation of modal state regulators is developed in view of the influence of variable pure delay in a microprocessor system with several periods of discreteness;

– the account of variable character of pure delay allows to remove subharmonic oscillations typical for high-speed closed digital systems;

– there is an opportunity to decrease a static mistake (to its full compensation) and to reduce a dynamic mistake significantly if disturbing influence in a system can be supervised and measured.

References

1. Volkov A. I. Algorythms and structures of microprocessor systems of high-dynamic electric drives control // Electrical engineering. 1999. № 8. P. 10–16.

2. Synthesis of microprocessor systems of asynchronous electric drives control applying the method of polynomial equations / I. Ya. Braslavsky, A. M. Zyuzev, Z. Sh. Ishmatov, S. I. Shilin // Electrical engineering. 1998. № 6. P. 20–24.

3. Zalyaleyev S. R. Design of microprocessor regulators of industrial electric drives : Tutorial. Krasnoyarsk : KSTU, 1995.

4. Zalyaleyev S. R., Pakhomov A. N. Equations of the state of a direct curent electric drives microprocessor system taking into account pure time delay in a control channel // Proceedings of higher educational institiutions. Electromechanics. 2002. № 3. P. 35–41.

5. Zalyaleyev S. R., Pakhomov A. N. Mathematical description of a control object of a digital system with two periods of discreteness and delay // Optimization of an operating mode of electric drives systems : Interacademic collection of scientific papers / editor-in-chief S. R. Zalyaleyev . Krasnoyarsk : KSTU, 2002. P. 157–168.

6. Kuo B. Theory and designing of digital control systems : translate from English. M., 1986.

I. N. Palamar, P. V. Sizov
Rybinsk State Aviation Technological Academy named after P. A. Solovyev, Russia, Rybinsk

A METHOD OF IMAGE SEGMENTATION WITH THE HELP OF AREAS GROWING AND MULTISCALE ANALISYS

In this article we analyzed advantages and disadvantages of existing methods of image segmentation. The development of an original algorithm of segmentation which uses the method of areas growing and multi scale analysis is presented. The capabilities of this method in different images segmentation are researched.

Keywords: segmentation, image, growing, wavelet.

Segmentation means selection of homogeneous fields in original digital images. It is one of the most important problems in modern systems of computer vision, which are used in many scientific and industrial spheres: medicine, metallography, air-photography, robotics, safety systems and others. It is at this stage of processing that conversion of an image from a set of pixels into a set of segments suitable for further recognition of scene objects takes place.

Nowadays there is a number of formalized methods of segmentation [1], which can be divided into two groups according to the basic principle of working:

– methods which initially select area boundaries (contours) as the drops of some feature of an image;

– methods which select segments having a homogeneous feature.

The first group includes methods which calculate the first and the second derivatives of the image function with the help of different masks (Roberts operator, Previtt operator, Laplace operator, Marr-Hildreth operator, etc.), supplemented with methods of contour binding (local binding, Hough's transformation, analysis with the help of graph theory). The main problem of these contours determination methods is that the derivable borders of an object are disconnected. It is understandable because the basic algorithms of segmentation are not set up to produce connected closed contours, and methods of binding are a superstructure for these algorithms and can solve this problem only at the expense of great computational burden.

The second group of segmentation methods includes first of all the method of threshold transformation. This method is rather wide-spread because it is quite easy and has simple characteristics. But though it was improved several times it has a defect which is also present in the above mentioned methods of contours selection. All of them do not use information about segments connection. This defect is absent in methods of segments selection because initially they were formed to produce connected areas meeting some requirement of homogeneity, that is, areas growing, division and fusion of areas, segmentation according to morphologic divide.

The method of areas growing in its basic version can be described as following:

– some points (cores of crystallization) are chosen on the original image; it is supposed that they belong to the selected areas, for example, these could be points with the highest level of brightness;

– then the areas growing starts from these points, that is, surrounding points are added to the selected points according to a criterion of their closeness, for example, difference in brightness;

– the growing stops when some condition is satisfied, for example, the peak declination of new point's brightness from the brightness of crystallization center.

It is obvious from the description that this algorithm takes into account the connection of areas and in practice can solve many problems much more successfully than the algorithms that use threshold transformation. The only drawback is an increased calculative complexity and the main difficulty is to find centers of crystallization and define the moment when an area stops growing.

From the given presentation of segmentation methods we can draw a conclusion that it is necessary to use the information about elements (contours) connection during image segmentation. So the modification of areas growing method seems to be the most perspective idea. Its aim is to get full segmentation rather than separate chosen objects from the background.

Let's define the requirements to a new method of segmentation:

– it should be very precise and reliable in image segmentation and should be able to work with rather complex images;

– it should implement the full image segmentation;

– it should pick out connected homogeneous areas;

– it should use information about textures, not only about brightness;

– it should be highly automated; operator's work during segmentation should be minimal.

First of all when describing the method of segmentation we should define the way of presenting the original data. Let's specify the original image as a vector-function $\overline{f}(x,y)$ that is defined on the two-dimensional space of integer as $\overline{f}(x,y)=\overline{p}_{x,y}$, where vector $\overline{p}$ = $[p^R,\ p^G,\ p^B]$ defines pixel's color in three-dimensional space RGB.

First of all, it's necessary to solve the problem of crystallization centers searching globally. The property of growth centers to be inside homogeneous areas prompts the way to find them – the method of differential transformation. Then in the transformed image the points of local minimums will correspond to the points with the lowest change of brightness in the neighborhood, that is, to the crystallization centers. For the majority of transformations used in this case (Previtt, Sobell, etc.) the radius of differentiation will not be more than one pixel. But in order to get more information about pixel's value variation near a scanned point one can use multi scale representation of immediate neighborhood with the help of wavelet-transformation.

It's known that wavelet analysis presents the original image as two pyramids. These are approximations and details representing the local changes of an image in different scales. Hence it follows that using wavelet-details it is possible to select the information about the behavior of an image function in the neighborhood of the explored point in different scales (frequencies) instead of selecting it only on the highest level as in differential operators. At the same time the method of filtering with a floating mask should be used, because it calculates wavelet-details of neighborhood for every point of an image.

So, any pixel can be connected with a neighborhood on the original image (fig. 1), which is used to calculate coefficients of multi scale analysis.

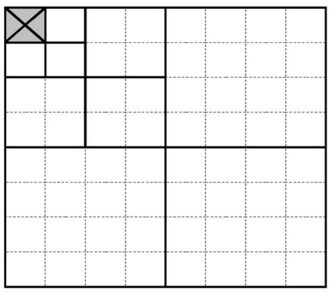


Fig. 1

Then for a chosen point of an image one can calculate several values of specification (in fig. 1 there are three of them). Their number depends on the level of transformation. For a deeper analysis level a bigger neighborhood is used. But as it's seen from the example, this neighborhood covers only one square around the pixel. The rest of the squares will be covered by reflection of the transformation template. As a result we'll get a quantitative measure of variability for an image function on a definite scale (let us call it wavelet-statistics) as a sum $W^l_{x,y} = V^{I,l}_{x,y} + V^{II,l}_{x,y} + V^{III,l}_{x,y} + V^{IV,l}_{x,y}$, where l is the level of analysis; $V^{I,l}_{x,y}$, $V^{II,l}_{x,y}$, $V^{III,l}_{x,y}$, $V^{IV,l}_{x,y}$ are components of wavelet-statistics which are calculated if four quadrants around the basic point. These summands differ by the vector of their calculation on the plane of an image. To define them let's introduce an auxiliary vector $\overline{mw}^{Q,l}_{x,y}$ of the average value on the l-level for a square of pixels with side 2^l. This vector is calculated with the help of recursive formulas:

$$\overline{mw}^{IV,l}_{x,y} = \overline{M}\left(\left\{\overline{mw}^{IV,l-1}_{x,y},\ \overline{mw}^{IV,l-1}_{x+2^{l-1},y},\ \overline{mw}^{IV,l-1}_{x,y+2^{l-1}},\ \overline{mw}^{IV,l-1}_{x+2^{l-1},y+2^{l-1}}\right\}\right),$$

$$\overline{mw}^{III,l}_{x,y} = \overline{M}\left(\left\{\overline{mw}^{III,l-1}_{x,y},\ \overline{mw}^{III,l-1}_{x+2^{l-1},y},\ \overline{mw}^{III,l-1}_{x,y-2^{l-1}},\ \overline{mw}^{III,l-1}_{x+2^{l-1},y-2^{l-1}}\right\}\right),$$

$$\overline{mw}^{II,l}_{x,y} = \overline{M}\left(\left\{\overline{mw}^{II,l-1}_{x,y},\ \overline{mw}^{II,l-1}_{x-2^{l-1},y},\ \overline{mw}^{II,l-1}_{x,y-2^{l-1}},\ \overline{mw}^{II,l-1}_{x-2^{l-1},y-2^{l-1}}\right\}\right),$$

$$\overline{mw}^{I,l}_{x,y} = \overline{M}\left(\left\{\overline{mw}^{I,l-1}_{x,y},\ \overline{mw}^{I,l-1}_{x-2^{l-1},y},\ \overline{mw}^{I,l-1}_{x,y+2^{l-1}},\ \overline{mw}^{I,l-1}_{x-2^{l-1},y+2^{l-1}}\right\}\right),$$

$$\overline{mw}^{Q,0}_{x,y} = \overline{p}_{x,y},\quad Q = \{I, II, III, IV\},$$

where $\overline{M}$ is an operator for calculating the middle point in three dimensional RGB-area in the set $P=\{\overline{p}_1, \overline{p}_2, \ldots \overline{p}_N\}$.

Now with the help of these formulas we can describe the calculation of wavelet-statistic components for l scale as

$$V^{IV,l}_{x,y} = D\left(\overline{mw}^{IV,l}_{x,y},\ \overline{mw}^{IV,l-1}_{x+2^{l-1},y}\right) + D\left(\overline{mw}^{IV,l}_{x,y},\ \overline{mw}^{IV,l-1}_{x,y+2^{l-1}}\right) + D\left(\overline{mw}^{IV,l}_{x,y},\ \overline{mw}^{IV,l-1}_{x+2^{l-1},y+2^{l-1}}\right),$$

$$V^{III,l}_{x,y} = D\left(\overline{mw}^{III,l}_{x,y},\ \overline{mw}^{III,l-1}_{x+2^{l-1},y}\right) + D\left(\overline{mw}^{III,l}_{x,y},\ \overline{mw}^{III,l-1}_{x,y-2^{l-1}}\right) + D\left(\overline{mw}^{IV,l}_{x,y},\ \overline{mw}^{IV,l-1}_{x+2^{l-1},y-2^{l-1}}\right),$$

$$V^{II,l}_{x,y} = D\left(\overline{mw}^{II,l}_{x,y},\ \overline{mw}^{II,l-1}_{x-2^{l-1},y}\right) + D\left(\overline{mw}^{II,l}_{x,y},\ \overline{mw}^{II,l-1}_{x,y-2^{l-1}}\right) + D\left(\overline{mw}^{II,l}_{x,y},\ \overline{mw}^{II,l-1}_{x-2^{l-1},y-2^{l-1}}\right),$$

$$V^{I,l}_{x,y} = D\left(\overline{mw}^{I,l}_{x,y},\ \overline{mw}^{I,l-1}_{x-2^{l-1},y}\right) + D\left(\overline{mw}^{I,l}_{x,y},\ \overline{mw}^{I,l-1}_{x,y+2^{l-1}}\right) + D\left(\overline{mw}^{I,l}_{x,y},\ \overline{mw}^{I,l-1}_{x-2^{l-1},y+2^{l-1}}\right),$$

where D is an operator for calculating Euclidean distance in RGB-area.

As a result after the calculation we'll get an array of numbers where a sequence of values $W^l_{x,y}$ reflecting the variation degree of an original function for different scales l in the nearest neighborhood the size of which is not constant and depend on the analysis level will correspond to every point $\overline{p}_{x,y}$. By adding all numbers from the sequence one can get a single value and then count a two-dimensional function of total wavelet-statistics:

$$UW^Z(x, y) = \sum_{l=1}^{Z} W^l_{x,y},\ Z \le \lfloor \log_2(\min\{X, Y\}) \rfloor,$$

where Z – is the depth of wavelet-statistics calculation; X, Y – are the total number of lines and columns on the original image. When the function is ready, one should start defining the centers of crystallization by way of searching for local minimums on it.

The use of common methods of optimization for a function with several variables is inefficacious because the studied function $UW^Z(x, y)$ is determined on a two-dimensional discrete space. Here one can use the method of suppression of non-maximums [2] as applied to the problem of minimization. It means scanning the image with the help of mask 3 × 3 in order to find points surrounded by pixels with biggest values. But in this case the points of inexact local minimums will be lost. And on the other hand according to the requirements imposed for the problem of crystallization centers searching in the area of equal minimums there should be one minimal point situated in the center of this area. So when the point is found one should start morphological filling of the connected area of the same level in order to define and save the middle point of the selected component as a center of crystallization. Usually morphological operations are described with the help of theory of sets [1]. In morphology dilatation ⊕ is used to determine the component of connection. Dilatation means widening the set on its borders. So, if point c is inside the researched area, the operation of defining the component of connection C_k for a set of values of function $UW^Z(x, y)$ looks like $C_k = (C_{k-1} \oplus B) \cap A$; $k = 0, 1, \ldots$; $A = \{UW^Z(x, y) \mid (UW^Z(x, y) = c)\}$; $C_0 = c$; where B is a dilatation primitive, in this case it is a square 3 × 3.

The filling is finished when $C_k = C_{k-1}$. Elements from the set C_k represent the image of the connected area of equal values for which it is necessary to determine a center of gravity. A center of gravity for the area with equal values can be determined as a point inside a rectangle that limits the area. Horizontal and vertical lines cross in this point in such a way that on the both sides they have the equal subsets of the studied area C_k. Then one should find the point closest to the center of gravity (because an area can be a ring, for example) belonging to C_k and define it as s new center of growth.

On determining the centers of crystallization one should start growing areas. In order to get the full and correct segmentation one should develop an integrated approach instead of using the described method many times for each segment. It means that the algorithm on each step should consider all the growing segments of the image, that is, it must be connected. Then, probably, it is worth choosing the full filling of the image by the segment as the best criterion to stop segmentation. But instead of adding a pixel to each area on each step one should add the only pixel closest to one of areas at the moment. By this the irregular segments growing can be reached. It corresponds to the real situation and thus we can solve the problem of choice of the criterion of the growing process stopping.

Of course, all the candidate-pixels, that can be included in a segment, should be adjacent to their areas. It means they should be situated on an outer border, which can be easily calculated with the help of morphological methods. The outer border *Gr* of area *Reg* (*Gr* and *Reg* are sets of pixels) can be determined as $Gr = (Reg \oplus B) \setminus Reg$.

Then one should search for only one pixel among the chosen boundary pixels. This pixel should be the closest to the neighboring area according to some feature. The closeness between a pixel and a chosen area can be determined by two methods:

– first one is the distance α between the value of a pixel's color and the average value of an area, $\alpha_{x,y} = D(\bar{g}_{x,y}, \bar{M}(Reg))$, $\bar{g}_{x,y} = \bar{p}_{x,y}$, $\bar{g}_{x,y} \in Gr$;

– second is the distance β in *Z*-dimensional area of wavelet-statistics $W^l_{x,y}$, $l = 1, 2, \ldots Z$ between vector $\bar{w}^Z_{Gr,x,y}$, that belongs to a boundary pixel and the average value of a corresponding area. Let's extend the operations of finding a mean and Euclidean distance in case of the *Z*-dimensional area :

$$\beta_{x,y} = D(\bar{w}^Z_{Gr,x,y}, \bar{M}(\{\bar{w}^Z_{Reg,x,y} \mid \bar{p}_{x,y} \in Reg\})),$$
$$\bar{w}^Z_{x,y} = [W^1_{x,y}, W^1_{x,y}, \ldots W^Z_{x,y}].$$

So, as a result of the analysis of the areas and adjoining pixels we'll get such point $\bar{g}_{x,y}$, that distance α or distance β for it will be minimal among all the pixels from the set *G*. This point should be united with the appropriate segment. The process of area growing stops when all the pixels of the image will be included into a set of pixels of all areas.

Probably in general case the described method will not let us have the correct segmentation, because there will be more centers of crystallization in the gradient image of wavelet statistics and, consequently, more segments than areas which a man can perceive. It will inevitably lead to excessive segmentation. This problem can be solved with the help of areas union on the segmented image. Let's use the same method as in area growing. That means we'll search for two closest areas on each step. Any two neighboring areas can be joined. Distance (α or β) be can be calculated with the help of the same equations as in growing. The only difference is that in both cases one should use average values of the areas instead of a parameter value for one pixel. Those two areas which will have the smallest distance d^k on *k*-step of union should be united. This process should be repeated until there is only one area left and it is equal to the whole image. But then there comes a question when to stop, which criterion should be used. When will the segmented image best correspond to a man's perception or give enough information that makes further distinction easier? One can figure out this moment while analyzing a sequence of allowed deviation of areas characteristics during their uniting. One can suggest that this function will steadily grow as in the process of areas uniting one has to allow greater and greater deviations of their parameters. In this case by differentiating this sequence we can define the moment after which in uniting the next pair one had to make an admission greatly differing from neighboring ones. In other words it is necessary to calculate the global maximum of the first derivative of the sequence of variations d^k. Then the moment *t* when an uniting should be stopped can be calculated as:

$$t = \arg_k(\max\{\Delta d^k, k = 1, 2, \ldots\}), \ \Delta d^k = d^{k+1} - d^k.$$

In the end it is necessary to go back to step t and finish the work.

Let's present the described method of segmentation as an algorithm (fig. 2).

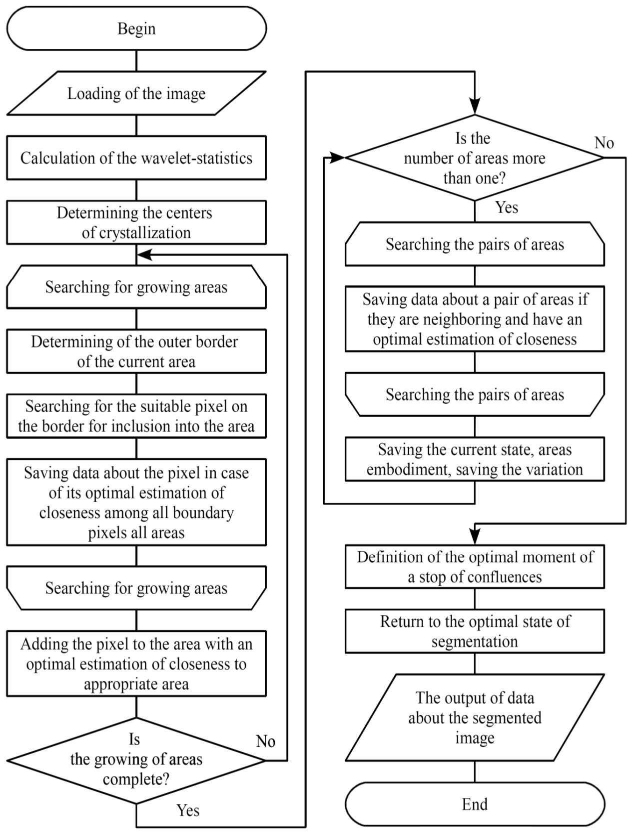


Fig. 2

Let's consider the results of the work of the suggested method for different images. At first let's take a simple two-colored image (fig. 3, left) in order to see how the method of searching for centers of crystallization works. There is also an image for wavelet-statistics (fig. 3, middle) and centers of crystallization (fig. 3, right).

It's obvious from fig. 3 that one center of crystallization (contrast crosses) was chosen for every homogeneous area. Every marked point is in the center of gravity of its area, though the value of wavelet-statistics (gradient) inside the area is zero. Such result was achieved due to the use of the morphologic algorithm of filling along with the minimization of wavelet statistic function $UW^{Z}(x, y)$.

Now let's apply the algorithm to a rather simple from the point of view of areas extraction image of blood cells (fig. 4, left).

The segmented images shown in fig. 4 (center) were got fully automatically. Every segment which is shown in fig. 4 on the right has an individual set of values of descriptors. It goes without saying that the main method settings (wavelet-statistics of the first level, growing and uniting according to color) were introduced. But they do not influence the result as much as, for example, the value of color in threshold transformation or in a simple areas growing. It is sufficient to say that variations of object's colors in fig. 4, left and small distortions will not influence the result of segmentation. Consequently the suggested method of segmentation has a high level of stability and automation.

Now let's check the capacity of the method in working with texture mapping (fig. 5, left). For this image one can successfully apply a deep multi scale analysis (for example, down to the fourth level) and wavelet-statistics in segmentation. In this case frequency information about the image will be used.

In this example (fig. 5, middle) automatic stop of uniting worked incorrectly. But the image that was got on an earlier step of areas uniting (fig. 5, right) shows the capacity of the method to work with texture mapping.

In conclusion let's figure out the progressive features of the suggested method of segmentation and its advantages:

– the method uses information about areas connection;

– it is a complex method; it combines dynamics, adaptivity of area growing and areas uniting with an opportunity to use different information about a studied scene;

– the method uses frequency information which gives an opportunity to work with texture mapping;

– high level of automation.

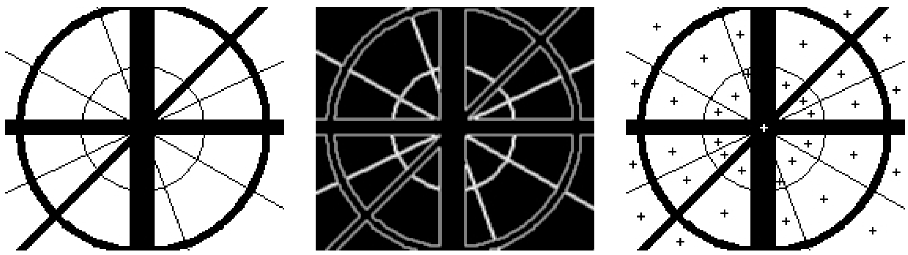


Fig. 3

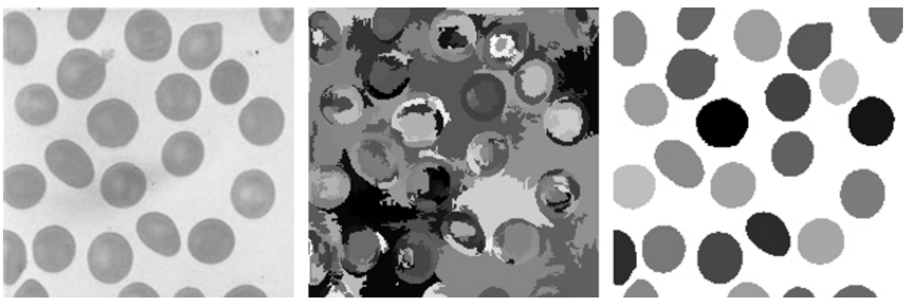


Fig. 4

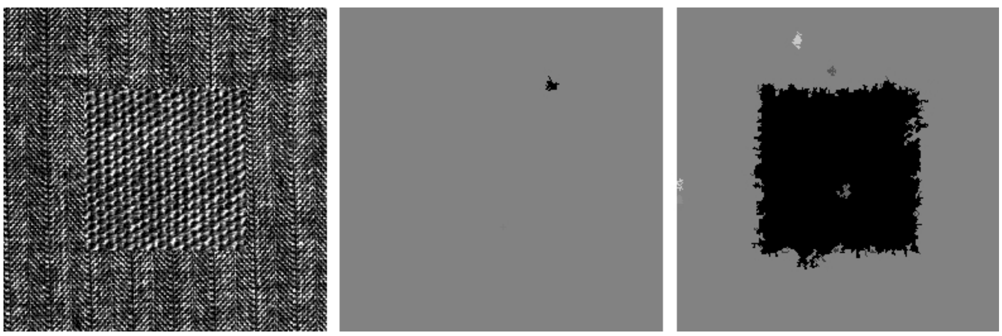


Fig. 5

Disadvantages:

– algorithms are complicated;

– there are high system requirements.

Except for the algorithm of segmentation itself in this work there were described some other new ideas:

– the method of multi scale analysis with extraction of frequency information;

– the algorithm of two-dimensional function minimization which uses morphological filling;

– combination of areas growing and areas uniting;

– the criterion of the definition of the optimal moment to stop uniting.

References

1. Gonzalez R. C., Woods R. E. Digital Image Processing. N. Y. : Prentice Hall, 2002.

2. Steinbrecher R. Bildverarbeitung in der Praxis [Electronic resource]. URL: http://www.rst-software.de/dbv/download.html.

Y. S. Petrov, V. E. Raspopov
Siberian Federal University, Russia, Krasnoyarsk

RESULTS OF COMPUTING EXPERIMENTS FOR WATER ECOLOGICAL SYSTEM MATHEMATICAL MODELING

The point-wise imitation and one-dimensional mathematical models of aquatic ecosystems have been overlooked. The developed models are intended for studying ecosystems in the Krasnoyarsk aquatic basin and in separate locations on the Yenisei River. The results of the computing experiments are presented.

Keywords: mathematical model, mathematical modeling of aquatic ecosystems, computing experiment.

Environmental issues have a designated place in the general list of issued for which mathematical modeling is used. The increase of the anthropogenic environmental impact, caused by intense exploitation of natural resources and the growth of industry leads to an ecological balance disruption. This is happening both on local (in separate areas of globe) and on planetary scale. The importance of struggling against anthropogenic eutrophication of reservoirs and their pollution is understood everywhere in the world. There had been a great amount of researches in limnology, mathematical modeling, and economy, connected with problems of preservation, restoration, and the effective exploitation of natural resources, such as lakes and manmade reservoirs. The ecological condition of the water bodies depends on a number of various factors and processes: hydrophysical, hydrobiological, hydrochemical, meteorological, and anthropogenic. Hydrophysical processes appreciably form a habitat of hydrobionts, define the transferred and sedimentation of substances, the intensity of pollution, and the self-cleaning of reservoirs.

The problem of water quality is complicated. Water bodies are complex physical, biochemical and ecological systems. To be able to predict the consequence of one decision or another, the corresponding tool by dint of which it is possible to analyze the sufficiency of information is required. Such a tool is the computing experiment based on mathematical modeling and numerical methods. An effective means of the arising problem objective analysis in the field of hydrobiology problems are the methods based on constructing and studying mathematical models of water ecosystems. The using of mathematical modeling and carrying out computing experiments allows us to predict the dynamics of water ecosystem development, and also to estimate the consequences of realizing various projects, connected with influence on the ecosystem.

A number of general claims to each mathematical model are known: the corresponding system of the equations should be closed and consistent; the model should describe a variety of physical phenomena and suppose the designing of realized numerical algorithm.

In the given work, some results of the calculations, carried out with a mathematical model of the water ecosystem (being an improvement of the model considered in [1]) are presented. The model is modified by the separation of green algae as independent components of a mathematical model and the introduction of an additional equation, describing the change in algae concentration.

As dynamic variables of model, the concentrations of green algae (CA0), of blue-green algae (CA1), of diatoms (CA2), of zooplankton (CZ), of bacteria (CB), of detritus (CD), of the inorganic phosphorus dissolved in water (PS), of the inorganic nitrogen dissolved in water (NS), of the organic matter dissolved in water (POB), and of the oxygen dissolved in water (O2) are taken.

In model the following processes are considered:

– growth of microorganisms;

– outflow of products of a metabolism;

– death rate of microorganisms;

– processes of settling;

– transitions on a trophic chain;

– decomposition processes;

– atmospheric reaeration (isolation of oxygen from water);

– denitrification (process of restoration of nitrates to the molecular nitrogen, caused by bacteria);

– limiting factors (illumination, temperature);

– water aeration (saturation of water by oxygen of air).

The main feature of the given model is the division of blue-green algae into two species: greens and blue-greens; this isn’t presented in many models, but is of great importance for the research of the reservoirs’ ecology, for their development is various.

The model allows predicting the dynamics of water ecosystem development; including the transformations of nitrogen and phosphorus, as basic biogenous elements, defining the efficiency and water quality in reservoirs.

The structure of model describing the functioning of an ecosystem is given in the flow chart (fig. 1).

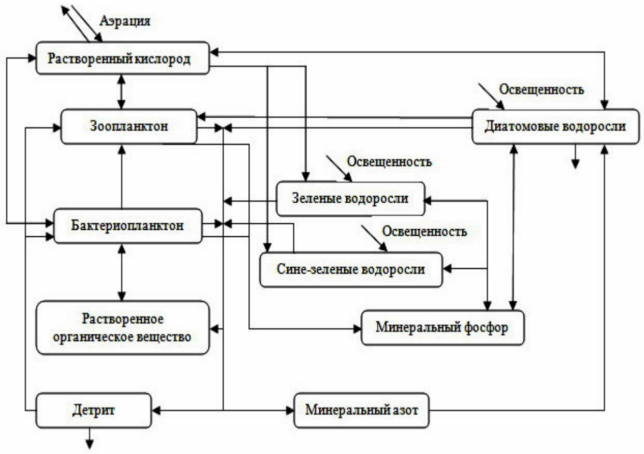


Fig. 1. Flow chart of the model. Arrows designate directions of substance streams between ecosystem components

On the basis of the flow chart the imitating model, describing the considered processes in the ecosystem is constructed. The mathematical model represents the following system of the ordinary differential equations with corresponding initial data:

$$\frac{dCA0}{dt} = (mA0 - RA0 - MA0)\cdot CA0 + \alpha_0 \cdot CA1 \cdot CA0,$$

$$\frac{dCA1}{dt} = (mA1 - RA1 - MA1)\cdot CA1 - \alpha_1 \cdot CA1 \cdot CA0,$$

$$\frac{dCA2}{dt} = (mA2 - RA2 - SA2 - MA2)\cdot CA2 - \frac{mZ \cdot CZ}{Y1},$$

$$\frac{dCZ}{dt} = (mZ - RZ - MZ)\cdot CZ,$$

$$\frac{dCB}{dt} = (mB - RB - MB)\cdot CB - \frac{mZ \cdot CZ}{Y2},$$

$$\frac{dCD}{dt} = MA0 \cdot CA0 + MA1 \cdot CA1 + MA2 \cdot CA2 + MZ \cdot CZ +$$

$$+ MB \cdot CB - SA3 * CD - \frac{mB \cdot CB}{Y3} - \frac{mZ \cdot CZ}{Y4}, \qquad (1)$$

$$\frac{dPS}{dt} = -(mA0 - RA0)\cdot PP0 \cdot CA0 - (mA1 - RA1) \times$$

$$\times PP1 \cdot CA1 - (mA2 - RA2)\cdot PP2 \cdot CA2 +$$

$$+ RZ \cdot CZ \cdot PP3 + RB \cdot CB \cdot PP4,$$

$$\frac{dNS}{dt} = RA0 \cdot PN0 \cdot CA0 + RA1 \cdot PN1 \cdot CA1 - (mA2 - RA2) \times$$

$$\times PN2 \cdot CA2 + RZ \cdot CZ \cdot PN3 + RB \cdot CB \cdot PN4,$$

$$\frac{dPOB}{dt} = -\frac{mB \cdot CB}{Y5} + h0 \cdot RA0 \cdot CA0 + h1 \cdot RA1 \cdot CA1 +$$

$$+ h2 \cdot RA2 \cdot CA2 + h3 \cdot RZ \cdot CZ + h4 \cdot RB \cdot CB,$$

$$\frac{dO2}{dt} = K1 \cdot (O2O - O2) + K_{acc} \cdot (mA0 \cdot CA0 + mA1 \times$$

$$\times CA1 + mA2 \cdot CA2) - alf \cdot (RA1 \cdot CA1 + RA2 \cdot CA2 +$$

$$+ RZ \cdot CZ + RB \cdot CB) - B1 \cdot mZ \cdot CZ.$$

where mAi are functions describing growth; coefficients RAi are breath; MAi are the death rate; SAi is the settling; Yi are proportionality coefficients; T in temperature in C^0; t is time.

In the live description the incoming and proceeding streams’ components are taken into account. Also is included the share of received resources (food) spent for growth and reproduction; proceeding is the consumption of species from given components; predators and death rate depending on every other possible reason. Meanwhile, the influence on the stream speed of the environment (temperature, etc.) is considered.

In microbiological systems as a rule, the growth rate is limited by a concentration of substrates. We have applied the hyperbolic dependence offered by Z mono for the description of the limitation process.

It is supposed, that the growth of green and blue-green algae is limited by phosphorus, while the growth of diatoms – by nitrogen and phosphorus. The growth functions, death rate, illumination, and temperature dependence, as well as all entrance data are included according to researches [1–3].

The constructed mathematical model represents the Cauchy problem for a system of ten ordinary differential equations. For the numerical solution of the Cauchy problem, the Runge–Kutta method of the fourth approximation order is applied:

$$\vec{y}_{n+1} = \vec{y}_n + \frac{1}{6}\tau(\vec{K}_1 + 2\vec{K}_2 + 2\vec{K}_3 + \vec{K}_4),$$

$$\vec{K}_1 = \vec{F}(t_n, \vec{y}_n),$$

$$\vec{K}_2 = \vec{F}(t_n + \frac{\tau}{2}, \vec{y}_n + \tau\frac{\vec{K}_1}{2}),$$

$$\vec{K}_3 = \vec{F}(t_n + \frac{\tau}{2}, \vec{y}_n + \tau\frac{\vec{K}_2}{2}),$$

$$\vec{K}_4 = \vec{F}(t_n + \tau, \vec{y}_n + \tau\vec{K}_3),\ n = 0, 1, \ldots,$$

where $\vec{y}$ is a vector function of unknown; $\vec{F}$ is the right part of system (1); τ is a step in time; $\vec{y}_0$ is specified.

Let’s note that the set of components in the model considerably complicates the problem, both the modeling, and in studying the model; as it is required to specify its value for each coefficient (fig. 2).

A complex of the programs is written, allowing the inputting of entrance data in an interactive mode. The calculation results can be received numerically, presented graphically, and transferred outside for subsequent processing. For the management of graphic representation of calculation results, a corresponding menu is provided.

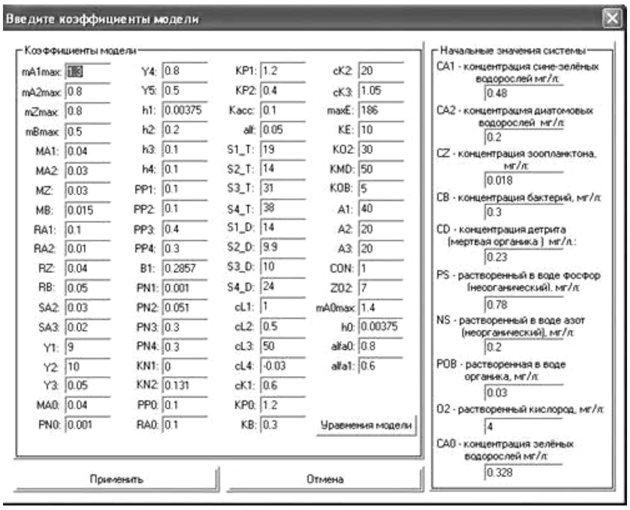


Fig. 2. Coefficients of modeling and initial system values

The program's complex is realized in Visual C ++ 6,0 with the use of MFC (Microsoft Foundation Classes) which is one of the most convenient and powerful tools among Windows' applications. The software product has a friendly interface, it is convenient in work, and is intended not only for mathematicians, but also for researchers who are not experts in programming.

The first numerical experiments have been devoted to the comparative analysis of results, obtained by means of a working mathematical model [1] and by its updating means (1) with same input data [4]. The calculation results have shown that the received concentrations of diatoms, bacteria, and detritus for a working model [1] poorly correspond to experimental data, in comparison with the results obtained from the model aforementioned.

Thus, the computing experiments that have been carried out have shown the effectiveness of separating green algae as an independent variable of the mathematical model for the reservoir ecosystem.

The following calculations with an improved mathematical model are meant for researching general tendencies of seasonal dynamic variable change for a model, using field data from the Novoselovsky reach of the Krasnoyarsk impoundment for 1998–2000. Notice that for the comparison of calculations results, we have used only the field data, the time moments of which are precisely known. During other time periods, due to the incompleteness of existing information, the comparative analysis of average data also shows qualitative calculation coincidence.

Particularly seasonal dynamics of diatoms demonstrate a qualitative picture of two "flowering" peaks: summer – with the maximum biomass of 5.9 mg/l, and autumn with the maximum biomass of 2.27 mg/l; this corresponds with the supervision data [5]. The total biomass of diatoms according to supervisions in July and August decreases to 1.2–2 mg/l. Model calculation has also shown a falling in values of biomass during this period (fig. 3).

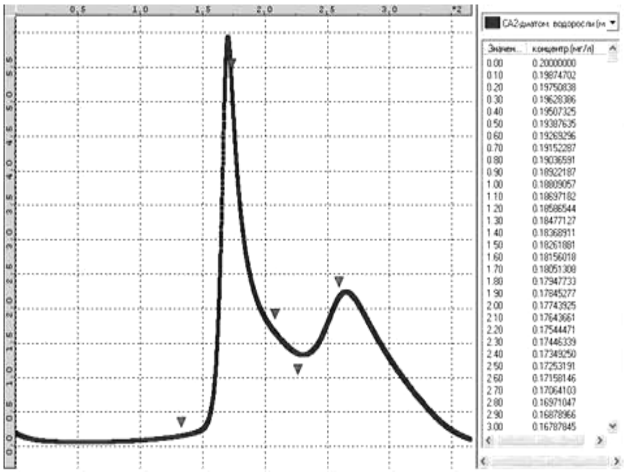


Fig. 3. Comparison of experimental data with numerical results for diatoms. Natural data is marked by triangles

The comparison of a seasonal course model for green and blue-green algae and experimental data [6], has shown that in a general understating of model concentration (approximately by 1.5 times) the relative time course had been precisely reconstructed (fig. 4). The annual course of biomass for zooplankton has a single peak and corresponds to the maximum values for green and blue-green algae biomasses, which also correspond with theoretical representations.

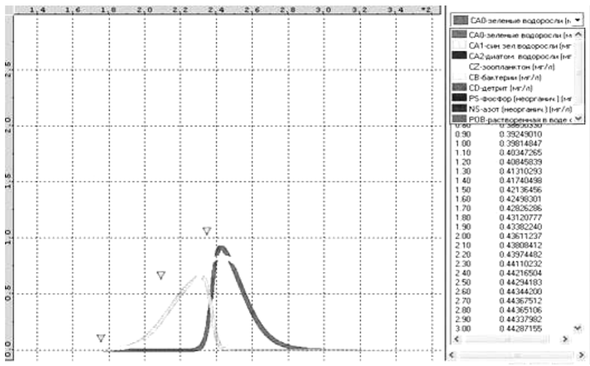


Fig. 4. Comparison of experimental data with numerical results for blue-green algae. Natural data are marked by triangles

For a seasonal course of chemical combinations of nitrogen and phosphorus concentrations, the calculation has shown the falling of values of nitrogen compound concentration during seasons, when intensive plankton growth occurs, and the maximum values at minima of plankton biomass.

The annual course of organic matter concentration has two expressed maxima with values of 0.29 mg/l in the beginning of the summer; and 0.54 mg/l in the autumn. These periods correspond to the maximum values of phytoplankton and zooplankton biomasses. For seasonal dynamics of bacteria and detritus, two peaks of development are typical: the first falls in the middle of the summer, the second – in the beginning of autumn. Thus, detritus influence the growth of bacteria and stimulate their productivity, which also corresponds to natural data.

The obtained calculation model coincides with experimental data, which testifies the adequacy of the examined model.

The one-dimensional model of the aquatic ecosystem. Along with point-wise model the mathematical model allowing mass transfer along the length of a reservoir (one-dimensional model in a horizontal plane) is examined. The mathematical model represents a differential equation system in partial derivatives of the first degree:

$$\frac{\partial \vec{U}}{\partial t} + V\frac{\partial \vec{U}}{\partial x} = \vec{F}(t, x, \vec{U}) \qquad (2)$$

with corresponding initial and edge conditions:

$$\vec{U}(t,0) = \vec{U}_0(t),$$

$$\vec{U}(0,x) = \vec{U}_1(x).$$

where V is the current speed of the reservoir; x is the spatial variable corresponding to the length of a reservoir; t is time; the right part of the equations (2) corresponds to the right part (1). In such modeling it is supposed, that the substance is evenly distributed along the width of the stream and moves with the average speed of the stream. So, the data is averaged for the depth and width of a reservoir. We will notice that the given model is expedient for using in the case when the length of a reservoir is greater than its width.

The introduced mathematical model is also realized numerically by means of an implicit difference scheme:

$$\frac{\vec{y}_j^{\,n+1} - \vec{y}_j^{\,n}}{\tau} + V\frac{\vec{y}_j^{\,n+1} - \vec{y}_{j-1}^{\,n+1}}{h} = \vec{F}_j^{\,n}.$$

For initial data, the data from the point-wise model is taken. For edge conditions – the solutions obtained from the point-wise model are used.

Initial distribution of all system components is considered uniform. The calculation was carried out for time from t = 0 to t = 365 (one year), for the Novoselovsky reach of the Krasnoyarsk impoundment, and also for sites on the Yenisei River downstream from the Krasnoyarsk Hydroelectric Power Plant with average current speed of the Yenisei of 1.2 km/h on a distance from ten to one hundred kilometers. In fig. 5 particularly, the results for bacteria calculations on a river site are given. It is visible, that under the specified conditions a concentration of bacteria in the chosen part of a reservoir changes considerably (we suppose that this difference of values is caused by the current).

Notice that the ecosystem of the Yenisei on the site adjoining the power plant is strongly impoverished because of the destroying action of the plant's turbines and the low water temperature. The self-cleaning process of water in this heavily polluted site is weakened. The zooplankton in process of substance decomposition plays an insignificant role. Bacterial mass cumulates which here, undergoes intense decomposition only at the inlet stream of the Angara.

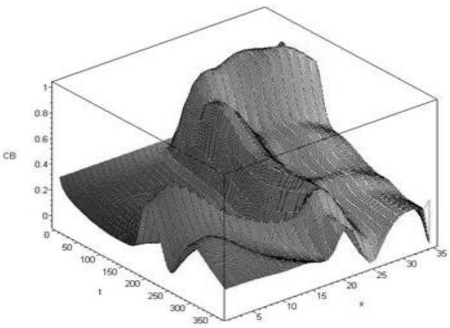


Fig. 5. Results of calculations for the bacteria carried out in a one-dimensional mathematical model

Calculations have shown that the further the distance is from the Krasnoyarsk Hydroelectric Power Plant, the more change occur in concentration, selected as a component; in particular, there is a shift of peaks for diatoms, bacteria, and detritus. At a distance from the power plant of over one hundred kilometers, the model depicts the dynamics of water ecosystem development in space less adequately. This is caused by the studying of quality water characteristics; it is necessary to take in account the more difficult and full processes: hydrodynamic (convective stirring, pressure, wind, and deep currents), heat transfer and illumination. The use of the models based on various variants of mechanic equations for liquid and heat transfer, and corresponding boundary conditions is necessary [7].

Let's note the basic results obtained in the work:

Point-wise and one-dimensional mathematical models of water ecosystems consisting of ten differential equations had been constructed. The use of models allows revealing of processes progressing dynamics in difficult ecological water systems, to predict a system status in time and in space (the one-dimensional model in a horizontal plane) for a distance up to one hundred kilometers, and to analyze problematic situations. The models in particular, make it possible to describe the change of hydrobionts and the basic biogenic elements, and also to reproduce occurrence situations, both for one and for two peaks of phytoplankton flowering during the vegetative season, depending on external conditions. It is necessary to notice, however, that the offered mathematical models are very sensitive to changes in parameters and demand a meticulous selection of coefficients for each specific aquatic ecosystem.

The results obtained by means of the models described above, can as well be used as well for estimating ecological risks.

A complex of computer programs has been produced, and the numerical simulation of some processes progressing in the ecosystem of the Novoselovsky reach of the Krasnoyarsk impoundment and in the Yenisei River had been carried out.

The authors express their gratitude to Professor Z. G. Gold for his useful consultations.

References

1. Mathematical simulation of reaches of the Krasnoyarsk impoundment / V. A. Sapozhnikov [et al.] // Association of subjects of the Russian Federation and a wildlife management problem in Prienisejsky Siberia : theses and materials of reports of inter-regional scientifically-practical conference / KSU. Krasnoyarsk, 2005. P. 296–298.

2. Borodin, A. L., Raspopov V. E. Numerical identification of coefficients of mathematical model of an ecosystem of a reservoir // Joint issue. Computing technologies. T. 13. Herald of the KazSU of AL-FARABI. A series of the mathematics, the mechanics, the informatics. Vol. 3 (58). Almaty–Novosibirsk, 2008. P. 302–306.

3. Gubanov V. G. Biotic circulation and interaction of trophic links in artificial and natural biosystems : diss. dr. phys.-math. sciences. Krasnoyarsk, 2004.

4. Petrov J. S. Special-purpose software for carrying out of computing experiments at mathematical modeling of water ecosystems // YOUTH AND THE SCIENCE: the XXI-st century BEGINNING : Materials of the All-Russia scientific and technical conf. of students, post-graduate students and young scientists. In 4 p. P. 1 / SFU. Krasnoyarsk, 2009. P. 78–80.

5. Kozhevnikova N. A., Phytoplankton of a deep-water part of a Krasnoyarsk impoundment // Alkologia. № 2. 2002. P. 39–40.

6. ShChur L. A. Structure and functional characteristics of bacterial plankton and phytoplankton in ecosystems of reservoirs of different type : diss. dr. biol. sciences. Krasnoyarsk, 2006.

7. Belolipetsky V. M., Genova S. N., Gurevich K. J. Platform for research of dynamics of hydrophysical and radio ecological characteristics of river system // Computing technologies / the Siberian Branch of the Russian Academy of Science. Vol. 6, № 2. 2001. P. 14–24.

V. V. Podkolzin, V. O. Osipyan
Kuban State University, Russia, Krasnodar

ON PROPERTIES OF KNAPSACK SYSTEMS OF INFORMATION PROTECTION WITH THE OPEN KEY IN Z_p

Properties of sequences of numbers expressed through components of a knapsack vector are investigated. Properties of isomorphic and similar knapsack systems of information protection are analyzed. Methods of increasing cryptographic security of knapsack systems of information protection with an open key are presented.

Keywords: a knapsack vector, isomorphism, cryptoanalysis, density, injectivity.

Let's express a set of natural numbers {0, 1, …, p–1} through Z_p and a set of all numerical sets of length n with components from Z_p. through Z_p^n .

A knapsack problem for set $w \in N$ and vector $A = (a_1, a_2, ..., a_n)$, where $a_i \in N$, $I = 1…n$, has the solution in Z_p if there is an equation solution

$$Ax^T = w, x \in Z_p^n \quad (1)$$

we will call vector A of equations (1) a knapsack vector.

A knapsack vector $A = (a_1, a_2, ..., a_n)$ is an injective one if for any natural w the equation (1) has not more than one solution. A knapsack vector which has two elements $a_i = a_j$, $I \neq j$, is not injective. Injectivity of a knapsack vector allows to speak about uniqueness of restoration of the original text according to the cryptogram. Supergrowing knapsack vectors are the simplest injective knapsack vectors from the point of view of understanding and algorithmization. For their components in Z_p the following relationships are carried out:

$$a_j > \sum_{i=1}^{j-1}(p-1)a_i, j = 2...n \quad (2)$$

A knapsack vector $A = (a_1, a_2, ..., a_n)$ is a nondecreasing one if its components are ordered according to the rule $a_{i-1} \le a_i$, $I = 2…n$. Accordingly, the vector is increasing if its components are ordered according to the rule $a_{i-1} < a_i$, $I = 2…n$.

Definition. Let's call vector $\Delta A = (\delta_1, \delta_2, ..., \delta_n)$ a variation of vector $A = (a_1, a_2, ..., a_n)$ ($a_i \in N$, $I = 1…n$) in Z_p, For its components the following correlations are carried out:

$$\delta_1 = a_1,\ \delta_j = a_j - \sum_{i=1}^{j-1}(p-1)a_i, j = 2...n. \quad (3)$$

On the basis of vector ΔA it is possible to define a knapsack vector A in Z_p corresponding to it:

$$a_1 = \delta_1,$$
$$a_i = \delta_i + (p-1)\sum_{j=1}^{i-1} a_j = \delta_i + (p-1)\sum_{j=1}^{i-1} p^{i-j-1}\delta_j,$$
$$I = 2...n. \quad (4)$$

Let's express a set of various values w for which equation (1) has the solution through $\mu(p, A)$. Capacity $\mu(p, A)$ does not exceed p^n since the quantity of various vectors in Z_p^n is equal to p^n. Value | μ (p, A) | reaches the upper boundary, if

$$\forall x_1, x_2 \in Z_p^n \; x_1 \neq x_2 \Rightarrow Ax_1^T \neq Ax_2^T. \quad (5)$$

Thus, capacity μ (p, A) reaches the upper boundary only when vector A is injective. Really, if vector A is injective, then correlations (5) are carried out and the number of various values $Ax^T(x\in Z_p^n)$ is equal to the number of various elements in Z_p^n, i. e. p^n. On the other hand, if $|\mu(p, A)| = p^n$, then there is a one-to-one dependentanizer between elements μ (p, A) and Z_p^n, and hence equation (1) for any $w \in \mu(p, A)$ has only one solution. From the latter follows an injectivity of knapsack vector A.

Definition. Let's call the value

$$d_p(A) = \frac{|\mu(p, A)|}{\sum_{i=1}^{n}(p-1)a_i} \quad (6)$$

density of a knapsack vector A in Z_p.

The density defines the relation of capacityμ (p, A) to the length of a cut $[0, \sum_{i=1}^{n}(p-1)a_i]$. It is obvious that $\forall\; x\in Z_p^n$ is a value $Ax^T \in [0, \sum_{i=1}^{n}(p-1)a_i]$. Thus, $0 < d_p(A) \leq 1$. Moreover for injective knapsack vectors the density is equal to 1 only when all components of a variation of vector A are equal to unit [1], and cryptoanalysis of such knapsack systems consists in finding p.

$W_x = Ax^T$, $w_x \in \mu(p, A)$ corresponds to each set $x = (\alpha_1, \alpha_2, ..., \alpha_n) \in Z_p^n$. We will write out the sequence $W_{\mu(p, A)} = (w_0, w_1, w_2, ..., w_k)$, where $w_i = Ax_i^T$, $x_i = (\alpha_1, \alpha_2, ..., \alpha_n)$, $i = \sum_{i=1}^{n}\alpha_i p^{n-i}$, $I = 1 \dots k$, $k = p^n-1$. If vector A is not injective in $W_{\mu(p, A)}$ there are two values $w_i = w_j$, $I \neq j$. We will designate sequence $\Delta W_{\mu(p, A)} = (m_1, m_2, ..., m_k)$, where $m_i = w_i - w_{i-1}$ ($I = 1\dots p^n-1$).

The sequence $\Delta W_{\mu(p, A)}$ is symmetric with respect to the middle and can be defined recursively relative to the dimension of a knapsack vector A.

Let $A_n = (a_1, a_2, ..., a_n)$ ($a_i \in N$, $I = 1\dots n$) be a knapsack vector. Vector $A_{n+1} = (a_1, a_2, ..., a_n, a_{n+1})$ is received from A_n by adding the component $a_{n+1}\in N$. Then

$$\Delta W_{\mu(p, A_{n+1})} = (\Delta W_{\mu(p, A_n)}, \delta_{n+1}, \Delta W_{\mu(p, A_n)}, \delta_{n+1}, \Delta W_{\mu(p, A_n)}, \dots, \delta_{n+1}, \Delta W_{\mu(p, A_n)}),$$

where δ_{n+1}, $\Delta W_{\mu(p, A_n)}$ is repeated p–1 times.

The sequence $\Delta W_{\mu(p, A)}$ describes distances between the elements of sequence $W_{\mu(p, A)}$, i. e. its "sparseness", and, hence, is the characteristic of μ (p, A).

From symmetry $\Delta W_{\mu(p, A)}$ it follows that any $w\in W_{\mu(p, A)}$ can be presented in two ways:

$$w = \sum_{j=1}^{n}\alpha_j a_j = \sum_{k=1}^{n}(p-1)a_k - \sum_{i=1}^{n}\beta_i a_i, \quad (7)$$

where $\alpha_i, \beta_i \in Z_p$, $I = 1\dots n$.

Lemma 1. $A_n = (a_1, a_2, ..., a_n)$ is an injective knapsack vector, where $a_i \in N$, $I = 1\dots n$. A vector $A_{n+1} = (a_1, a_2, ..., a_n, a_{n+1})$ is received from A_n by adding the component $a_{n+1}\in N$, $\Delta A_{n+1} = (\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})$ is a variation of vector A_{n+1} and $\delta_{n+1} > 0$. Then $A_{n+1} = (a_1, a_2, ..., a_n, a_{n+1})$ is an injective knapsack vector.

The proof.

Let's show that $\forall\; w_x \in\mu(p, A_{n+1})$ equation (1) has only one solution.

As w_x belongs to set $\mu(p, A_{n+1})$ it follows that $\exists x = (\alpha_1, \alpha_2, \dots, \alpha_n, \alpha_{n+1}) \in Z_p^{n+1}$ for which $w_x = A_{n+1}x^T$ is carried out.

1. If $\alpha_{n+1} = 0$, then $w_x \in\mu(p, A_n)$ and (1) has the only solution because of injectivity of A_n;

2. Let $0 < \alpha_{n+1} < p$. As $\delta_{n+1} > 0$ then any element $\mu(p, A_n)$ is less than a_{n+1}. Thus, if there is unique α_{n+1} and $w'_x \in \mu(p, A_n)$ then $w_x = \alpha_{n+1}a_{n+1} + w'_x$ and consequently equation (1) has the only solution.

From randomness $w_x \in\mu(p, A_{n+1})$ it follows that A_{n+1} is an injective knapsack vector.

Lemma 2. $A_n = (a_1, a_2, ..., a_n)$ is an injective increasing knapsack vector, where $a_i\in N$, $I = 1\dots n$. Vector $A_{n+1} = (a_1, a_2, ..., a_n, a_{n+1})$ is received from A_n by adding the component $a_{n+1}\in N$, $\Delta A_{n+1} = (\delta_1, \delta_2, ..., \delta_n, \delta_{n+1})$ is a variation of vector A_{n+1} and $\delta_{n+1} < 0$.

Vector $A_{n+1} = (a_1, a_2, ..., a_n, a_{n+1})$ is an injective increasing knapsack vector if the following equation is carried out:

$$(a_n - \sum_{j=1}^{n}(p-1)a_j < \delta_{n+1}) \;\&\; (|\delta_{n+1}| \notin W_{\mu(2p-1, A_n)}).$$

The proof.

First of all we will define a condition at which A_{n+1} will be increasing. Since A_n is an increasing vector, it is necessary to follow the condition

$$a_n < a_{n+1} = \sum_{j=1}^{n}(p-1)a_j + \delta_{n+1}.$$

Hence

$$a_n - \sum_{j=1}^{n}(p-1)a_j < \delta_{n+1}.$$

Let $A_{n+1} = (a_1, a_2, ..., a_n, a_{n+1})$ be increasing, but not injective, i. e. let there exist $\omega_x \in\mu(p, A_{n+1})$, then the equation (1) does not have only one solution. From the injectivity of A_n and properties of sequences $W_{\mu(p, A_n)}$ and $W_{\mu(p, A_{n+1})}$ it follows that all such ω_x belong to cuts $[a_{n+1} + k\,a_{n+1}, \sum_{j=1}^{n}(p-1)a_j + k\,a_{n+1}]$, where $k = 0... p$–2.

Also, if

$$a_{n+1} = \sum_{j=1}^{n}(p-1)a_j + \delta_{n+1} \le \omega_x \le \sum_{j=1}^{n}(p-1)a_j \quad (8)$$

and equation (1) has more than one solution for ω_x, then the equation (1) also has more than one solution for $\omega_x + k\, a_{n+1}$, where $k = 0\ldots p–2$, and on the contrary.

On the basis of the above-stated information we will consider ω_x satisfying (8), then $\omega_x \in \mu\,(p, A_n)$ and $\omega_x \in \mu\,(p, A_{n+1})$.

As ω_x belongs to set $\mu\,(p, A_{n+1})$ we have:

$$\omega_x = a_{n+1} + \sum_{j=1}^{n}\beta_j a_j = \left(\sum_{k=1}^{n}(p-1)a_k + \delta_{n+1}\right) + \sum_{j=1}^{n}\beta_j a_j,$$

where $\beta_i \in Z_p$, $I = 1\ldots n$, $0 < \alpha < p–1$.

As ω_x belongs to set $\mu\,(p, A_n)$ and validity (7) we have:

$$\omega_x = \sum_{j=1}^{n}\gamma_j a_j = \sum_{k=1}^{n}(p-1)a_k - \sum_{j=1}^{n}\varphi_j a_j,$$

where γ_i, $\varphi_i \in Z_p$, $I = 1\ldots n$.

Thus, there is an equality:

$$\sum_{k=1}^{n}(p-1)a_k - \sum_{j=1}^{n}\varphi_j a_j = \sum_{k=1}^{n}(p-1)a_k + \delta_{n+1} +$$

$$+ \sum_{j=1}^{n}\beta_j a_j - \delta_{n+1} = \sum_{j=1}^{n}(\beta_j + \varphi_j)a_j.$$

From the latter equality it follows that $–\delta_{n+1} \in W_{\mu\,(2p–1,\, An)}$. Hence, for injectivity of vector A_{n+1}, $|\delta_{n+1}| \notin W_{\mu\,(2p-1,\, An)}$ is necessary.

Then we we will define an addition operation $\oplus$ on set $\mu\,(p, A)$ of knapsack vector $A = (a_1, a_2, ..., a_n)$ as follows:

$$\forall w_1, w_2 \in \mu\,(p,)\; w = w_1 \oplus w_2 =$$
$$= \sum_{i=1}^{n}\alpha_i a_i \oplus \sum_{i=1}^{n}\beta_i a_i = \sum_{i=1}^{n}\gamma_i a_i, \quad (9)$$

where $\gamma_i = (\alpha_i + \beta_i)\, mod\, p$; α_i, $\beta_i \in Z_p$, $I = 1\ldots n$.

The set $\mu(p, A)$ with an addition operation $\oplus$ forms an additive finite Abelian group $(\mu(p, A), \oplus)$.

Definition. Two knapsack vectors $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_k)$, whose variation vectors ΔA and ΔB differ only in the value of the first component are isomorphic ones. We will denote them as $A \approx B$ if there is an isomorphism $f{:}\ \mu(p, A) \to \mu(p, B)$.

Two knapsack vectors can be isomorphic only when they have identical dimension and $|\mu\,(p, A)| = |\mu\,(p, B)|$.

Let's consider two isomorphic knapsack vectors $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_k)$. From (4) we have:

$$a_1 = \delta_1,\; a_i = \delta_i + (p-1)\sum_{j=1}^{i-1}p^{i-j-1}\delta_j,$$

$$b_1 = \delta'_1,\; b_i = \delta_i + (p-1)\left(p^{i-2}\delta'_1 + \sum_{j=2}^{i-1}p^{i-j-1}\delta_j\right),\; I = 2\ldots n.$$

Let's call value $\varepsilon\,(A, B) = \delta'_1 - \delta_1$ a coefficient of isomorphism of two vectors A and B.

Then

$$b_1 = \delta_1 + \varepsilon,\; b_i = \delta_i + (p-1)\left(p^{i-2}\varepsilon + \sum_{j=1}^{i-1}p^{i-j-1}\delta_j\right),$$

$$b_1 = a_1 + \varepsilon,\quad b_i = a_i + (p-1)\,p^{i-2}\,\varepsilon,$$
$$I = 2\ldots n,\; \varepsilon = \varepsilon\,(A, B). \quad (10)$$

And the following correlation is valid :

$$\sum_{i=1}^{j-1}(p-1)b_i = (p-1)(a_1+\varepsilon) + \sum_{i=2}^{j-1}(p-1)(a_i + (p-1)p^{i-2}\varepsilon) =$$
$$= \sum_{i=1}^{j-1}(p-1)a_i + (p-1)\varepsilon(1 + \sum_{i=2}^{j-1}p^{i-2}) =$$
$$= \sum_{i=1}^{j-1}(p-1)a_i + (p-1)\varepsilon p^{j-2}. \quad (11)$$

On the basis of properties of sequences $W_{\mu\,(p,\,A)}$ and $W_{\mu\,(p,\,B)}$ it is possible to draw a conclusion that $W_{\mu\,(p,\,B)}$ is received from $W_{\mu\,(p,\,A)}$ by “recursive scaling” on ε relative to nodal values $(a_2, ..., a_n)$, and each value a_i is displaced according to (10). Sequence $\Delta W_{\mu\,(p,\,B)}$ is received from $\Delta W_{\mu\,(p,\,A)}$ by replacement of all occurrences δ_1 on $\delta_1 + \varepsilon$.

If for knapsack vectors $A = (a_1, a_2, ..., a_n)$, $B = (b_1, b_2, ..., b_n)$ and $C = (c_1, c_2, ..., c_n)$ $A \approx B$ and $B \approx C$ are carried out then $A \approx C$. Really, due to bijectivity $f{:}\mu\,(p, A) \to \mu\,(p, B)$ and $g{:}\mu\,(p, B) \to \mu\,(p, C)$ it follows that $h = g\,f$: $\mu\,(p, A) \to \mu\,(p, C)$ is bijective and $\varepsilon\,(A, C) = \varepsilon\,(A, B) + \varepsilon\,(B, C)$.

Isomorphism of knapsack vectors is an equivalence relation, and, hence, a set of isomorphic vectors forms an equivalence class. In each class there is a vector for which the coefficient of isomorphism with any other vector of this class is non-negative. Let's call such a vector a base vector of an equivalence class.

Let $\Theta = (\theta_1, \theta_2, \ldots, \theta_n)$ be a base vector of some equivalence class and $A = (a_1, a_2, ..., a_n)$ be an arbitrary element of the same class, i. e. $\Theta \approx A$, $\varepsilon\,(\Theta, A) > 0$. As $|\mu\,(p, A)| = |\mu\,(p, \Theta)|$ from density definition of a knapsack vector in Z_p we have:

$$|\mu\,(p,)| = d_p(A)\sum_{i=1}^{n}(p-1)a_i =$$
$$= d_p(\Theta)\sum_{i=1}^{n}(p-1)\theta_i = |\mu\,(p, \Theta)|.$$

Owing to (11) it follows that:

$$d_p(A)\sum_{i=1}^{n}(p-1)a_i = d_p(A)\left(\sum_{i=1}^{n}(p-1)\theta_i + \varepsilon(p-1)p^{n-2}\right) =$$
$$= d_p(\Theta)\sum_{i=1}^{n}(p-1)\theta_i.$$

From the latter we will express $d_p(\Theta)$:

$$d_p(\Theta) = d_p(A)\left(1 + \frac{\varepsilon p^{n-2}}{\sum_{i=1}^{n}\theta_i}\right),\ \text{where } \varepsilon = \varepsilon\,(\Theta, A).$$

$$d_p(\Theta) = d_p(A)(1+k\,\varepsilon(\Theta, A)),$$

where

$$k = \frac{p^{n-2}}{\sum_{i=1}^{n} \theta_i} = \text{cont.} \quad (12)$$

Thus, the basic vector has the greatest density among all vectors of its equivalence class.

In case if the basic vector Θ is supergrowing then vector A is also supergrowing. Really from (2) and (10) we have:

$$\sum_{i=1}^{j-1}(p-1)a_i = (p-1)(\theta_1+\varepsilon) + \sum_{i=2}^{j-1}(p-1)(\theta_i + (p-1)p^{i-2}\varepsilon) =$$

$$= \sum_{i=1}^{j-1}(p-1)\theta_i + (p-1)\varepsilon(1+\sum_{i=2}^{j-1}p^{i-2}) <$$

$$< \theta_j + (p-1)\varepsilon p^{j-2} = a_j,\ \varepsilon = \varepsilon(\Theta, A).$$

From the latter inequality it follows that for any equivalence class with a basic supergrowing vector there is a knapsack vector from the given class for any positive coefficient of isomorphism. Generally the given statement is not true. For example, for an injective vector (15, 42, 51, 83) there is no isomorphic vector in Z_2 with an isomorphism coefficient equal to 10 since vector (25, 52, 71, 123) is not injective.

Thus, KSPI with knapsack vector A is possible to transform into equivalent KSPI with a knapsack vector Θ, where Θ is a basic vector of an equivalence class of vector A. The expediency of the given transformation is caused by smaller volume of calculations μ (p, Θ) and memory expenses. For example, to store each element μ (2, A) of supergrowing knapsack vector A = (45, 69, 218, 415, 796, 1752, 3588, 7375, 17897, 36073) 17 bits of memory are necessary, and to store corresponding values of a basic vector Θ = (1, 25, 130, 239, 444, 1048, 2180, 4559, 12265, 24809) 16 bits for each are enough. If values of a knapsack vector components are great and if there is corresponding dimension then the memory capacity necessary to store elements μ (p, A) can exceed the sizes of standard types of programming languages and consequently will demand additional procedures for storage and performance of operations with such "big" numbers which, naturally, causes the increase in time and memory expenses. In particular for the above-stated example to store values μ (2, B) of supergrowing vector B = (444444444, 444444468, 888889016, 1777778011, 3555555988, 7111112136, 14222224356, 28444448911, 56888900969, 11377780227) belonging to the same class of equivalence already 38 bits are necessary for each.

Theorem. Let $A = (a_1, a_2, ..., a_n)$ be an injective knapsack vector with dimension n and $t \neq 0$ be an integer value. Then, an injective knapsack vector with dimension n by means of whose components in Z_p all elements of a set are expressed $\{w + t | w \in \mu(p,)\}$ does not exist.

The proof.

Let's assume that an injective knapsack vector $B = (b_1, b_2, ..., b_n)$ exists. Then $\{w+t | w \in \mu(p,A)\} \subseteq \mu(p, B)$.

1. $t > 0$. Then $|\mu(p, B)| \geq |\mu(p, A)| + 1$ since zero is included in μ (p, B), but is not included in $\{w + t | w \in \mu(p,)\}$. But due to injectivity of vectors A and B $|\mu(p, B)| = |\mu(p, A)|$ is carried out. As we can see there is contradiction.

2. $t < 0$. Since $0 \in \mu(p,A)$, $t \in \mu(p, B)$ that contradicts $b_i \in N$, i=1, ..., n.

Thus, updating of KSPI by way of changing the numerical value of a crypto text leads to increase in the complexity of its crypto analysis.

Definition. Two knapsack vectors $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_n)$ are similar, we will denote them A≅B only when there is a mutually single-valued transformation $f: A \to B$ such that:

– $\forall a \in A\ f(Ca) = Cf(a)$, where $C \in Z$;

– $\forall a', a'' \in A$, $f(a' + a'') = f(a') + f(a'')$ is carried out.

Two vectors one of which is received from another by strong modular multiplication can serve as an example of two similar injective knapsack vectors.

Let us investigate the properties of two similar injective knapsack vectors $A = (a_1, a_2, ..., a_n)$ and $B = (b_1, b_2, ..., b_n)$ the transformation of which is defined by function $f(x) = cx$ in some field where c is some constant:

$$f(a_i) = ca_i = b_i,\ I = 1\ldots n,$$

$$\forall w_a \in \mu(p,)\ f(w_a) = f(\sum_{i=1}^{n}\alpha_i a_i)$$

$$= \sum_{i=1}^{n}\alpha_i f(a_i) = \sum_{i=1}^{n}\alpha_i(ca_i) = \sum_{i=1}^{n}\alpha_i b_i\,.$$

Densities of such vectors are connected by a correlation:

$$d_p(B) = \frac{|\mu_p(B)|}{\sum_{i=1}^{n}(p-1)b_i} = \frac{|\mu_p(A)|}{\sum_{i=1}^{n}(p-1)ca_i} = \frac{|\mu_p(A)|}{c\left(\sum_{i=1}^{n}(p-1)a_i\right)},$$

$$d_p(A) = c\, d_p(B). \quad (13)$$

Sequences $W_{\mu(p,\ A)}$ and $W_{\mu(p,\ B)}$ possess properties defined by a correlation (10). The elements of sequences $\Delta W_{\mu(p, A)}$ and $\Delta W_{\mu(p, B)}$ are connected as follows:

$$m_i = ch_i,\ I = 1\ldots n, \text{ where } m_i \in \Delta W_{\mu(p, B)},\ h_i \in \Delta W_{\mu(p, A)}$$

The most widely known are systems of information protection with an open key and with a knapsack on the basis of a secret key [2] in which a vector received from a knapsack vector by strong modular multiplication by values of a secret key is used as an open key. It is possible to perform the crypto analysis of such systems by analytical or statistical methods, or by means of the analysis of an open key.

Analytical methods are based on methods of decisions of equation (1) on the basis of known values from μ(p,). Applicability of the given methods is based on volumes of done calculations. The upper boundary of a number of solutions (1) is presented in [3] and generally is a NP-full problem.

Statistical methods are based on statistical characteristics of elements of a natural language or other language of the original text and the statistics of crypto text elements. The main objective of such methods is to find a mutually single-valued correspondence between the elements of an original text and a cipher text rather than to find a knapsack vector. They are applicable only in the presence of statistical volumes of cipher texts.

Methods of crypto analysis of an open key consist in restoration of a KSPI knapsack vector according to an open key vector. In particular, for two supergrowing knapsack vectors, received one from another by means of strong modular multiplication, A. Shamir offers an algorithm of finding a knapsack vector *A* KSPI if vector *B* [2] is known.

On the basis of knapsack vectors properties described above it is possible to formulate the following results:

1. Crypto analysis of KSPI can be made not only on the basis of statistics of cipher texts elements values, but also on distribution of values. As the probability of occurrences of elements $\Delta W_{\mu (p, A)}$ sequences of knapsack vector $A = (a_1, a_2, ..., a_n)$ in Z_p is a constant value for the set dimension n, the table of probabilities is calculated at the stage of preliminary preparation of crypto analysis. The analysis of cipher texts is made on the basis of differences between pairs of values of its elements. In this case a number of various values of a cipher text elements is more important than the volume of known cipher texts. The construction of an injective knapsack vector is carried out on the basis of properties $W_{\mu (p, A)}$ and Lemma 1.

2. The applicability of statistical methods of cipher texts analysis is based on its volume. Therefore if volumes of such information are small then the given methods are practically inapplicable. Updating KSPI with one knapsack vector into a system with dynamically generated knapsack vectors [4; 5] leads to practical inapplicability of statistical methods of cipher texts analysis.

To increase the cryptographic security of classical systems of information protection with an open key and with a knapsack it is necessary not only to use isomorphic and similar knapsack vectors, but also to change values of exits of the enciphering block of KSPI by value of some constant. For example, having altered a classical system of information protection with an open key and with a knapsack on the basis of a secret key (m, t) [2], it is possible to raise the system cryptographic security essentially.

Let's consider a simple example. Let $A = (2, 5, 6)$ be an injective increasing knapsack vector. Before the definition of an open key – vector B, we will apply function $f(x) = x^2 - x$ to the elements of vector A and considering that $f(2) = 2$, $f(5) = 20$, $f(6) = 30$, we will receive $A' = (2, 20, 30)$. Using pair $m = 220$ and $t = 17$ as a secret key [2] we will receive open key $B = (34, 120, 70)$ by strong modular multiplication [2]. A crypto analysis of vector B according to A. Shamir's algorithm can lead only to reception of a supergrowing vector A' [2] in which cipher texts $w = 7$ is inadmissible. Thus, the use of a secret key (m, t, $f(x)$) leads to the fact, that known methods of the analysis of an information protection system with an open key, in particular, those using strong modular multiplication, are inapplicable or demand additional expenses concerning transformation search $f(x)$.

References

1. Osipyan V. O. Development of methods of information transmission and security systems construction. Krasnodar, 2004.

2. Salomaa A. Cryptography with an open key. M. : World, 1995.

3. Podkolzin V. V., Osipyan V. O. Upper boundary of a number of solutions of a generalized task of a knapsack on a set point // Actual problems of information technologies safety : materials of III International theoretical and practical conf. / edited by O. N. Zhdanov, V. V. Zolotarev ; Siberian state aerospace university. Krasnoyarsk, 2009. P. 30–33.

4. Podkolzin V. V. A model of information security system with an open key on the basis of dynamic generation of a knapsack vector. M. : OPandPM, 2009. Vol. 16. Issue 5. P. 913–914.

5. Podkolzin V. V., Osipyan V. O. Of one modification of information security task with an open key on the basis of a generalized knapsack point. M. : OPandPM, 2009, Vol. 16. Issue 5. P. 905.

V. V. Prudkov
Siberian State Airspace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

THE AUTOMATION PROCESS PROCEDURES OF THE SUBSYSTEMS BLOCK VERIFICATION FOR SPACE VEHICLES MANAGEMENT

Described are the automation procedures for the software of logic verification in subsystem of the management perspective space vehicle block functioning. The achieved results by the implementation of these procedures are presented.

Keywords: management block, central-processor module, interfacing of the management block, independent working off, automation.

Modern requirements in the field of space services and technologies demand from space vehicle (SV) and onboard equipment manufacturers the reduction of terms and expenses from design, qualifications, the manufacturing and tests of onboard equipment and space vehicles, and the increase in reliability and the quality of the products. This is required at constant minimization of their energy and mass features. Thus, it is necessary to achieve an essential economization of resources: human, financial, and material during simultaneous reduction of design terms for spacecraft.

Modern radio-electronic equipment (REE) based on programmed logic integrated schemes (COTT software VELVET), microcontrollers, built-in computing modules etc. demand special debugging facilities. The designing of the REE is organized as a constant interaction between the software developer and the developer of schemes. Flexibly organized equipment, which gives a chance at any moment to spend fast reconfiguration of test apparatus and interesting experiments, allows spending optimally on the REE debugging works.

Currently at JSC Information Satellite Systems-Reshetnev Company – there is an intensity of schedules for manufacturing REE of modern space vehicles for carrying out working off tests at a scale of Earth-based experimental works; there is also a complete set of exclusively high regular products. The necessity to increase the efficiency of the working REE and to reduce its terms for restored products, involves the creation of unified and automated workplaces for experimental Earth-based REE works, which concern management blocks.

The blocks of management, which are part of an onboard complex of management developed at the enterprise (BM OCM) for modern and perspective space vehicles, are designed according to a modular principle. Inside the structure of the management block are: the central-processor module (CPM) and the interface modules of linking (IML or subsystems). The CPM allows realizing all logical functions for concrete equipment using software. The IML carries out the managements of the SV systems. The information logical communication CPM and IML is carried out by the consecutive peripheral interface (PPI). The BM management is carried out by an onboard computer complex (OCC) on the multiplex channel of exchange (MCE) (GOST P 52070–2003). The CPM accepts management commands of the MCE from the OCC, decodes them, and gives out words of data (WD), containing management commands (MC), in corresponding subsystems – the BM (IML). The OCC reads out WD from the CPM, which contain telemeteric information.

The REE on the base of programmed logic integrated schemes possesses specific features of test processing, inherited in the process of developing software products. The emphasis is on system integrated decisions. The scheme and the solution become simpler, but the architecture of the device since additional non-material makings has become part of its structure, and the software becomes complicated. It is necessary to provide the developer (the circuitry-designer-programmer) with special conditions of debugging and rapidly changing software tests.

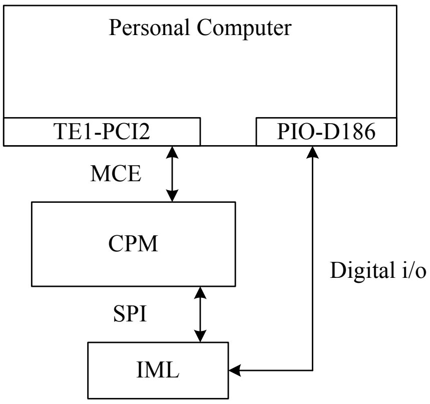
Today the service time and the active space vehicles existence are 10–15 years. For its maintenance, which is one of problems, the verification of reliability onboard REE is one of the issues is the verification of the logical functioning IML or subsystems and their components. The IML it is hardware that represent blocks consisting of several schemes executed in the form of a universal constructive. They carry out the specialized problems, carrying out a link role between the onboard computer complex and the SV systems.

During the verification of logical IML functioning, the put-in-pawn circuit and technical decisions on the conformity to the technical project (T3) should be confirmed. The given stage is called the laboratory-based work off IML tests and consists in the revelation of errors in their designing, manufacturing, and as errors in projects COTT software VELVET IML.

For the realization of this stage at a designing department and working out the onboard REE, the workplace (PM) we have created an independently working IML [1] (sce figure), on the base of a laboratory work off complex [2].

The workplace of the independent working IML consists of the personal computer (PC), the CPM and the connected IML. For information exchange between the personal computer and the CPM, circuit board TE1-PCI2 supporting report MCE manufactured by the firm “Elkus” is used. For the communication between the personal computer and the IML circuit board PIO-D186 with a digital input/output is used.

The hardware of such a workplace realizes the structure, and establishes information communications for its various elements. The basic tool for the achievement of the general purpose of such a system is the software.



The Workplace of an independent working IML

Such software should realize uniform information and logical space between the equipment of the control and fulfilled subsystems, to provide flexible and full verification of all logic functions of the fulfilled subsystems, and meet independent work with the deadlines at with the least expenditures of labour.

For the solution of the given problems, the creation of an automation process for the working IML is necessary. Introduction in software automation means it will allow reducing the human factor to a minimum, thereby raising the reliability of the subsystems working process. The performance of the majority of the functions in an automatic mode will allow to reduce time and the expenditures of labour demanded for the working subsystems.

In result of the detailed software analysis [2; 3] we have created a solution to the tasks in view of the independently working IML.

The developed software has following basic procedures of automation:

– the generation of tests;
– the preliminary analysis of the test;
– the automatic change of test data;
– the analysis of the received data for reliability;
– the transfer of the received WD values to corresponding physical scales.

The generation of the tests. The basic stage of any REE test is the creation of a set of tests at which there is final completeness and the degree of working REE will depend on. The automated procedure of test generation allows facilitating the procedures of their creation for the operator and bringing possible errors for manual input to zero.

The procedure of automatic test generation allows creating tests for any subsystem. For the generation of the test it is necessary for the operator to enter the initial parameters for a concrete subsystem. On the basis of the entered data the software builds the time diagram of performance for the test. Even before the direct development of the test, the operator can visually see its structure with the time delays between various operations. The given decision allows him to manually replace the creation of a set of packages with some manipulations of the operator in the program.

The preliminary analysis of the test. The preliminary analysis procedure of the test analyzes the accomplished sequence of operations on its completeness and correctness of figures. In case of default for these conditions, corresponding preventions that excludes possible errors brought by the operator stand out, during operations updating of the test. For example: if there is an operation of delivery for a 2nd IML circuit board command, there should be an operations of reading for the 2nd board and an operation for the execution of the given command. In a case when there is not a single one of the specified, there are negative results of the analysis with the list of missing operations.

The given procedure automates the search for missing operations for the manual task of tests and for their editing that in turn, reduces the probability of not completing the working IML.

Automatic change of test data. The carried out test is limited by static values of data which are set in them by the operator. For the fast change of the information in tests, a procedure of automatic data change has been created. Its essence consists in the following: after the performance of the test with single data elements, there is a change in data and the performance of the current test is already different. Change of data occurs in advance of set masks. Thus, the given procedure allows starting the set tests with an automatic data change for each cycle of their performance; the volume of the changeable data is set preliminary by the operator.

The analysis of the received data for reliability. While carrying out of any test, a great volume of the targeted information, which is necessary for analyzing turns up. For a person, a similar analysis represents a long and time-consuming process and the chance of error increases.

The procedure of the received data analysis makes the analysis of the exchange report, revealing errors in the functioning IML. On the base of the analysis, displayed to the operator is the detailed information by the revealed negative results. Such results can be: the absence of the corresponding bit in the WD on the given out command; occurrence of additional bits in the WD, on a command on which they should not appear; the absence of various signs of subsystems in the WD, etc.

The transference of received WD values in corresponding physical sizes. IML developers write the containing codes of values for various parameters in SD devices; these are temperature, pressure, resistance etc. During independent IML workings it is necessary for the operator to translate the received codes in corresponding physical sizes, for the further analysis of the data.

The machine translation procedure is adjusted depending on the fulfilled IML. As an options the price of the younger category for a code, various factors of formulas of recalculation, etc can be specified. The given procedure allows the reduction of time spent by the operator to translate the codes of WD values read from the IML.

Introduction results of the automation procedures. The developed software has completely solved the tasks in view of independent working IML, and thanks to the automation procedures, allowing:

– to automate the process of independent working IMLs to 90 %, leaving the operator only the performance and the analysis of specific checks;

– to reduce time spent for working out concrete IMLs, from several weeks to 1–2 working days;

– to spend simultaneous working to 8 IMLs as a part of a workplace, connected among each other on interblock sockets and connected to the CPM;

– to check the working IML capacity during irregular situations, by their modelling;

– to independently fulfill each complete the IML set (basic/reserve) connected to each complete the CPM set (basic/reserve);

– to fulfill the BM in gathering, with use regular cables as the IML connections;

– to use the PM at any stage of REE tests thanks to flexibility and universality.

Currently, the given software – the independent working IML is used in space vehicle management blocks – “Monsoon”, “Glonass-to”, “Amos-5”; and is used at the working SV “Luch-5”. During the tests, the correctness of the construction of the software and correctness of the approach connected with design of the automated procedures have been confirmed.

Thus, the developed software has proved its reliability, universality, and simplicity in use, thanks to what it is applicable for the working of the subsequent IML management block of perspective SVs. The procedures of the automated software and their algorithms are applicable for designing the software of workings REEs.

References

1. Prudkov V. V. The working place of autonomous processing of interfacial modules in the conjugation bloc of perspective SV control // Modern instrumental systems, information technologies and innovations: Materials of the VII International (distance) scientific conf. / Ozerski technological institute MIFI. Kursk, 2010. P. 150–152.

2. Pichkalev A. V. Testing radio-electric equipment at a workout laboratory complex // Reshetnev Conference : materials of the XII International science conf. ; SibSAU. Krasnoyarsk, 2008. P. 158–159.

3. Prudkov V. V. Particularities of constructing software for the autonomous performance of subsystems in the control bloc of perspective SV // Reshetnev Conference : Materials of the XII International science conf. ; SibSAU. Krasnoyarsk, 2009. P. 531–532.

A. V. Starovoitov
Siberian Federal University, Russia, Krasnoyarsk

A MULTIDIMENSIONAL ANALOG OF THE COOLEY-TUKEY FFT ALGORITHM

In this article a recurring sequence of orthogonal basis in the n-dimensional case has been applied to derive formulas of n-dimensional fast Fourier transform algorithm, which uses $\frac{2^n-1}{2^n}N^n\log_2 N$ *complex multiplication and* $nN^n\log_2 N$ *complex addition; where* $N=2^s$ *– is a number of counts on one of the axes.*

Keywords: space of signals, orthogonal basis sequence, multidimensional discrete Fourier transform.

Recurrent sequence of orthogonal bases in space of signals is well studied [1] and has numerous applications, including the derivation of Fourier’s formulas of fast transformation.

In this article the recurrent sequence of orthogonal bases to a *n*-dimensional case is applied in order to derive formulas of a fast *n*-dimensional Fourier transformation variant, using $\frac{2^n-1}{2^n}N^n\log_2 N$ complex multiplication and $nN^n\log_2 N$ complex addition, where $N=2^s$ – is a number of counts on one of the axes (known in studies as in [2]). This variant n FFT contains a smaller number of complex multiplication operations than other algorithms, where the multidimensional Fourier transformation is carried out by repeated application of one-dimensional FFT (for example, see [3; 4]).

Furthermore, we give definitions and basic statements from the theory of multidimensional signals, which are used in the article.

To construct *n*-dimensional recurrent sequence of orthogonal bases we use the scheme of the statement, given in [1] for a one-dimensional case.

1. The space of periodic n-dimensional signals.

Definition 1. With a fixed N, the n-dimensional periodic signal shall be a periodic complex function of integer argument, with the period N on each variable.

Define operations of adding the two signals x_1, x_2 and multiplying the signal x by a complex number c:

$$y(j)=x_1(j)+x_2(j);$$
$$y(j)=c\cdot x(j),$$

where $x(j)$ – is the count of a signal x at point $j\in \mathrm{Z}^n$. Then, a set of signals C_N^n becomes a linear complex space. A zero element in C_N^n is the signal O such, that $\mathrm{O}(j)=0$ for all is $j\in \mathrm{Z}^n$. Scalar produce and norm of space C_N^n are:

$$\langle x,y\rangle=\sum_{j\in B_n(N)} x(j)\overline{y(j)},$$
$$\| x\| =\langle x,x\rangle^{1/2},$$

where $B_n(N)$ – is a set of integer vectors from $[0,N-1]^n$.

Definition 2. The unit n-dimensional periodic impulse, with the period N on each variable, is a signal δ_N^n such, that $\delta_N^n(j)=1$, if each coordinate of a vector j divided by N and $\delta_N^n(j)=0$ otherwise.

The Following statements are true for a unit impulse.

1) $\delta_N^n(j_1,...,j_n)=\delta_N^n(|j_1|,...,|j_n|)$;

2) $\delta_N^n(j_1,...,j_n)=\delta_N^1(j_1)\cdot ...\cdot \delta_N^1(j_n)$;

3) For $x\in \mathrm{C}_N^n$ the equality is true:

$$x(j)=\sum_{t\in B_n(N)} x(t)\delta_N^n(j-t), \quad (1)$$

for any $j\in B_n(N)$.

Let $w_N=\exp(\frac{2\pi i}{N})$.

Lemma 1. Then

$$\delta_N^n(j)=\frac{1}{N^n}\sum_{t\in B_n(N)} w_N^{(j,t)}, \quad (2)$$

where (j,t) – is the scalar product of vectors j and t.

Equation is checked (2) by direct calculation.

Definition 3. The n-dimensional discrete Fourier transform is called a depiction: $F_N:\mathrm{C}_N^n\to \mathrm{C}_N^n$ take each signal x to a signal X, where:

$$X(j)=\sum_{t\in B_n(N)} x(t)w_N^{-(j,t)},\ j\in B_n(N).$$

Note that, for DFT the formula of inversion is true:

$$x(t)=\frac{1}{N^n}\sum_{j\in B_n(N)} X(j)\cdot w_N^{(j,t)}$$

and the Parseval identity:

$$\text{if } X=F_N(x),\ Y=F_N(y),$$
$$\langle x,y\rangle=\frac{1}{N^n}\langle X,Y\rangle.$$

2. The recurrent sequences of orthogonal bases.

Let $N=2^s$, $N_\nu=2^{s-\nu}$, $\Delta_\nu=2^{\nu-1}$. We shall construct recurrent sequence of bases $f_0,f_1,...,f_s$, where f_t – t-th basis, consisting of N^n signals $f_t(k)$, $k\in B_n(N)$. We will denote a value of a signal $f_t(k)$ at count $j=(j_1,...,j_n)$, $j\in B_n(N)$ by $f_t(k;j)$.

Let $B_n^1(N)$ by a set of integer vectors from $[0,N_\nu-1]^n$ and $B_n^2(N)$ by a set of integer vectors from $[0,\Delta_\nu-1]^n$. We will define the sequence of orthogonal bases as:

$$f_0(k;j)=\delta_N^n(j-k)=\delta_N^1(j_1-k_1)\times$$
$$\times\delta_N^2(j_2-k_2)\ ...\ \delta_N^n(j_n-k_n),k,j\in B_n^{(N)}.$$

$$f_\nu(l_1+\sigma_1\Delta_\nu+p_1\Delta_{\nu+1},l_2+\sigma_2\Delta_\nu+p_2\Delta_{\nu+1},...,l_n+\sigma_n\Delta_\nu+p_n\Delta_{\nu+1})=$$
$$=\sum_{\tau_1=0}^{1}...\sum_{\tau_n=0}^{1} w_{\Delta_{\nu+1}}^{\sum_{i=0}^{n}\tau_i(l_i+\sigma_i\Delta_\nu)} f_{\nu-1}\times$$
$$\times(l_1+2\Delta_\nu p_1+\tau_1\Delta_\nu,...,l_n+2\Delta_\nu p_n+\tau_n\Delta_\nu), \quad (3)$$

where $p=(p_1,...,p_n)\in B_n^1(N), l=(l_1,...,l_n)\in B_n^2(N),\sigma_i$ is equal to 0 or 1 for all $i=1,...,n,\nu=1,...,s$.

For studying the properties of recurrent sequence of bases, we can use reverse rearrangement [1].

Let j by an integer from set $J=\{0,1,...,2^\nu-1\}$ be equal to $j_{\nu-1}2^{\nu-1}+...+j_1 2+j_0$ in a binary system, where $j_i=0,1$ for all $i=0,...,\nu-1$. A vector $(j_{\nu-1},...,j_1,j_0)_2$ is called a binary code of number j. We compare number $j_1\in J$ with number j, which is set by a binary code $(j_0,j_1,...,j_{\nu-1})_2$. Rearrangement $rev_\nu(j)=j_1$ for set J is called reverse rearrangement. For reverse rearrangements the following equalities are true:

$$\begin{aligned}2rev_{\nu-1}(q)&=rev_\nu(q);\\ 2rev_{\nu-1}(q)+1&=rev_\nu(\Delta_\nu+q).\end{aligned} \quad (4)$$

Using a reverse rearrangement we can prove that:

$$f_\nu(l_1+p_1\Delta_{\nu+1},...,l_n+p_n\Delta_{\nu+1})=$$
$$=\sum_{q_1=0}^{\Delta_{\nu+1}-1}...\sum_{q_n=0}^{\Delta_{\nu+1}-1} w_{\Delta_{\nu+1}}^{\sum_{i=0}^{n} l_i rev_\nu(q_i)} f_0(q_1+p_1\Delta_{\nu+1},...,q_n+p_n\Delta_{\nu+1}),$$

where $p=(p_1,...,p_n)\in B_n^1(N)$, $l=(l_1,...,l_n)\in B_n^2(N)$, $\nu=1,...,s$.

In particular, if $\nu=s$ we have:

$$f_s(l;j)=\sum_{q_1=0}^{N-1}...\sum_{q_n=0}^{N-1} w_N^{\sum_{i=1}^{n} l_i rev_\nu(q_i)}\delta_N^n(j_1-q_1,...,j_n-q_n)=w_N^{\sum_{i=0}^{n} l_i rev_\nu(j_i)}.$$

Theorem 1. For all $\nu = 0,...,s$, $k \in B_n(N)$, a set of signals $f_\nu = f_\nu(k)$ is orthogonal and $\| f_\nu(k) \|^2 = 2^{n\nu}$.

The solution. Let $\nu = 0$. Then:

$$\langle f_0(k), f_0(k') \rangle = \sum_{j \in B_n(N)} \delta_N^n(j-k) \cdot \delta_N^n(j-k')$$

the last sum can be distinct from zero only when $k = k'$, in other case it is equal to 1 and the theorem is proved.

Let's now $\nu = 1,...,s$ and $k, k' \in B_n(N)$, which are presented in the following way:

$$k = (k_1,...,k_n) = (l_1 + p_1\Delta_{\nu+1},...,l_n + p_n\Delta_{\nu+1}),$$
$$k' = (k'_1,...,k'_n) = (l'_1 + p'_1\Delta_{\nu+1},...,l'_n + p'_n\Delta_{\nu+1}),$$

where $l = (l_1,...,l_n)$, $l' = (l'_1,...,l'_n)$ belongs to $B_n^2(N)$, and $p = (p_1,...,p_n)$, $p' = (p'_1,...,p'_n)$ belongs to $B_n^1(N)$. Then:

$$\langle f_\nu(k), f_\nu(k') \rangle = \langle f_\nu(l_1 + p_1\Delta_{\nu+1},...,l_n + p_n\Delta_{\nu+1}), f_\nu(l'_1 + p'_1\Delta_{\nu+1},...,l'_n + p'_n\Delta_{\nu+1}) \rangle =$$

$$= \langle \sum_{q_1=0}^{\Delta_{\nu+1}-1} ... \sum_{q_n=0}^{\Delta_{\nu+1}-1} w_{\Delta_{\nu+1}}^{\sum_{i=0}^{N} l_i rev_\nu(q_i)} f_0(q_1 + p_1\Delta_{\nu+1},...,q_n + p_n\Delta_{\nu+1}),$$

$$\sum_{q'_1=0}^{\Delta_{\nu+1}-1} ... \sum_{q'_n=0}^{\Delta_{\nu+1}-1} w_{\Delta_{\nu+1}}^{\sum_{i=0}^{N} l'_i rev_\nu(q'_i)} f_0(q'_1 + p'_1\Delta_{\nu+1},...,q'_n + p'_n\Delta_{\nu+1}) \rangle =$$

$$= \sum_{q_1=0}^{\Delta_{\nu+1}-1} ... \sum_{q_n=0}^{\Delta_{\nu+1}-1} \sum_{q'_1=0}^{\Delta_{\nu+1}-1} ... \sum_{q'_n=0}^{\Delta_{\nu+1}-1} w_{\Delta_{\nu+1}}^{\sum_{i=1}^{n} l_i rev_\nu(q_i) - l'_i rev_\nu(q'_i)} \times$$
$$\times \delta_N^n(q_1 - q'_1 + (p_1 - p'_1)\Delta_{\nu+1},...,q_n - q'_n + (p_n - p'_n)\Delta_{\nu+1}).$$

Arguments of a unit impulse δ_N^n on the module do not exceed $N-1$. For $p_t = p'_t$ and at the some t arguments are distinct from zero for all $q_i, q'_i \in 0 : \Delta_{\nu+1} - 1$, $i = 1,...,N$, as $| q_i - q'_i | \le \Delta_{\nu+1} - 1$. Therefore $\langle f_\nu(k), f'_\nu(k') \rangle = 0$, if $p_j \ne p'_j$.

Let $p_j = p'_j$. For all $j = 1,...,n$ then:

$$\langle f_\nu(k), f_\nu(k') \rangle = \sum_{q_1=0}^{\Delta_{\nu+1}-1} ... \sum_{q_n=0}^{\Delta_{\nu+1}-1} w_{\Delta_{\nu+1}}^{\sum_{i=1}^{n} (l_i - l'_i) rev_\nu(q_i)} =$$

$$= \sum_{q_1=0}^{\Delta_{\nu+1}-1} ... \sum_{q_n=0}^{\Delta_{\nu+1}-1} w_{\Delta_{\nu+1}}^{\sum_{i=1}^{n} q'_i} = \Delta_{\nu+1} ... \Delta_{\nu+1} \delta_{\Delta_{\nu+1}}^n (l_1 - l'_1,...,l_n - l'_n).$$

From the last formula it is concluded, that the scalar product $\langle f_\nu(k, f_\nu(k') \rangle$ is distinct from zero only, if $p_j = p'_j$, $l_i = l'_i$, where $i, j = 1,...,n$. In the last case $\| f_\nu(k_1,...,k_n) \| = \Delta_{\nu+1} ... \Delta_{\nu+1} = 2^{n\nu}$ for all $k_1,...,k_n = 0 : N-1$. The theorem is now proved.

3. Sequence application of orthogonal bases to denote the fast discrete Fourier transform.

Let $x(j) = x(j_1,...,j_n) \in \mathrm{C}_N^n$, $j \in B_n(N)$. We compare a signal $x(j)$ a to signal $x_0(j) = x(rev_s(j_1),...,rev_s(j_n))$ and we will spread out $x_0(j)$ on basis f_ν:

$$x_0 = \frac{1}{2^{n\nu}} \sum_{k \in B_n(N)} x_\nu(k) f_\nu(k) \quad (5)$$

($\frac{1}{2^{n\nu}}$ – is a normalizing multiplier). Multiplying both parts (5) by $f_\nu(l)$ scalar $l \in B_n(N)$. Then

$$\langle x_0, f_\nu(l) \rangle = x_\nu(l),$$
$$x_\nu(k) = \sum_{j \in B_n(N)} x_0(j) f_\nu(k) =$$
$$= \sum_{j \in B_n(N)} x(rev_s(j_1),...,rev_s(j_n)) f_\nu(k; j),$$

and coefficients $x_\nu(k)$ in (5) are determined.

In particular, for $\nu = 0$ we have from (1):

$$x_0(k) = \sum_{j \in B_n(N)} x(rev_s(j_1),...,rev_s(j_n) \delta_N^n(j-k) = x(rev_s(k_1),...,rev_s(k_n)). \quad (6)$$

From (3) we get the following:

$$x_\nu(l_1 + \delta_1\Delta_\nu + p_1\Delta_{\nu+1},...,l_n + \delta_n\Delta_\nu + p_n\Delta_{\nu+1}) =$$
$$= \langle x_0, f_\nu(l_1 + \delta_1\Delta_\nu + p_1\Delta_{\nu+1},...,l_n + \delta_n\Delta_\nu + p_n\Delta_{\nu+1}) \rangle =$$
$$= \sum_{\tau_1=0}^{1} ... \sum_{\tau_n=0}^{1} w_{\Delta_{\nu+1}}^{\sum_{i=1}^{n} \tau_i(l_i + \sigma_i\Delta_\nu)} \langle x_0, f_{\nu-1}(l_1 + 2\Delta_\nu p_1 + \tau_1\Delta_\nu,...,l_n + 2\Delta_\nu p_n + \tau_n\Delta_\nu) \rangle = \sum_{\tau_1=0}^{1} ... \sum_{\tau_n=0}^{1} w_{\Delta_{\nu+1}}^{\sum_{i=1}^{n} \tau_i(l_i + \sigma_i\Delta_\nu)} \times$$
$$\times x_{\nu-1}(l_1 + 2\Delta_\nu p_1 + \tau_1\Delta_\nu,...,l_n + 2\Delta_\nu p_n + \tau_n\Delta_\nu), \quad (7)$$

where $p = (p_1,...,p_n) \in B_n^1(N)$, $l = (l_1,...,l_n) \in B_n^2(N)$, $\nu = 1,...,s$ and $\sigma_1,...,\sigma_n$ are equal to 0 or 1.

As

$$x_s(k) = x(k_1,...,k_n) =$$
$$= \sum_{j \in B_n(N)} x(rev_s(j_1),...,rev_s(j_n)) \cdot w_N^{-\sum_{i=0}^{n} k_i rev_s(j_i)} =$$
$$= \sum_{j \in B_n(N)} x(j) w_N^{-\sum_{i=0}^{n} k_i j_i} = X(k),$$

where $k \in B_n(N)$ and coefficients $x_s(k)$ define components of a spectrum for a signal x on a basic period.

From (6) and (7) we have received the recurrent scheme for the calculation of a spectrum for a signal $x \in \mathrm{C}_N^n$:

$$x_0(k) = x(rev_s(k_1),...,rev_s(k_n));$$

$$x_\nu(l_1+\sigma_1\Delta_\nu+p_1\Delta_{\nu+1},...,l_n+\sigma_n\Delta_\nu+p_n\Delta_{\nu+1}) = \qquad (8)$$

$$= \sum_{\tau_1=0}^{1}...\sum_{\tau_n=0}^{1} w_{\Delta_{\nu+1}}^{\sum_{i=0}^{n}\tau_i(l_i+\sigma_i\Delta_\nu)} \cdot x_{\nu-1} \times$$

$$\times(l_1+2\Delta_\nu p_1+\tau_1\Delta_\nu,...,l_n+2\Delta_\nu p_n+\tau_n\Delta_\nu),$$

where $p=(p_1,...,p_n)\in B_n^1(N)$, $l=(l_1,...,l_n)\in B_n^2(N)$, $\nu=1,...,s$ and $\sigma_1,...,\sigma_n$ are equal to 0 or 1.

Let's find a number of complex addition and multiplication operations necessary for finding a spectrum of the signal for scheme (8).

Lemma 2. For some r vectors $t=(t_1,...,t_r)$ and $\sigma=(\sigma_1,...,\sigma_r)$, where $t_i,\sigma_i\in\overline{0,1}$. Then the calculation of all the values of some function:

$$S(\sigma)=\sum_t f(t)(-1)^{\langle\sigma,t\rangle}$$

requires $r\cdot 2^r$ additions (subtractions).

The solution. To prove we apply an induction on r.

Let $r=2$. Then:

$$S(\sigma)=S(\sigma_1,\sigma_2)=\sum_{t_1=0}^{1}\sum_{t_2=0}^{1} f(t_1,t_2)\cdot(-1)^{\sigma_1t_1+\sigma_2t_2} =$$

$$= f(0,0)+f(1,0)(-1)^{\sigma_1}+f(0,1)(-1)^{\sigma_2}+$$

$$+f(1,1)(-1)^{\sigma_1+\sigma_2}.$$

Let's define:

$$S_1(\sigma)=S(0,0)=f(0,0)+f(1,0)+f(0,1)+f(1,1)=$$

$$=(f(0,0)+f(1,0))+(f(0,1)+f(1,1))=S_1^*+S_3^*;$$

$$S_2(\sigma)=S(1,0)=f(0,0)-f(1,0)+f(0,1)-f(1,1)=$$

$$=(f(0,0)-f(1,0))+(f(0,1)-f(1,1))=S_2^*+S_4^*;$$

$$S_3(\sigma)=S(0,0)=f(0,0)+f(1,0)-f(0,1)-f(1,1)=$$

$$=(f(0,0)+f(1,0))-(f(0,1)+f(1,1))=S_1^*-S_3^*;$$

$$S_1(\sigma)=S(0,0)=f(0,0)-f(1,0)-f(0,1)+f(1,1)=$$

$$=(f(0,0)-f(1,0))-(f(0,1)-f(1,1))=S_2^*-S_4^*,$$

where:

$$S_1^*=f(0,0)+f(1,0), S_2^*=f(0,0)-f(1,0),$$

$$S_3^*=f(0,1)+f(1,1), S_4^*=f(0,1)-f(1,1).$$

For calculating S_i^*, $i=1,2,3,4$ it is required to apply 4 additions (subtractions); to calculate all values S it requires 8 such operations and then the statement of the lemma is correct.

Let the statement of the lemma be correct, if $r=k$; for any function $g(t)$ i. e. all values of the function:

$$S(\sigma_1,...,\sigma_k)=\sum_t g(t)(-1)^{\langle\sigma,t\rangle}$$

are calculated by $2\cdot 2^k$ additions (subtractions), where $t=(t_1,...,t_k)$.

Let's consider $r=k+1$.

$$S(\sigma_1,...\sigma_{k+1})=\sum_{t_1=0}^{1}...\sum_{t_{k+1}=0}^{1} f(t_1,...,t_{k+1})\cdot(-1)^{\sigma_1t_1+...+\sigma_{k+1}t_{k+1}} =$$

$$=\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f(t_1,...,t_k,0)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}+$$

$$+\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f(t_1,...,t_k,1)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k+\sigma_{k+1}}.$$

Let's denote:

$$S(\sigma_1,\sigma_2,...,\sigma_k,0)=S_1(\sigma_1,\sigma_2,...,\sigma_k)=$$

$$=\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f(t_1,...,t_k,0)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}+$$

$$+\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f(t_1,...,t_k,1)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}=$$

$$=\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1}(f(t_1,...,t_k,0)+f(t_1,...,t_k,1))\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k};$$

$$S(\sigma_1,\sigma_2,...,\sigma_k,1)=S_2(\sigma_1,\sigma_2,...,\sigma_k)=$$

$$=\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f(t_1,...,t_k,0)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}-$$

$$-\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f(t_1,...,t_k,1)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}=$$

$$=\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1}(f(t_1,...,t_k,0)-f(t_1,...,t_k,1))\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}.$$

Let's define how many operations of addition (subtraction) are required for calculating S_1 and S_2.

Let's denote $f^+(t_1,...,t_k)=f(t_1,...,t_k,0)+$, $+f(t_1,...,t_k,1)$ $f^-(t_1,...,t_k)=f(t_1,...,t_k,0)-f(t_1,...,t_k,1)$. For calculating all values f^+ required are 2^k additions and to calculate $S_1=\sum_{t_1=0}^{1}...\sum_{t_k=0}^{1} f^+(t_1,...,t_k)\cdot(-1)^{\sigma_1t_1+...+\sigma_kt_k}$ required are $k2^k$ according to the induction assumption.

Then the total sum of addition (subtraction) operations for calculation S_1 requires $k2^k+2^k$. Such a number of operations it is necessary for S_2.

Therefore, to calculate of all values S $(k+1)2^{k+1}$ additions (subtractions) are required; this needed to be proved.

Theorem 2. To calculate the recurrent scheme for a signal spectrum $x\in C_N^n$ ($N=2^s$)

$$x_0(k)=x(rev_s(k_1),...,rev_s(k_n));$$

$$x_\nu(l_1+\sigma_1\Delta_\nu+p_1\Delta_{\nu+1},...,l_n+\sigma_n\Delta_\nu+p_n\Delta_{\nu+1}) = \qquad (9)$$

$$= \sum_{\tau_1=0}^{1}...\sum_{\tau_n=0}^{1} w_{\Delta_{\nu+1}}^{\sum_{i=0}^{n}\tau_i(l_i+\sigma_i\Delta_\nu)} \cdot x_{\nu-1} \times$$

$$\times(l_1+2\Delta_\nu p_1+\tau_1\Delta_\nu,...,l_n+2\Delta_\nu p_n+\tau_n\Delta_\nu),$$

demands $\frac{2^n-1}{2^n}N^n\log_2 N$ complex multiplications and $nN^n\log_2 N$ complex additions.

The solution. First, we will find a number of complex multiplications. Complex multiplication is multiplication only by $w_N^{\sum_{i=0}^{n}\tau_i(l_i+\sigma_i\Delta_\nu)}$. We shall define the number of products required in (9) for all parameters $\nu^*, l^*=(l_1^*,...,l_n^*)$ and $p^*=(p_1^*,...,p_n^*)$. For this purpose we shall consider products:

$$w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i(l_i^*+\sigma_i\Delta_{\nu^*})}\cdot x_{\nu^*-1}\times$$
$$\times(l_1^*+2p_1^*\Delta_\nu+\tau_1\Delta_{\nu^*},...,l_n^*+2p_n^*\Delta_\nu+\tau_n\Delta_{\nu^*}),$$

where $\sigma_i\in\overline{0,1}$.

We have to notice that:

$$w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i(l_i^*+\sigma_i\Delta_{\nu^*})}=w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i l_i^*+\tau_i\sigma_i\Delta_{\nu^*}}=$$
$$=w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i l_i^*}\cdot w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i\sigma_i\Delta_{\nu^*}}=(-1)^{\sum_{i=0}^{n}\tau_i\sigma_i}\cdot w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i l_i^*}. \quad (10)$$

That is product:

$$w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i(l_i^*+\sigma_i\Delta_{\nu^*})}\cdot x_{\nu^*-1}(l_1^*+2p_1^*\Delta_\nu+\tau_1\Delta_{\nu^*},...,l_n^*+2p_n^*\Delta_\nu+\tau_n\Delta_{\nu^*}).$$

Let parameters $\nu^*, l^*=(l_1^*,...,l_n^*)$ and $p^*=(p_1^*,...,p_n^*)$ by fixed; then product (10) can possibly be replaced with complex product:

$$w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i l_i^*}\cdot x_{\nu^*-1}(l_1^*+2p_1^*\Delta_\nu+\tau_1\Delta_{\nu^*},...,l_n^*+2p_n^*\Delta_\nu+\tau_n\Delta_{\nu^*}) \quad (11)$$

as $(-1)^{\sum_{i=0}^{n}\tau_i\sigma_i}$ can accept values only ±1.

The number of complex products of a kind (11) is equal to a number of every possible vector $\tau=(\tau_1,...,\tau_n)$, where $\tau_i\in\overline{0,1}$, i. e. 2^n-1, since we have a product of real numbers for $r=0$. As parameter $\nu\in1:s$, vectors $l=(l_1,...,l_n)$ and $p=(p_1,...,p_n)$ (where $p_i=0,1,...,N^\nu-1$ ($N^\nu=N/2^\nu$)), $l_i=0,1,...,\Delta_\nu-1$ ($\Delta_\nu=2^{\nu-1}$), then the total complex multiplications is equal to: $s(\frac{N}{2^\nu}\cdot2^{\nu-1})^n\cdot(2^n-1)=\frac{2^n-1}{2^n}N^n\log_2 N$.

Let's find the amount of complex addition in the algorithm. For fixed ν^* and $l^*=(l_1^*,...,l_n^*)$, $p^*=(p_1^*,...,p_n^*)$ we have:

$$x_\nu^*(l_1^*+\sigma_1\Delta_{\nu^*}+p_1\Delta_{\nu^*+1},...,l_n^*+\sigma_n\Delta_{\nu^*}+p_n^*\Delta_{\nu^*+1})=$$
$$=\sum_{\tau_1=0}^{1}...\sum_{\tau_n=0}^{1}w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i(l_i^*+\sigma_i\Delta_\nu)}\cdot x_{\nu^*-1}\times \quad (12)$$
$$\times(l_1^*+2p_1^*\Delta_\nu+\tau_1\Delta_{\nu^*},...,l_n^*+2p_n^*\Delta_\nu+\tau_n\Delta_{\nu^*}).$$

From (10) follows, that for calculation (12) we need to calculate expressions:

$$f(\tau)=w_{\Delta_{\nu^*+1}}^{\sum_{i=0}^{n}\tau_i l_i^*}x_{\nu^*-1}(l_1^*+2p_1^*\Delta_\nu+\tau_1\Delta_{\nu^*},...,l_n^*+2p_n^*\Delta_\nu+\tau_n\Delta_{\nu^*})$$

which depend only on $\tau=(\tau_1,...,\tau_n)$, where $\tau_i\in\overline{0:1}$. Then (12) can be presented as:

$$x_\nu^*(l_1^*+\sigma_1\Delta_{\nu^*}+p_1\Delta_{\nu^*+1},...,l_n^*+\sigma_n\Delta_{\nu^*}+p_n^*\Delta_{\nu^*+1})=$$
$$=\sum_{\tau_1=0}^{1}...\sum_{\tau_n=0}^{1}(-1)^{\sum_{i=0}^{n}\tau_i\sigma_i}\cdot f(\tau).$$

This way, complex addition is required from Lemma 2 to calculate:

$$x_\nu^*(l_1^*+\sigma_1\Delta_{\nu^*}+p_1\Delta_{\nu^*+1},...,l_n^*+\sigma_n\Delta_{\nu^*}+p_n^*\Delta_{\nu^*+1})\ \ n2^n.$$

As parameter $\nu\in1:s$, vectors $l=(l_1,...,l_n)$ and $p=(p_1,...,p_n)$, where $p_i=0,1,...,N^\nu-1$ ($N^\nu=N/2^\nu$), $l_i=0,1,...,\Delta_\nu-1$ ($\Delta_\nu=2^{\nu-1}$), then the sum of complex additions are equal to: $s(\frac{N}{2^\nu}\cdot2^{\nu-1})^n\cdot(n2^n)=nN^n\log_2 N$. The theorem is now proved.

References

1. Malozemov V. N., Macharskii S. M. Basises of diskrete harmonic analysis. S. 2. St. Petersburg : NIIMM. 2003.

2. Dudgeon D. E., Mersereau R. M. Multidimensional Digital Signal Processing : Transl. from English. M. : Mir, 1988.

3. Blahut R. Fast Algorithms for Digital Signal Processing : Transl. from English. M. : Mir, 1989.

4. Oppenheim A. V., Schafer R. W. Digital Signal Processing : Transl. from English. M. : Sviaz, 1979.

I. V. Toupitsyn
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

THE FAST ALGORITHM OF INTERMEDIATE PERSPECTIVES RECONSTRUCTION ON A STEREOPAIR

Here we have considered the effective A. A. Lukianitsa algorithm of stereo-reconstruction. We have offered a modification of the particular algorithm, which allows increase in the speed of the algorithm functioning. We have also presented the results of the conducted experiments.

Keywords: stereo image, epipolar geometry, corresponding points, disparity.

One of the up-to-date trends of video information ideas is supposed to be the task of presenting dimensional data. Due to this, devices which are able to produce volumetric images (multiperspective monitors, projection systems) are becoming more and more widespread. For the proper functioning of these devices it is necessary to obtain a certain number of perspectives of the reproduced scene; the number of which is defined by a certain task. For instance, for watching videos 4–6 perspectives are usually enough for comfort. For serious applications such as 3D-tomography and radiography, graphical work stations CAD/CAM, the representation of presented settings (control tower, search, and rescue services), etc, one could require from 40 to 150 perspectives. There are some special devices providing from 2 (stereo-video cameras) to 8 perspective settings. 8 perspectives video cameras were used in a prototype multiperspective *NIKFI* television system, and it is quite difficult to imagine a video camera with a large number of perspectives.

Equally difficult is considered the recording and transmitting of such a signal through connecting channels. Consequently, there is a need to reconstruct a required number of perspectives – i.e. images which have been obtained with the help of a similar camera, if it had an intermediate position between cameras used in stereo filming. The majority of algorithms solve this problem on two levels: searching for the corresponding points, i.e. points at the images fitting to the same spot of a three dimensional space, and the approximation of intermediate perspectives on the basis of the reveled appropriate points. Usually, if all the corresponding points, found at the approximation of the intermediate perspectives cause no difficulties the primarily object of the intermediate perspective reconstruction is the search for corresponding points. As an example let us examine the functioning of a simple stereovision algorithm. It consists out of several steps:

1. We have selected the first (upper) epipolar line as a current line.
2. For the current epipolar line we have selected the pixel most left of the current pixel.
3. We have matched the current pixel of the left image to every pixel of the right image.
4. We have selected the most similar pixel from the right image and wrote it in the array of the corresponding pixels. The array of corresponding pixels represents a 2D integer-valued array; the index elements of which are the positions of the pixels on the left image and the meanings of the elements – are the positions of the corresponding pixels of the right image.
5. If the current pixel is not the last in the active row then the following pixel is getting the current pixel and the algorithm goes to item 3. Otherwise, to item 6.
6. If the active epipolar line is not the last line, then the following epipolar line becomes active and goes to item 2. Otherwise the algorithm stops functioning.

The functioning of a simple stereovision algorithm is presented as a logical diagram (as shown in fig. 1). The result of the algorithm functioning is an array of corresponding pixels. Then, using this array one can build a depth card or reconstruct the intermediate perspectives, i. e. images which could be obtained with a similar camera in case it took intermediate positions between the cameras, used in stereo filming.

The reconstruction of intermediate views according to the A. A. Lukianitsa method. Setting the task. Suppose we have a stereo pair – i.e. two images of the same setting fixed done with a stereo camera. Let's mark the image taken by the left camera, with letter *L*, and the image of the right camera – with letter *R*. Let's introduce the orthogonal coordinate system coordinated with the cameras: we direct axis X to the right, through the image centers, axis *Y* – downwards along the vertical image axis and axis *Z* adds them to the right three. The starting point of the coordinate shall be chosen in such a way that the images are in the *Oxy* plane, and image centers are symmetrical according to plane *Oyz* (fig. 2). This setting is registered by two cameras: *L* and *R*. We need to build an intermediate perspective *M*.

Say the center of the left image in the chosen coordinate system has the coordinates (x_L ,0 ,0), and the center of the right is – (x_R ,0 ,0). The task is that an image should be built in the plane *Oxy* with its center in a point (x_M, 0, 0), closely situated to the image which was received by the camera the optical axis of which would pass through point (x_M, 0, 0). The suggested method consists of two levels:

– the location of the corresponding points on the left and the right images;

– the approximation of the intermediate image on the base of conjugated points and the filling of the fields for which the correspondence is not single-digit [1; 2].

The search algorithm for corresponding points. Say every image contained $K \times N$ pixels on the horizontal and vertical lines correspondingly. The task of finding the conjugated points is the following: for every pixel of the left image L_{ki}, $k = 1, \ldots, K$, $i = 1, \ldots, N$. We need to find pixel R_{kj}, $j = 1, \ldots, N$ at the right image corresponding with the same point of the registered settings.

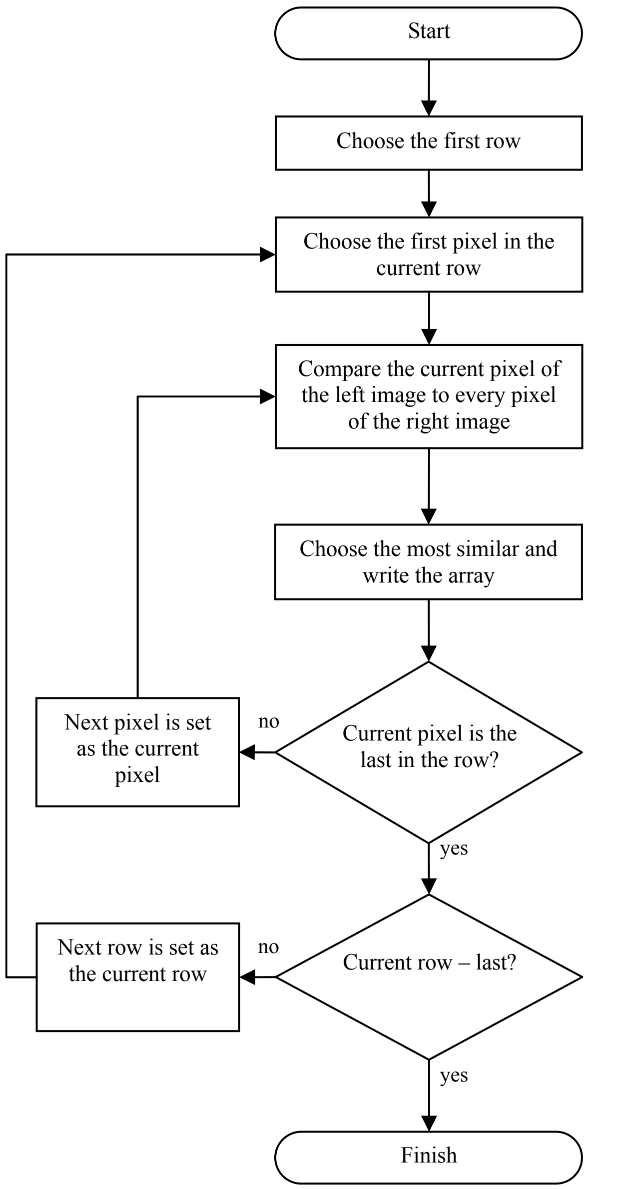


Fig. 1. The logical diagram of the stereovision functioning for a simple algorithm

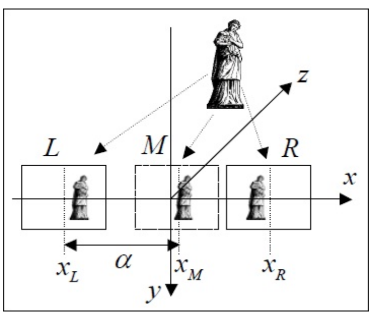


Fig. 2. Geometry of stereo filming

Suppose the cameras were fixed at the same horizontal line and have the same gain constant. Then the horizontal lines of the photo detectors coincide with the epipolar lines; that's why the conjugated point should be situated on the lines with the same number. This allows us to write down the following expression for calculation the epipolar lines:

$$L(k,i) \approx R(k, j + \Delta_j),$$

where Δ_j – is a shift, depending on the pixel number.

The finding of the corresponding points consists in the calculation of a displacement for all pixels. Further, for simplicity we shall omit the first index, because it has the same value for the compared images, and state that the process described below is executed for each pair of corresponding lines in the left and right images. To find expedient candidates for corresponding points we should compare some areas Ω in the environment of these pixels. It essentially increases the stability of the algorithm. Indeed, there may be differences in the images with a similar section of the scene under different angles, because of the mutual position of light sources and scenic elements. Besides, some areas can be visible for one camera and are invisible for another.

Let the compared area Ω have the following sizes $k \times n$, then in the environment of the j-th pixel on the right image we should scan a bigger sized area: $m \times l$, $m \geq k$, $l > n$. The distance function $d(i, j)$ is calculated during the scanning process. This function characterizes the proximity of images in area Ω_i of pixel $L(i)$ and in area Ω_j of pixel $R(j)$. The distance function is usually taken as a sum of differences for corresponding pixels intensities from Ω_i and Ω_j:

$$d(i,j) = \sum_{m=-k/n}^{k/2} \sum_{l=-n/2}^{n/2} \left| L(k+m, i+l) - R(k+m, j+l) \right|,$$

or as the comparing of the images from Ω_i and Ω_j gradients.

Practically, it is better to use a combination of these methods, because they have various efficiency for different types of images. Often, various researchers use a method of intensity comparison; however; when image processing with homogeneous vast objects, for example, a house wall, this method leads to the appearance of a plateau for function $d(i, j)$ that complicates a selection of corresponding points. In this case the gradient difference allows the considering of "thinner" image details.

To select the corresponding points from all possible pretenders we should meet two conditions. First of all, it's obvious that the index numbers of the corresponding points should minimize the total difference of the compared areas on the left and right images. Secondly, the offset value cannot decrease in the process of index j increase, i. e. $\Delta_j \geq \Delta_{j-1}$. This condition follows from the geometric properties of the stereo pair. Th next algorithm based on the dynamic programming method is offered to satisfy these conditions.

Let's introduce two accessory functions: real ψ and integer P. The first of these is necessary for the accumulation of total differences for compared areas. The second is reserved for the storage of appropriate indexes.

The dynamic programming algorithm consists of three steps:

a) Initialization:

$\psi(i,1) = d(i,1),\ i = 1, ..., N;$

$P(i,1) = i, \qquad i = 1, ..., N.$

b) Inductive passage:

$\psi(i, j) = \min_{1\le k\le i}\psi(k, j-1) + d(i, j),\ i = 2, ..., N,\ j = 1, ..., N,$

$P(i, j) = \arg\min_{1\le k\le i} \psi(k, j-1), \quad i = 2, ..., N,\ j = 1, ..., N.$

c) Reverse move:

$J(N) = \arg\min_{1\le i\le N} \psi(i, N),$

$J(j) = P(J(j+1), j+1),\ j = N-1, ..., 1.$

The array of indexes for corresponding points J at which for any pixel of the left image $L(i)$ the corresponding pixel of the right image is described as $R(J(i))$ will be constructed in result. The dynamic programming algorithm is presented in fig. 3.

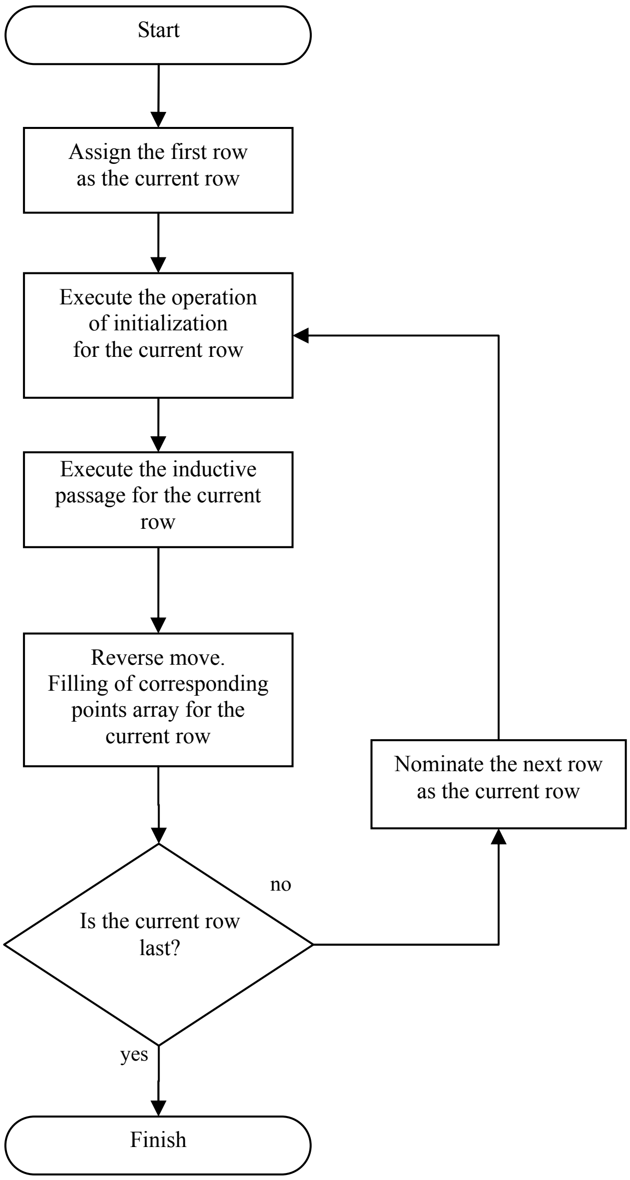


Fig. 3. The dynamic programming algorithm

This algorithm executes image processing line-by-line; this allows parallelizing calculations. It considerably speeds up stereo reconstruction procedures, but a consequence of progressive processing may be the effect of “combing”. Unfortunately this algorithm does not offer any methods of reducing these effects.

Let’s consider each step of the presented algorithm in detail. The initialization procedure serves for the definition of initial values of functions ψ and P. Function ψ is calculated as d function between the first pixel of the left image in the current line and each pixel of the right image in the current line. P function is calculated as pixel indexes of the right image.

The next step of the dynamic programming algorithm is the inductive passage. Here calculated are the subsequent values of functions ψ and P. Function ψ accumulates the total values of the compared areas. The total difference accumulation of compared areas increases the accuracy of finding corresponding points.

The last step of the dynamic programming algorithm is a reverse move. At this step the corresponding array point J is filled. The pixels of the right image, which have a minimal difference between each other, are selected as corresponding points for the pixels on the left image.

Intermediate view approximation

Let’s introduce parameter $\alpha = \frac{x_M - x_L}{x_M - x_R}$, which characterizes the relative degree of proximity for the reestablished view on the left image. Further for all values of $i = 1,\ldots,N, j = 1,\ldots,N$ we will calculate:

$$k = [\alpha j + (1 - \alpha)J(j)],$$
$$M(i,k) = \alpha L(i, j) + (1 - \alpha)R(i, J(j)).$$

As we can see from these correlations some points of the intermediate view can be unfilled. This happens because the corresponding fragments of the three-dimensional object are visible for one camera, but invisible for the other. Let it be that in the reconstructed image in any i-th row pixels $j_0, \ldots, j_0 + m$ are unfilled. There are two cases:

– if $J(j_0 - 1) = J(j_0 + m + 1)$ then these pixels are visible only for the left camera and they are filled by proper values from the left image:

$$M(i, j) = L(i, j),$$
$$j = j_0, \ldots, j_0 + m;$$

– if $J(j_0 - 1) < J(j_0 + m + 1)$ then these pixels are visible only for the right camera and they are filled by respective values from the right image:

$$M(i, j) = L(i, J(j)),$$
$$j = j_0, \ldots, j_0 + m.$$

Increasing the speed of the algorithm. Though the speed of this algorithm is high enough, it can be insufficient for the solution of some tasks. Consequently, it’s necessary to find a way of increasing the presented algorithm speed.

As a rule, in all stereo algorithms the CPU time for most of the part is occupied by the calculation of the distance function for the row pixels. To minimize the time spent on the distance function calculation it is necessary to simplify the function or decrease the quantity of pixels for which it is calculated. One of the possible variants consists in decreasing the size of the frame around the selected pixels but this does not always give good results, and is not applicable for all images.

Apparently, from the aforesaid, the simplification of the distance function is not possible; therefore, it is necessary to reduce the number of computations for this function. The improvement of the discussed algorithm is linked with the introduction of an additional variable: maximal disparity D_{max}. Maximal disparity is the maximal difference between the offset of a point in three-dimensional space on the left and right images. This variable is set by the user for each stereo pair. The value of this variable can be calculated if the location of the cameras and the distance from the stereo photography plane to the maximal distant object of the scene is known. Let's rewrite the expressions used on steps of the initialization, the inductive passage, and the reverse move; given the setting of maximum disparity in the following form:

1. Initialization:

$$\psi(i,1) = d(i,1), \quad i = 1,...,N;$$

$$P(i,1) = i, \qquad i = 1,...,N.$$

2. Inductive passage:

$$\psi(i,j) = \min_{D_{\max} \le k \le i} \psi(k, j-1) + d(i,j), \; i = 2,...,N, \; j = 1,...,N,$$

$$P(i,j) = \arg\min_{D_{\max} \le k \le i} \psi(k, j-1), \quad i = 2,...,N, \quad j = 1,...,N.$$

3. Reverse move:

$$J(N) = \arg\min_{D_{\max} \le i \le N} \psi(i,N),$$

$$J(j) = P(J(j+1), j+1), \quad j = N-1,...,1.$$

In result of the entered changes, the speed of the algorithm considerably increases because now there is no necessity to compute the value of the distance function for each pair of pixels in a row. The presented algorithms were realized in *C*++ language. The realized experimental algorithms have been carried out (sce table). The time for locating corresponding points of one line has been measured. The distinctions in the results of algorithm operating have been evaluated. Comparison has been carried out on a test stereo pair with a 10 pixel width (fig. 4).

From table it is apparent that the modified algorithm has a higher speed. However, the quantity metric of truly defined corresponding pixels worsens by 2–6 % in comparison to A. A. Lukjanitsy's base algorithm [2].

In this article the stereo reconstruction algorithm of Lukianitsa A. A. had been overlooked. The speed of this algorithm was also estimated. Modifications, which increase the speed of this algorithm, were offered. The modified algorithm was realized in high-level programming language (*C*++). In result of the conducted experiments, the operation speed of the modified algorithm at insignificant deterioration of stereo reconstruction was registered.

References

1. Lukianitsa A. A. Effective Algorithm for Reconstruction of Intermediate Views from Stereo pair // Graphicon 2006 : Materials of Intern. conf. Novosibirsk, 2006. P. 15–18.

2. Bobick A. F., Intille S. S. Large occlusion stereo // Int. Journal of Computer Vision. 33(3). 1999. P. 181–200.

Experiment results

Stereo pair images width	Frame of compared areas size	Maximal disparity	Selected algorithm	Time of algorithm
472	5x5	–	Lukianitsa	0.343 s
		10	Modified	0.047 s
		20		0.047 s
716		–	Lukianitsa	1.311 s
		20	Modified	0.156 s
		40		0.265 s
3295		–	Lukianitsa	93.102 s
		40	Modified	5.304 s
		100		11.949 s

a *b*

Fig. 4. Test stereo pair :
a – left image; *b* – right image

E. N. Vasilyev, V. V. Derevyanko
Institute of Computational Modelling Russian Academy of Sciences, Siberian Branch,
Russia, Krasnoyarsk

MATHEMATICAL MODEL OF HEAT EXCHANGE PROCESSES IN HONEYCOMB PANELS WITH HEAT PIPES

This article presents the analysis results of heat exchange processes in the honeycomb panel and depicts the results of temperature modelling modes for intensive heat loading modes.

Keywords: Honeycomb panel, heat exchange, mathematical modeling.

Nowadays honeycomb panels (HP) are widely used in the design of spacecrafts. Honeycomb panels are characterized by high mechanical durability and very small density. They are used as spacecraft construction units. Electronics packages, assemblies, and heat pipes are placed on them. Normal functioning of electronics requires a special temperature mode corresponding to the external environment. The small density volume of the honeycomb panel construction leads to low heat-transmitting properties of the panel. As a result it's difficult to organize an effective heat rejection from electronics and to secure optimal thermal modes for them. The heterogeneous structure of a honeycomb panel makes the process of computations of the heat exchange and analysis of heat modes of electronics complicated.

Different mathematical models are used for computing spacecraft temperature conditions. One of the most widespread approaches for spacecraft heat exchange process description is the use of heat-balance equations. Unsteady heat mathematical models for this approach had been considered in [1]. Research [2] contains a description of a mathematical model for thermal conditions of devices located on a honeycomb panel. This paper presents a mathematical model of the heat exchange in the honeycomb panel with heat pipes. The model is based on the numerical solution of unsteady heat conduction equations using a finite-difference space splitting scheme. The high efficiency of the methods permits to increase the level of detailing during temperature fields computations. The model is intended for computations of unsteady heat modes of electronics, optimization of the composition, and properties of the honeycomb panel, and also for the optimization of the quantity and arrangement of the heat pipes on the honeycomb panel.

Heat exchange process analysis. A honeycomb panel is a plain panel which consists of two parallels plates (hems *1* and *2*). The space between the plates is filled by honeycombs *3* made of metallic foil (honeycomb-filling) (fig. 1). Honeycombs have low heat-exchange abilities because of small density volume and therefore, heat pipes are used to improve the heat transfer. The highly effective heat-exchange in the heat pipes is reached due to circulation and phase changes of the coolant in the internal duct *4*. Heat pipes are located internally in honeycomb. They are fixed to the hem using landing ground *5*. Heat exchange in the honeycomb is significantly different from the heat transfer in a solid metal panel. First of all, the heat exchange in honeycombs is three-dimensional and anisotropic due to the presence of honeycomb-filling and heat pipes. Secondly, heat transfer in the honeycomb can be performed both by thermal conductivity and radiation.

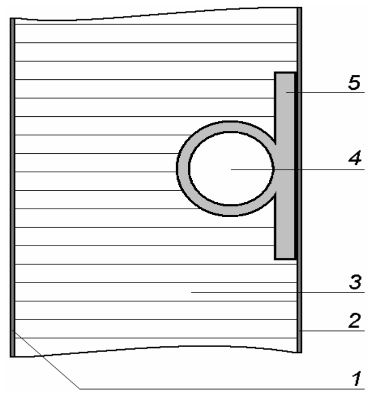


Fig. 1. Fragment of a honeycomb panel with a heat pipe

To determine the process features, let us consider the importance of heat exchange in the volume of a honeycomb. Typically, the density of a honeycomb-filling is approximately 100 times less than the density of material from which the filling is made. Therefore, effective value of heat conduction coefficient λ^* in the transverse location ("hem *1*–hem *2*") is only 0.01 of the heat conduction coefficient value of solid material. If the value of thermal flux is 400 Wt/m^2 and the honeycomb-filling thickness is d = 0,03 m, then the value of the temperature drop is $\Delta T = qd / \lambda^* \approx$ 8K. The density of radiant flux between hem *1* and hem *2* can be estimated using the following formula: $q_R = \varepsilon\sigma_{SB}(T_1^4 - T_2^4)$; where σ_{SB} is the Stefan-Boltzmann constant. Assuming that ε = 0,2, T_1 = 308 K, and T_2 = 300 K, we get $q_R \approx 1$ Wt/m^2. This value is a negligible quantity in comparison with the value of the heat flux which is transferred by the thermal conductivity.

The insignificant hem thickness and the low density of honeycomb-filling leads to low honeycomb-filling thermal conductivity in longitudinal directions. At the same time the effective heat conduction coefficient of honeycomb-filling is 5–10 times less than the one on the hems. Thus, one may neglect heat transfer in longitudinal direction for approximate computations. In more exact computations it should be considered as an addition to the heat conductivity of the hems. In general, the addition to the head conduction coefficient depends on direction of the honeycomb-filling's anisotropy. The size of honeycomb has an order of several millimeters, so in the model it is a continuum, which has thermal resistance dependant on honeycomb's parameters.

The estimations witness that heat transfer in the longitudinal direction in honeycombs is carried out mainly on the hems due to the thermal conductivity. The heat transfer in crosswise direction between hems is achieved by thermal conductivity of honeycomb-filling's material. In the presented computational model, the honeycomb panel is considered as two plain solid planes. These planes exchange thermal energy between each other through thermal resistivity of the honeycomb-filling. Radiation is taken into account on the external plane of the honeycomb in the process of heat exchange with the environment.

Mathematical problem definition and computational algorithm. In order to examine the temperature mode of a honeycomb panel, a mathematical model represents a system of two two-dimensional unsteady equations with variable coefficients, which give heat conduction in both hems:

$$c_m(x,y)\frac{\partial T_i}{\partial t}=\frac{\partial}{\partial x}\left[\lambda(x,y)\frac{\partial T_i}{\partial x}\right]+ \\ +\frac{\partial}{\partial y}\left[\lambda(x,y)\frac{\partial T_i}{\partial y}\right]+q_v(x,y,t)-\alpha_v(x,y,t)T_i, \quad (1)$$

where c_m is the material specific volume heat, T_i is the temperature, i = 1, 2 is the index corresponded to the number of the hem, λ is the heat conduction coefficient, t is time, x, y are space coordinates, q_v is the volume power of heat sources and sinks, α_v is coefficient of heat exchange with environment.

Equations (1) are completed by boundary conditions:

$$\left[\lambda\frac{\partial T}{\partial l}+\alpha T\right]_{l=0,L}=q\big|_{l=0,L},$$

where $l = x, y$, and entry conditions:

$$T\big|_{t=0}=T_0(x,y).$$

The value of the heat flux, which proceeded from one hem to another one, is defined as:

$$q(x,y)=\frac{T_1(x,y)-T_2(x,y)}{R}, \quad (2)$$

where T_1 and T_2 are temperatures of hems *1* and *2*, R is the specific thermal resistance of the honeycomb-filling. Equation (2) describes the heat exchange between hems; this has been taken in account in equation (1) with the help of last two members. For example, in order to compute the temperature field of the first hem, the first member is computed as $q_v(x,y)=T_2(x,y)/Rd_1$, the second member as $\alpha_v(x,y)=1/Rd_1$, where d_1 is the thickness of the first hem. Similar expressions are used for the second hem. This algorithm of heat exchange computations between honeycomb panel hems provides stability of the whole computational algorithm. Moreover, thermal flux from electronics, external thermal fluxes and radiation from hems' surfaces is also included in q_v.

The mathematical model takes into account the heat pipes on the hems of honeycomb as zones with high unilateral thermal conductivity and its value of heat conduction coefficient is set for the axial direction, which is selected to be equivalent to the heat transport ability of the heat pipe. The thermal conductivity of the heat pipes' landing ground is also taken into account because it can significantly influence the heat transfer in a crosswise direction.

One of the main criteria when selecting a computational method to solve the problem is the efficiency and stability of the algorithm. Moreover, the nature of the task requires an algorithm which allows using one directional component of heat conduction coefficients λ_x and λ_y which are different because of the heat pipes presence. The algorithm of summary approximation [3] suits these requirements using task spatial coordinates splitting. A two-dimensional unsteady problem is solved in two steps at each time step. At every step a local one-dimensional problem is solved with the help of the implicit difference scheme.

A grid function $T^j_{n,m}$ and intermediate function $U^j_{n,m}$ are introduced. They correspond to the values of the temperature in the nodes of the computational coordinate grid:

$$x_n=(n-1)h_x,\; y_m=(m-1)h_y,\; t_j=(j-1)\tau.$$

At the first step, a local one-dimensional problem is solved. It takes into account only the λ_x component of the heat conduction coefficient. The problem is solved for each y_m in the direction x. In result, all values of the intermediate function $U^j_{n,m}$ are determined.

For $n = 1$:

$$U^j_{2,m}-\left(1+\frac{\alpha_{0,x}h_x}{\lambda_{3/2,m}}+\frac{(c_v\rho)_{1,m}h_x^2}{2\lambda_{3/2,m}\tau}+\frac{\alpha_{V1,m}h_x^2}{4\lambda_{3/2,m}}\right)U^j_{1,m}+\frac{q_{0,x}h_x}{\lambda_{3/2,m}}+ \\ +\frac{h_x^2}{2\lambda_{3/2,m}}\left(\frac{q_{V1,m}}{2}+\frac{(c_v\rho)_{1,m}}{\tau}T^{j-1}_{1,m}\right)=0, \quad (3)$$

For $n = 2, ..., N–1$:

$$U^j_{n+1,m}-\left(1+\frac{\lambda_{n-1/2,m}}{\lambda_{n+1/2,m}}+\frac{(c_v\rho)_{n,m}h_x^2}{\lambda_{n+1/2,m}\tau}+\frac{\alpha_{Vn,m}h_x^2}{2\lambda_{n+1/2,m}}\right)U^j_{n,m}+ \\ +\frac{\lambda_{n-1/2,m}}{\lambda_{n+1/2,m}}U^j_{n-1,m}+\frac{h_x^2}{\lambda_{n+1/2,m}}\left(\frac{q_{Vn,m}}{2}+\frac{(c_v\rho)_{n,m}}{\tau}T^{j-1}_{n,m}\right)=0. \quad (4)$$

For $n = N$:

$$-\left(1+\frac{\alpha_{0,x}h_x}{\lambda_{N-1/2,m}}+\frac{(c_v\rho)_{N,m}h_x^2}{2\lambda_{N-1/2,m}\tau}+\frac{\alpha_{VN,m}h_x^2}{4\lambda_{N-1/2,m}}\right)U^j_{N,m}+U^j_{N-1,m}+ \\ +\frac{q_{L,x}h_x}{\lambda_{N-1/2,m}}+\frac{h_x^2}{2\lambda_{N-1/2,m}}\left(\frac{q_{VN,m}}{2}+\frac{(c_v\rho)_{N,m}}{\tau}T^{j-1}_{N,m}\right)=0. \quad (5)$$

At the second step a similar local one-dimensional problem, which takes into account only the λ_y component of heat conduction coefficient, is solved. The problem is to solve x_n in each the direction of y. In result, all values of the grid function $T_{n,m}^{j+1}$ on the next time step are determined.

Algebraic equations (3)–(5) are solved by a sweep method. Entry parameters of the program include: the geometrical characteristics of honeycomb panel, materials, their thermo physical properties, and electronic heat generation parameters.

Results of computations. The choice of differential steps sizes is very important in the computations. In order to obtain adequate computational results without significant smoothing of the temperature gradients, the step size shouldn't exceed the typical scale of temperature fields' heterogeneousness. In the present task, the typical scale is defined by heat pipes. So, the lateral dimension of the heat pipes, which is approximately equal to 10 mm, defines the value of the differential spatial steps as $h_x = h_y = 10$ mm.

First, specificities of the heat exchange in a honeycomb panel had been examined on the model task in which the honeycomb panel had the following parameters: lateral dimensions of the honeycomb panel were 600 × 300 mm^2, hem thickness was 0.4 mm, the distance between hems was 30 mm, specific heat resistance of the honeycomb-filling was R = 10 m/Wt. The external hem of the honeycomb panel was a radiant surface. This surface emits a heat flux which corresponds to the radiation of outer space with emissivity equal to $\varepsilon = 0.7$. There is a heat-generating device on the left half of the internal hem. Its slot sizes are 300 × 300 mm^2. The power of the device is 50 Wt. It's assumed that the heat flux from the device is spread uniformly in the area of the slot. In order to improve the heat transfer, two heat pipes had been installed on the honeycomb panel.

The duration of the modeling process had been selected in such a way that the process had time to achieve a steady-state regime. The computation results are presented in fig. 2 and 3. The temperature fields for the internal and external honeycomb panel hems are marked by isolines and gradations of a gray color. The temperature interval between the isolines is 1 $^{\circ}$C. As shown in fig. 2, the heterogeneity of temperature distribution in the internal hem is high. Maximal temperature drop in the hem is about 26 $^{\circ}$C. This is due to low heat-exchange abilities of the hem in the normal direction of the heat pipes. The temperature range in the external honeycomb panel hem is more even (fig. 3). Its maximal temperature drop is approximately equal to 20 $^{\circ}$C.

Significant heterogeneities in temperature distributions make it almost impossible to ensure reliable control of optimal electronic mode temperature. In order to make the temperature range more even, special heat stabilized plates are placed on the honeycomb panel hem. The plates are made from a metal with a high heat conductivity coefficient. Temperature drop is decreased proportionally to the plate thickness. For example, using aluminum plates with a thickness of 6 mm, leads to a decrease in temperature drop to 2 $^{\circ}$C.

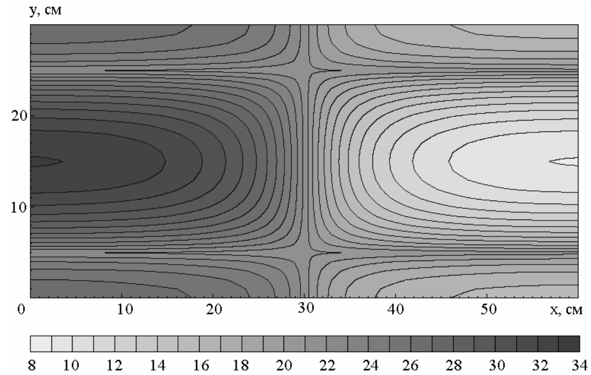


Fig. 2. Temperature field of a honeycomb panel internal surface

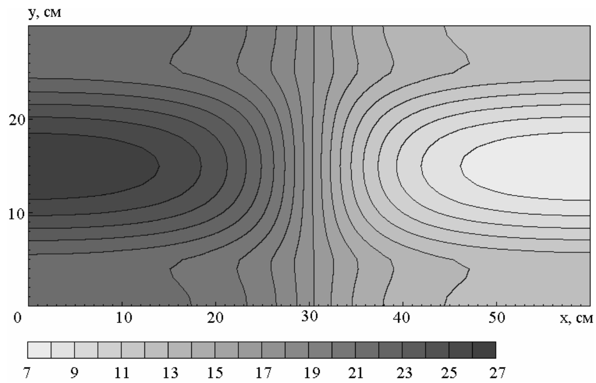


Fig. 3. Temperature field of a honeycomb panel external surface

The increase of the plates' thickness leads to a significant increase of the whole construction's weight. In the future, it seems reasonable to use hyper heat-conducting materials for increasing heat-transmitting capabilities, instead of heat stabilized plates. The results of the computations show, that hyper heat-conducting plates with a thickness of 2 mm and λ_{ef} = 5 000 Wt/(m×K) provide almost homogeneous temperature ranges in the internal hem, with temperature drops of less than 0,5 $^{\circ}$C, because the high heat-transmitting ability of hyper heat-conducting plate effectively levels the influence of heat from devices and external heat fluxes on radiating surfaces.

Real honeycomb panel constructions can contain a sufficiently larger number of heat-generating devices and heat pipes. An example of the computations is presented on the fig. 4; in it the distribution of temperature on the internal hem of honeycomb panel is presented. The boundaries of device slots are outlined in a white color. Given temperature ranges with heterogeneous "spotty" structure is a result of joint action of different mechanisms: the device's heat generation, heat transfer in the hems, and heat pipes, heat-exchange between the hems and radiation from the external hem. In our opinion, the applied mathematical models are the most actual for the computation of the modes in the complex design of honeycomb panels, because these computational experiments permit to obtain the most detailed information about thermal mode results in various external conditions.

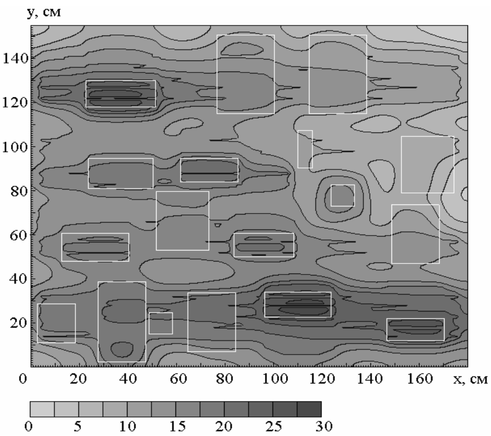


Fig. 4. Temperature field of a thermal-loaded honeycomb panel internal surface

It should be noticed that computational experiments for honeycomb panels with complex construction require a significant volume of different input information like thermophysical characteristics and geometric parameters of the constructions (sizes, coordinates of the device slots), mass, etc. The collection of such information requires much time. The time consumption increases significantly if it's necessary to make a number of computational experiments for different honeycomb panel constructions in order to find the optimal temperature mode. To decrease the consumption it's reasonable to automatize the process of input information gathering. The information should be read from the CAD-system database. To realize this approach, we are currently developing a software complex for the computation of honeycomb panel thermal modes on the base of the presented mathematical model. This complex is integrated with a CAD-system. It will allow significant simplifying of the input procedure, output, edition, and visualization of the input and calculated data.

The developed computational model allows the conducting of computations for honeycomb panel hems' temperature fields, considering the detailed information on the honeycomb panel design, heat generation of electronics, and external thermal conditions. In order to find the optimal thermal modes for electronics, it's possible to use this model to optimize the honeycomb panel's construction, and to select the most appropriate arrangements of the devices and heat pipes in the panel's hems.

References

1. Unsteady thermal conditions of spacecrafts of satellite system / M. V. Kraev, O. V. Zagar, V. M. Kraev, K. F Golikovskaja ; SibSAU. Krasnoyarsk, 2004.

2. Alekseev N. G., Zagar O. V., Kasjanov A. O. The system of maintenance of a thermal mode of the device with regulation of temperature in a narrow range // Reshetnev's readings: Proc. of XI Intern. scientific conf. Krasnoyarsk, 2007. P. 213.

3. Samarskij A. A. Theory of difference scheme. M. : Nauka, 1989.

S. V. Vohmyanin
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

TESTING THE ALGORYTHM OF THE "CATERPILLAR"-SSA METHOD FOR TIME SERIES RECOVERY

The basic algorithm of the "Caterpillar"-SSA method is considered and tested.

Keywords: trend allocation, periodicals finding, silencing, decomposition of time series into components.

One of the significant problems in the analysis of time series is the separation of trend and periodicals presses from the noise. This research is about a robust method of time series analysis: "Caterpillar"-SSA, which is currently being developed.

Let's investigate the functioning of this algorithm and state, in what its specificity is exactly. The variant of the algorithm described below doesn't essentially differ from the basic one [1], it has only been simplified without any changes in result.

We consider the given time series F:

$$f_0, f_1, ..., f_{N-1}, \tag{1}$$

where N is its length. Further we assume that F is a nonzero series.

The algorithm consists of four consistent steps: investment, singular decomposition, grouping, and diagonal averaging.

The investment procedure converts the time series F into a sequence of multidimensional vectors called the trajectory matrix.

To analyze the time series we select parameter L called "the length of period", which is in the open interval $1 < L < N$. Thus $K = N-L-1$ investment vectors are created:

$$X_i = (f_{i-1}, f_i, ..., f_{i+L-2})^T,\ 1 \le i \le K. \quad (2)$$

These vectors form the trajectory matrix of the series F the columns of which are the sliding parts of the series with length L: from the first point to L-th, from the second to $(L + 1)$-th and so on:

$$X = [X_1 : X_2 : ... : X_K] = \begin{pmatrix} f_0 & f_1 & ... & f_{K-1} \\ f_1 & f_2 & ... & f_K \\ ... & ... & ... & ... \\ f_{L-1} & f_L & ... & f_{N-1} \end{pmatrix}. \quad (3)$$

It's known that univocal conformity exists between matrixes of dimension $L \times K$ like (2) and the series (1) of length $N = L + K-1$ [1].

The result of the following step will be a singular decomposition of the trajectory matrix (2) in the sum of elementary matrixes.

Let $S = X \cdot X^T$. We will assign the eigenvalues of matrix S taken in nondecreasing order as $\lambda_1, \lambda_2, \ldots, \lambda_L$, and the orthonormal system of eigenvectors of matrix S, corresponding to ordered eigenvalues, such as $U_1, U_2, ..., U_L$. Then the singular decomposition of trajectory matrix X is to be written as the following expression:

$$X = \sum V_i, \quad (4)$$

where $V_i = U_i \cdot U_i^T \cdot X$, $I = 1, \ldots, L$. Considering that each of the matrixes V_i to have rank 1, they can be denoted as elementary matrixes.

The initial time series is assumed to be a sum of several series. The results allow us under certain conditions, to define, according to the form of the eigenvalues and the eigenvectors, what kind of items they are and what combination of elementary matrixes corresponds to each of them.

At the next stage there is a grouping, by decomposition (3) the set of indexes $\{1, 2, \ldots, L\}$ is divided into m non-crossing subsets $I_1, I_2, ..., I_m$. Thereby the decomposition (3) can be written down as:

$$X = \sum_{i=1}^{m} Y_i, \quad (5)$$

where $Y_i = \sum_{k \in I_i} V_k$ are the resultant matrixes for each subset I_i, $I = 1, \ldots, m$.

Actually, precisely at the grouping stage, the initial time series is divided into periodicals, noise, and trend. The basic criterion of the grouping is the importance of each elementary matrix V_k, to be corresponding to its eigenvalue λ_k.

At the last stage of the algorithm each matrix of grouped decomposition (4) is converted into a series of length N.

Let $L^* = \min(L, K)$, $K^* = \max(L, K)$. Also let $y^*_{ij} = Y_{ij}$, if $L < K$ and $y^*_{ij} = Y_{ji}$, if $L > K$. Diagonal averaging converts each resultant matrix $Y^{(s)}$, $s = 1, 2, \ldots, m$, into series $\tilde{f}^{(s)}$ with the help of the following formula:

$$\tilde{f}_k = \begin{cases} \dfrac{1}{k+1} \sum_{n=1}^{k+1} y^*_{n,k-n+2}, & 0 \le k < L^* - 1, \\ \dfrac{1}{L^*} \sum_{n=1}^{L^*} y^*_{n,k-n+2}, & L^* - 1 \le k < K^*, \\ \dfrac{1}{N-k} \sum_{n=k-K^*+2}^{N-K^*+1} y^*_{n,k-n+2}, & K^* \le k < N. \end{cases} \quad (6)$$

This formula corresponds to the averaging of the elements along "diagonals" $I + j = k + 2$.

Thus, applying diagonal averaging (5) to resultants matrixes $Y^{(s)}$, we get a series $\tilde{F}^{(s)} = (\tilde{f}_0^{(s)}, \tilde{f}_1^{(s)}, ..., \tilde{f}_{N-1}^{(s)})$. The initial time series F is decomposed into the sum of m series:

$$F = \sum_{n=1}^{m} \tilde{F}^{(s)},\ f_i = \sum_{n=1}^{m} \tilde{f}_i^{(s)},$$
$$n = 0, 1, \ldots, N-1,\ s = 1, 2, \ldots, m. \quad (7)$$

So, the result of the algorithm is the decomposition of the time series into interpreted additive components. For all this it doesn't require stationarity from the series, knowledge of the trend modelor, or any data about the presence of periodicals in the series and their periods. With such simple assumptions, the "Caterpillar"-SSA method is able to solve various tasks, such as trend allocation, detection of periodical presses, number smoothing and the construction of the full decomposition of the series into the sum of trend, periodicals and noise [2].

Certainly, the given method also has some disadvantages. First of all, there isn't an automatic grouping of the components of singular decomposition of the trajectory matrix to get the components of the initial series. At the same time successful decomposition depends on the correct grouping. Secondly, the absence of a model doesn't allow to prove the hypothesis about the presence of this or that component in the time series (this disadvantage is objectively inherent in non-parametric methods). We should also state that the considered non-parametric method in certain situations permits us to obtain the results, which frequently slightly differ in accuracy from many parametrical methods in the analysis of the series with the defined model [3].

Let's look at the algorithm work on three various examples to investigate its advantages and disadvantages. There is a time series in the each example, consisting of the sum of the generated interferences R_i and given required function x_i:

$$f_i = x_i + R_i.$$

Further, we define the criterion of efficiency by the formula:

$$W = \frac{\sum (A_i - x_i)^2}{\sum (R_i)^2} \cdot 100\ \%,$$

where A_i is a restored (cleared of noise) series achieved with the help of the algorithm. In (7) the numerator is the sum of squares of deviations between restored series and "clear" series, and the denominator is the sum of squares of interferences. So, formula (7) shows the parts of the interferences are not separated after the application of the algorithm; we shall call it "silencing".

Example 1. A simple time series; weak interferences:

$$x_i = I + 10;\ I = 0, 1, \ldots, 49;\ N = 50;\ L = 25.$$

R_i is a random value with uniform distribution from the interval [–2; 2].

Matrix S has dimensions 25 × 25 and 25 eigenvalues λ_i (tabl. 1).

The grouping of indexes 24-th and 25-th is chosen, as corresponding to the most significant components. Elementary matrixes V_{24} and V_{25} correspond to them. Calculating diagonal averaging for resultant matrix $Y^o = V_{24} + V_{25}$, we get the restored series (fig. 1).

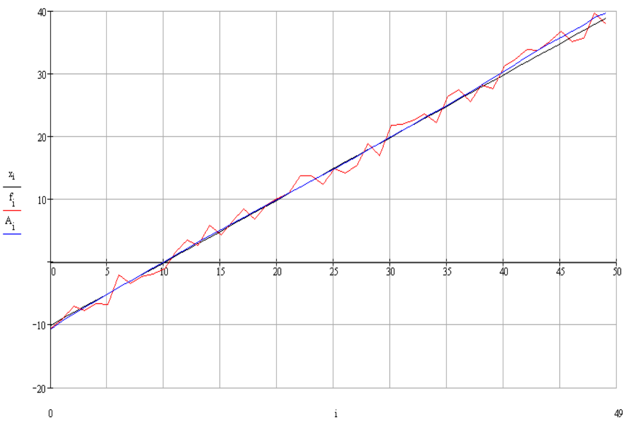


Fig. 1. Graphs of series: "clear", with noise and restored

Noise clearing is $W = 11.4\ \%$ of the initial interferences.

Example 2. A series with periodicals, average interferences:

$$x_i = \frac{i(i-60)}{100} + 5\sin(i);$$
$$I = 0, 1, \ldots, 59;\ N = 60;\ L = 30.$$

R_i is a random value with uniform distribution from the interval [–3; 3].

Matrix S has dimensions of 30×30 and 30 eigenvalues λ_i (tab. 2).

Grouping of ones indexes from 27-th to 30-th is chosen, as corresponding to the most significant components. Elementary matrixes V_{27}, V_{28}, V_{29} and V_{30} correspond to them. Calculating the diagonal averaging for resultant matrix $Y^o = V_{27} + V_{28} + V_{29} + V_{30}$, we get the restored series (fig. 2).

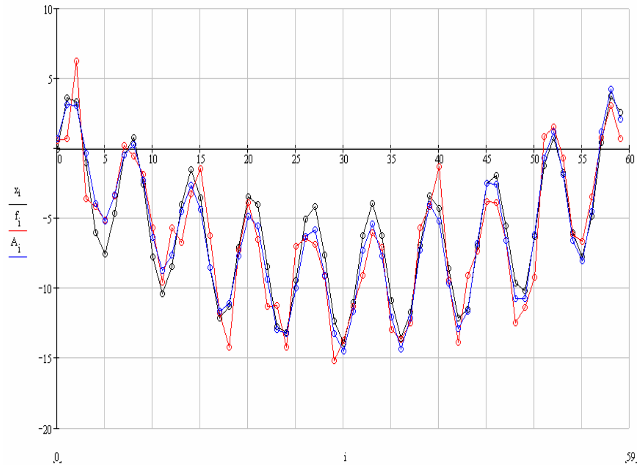


Fig. 2. Graphs of series: "clear", with noise and restored

Noise clearing is $W = 25.6\ \%$ of initial interferences.

Example 3. A series with several periodicals, high interferences:

$$x_i = 0.03i + 1.6\sin(0.3i + 0.17) + 1.3\sin(2i + 0.57);$$
$$I = 0, 1, \ldots, 49;\ N = 50;\ L = 15.$$

R_i is a random value with normal distribution, $\sigma = \sqrt{3}$.

Matrix S has dimensions 15 × 15 and 15 eigenvalues λ_i (tab. 3).

In this situation due to the high interferences the choice of components for the grouping is inconvenient, and to recognize a trend and periodicals is difficult. The analysis has shown that the increase in index quantity in a similar situation results in the restoring of not only an additive component, but also that of non-separated noise.

Table 1

The contribution of eigenvalues λ_i of matrix S, in percentage of their sum

i	1	2	3	4	5	6	7	8	9	10	11	12	13
λ_i, %	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,01	0,01	0,00	0,01	0,01	0,01
i	14	15	16	17	18	19	20	21	22	23	24	25	
λ_i, %	0,02	0,02	0,02	0,02	0,03	0,03	0,03	0,04	0,08	0,08	2,76	96,8	

Table 2

The contribution of eigenvalues λ_i of matrix S, in percentage of their sum

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
λ_i, %	0,03	0,03	0,03	0,03	0,04	0,05	0,02	0,07	0,07	0,07	0,08	0,01	0,00	0,00	0,12
i	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
λ_i, %	0,11	0,12	0,12	0,17	0,20	0,21	0,21	0,22	0,26	0,29	0,33	4,14	5,97	7,58	79,44

Table 3

The contribution of eigenvalues λ_i of matrix S, in percentage of their sum

i	1	2	3	4	5	6	7	8
λ_i, %	2,50	2,49	2,85	2,14	1,94	3,29	4,00	4,73
i	9	10	11	12	13	14	15	
λ_i, %	5,26	6,83	7,78	9,99	11,31	13,83	21,07	

Noise clearing with the 3 most significant components is $W = 21.8\,\%$, for 4 it's $W = 29.2\,\%$ and for 5 it's $W = 34.6\,\%$.

The results for 3 selected components are shown below as graphs (fig. 3).

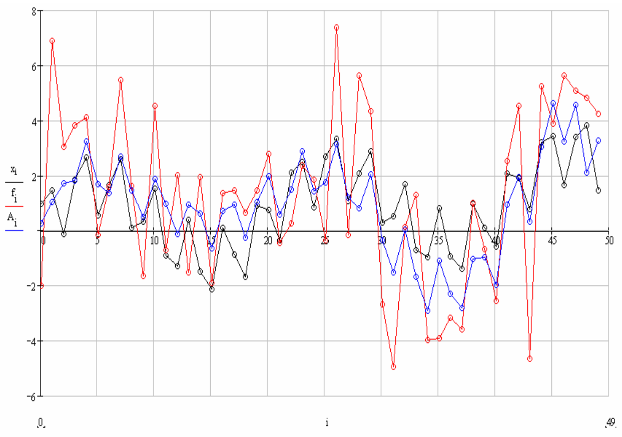


Fig. 3. Graphs of series: "clear", with noise and restored

Concluding the given examples we can state that the basic algorithm of the "Caterpillar"-SSA method copes with the assigned task: for time series it separates trend and periodicals from interferences, reducing noise level down to 2–3 times; although the types of significant components aren't defined, whether they are linear, periodic, logarithmic or other. This is an advantage of the method, which will make possible to create a powerful mechanism of non-parametric analysis of time series in the future, including computer programs.

The disadvantage of the basic algorithm is the necessity of manual intervention for the divided components analysis; also there is a problem in selecting the length of period and the quality of additive components division, depending on that. Further research will be dedicated to the automation of analyzing processes and other methods, improving the quality of the algorithm work results and reducing the manual aspect in this process.

References

1. Golyandina N. E. The method of "Caterpillar"-SSA: the analysis of temporal aisles : textbook. Saint-Petersburg, 2004.
2. The main components of temporal aisles: the "Caterpillar" method / under the editorial of D. L. Danilov, A.A. Zhigliavski. Saint-Petersburg : Presscom, 1997.
3. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of Time Series Structure: SSA and Related Techniques. London : Chapman& Hall/CRC, 2001.

Yu. Yu. Yakunin, A. A. Gorodilov
Siberian Federal University, Russia, Krasnoyarsk

ALGORITHMS FOR CALCULATING COMPLEX INDICATORS IN DYNAMIC STRUCTURES OF DATA REPRESENTATION

This paper presents algorithms for calculating complex indicators on set factual data, represented in dynamic structures with the application of the graph theory.

Keywords: dynamic structures of the data, tripartite graph, algorithm of graph's round, complex indicators.

The problem of rupture between scientific methods of representing (describing) real world objects and the storage of this information in information systems has existed for a long time and has not yet been solved satisfyingly. The database management system (DBMS) is the best system available today, which allows the storage of information in the form of objects [1; 2] (the object-oriented approach [3]) or globals [1] (hierarchical representation of the information in the form of a tree). However, even such an approach can capture only part of the variety represented in the information of modern scientific methods [4]. Such rupture substantially slows the development of science and engineering in the field of information technology.

According to this, the essential restriction for information system design is the standard way of data storage, which is based on static structures (i. e. for the description of a subjected field's objects in order to store information, a database of the data storage structure is created in advance). This results in the fact that such structures should be created by the designer of information system during a stage of its designing and it (this structure) cannot change during the development of this system and its maintenance. It is not necessary to speak about the expenses at which changes in the system come… It is obvious that if such changes are possible, even in an insignificant part of this structure, it would come at the same expense as the original production of the system, possibly even exceeding it. Any attempts of creating dynamical structures of data storage lead to problems of data transformation in one of the existing ways of their storage in DBMS (relational, objective, or hierarchical). Any of the listed ways is not capable of functioning directly with its dynamical structures. Frequently, such problems are solved individually in the process of their occurrence. As a result there are no general approaches for the solution of these problems; it is realized in some software products in the form of separate modules with similar problems, but these solutions are commercial secrets. Thus, there is a research problem in the description of dynamic data structures, information processing in these structures, and its storage in DBMS. There is also a separate problem of calculating complex indicators on a set of factual data in the dynamic structure, the solution of which is mentioned below.

For the description of dynamical data structure we have offered to allocate the following categories of the information [5]:

– indicators – quantitative characteristics of objects;

– qualifiers – structural formation of the data consisting of interconnected classes and their instance (objects);

– factual data – values of indicators concerning one or several qualifiers; to represent each category of information we will use corresponding graphs with descriptions.

The graph of indicators consists of two parts: the first – which represents a tree (fig. 1), containing nodes-categories (classification of indicators), and the second − a simple graph with any number of parts on the graph, the nodes of which represent indicators.

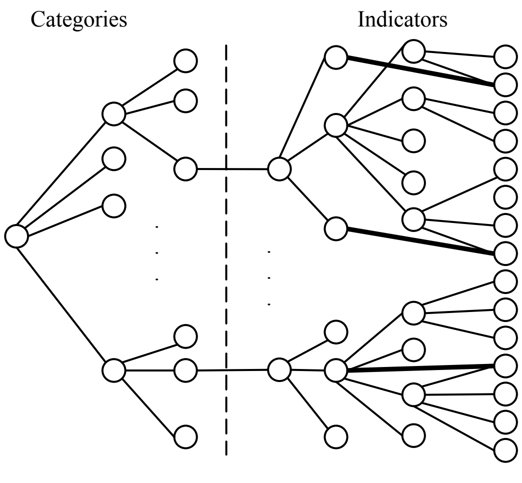


Fig. 1. The graph of indicators

In the rib section (from left to right) is the division of categories into subcategories. In the rib part indicator (from left to right) is the division of the aggregated indicators into more private and then to elementary indicators (indivisible). Fig. 1 shows that the aggregated indicators can consist not only of previous level indicators (k), but from indicators of the following level ($k + 1$), through level ($k + 2$), and so on.

The graph of qualifiers we shall consider is in the example of fig. 2. It can be clearly seen that the graph is divided into three levels, each of which carry own sense load. So level of categories is intended for splitting the qualifiers into integrated categories, i. e. the descriptions of a classification tree structure. At the qualifiers' level there are multiple nodes (qualifiers) with ribs between the category level, showing the qualifier attachment to a category; and ribs between qualifier nodes, showing the interaction between them. The node is characterized not only by name, but by a number of attributes.

The level of qualifier values, similarly to the previous level, contains multiple nodes; each of one has an obligatory rib depicting attachment of any qualifier's value to some of the unessential ribs, depicting the dependence between the values of different qualifiers. The nodes of the given level also are characterized by name and attribute values, corresponding to the nodes in the qualifier level with which they joined by ribs.

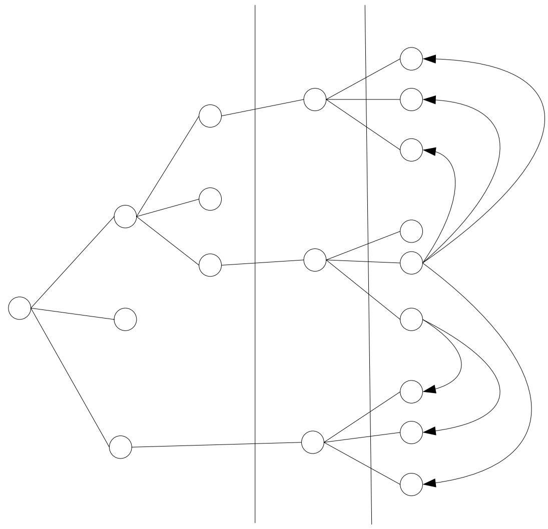


Fig. 2. Graph of qualifiers:
Cat – category of qualifiers; Q – qualifier; VQ – value of qualifiers

Let's draw an analogy graph representing the structure of qualifiers among with other ways of data presentation: for example, with the objective. As an example, we shall consider the organizational structure of some organization. We will allocate two essences in this example: management and department. The graph model of this example is presented in fig. 3, *a*, and the objective in fig. 3, *b*.

The factual data is the graph with a set of nodes, which do not have ribs between each other, but have ribs with qualifier and indicator nodes (fig. 4). Factual data nodes represent numbers, which are quantitative characteristics of one or several indicators in the relation with one or several qualifiers. For example, for the indicator "salary" and values of the qualifier "department of social work", the factual given value may equal X rubles a year.

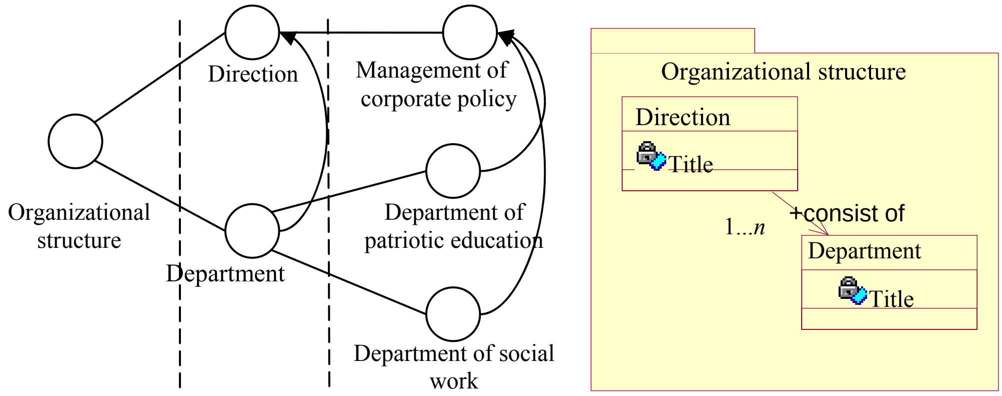


Fig. 3. Examples of representation structure:
a – graph view; *b* – objective view

Thus, the factual data nodes incorporating with nodes of qualifier and indicator graphs form a graph with difficult interrelations (fig. 5).

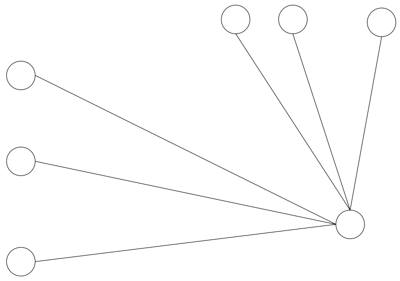


Fig. 4. Fragment from the factual data graph:
VQ – value of qualifier; I – indicator; FD – the factual data

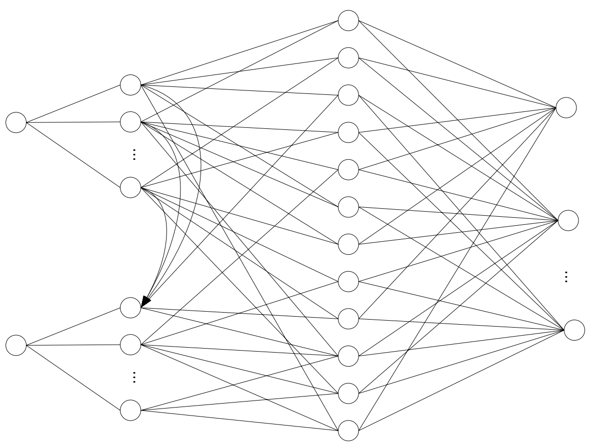


Fig. 5. The factual data graph:
Q – qualifier; VQ – value of qualifier; I – indicator;
FD – the factual data

In order to carry out analytical operations over the factual data for the purpose of calculating complex indicators from the presented graph (fig. 5), not all its nodes and ribs will need to be engaged. For example, in the organizational unit "management of corporate policy", necessary for calculating the total annual salary, we need the factual data sum values which are connected to nodes representing the management departments, and the node of the "salary" indicator.

Thus, from graph presented in fig. 5, only the factual data nodes and the qualifier counts nodes, along with indicators connected to them will be used. Besides, if for the initial data of the algorithm for calculating complex indicators will be used the set of qualifier value nodes and the set of indicator nodes for which it is necessary to make calculations; ribs between the qualifiers' value tops will be not necessary. Therefore, considering all the aforesaid, we shall simplify the factual data graph (fig. 5) and transform it into a tripartite graph (fig. 6). It represents three sets of nodes (X_1, X_2, and X_3); in each, the nodes have no ribs among each other, but have ribs between the nodes of the following sets [6].

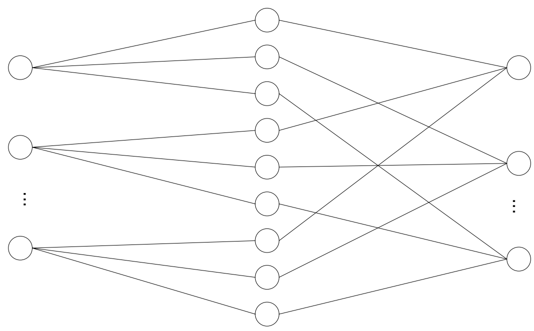


Fig. 6. Tripartite factual data graph:
VQ – value of qualifier; I – indicator;
X – set of nodes

The tripartite graph is a particular case of the *N*-partite graph (fig. 7). It has its own properties, distinctive from those of the simple graph for which there are already some algorithms of its traversal [7].

Such algorithms poorly work on *N*-partite graphs as they are focused on achieving different targets (such as, finding the shortest way, searching for certain values, etc.), instead of solving the problem of selecting nodes in one partite with the set conditions through node subsets in the adjacent partite. With such properties the algorithm of traversal count is required to calculate the complex indicators. In this article we have offered algorithms of traversal *N*-partite graph for the purpose of searching nodes of one share, having a full set* of ribs with set search conditions by node subsets in the following share of the graph.

*Conditions define subsets of nodes for (k – 1) and (k + 1) partite the graph if the search is carried out in k-th partite. If the node of k-th partite is connected to all nodes of the set subsets in (k – 1) and (k + 1) partite, such a node has a full set of ribs.

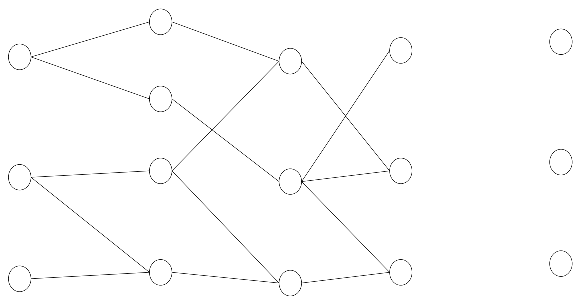


Fig. 7. N-partite graph
x – node of the graph; X – set of nodes

To realize the search of nodes in one partite of the *N*-partite graph it is necessary to define two subsets of nodes X'_{k-1} and X'_{k+1}, united by multiple search conditions U_k (1), where k – is the partite number in which the search will be fulfilled. When realizing the search of nodes in the k-th partite, a subset will be generated, where the nodes having a full set of ribs with subsets from search conditions will be selected (for each node $x'_{k,i} \in X'_k$, condition (2) will be satisfied):

$$U_k = X'_{k-1} \cup X'_{k+1}, \quad (1)$$

$$(\Gamma(x'_{k,i}) \cap X'_{k-1}) \cup (\Gamma(x'_{k,i}) \cap X'_{k+1}) = X'_{k-1} \cup X'_{k+1}, \quad (2)$$

where $\Gamma(x'_{k,i})$ – is the set of all nodes having ribs with node $x'_{k,i}$.

Two new search algorithms of the graph *N*-partite are described further: “traversal of one node vicinities” and “the mark of nodes”.

Traversal of one node vicinities:

– select all nodes from set X_k, which have ribs with one element of set U_k and to place them in set X'_k;

– check each node $x'_{k,i} \in X'_k$, whether it has a full set of ribs with nodes of set U_k. If the condition is not right, node $x'_{k,i}$ leaves set X'_k.

The analysis of the given algorithm has allowed us to reveal the dependence of the transitions C quantity, necessary for conducting the search, from the search conditions (3):

$$C = \sum_{i=1}^{n} \left(\left| \Gamma(x'_{k,i}) \right| - 1 \right), \quad (3)$$

where $\left| \Gamma(x'_{k,i}) \right|$ – is the quantity of nodes connected to element $x'_{k,i}$.

Mark of nodes:

– choose node $u_{k,i}$ from set U_k. We will introduce a set of labels M the dimension of which is equal to $\left| X_k \right|$ and for each element of this set we will appropriate “0”. $i = 1$;

– for node $u_{k,i}$ we find set $\Gamma(u_{k,i}) \subset X_k$. For each $x_{k,j} \in X_k$, where conditions are realized $x_{k,j} \in \Gamma(u_{k,i})$, the corresponding label $m_j \in M$ will increase the value by “1”;

– $i = i + 1$. If $i = n$ we pass on to the following step. Otherwise we pass to step 2;

– find the node in set U_k with the least quantity of ribs and move the nodes connected to it from X_k to set X'_k;

– for resultant set X'_{res} from set X'_k we select nodes having a full set of ribs with nodes from set U_k: $X'_{res} = \{\forall x_{k,i} \in X'_k \mid m_i = \left| U_k \right| \}$.

Let’s define the given algorithm dependency of transitions C quantity, necessary for conducting search, from search conditions (4).

$$C = \sum_{i=1}^{n} \left| \Gamma(u_{k,i}) \in X_k \right| + \min(A), \quad (4)$$

where $A = \left\{ \left| \Gamma(u_{k,i}) \in X_k \right|, i = \overline{1,n} \right\}$.

The analysis of formulas (3), (4) has shown that at a small number of search conditions (from 1 to 4) is more effective (from the position of quantity transition between graph nodes) than the algorithm “the mark of nodes”; and works more effectively during the increase of search conditions in the algorithm “traversal of one node vicinities”.

The complex indicator is calculated by the factual data multiplier X_2 of the tripartite graph G_3 (fig. 6) by allocating the subset $X'_2 \subset X_2$, according to conditions U_2 (1), and performing operations over this subset. Thus, the calculation function of the complex indicator represents an algorithmic search function on a tripartite graph with the set conditions (5) and analytical transformation function over the received set:

$$CI = F(G_3, U_2). \quad (5)$$

Let’s consider a calculating example of a complex indicator with the application of new search algorithms on the *N*-partite graph. In fig. 8 there is a factual data graph, in which it is required to calculate the quantity of turning graduates in 2007. In table there is a description of the account nodes (fig. 8) for the resulted example.

Symbols of qualifier and indicator values

Node	The node characteristic	Explanation
$x_{1.1}$	TC-1	Name of educational institution
$x_{1.2}$	TC-2	Name of educational institution
$x_{1.3}$	2007	Period (all 2007 year)
$x_{1.4}$	2008	Period (all 2008 year)
$x_{1.5}$	Turner	Occupation
$x_{1.6}$	Driver	Occupation
$x_{1.7}$	Cook	Occupation
$x_{2.1}$	10	The quantitative characteristic factual data
$x_{2.2}$	20	The quantitative characteristic factual data
$x_{2.3}$	5	The quantitative characteristic factual data
$x_{2.4}$	7	The quantitative characteristic factual data
$x_{2.5}$	13	The quantitative characteristic factual data
$x_{2.6}$	3	The quantitative characteristic factual data
$x_{2.7}$	7	The quantitative characteristic factual data
$x_{2.8}$	4	The quantitative characteristic factual data
$x_{2.9}$	11	The quantitative characteristic factual data
$x_{2.10}$	6	The quantitative characteristic factual data
$x_{3.1}$	Graduating students	Indicator name
$x_{3.1}$	Set of students on the first course	Indicator name

Before the search has begun, we will set a number of conditions, according to (1):

$$U_2 = \{x_{1.3}, x_{1.5}\} \cup \{x_{3.1}\} = \{x_{1.3}, x_{1.4}, x_{3.1}\}.$$

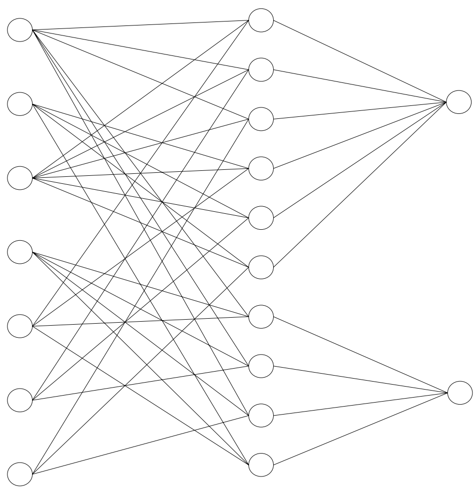


Fig. 8. Example of the factual data on graduates: x – graph node; X – set of nodes

The application of the “traversal of one node vicinities” algorithm:

– select all nodes from set X_2 which have ribs with node $x_{1.3}$ from set U_2 and place them into set:

$$X_2' = \{x_{2.1}, x_{2.2}, x_{2.3}, x_{2.4}, x_{2.5}, x_{2.6}\};$$

– check each node $x'_{2,i} \in X'_2$ for possessing a full set of ribs with nodes of set U_2. If the condition is not fulfilled, the node $x'_{2,i}$ is removed from set X'_2. As a result, in set X'_2 there are the following elements: $\{x_{2.1}, x_{2.3}\}$.

Element values of set X'_2 will be taken from table. To calculate the complex indicator – “quantity of graduates in 2007” we will sum the elements of the obtained subset and receive the required value – “15”.

Applying the algorithm: “Mark of nodes”:

– $i = 1$. Choose node $u_{2,i}$ from set U_2: ($u_{2,1} = x_{1,3}$). We will introduce a set of labels M the dimensions of which are equal to $|X_2| = 10$; and we will appropriate “0” to each element of this set;

– for node $u_{2,i}$ we find set $\Gamma(u_{2,i}) \subset X_2$ for $\forall i$. For each $x_{2,j} \in X_2$ for which $x_{2,j} \in \Gamma(u_{2,i})$ the condition is satisfactory, we will increase corresponding label $m_j \in M$ value by “1”.

$i = 1$: $u_{2,1} = x_{1,3}$, $\Gamma(u_{2,1}) = \{x_{2.1}, x_{2.2}, x_{2.3}, x_{2.4}, x_{2.5}, x_{2.6}\}$,
M = {1, 1, 1, 1, 1, 1, 0, 0, 0, 0}
$i = 2$: $u_{2.2} = x_{1.5}$, $\Gamma(u_{2.2}) = \{x_{2.1}, x_{2.3}, x_{2.10}\}$,
M = {2, 1, 2, 1, 1, 1, 0, 0, 0, 1}
$i = 3$: $u_{2.3} = x_{3.1}$, $\Gamma(u_{2.3}) = \{x_{2.1}, x_{2.2}, x_{2.3}, x_{2.4}, x_{2.5}, x_{2.6}\}$,
M = {3, 2, 3, 2, 2, 2, 0, 0, 0, 1};

– find a node in set U_2 with the least quantity of ribs, in case $x_{1.5}$ and move the nodes connected to it from X_2 to set $X'_2 = \{x_{2.1}, x_{2.3}, x_{2.10}\}$;

– in the resultant set X'_{res} from set X'_2 we select the nodes having a full set of ribs with nodes from set U_2:

$$X'_{res} = \{\forall x_{2,i} \in X'_2 \mid m_i = |U_2|\} = \{x_{2.1}, x_{2.3}\}.$$

The complex indicator is calculated similarly to the previous example and equals 15.

The presented approach for representing dynamic data structures allows designing information systems with changeable information structures of the subject domain objects, and its processing on region level. The algorithms

$x_{1.1}$

$x_{2.}$

$x_{1.2}$ $x_{2.2}$ $x_{3.1}$

described in article, can be used to project automated systems with dynamic structures, which will be built on the basis of the proposed approach; for calculating complex indicators when processing statistical information.

The development of the approach is planned in the following ways:

– the revealing of the structure representing the indicator graph, features of its construction and the traversal algorithms;

– researching invariable database control system storage methods for graph indicator, qualifiers, and factual data; working out techniques to work with account elements in these systems;

– researching possible automatic processing ways for factual data: in order to reveal doubtful data, new complex indicators, and new data classes for further analysis.

References

1. Post-relational SUBD Cache 5. Objective-orientated development of applications / V. Kirsten, M. Iringer, M. Kjun, B. Rerig. 2 ed. Moscow : LLC «Binom-Press», 2005.

2. Kite T. Oracle for professionals. Saint-Petersburg : DiaSoftUP, 2005.

3. Faulmer M., Scot K. UML. Basics. Saint-Petersburg : Symbol-Plus, 2002.

4. Volkov V. N., Denisov A. A. System theory : textbook. Moscow : Higher School, 2006.

5. Shovkun A. V. Constructing a corporative informative-analytical system in conditions of constantly changing business // Scientific-technical information. Moscow : VINITI, 2004. № 9. P. 1–6. Series 1.

6. Harari F. Graphs theory. Moscow : Editorial URS, 2003.

7. Ananiy V., Levitin A. Algorothms: Introduction to the design and analysis of algorithms. Moscow : Williams, 2006. P. 189–195.

O. V. Zaitsev
JSC Russian Space Systems, Russia, Moscow

THE APPLIED METHOD FOR CARRIER FREQUENCY RESTORATION OF THE TELEMETRY SYSTEMS SIGNAL BY DIGITAL PROCESSING

The paper considers the issue of restoring the level of the telemetry signal carrier frequency at digital processing in the automatic carrier control tract and the calculation of the threshold for taking the decision about the validity of received information symbol from the spacecraft and carrier rocket. We have described the applied method and algorithm for calculating the level of the carrier frequency and the level of threshold for making a solution based on histogram processing of the signal from the output of the frequency detector.

Keywords: control system, signal processing, telemetry.

The control of flight task execution by class SC/CR onboard systems is performed based on the processing of telemetric data (TMD) on the status of most units and devices of the object [1]. For a satisfactory receipt and processing of TMD achieving proper ground means that the following parameters are required for positive detection of received char (“0” or “1”): the restoration of radio signal carrier frequency and the calculation of decisions making the threshold.

Currently, the fast Fourier transformation (FFT) or data parameter linear filtering procedures are usually applied to repair carrier frequency of carrier-shift radio signal [2]. The use of FFT requires a large number of processing operations to get a necessary result, and consequently, a significant time to analyze the signal. The linear filtering (LF) usage leads to complicity for unbiased signal estimate under processing. Besides, the usage of conventional approaches (FFT and LF) for signal processing on a significant noise background does not allow positive detecting of carrier frequency radio-signals [3]; this considerably decreases the sensitivity of radio receiver digital channels.

Alternatives to FFT and LF correction carrier frequency procedures and algorithms are still insufficiently presented in Russian and foreign studies [4–6]. Therefore a development of procedures and algorithms (P&A) for carrier frequency correction (CFR) based on the histogram procedure is considered to be rather urgent, since this method would require considerably less computing sources than FFT.

The CFR procedure presented in this paper is developed based of the histogram procedure. The point of the procedure is the following: the range of possible signal levels at the detector’s output frequency is divided into an optimal number of control levels or intervals. The signal beyond the frequency detector output is transferred to histogram level construction units (HLCU). The amplitude of the signal is compared to the value of the control levels in HLCU; the number of values for the signal amplitude within the interval $\Delta_k = \frac{A_k - A_{k-1}}{2}$ is registered. A_k and A_{k-1} – values of neighbor control levels. A one-dimensional vector (vector of the amount values within interval $\Delta(k)$ is generated –

histogram $H(k)$. The storage vector $S(k)$ or sum vector of the signal levels within the interval $\Delta(k)$ is generated at the same time. Signal max and min levels are detected by the histogram values. The average value of the upper (or lower) signal level is calculated by the next step, based on $H(k)$ the upper limits' value and stored (accumulated) signal max value $S(k)$ (sce figure).

Figure presents an output signal of the frequency detector $A(t)$ for a typical binary receiver and the histogram corresponding to the signal. Two upper limits of the histogram $H(k)$ correspond to upper level A_1 and lower level A_0 signals. The values of the upper limits are identified by quantity of "0" and "1" chars in estimated interval of data message; if the number of chars "0" is equal to the number of chars "1" then the amplitude of the upper limits is also equal to each other. When noise level increase, the upper limits come into a single upper limit which corresponds to the rationing of average signal and noises levels. The interval between the upper limits is determined by a signal swing. The center $(A_1 + A_0)/2$ of the interval is a level of the carrier and depends on frequency mismatch only. Also the center of the interval is an optimal threshold for decision devices which quantizes analog signal from the frequency detector into binary levels "1" and "0". Signal $D(t)$ at output of decision, taking circuit, is a binary image of signal $A(t)$ (sce figure).

The amount of "0" and "1" in data message can be very different for each separate time period. Maximum number of "0" or "1" (as shown at the histogram $H(k)$ in figure) defines the average current value of "0" ("1"). The interval between the histogram's upper limit levels is divided into half; so, the value is considered as an average level of repaired carrier frequency [6; 7].

Thus, the CFR method, applying the histogram procedure includes the following stages:

1. The identification of the interval where "1" and "0" signal levels are estimated. It's defined that the best results occur if the estimated interval is equal to the interval of signal accumulation levels for an automatic frequency control circuit. This is related to the fact that signal level has minimum changing during this interval.

2. The calculation of a number of signal levels within control intervals and the accumulation (storage) of signal current levels for each row of the histogram at the interval under processing.

3. The determination of the histogram's upper limits and the computing of average current level of the carrier frequency.

4. Correction of the input signal.

The general features of the procedure are:

– provision of the carrier level linear dependence estimation from frequency mismatch;

– provision of the carrier average level unbiased estimate (its level does not depend on the amount of "1" and "0" in data message);

– high resistivity of the carrier level repair algorithm (at high noise levels) which is developed on the basis of the procedure suggested in this study.

The following algorithm is developed on the basis of the suggested procedure. The frequency detector output histogram $H(k)$ is plotted on a time interval equal to $\tau_{\text{АПЧ}}$. The number (k) of ranges or rows in the histogram is determined by required accuracy calculation and by the certainty of upper limits detection corresponded to char signal levels.

The sums $S(k)$ are calculated in parallel for the histogram's k ranges.

Then the following operations are performed:

1. Determination of $\max_0 = H_{(k=mx0)}$ and $\max_1 = H_{(k=mx1)}$ values which correspond to levels of "0" and "1" signals.

2. Estimation of average level: $\hat{A}_0 = S_{(mx0)} / n_{(mx0)}$ and $\hat{A}_1 = S_{(mx1)} / n_{(mx1)}$, where $n_{(mx0)}$ and $n_{(mx1)}$ is the number of accounts for the appropriate levels, $\hat{A}_H = (\hat{A}_0 + \hat{A}_1)/2$ – the estimation of the average carrier frequency level (CF) coming to a predictor.

Let's introduce the following signs:

$\hat{A}_H = Y(k)$ – the estimation of CF level for k moment of time;

$\tilde{A}_H = Y(k+1 \mid k)$ – the output value of the extrapolator, which is the optimal threshold for the taking decision circuit (the circuit quantizes frequency detector counts to chars "0" and "1").

In general case, the extrapolator is defined by an equation of the following class:

$$Y(k+1 \mid k) = F\left[Y(k), Y(k-1), ..., Y(k-(r-1))\right],$$

where F – a certain function of r indeterminate values; r – extrapolator's memory depth.

Tracker unit uses 2 predictors: linear extrapolator – for two points ($r = 2$) and quadric extrapolator – for 6 points ($r = 5$):

$$Y_1(k+1 \mid k) = 2 \cdot Y(k) - Y(k-1),$$

$$Y_2(k+1 \mid k) = \big(9 \cdot Y(k) - 4 \cdot Y(k-2) - - 3 \cdot Y(k-4) + 3 \cdot Y(k-5)\big)/5.$$

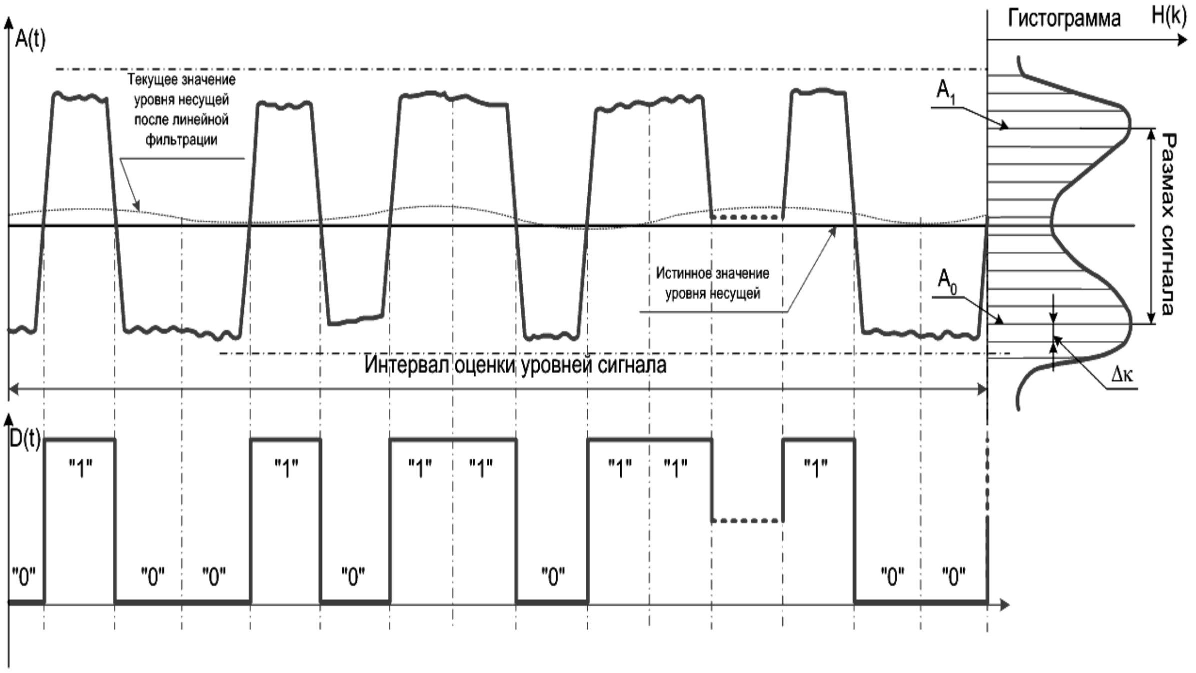
The results of the predictors computing are joined for collateral extrapolation of linear, quadric etc. dependence of parameter $Y(k)$:

$$Y_\Sigma(k+1 \mid k) = \sum\nolimits_j \alpha_j \cdot Y_j(k+1 \mid k),$$

where they must correspond to relation $\sum_j \alpha_j = 1$. Summation is performed for all j extrapolation variants.

Values α_j are calculated from a minimum value of extrapolation error:

$$\sigma^2\left[Y_\Sigma(k+1 \mid k)\right] = = \left[\sum\nolimits_j \alpha_j \cdot Y_j(k+1 \mid k) - Y(k+1)\right]^2.$$



Frequency detector output signal and the histogram of the signal; the repaired signal at making a solution for circuit output

For linear and quadric extrapolations:

$$Y(k+1|k) = \alpha_1 \cdot Y_1(k+1|k) + \alpha_2 \cdot Y_2(k+1|k),$$

$$\sigma^2\left[Y_\Sigma(k+1|k)\right] = $$

$$= \left[\alpha_1 \cdot Y_1(k+1|k) + \alpha_2 \cdot Y_2(k+1|k) - Y(k+1)\right]^2.$$

Minimum value of the relation above is obtained for the following values of coefficients:

$$\alpha_1 = \frac{\sigma^2\left[Y_2(k+1|k)\right]}{\left(\sigma^2\left[Y_1(k+1|k)\right] + \sigma^2\left[Y_2(k+1|k)\right]\right)},$$

$$\alpha_2 = \frac{\sigma^2\left[Y_1(k+1|k)\right]}{\left(\sigma^2\left[Y_1(k+1|k)\right] + \sigma^2\left[Y_2(k+1|k)\right]\right)}.$$

Therefore:

$$Y_\Sigma(k+1|k) = \frac{\sigma^2\left[Y_2(k+1|k)\right]\cdot Y_1(k+1|k) + \sigma^2\left[Y_1(k+1|k)\right]\cdot Y_2(k+1|k)}{\sigma^2\left[Y_1(k+1|k)\right] + \sigma^2\left[Y_2(k+1|k)\right]}.$$

Using the belief ratio, an output value of extrapolated parameter is calculated using the following formula:

$$Y_{\text{вых}}(k+1|k) = \left[1 - W(k)\right] \times$$

$$\times Y_\Sigma(k+1|k) + W(k)\cdot Y(k),$$

where the belief ratio is such that $0 \le W(k) \le 1$. So, if the belief ratio is high (normal tracking conditions), then the extrapolation of the device relies more on the current value of parameter $Y(k)$. Otherwise, the parameter is predicted by its values which were before the decrease of the belief ratio. The result of the experiments shows that the properties of this prediction procedure are the same as those of the Kalman filtering.

The results for the application of the described procedure can be stated briefly as:

1. The procedure and algorithm are developed for the correction of the average level of frequency-shift keyed signal carrier with a histogram procedure beyond frequency detector, without the usage of FFT or LF, which provides high noise immunity and performance.

2. The estimation of efficiency of the procedure and algorithm was done using real data flow. The estimation showed a considerable decrease in the number of computing operations during the correction of the carrier frequency average level and durability of the algorithm to noise. The dependence of $\hat{A}_H(\Delta f)$ from carrier frequency mismatch Δf is linear.

The estimation of the carrier frequency level does not depends on the number of "0" and "1" chars in data flow and does not require the setting of special messages or markers for the frequency level determination. The procedure allows the obtaining of unbiased estimates of the average signal level beyond the frequency detector when the carrier frequency deviates.

The procedure and algorithm of the CFR shows durability, linearity, and precision of the proposed estimation carrier level histogram procedure for solving the received char meaning.

References

1. Solovev U. A. Systems of satellite navigation. M. : Eko-Trends, 2000.

2. The onboard telemetry equipment of space flying devices / S. M. Perevertkin [et al.]. M., 1977.

3. Maks Z. Methods and equipment of signals processing at physical measurements : in 2 p. : Transl. from Fr. M. : Mir, 1983. P. 642.

4. Makklellan Dg., Reider Ch. Theory of number application in digital signals processing : Transl. from English. M. : Radio and communication, 1983. P. 376

5. Sverdlik M. B. Optimal discrete signals. M. : Sov. Radio, 1975. P. 200.

6. Radio systems of information transfer / B. A. Borisov [et al.]. M. : Radio and connection, 1990. P. 456.

7. Modern telemetry in the theory and in practice : A training course. SPb. : Science and Technics, 2007. P. 672.

U. V. Zakharov
Siberian State Technological University, Russia, Krasnoyarsk

A. U. Vlasov, R. V. Avakumov
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

MAGNETIZATION OF MULTILAYER FERROMAGNETIC FILM WITH A NONMAGNETIC INTERLAYER

The magnetization of magnetic film, consisting out of two ferromagnetic layers with a nonmagnetic interlayer applied to antiferromagnetic substrate is considered in this paper.

Keywords: ferromagnetic, antiferromagnetic, interlayer interaction.

Currently, the issue we consider is of great significance because there is an international interest in multilayer magnetic systems. Such systems are used in magnetoresistive sensors, components of magnetic send-receive, spin diodes. The magnetization of a multilayer magnetic system consisting of two ferromagnetic layers with a nonmagnetic interlayer applied to antiferromagnetic substrate is considered in the presented paper. The theoretical model used for the investigation of the stated process had been introduced in paper [1]. Inhomogeneous distribution takes place in this system because of substrate influence. This is why we use a stated model.

Physical properties of the film are defined by its boundaries. A bilayer system with the ferromagnetic-on-antiferromagnetic-substrate type is investigated in research paper [1] where the boundary condition of clamped magnetization vector type had been considered.

In paper [2] it had been demonstrated that the magnetization process of such a system has a threshold type. The author [2] points out an analogue between the magnetization process and the bending of elastic rod considered in paper [3]. It is also necessary to refer to paper [4].

Later the boundary condition of the clamped vector type was replaced by condition of elastically restrained vector type by introducing an effective interlayer at the ferromagnetic-antiferromagnetic boundary [5]. In later research of bilayer ferromagnetic film, the layers of which were rigidly bond with an antiferromagnetic substrate was overlooked in [6]; using an analog of a two-part elastic shank clamped at one edge and free at the other [7].

Today the presence of a nonmagnetic interlayer in ferromagnetic systems and its influence on threshold fields and the distribution of magnetization aren't being studied.

The potential energy of the magnetic system is given as expression [8]:

$$F\left(\mathbf{M}, \frac{\partial M_i}{\partial x_k}\right) = \frac{1}{2}\alpha_{ik}\frac{\partial \mathbf{M}}{\partial x_i}\frac{\partial \mathbf{M}}{\partial x_k} + w_a\left(\mathbf{M}\right) + f\left(M^2\right), \quad (1)$$

where, the first summand represents the quadratic form of derivatives, $w_a(\mathbf{M})$ is the magnetic anisotropy energy, $f(M^2)$ is some function of M^2. We only overlook isotropic ferromagnetic films where magnetization inhomogeneous is present along thickness of the object. Thereby, the second summand turns into zero and the first summand turns into the following expression:

$$\frac{1}{2}\alpha\left(\frac{d\mathbf{M}}{dz}\right)^2.$$

Axis z is directed perpendicularly to the layers. A variation of the third summand results in zero because the magnetization vector length doesn't change; therefore we shall not consider it. Thereby, the energy of the ferromagnetic layer may be represented as:

$$U = \int_0^{d_1}\left(\frac{1}{2}\alpha_1\left(\frac{d\mathbf{M}_1}{dz}\right)^2 - \mathbf{M}_1\mathbf{H}\right)dz + \int_{d_1+d_s}^{d_1+d_s+d_2}\left(\frac{1}{2}\alpha_1\left(\frac{d\mathbf{M}_1}{dz}\right)^2 - \mathbf{M}_1\mathbf{H}\right)dz, \quad (2)$$

where d_1, d_2 are the thicknesses, $\mathbf{M}_1$, $\mathbf{M}_2$ are the magnetization densities, α_1, α_2 are the exchange constants of first and second layers respectively; d_s is the interlayer thickness, $\mathbf{H}$ is the applied magnetic field.

The energy of interlayer is given in the expression:

$$U_s = -\frac{\alpha_s}{d_s}\mathbf{M}_1\mathbf{M}_2, \quad (3)$$

where α_s is the interlayer exchange constant.

The total energy of the ferromagnetic system is:

$$U = \int_0^{d_1} \left(\frac{1}{2}\alpha_1 \left(\frac{d\mathbf{M}_1}{dz} \right)^2 - \mathbf{M}_1\mathbf{H} \right) dz + \\ + \int_{d_1}^{d_2} \left(\frac{1}{2}\alpha_2 \left(\frac{d\mathbf{M}_2}{dz} \right)^2 - \mathbf{M}_2\mathbf{H} \right) dz - \frac{\alpha_s}{d_s}\mathbf{M}_1\mathbf{M}_2 \Big|_{d_1} . \quad (4)$$

The value of thickness d_s is negligible in comparison to thicknesses d_1 and d_2. It's rather convenient to proceed to the generalized coordinates, which represent the rotation angles between the magnetization vectors and x axis directed along the applied field. Then,

$$U = \int_0^{d_1} \left(\frac{1}{2}\alpha_1 M_1^2 \left(\frac{d\varphi_1}{dz} \right)^2 + M_1 H \cos \varphi_1 \right) dz + \\ + \int_{d_1}^{d_2} \left(\frac{1}{2}\alpha_2 M_2^2 \left(\frac{d\varphi_2}{dz} \right)^2 + M_2 H \cos \varphi_2 \right) dz - \\ - \frac{\alpha_s}{d_s} M_1 M_2 \cos\left(\varphi_2 - \varphi_1\right) \Big|_{d_1} . \quad (5)$$

The minimum condition of potential energy gives:

$$\delta U = \alpha_1 M_1^2 \frac{d\varphi_1}{dz}\delta\varphi_1 \Big|_0^{d_1} + \alpha_2 M_2^2 \frac{d\varphi_2}{dz}\delta\varphi_2 \Big|_{d_1}^{d_2} - \\ - \int_0^{d_1} \left(\alpha_1 M_1^2 \frac{d^2\varphi_1}{dz^2} + M_1 H \sin\varphi_1 \right) \delta\varphi_1 dz - \\ - \int_{d_1}^{d_2} \left(\alpha_2 M_2^2 \frac{d^2\varphi_2}{dz^2} + M_2 H \sin\varphi_2 \right) \delta\varphi_2 dz + \\ + \frac{\alpha_s}{d_s} M_1 M_2 \sin\left(\varphi_2 - \varphi_1\right) \delta\left(\varphi_2 - \varphi_1\right) \Big|_{d_1} = 0. \quad (6)$$

Differential equations set follows from (6) as long as variations $\delta\varphi_1$ and $\delta\varphi_2$ contained in integral are arbitrary:

$$\begin{cases} \alpha_1 M_1 \dfrac{d^2\varphi_1}{dz^2} + H \sin\varphi_1 = 0, \\ \alpha_2 M_2 \dfrac{d^2\varphi_2}{dz^2} + H \sin\varphi_2 = 0. \end{cases} \quad (7)$$

Matching of condition at point d_1 follows from (6):

$$\begin{cases} \dfrac{\alpha_1}{d_1} \dfrac{d\varphi_1}{dz} = \dfrac{\alpha_s}{d_s} \dfrac{M_2}{M_1} \sin\left(\varphi_2 - \varphi_1\right), \\ \dfrac{\alpha_1 M_1^2}{d_1} \dfrac{d\varphi_1}{dz} = \dfrac{\alpha_2 M_2^2}{d_2} \dfrac{d\varphi_2}{dz}. \end{cases} \quad (8)$$

Here z is a normalized variable. Then we will use designation $C = (\alpha_s M_2)/(d_s M_1)$. This quantity characterizes the degree of the layer's fixity. The infinitely large value of C gives $\varphi_1 = \varphi_2$, substituting for first equation of (8).

The boundary conditions are also given from (6):

$$\begin{cases} \alpha_1 M_1^2 \dfrac{d\varphi_1}{dz}\delta\varphi_1 \Big|_0 = 0, \\ \alpha_2 M_2^2 \dfrac{d\varphi_2}{dz}\delta\varphi_2 \Big|_{d_1+d_2} = 0. \end{cases} \quad (9)$$

Angle φ_1 equals zero at the antiferromagnetic boundary because the applied magnetic field we consider is solidly joined to the antiferromagnetic anisotropy field. Where the vacuum border is angle φ_2 shall be subject to free magnetic momentums. Mathematically, this looks like:

$$\begin{cases} \varphi_1 \big|_0 = 0, \\ \dfrac{d\varphi_2}{dz} \Big|_{d_1+d_2} = 0. \end{cases} \quad (10)$$

For example let's examine a magnetic film consisting of two ferromagnetic with equal thickness d and with a nonmagnetic interlayer:

$$\frac{d^2\varphi_i}{dz^2} + \frac{Hd^2}{\alpha_i M_i} \sin\varphi_i = 0, \quad i = 1, 2. \quad (11)$$

In these equations z is normalized variable. The solution of these equations could be represented as [1]:

$$\varphi_i\left(z\right) = 2\arcsin\left[k_i \operatorname{sn}\left(\sqrt{\frac{Hd^2}{\alpha_i M_i}} z + F_i, k_i \right) \right], \quad (12)$$

where $\operatorname{sn}(u, k)$ is the Jacobi sine. Equations (11) amplified with the matching condition:

$$\begin{cases} \alpha_1 M_1 \dfrac{d\varphi_1(1)}{dz} = \alpha_2 M_2 \dfrac{d\varphi_2(0)}{dz}, \\ \dfrac{\alpha_1}{d} \dfrac{d\varphi_1(1)}{dz} = C \sin\left(\varphi_2(0) - \varphi_1(1)\right), \end{cases}$$

and the border conditions:

$$\begin{cases} \dfrac{d\varphi_2(1)}{dz} = 0, \\ \varphi_1(0) = 0. \end{cases} \quad (13)$$

Fig. 1 represents the system of coordinates. A new coordinate system is introduced for each layer. The spiral emphasizes the interlayer interaction.

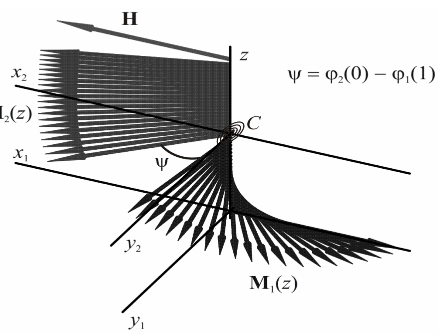


Fig. 1. Coordinate system

Applying the border conditions and matching the condition to the solution gives:

$$\begin{cases} k_1 \operatorname{cn} u_1 = \gamma k_2 \operatorname{cn} u_2, \\ \rho u_1 k_1 \operatorname{cn} u_1 = \left(k_2 \operatorname{sn} u_2 \operatorname{dn} u_1 - k_1 \operatorname{sn} u_1 \operatorname{dn} u_2 \right) \\ \left(\operatorname{dn} u_1 \operatorname{dn} u_2 + k_1 k_2 \operatorname{sn} u_2 \operatorname{sn} u_1 \right), \end{cases} \quad (14)$$

where:

$$u_1 = \frac{\pi}{4}\sqrt{\frac{h}{h_c}},\ u_2 = F_2 = K(k_2) - \frac{\pi}{4\gamma}\frac{M_2}{M_1}\sqrt{\frac{h}{h_c}},$$

$$h_c = \left(\frac{\pi}{2}\right)^2 \frac{\alpha_1}{d^2},\ \gamma = \sqrt{\frac{\alpha_2 M_2^3}{\alpha_1 M_1^3}},\ \rho = \frac{\alpha_1}{Cd},$$

cn(u, k), sn(u, k), dn(u, k) are the Jacobi cosine, sine and delta-function respectively. Constant F_1 equals zero. To define the threshold fields, elliptic modules k must be turned to zero in (14). Then, after some transformations, the transcendental equation derives from (14):

$$\rho\frac{\pi}{4}\sqrt{\frac{h^{tr}}{h_c}} + \mathrm{tg}\left(\frac{\pi}{4}\sqrt{\frac{h^{tr}}{h_c}}\right) = \frac{1}{\gamma}\mathrm{ctg}\left(\frac{\pi}{4\gamma}\sqrt{\frac{h^{tr}}{h_c}}\right), \quad (15)$$

where $h^{tr} = H^{tr}/M$ is the threshold field. The solution of (15) relative to h^{tr}/h_c by different values of parameters γ and ρ gives surface shown in fig. 2.

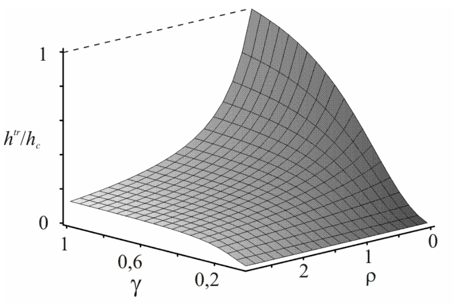


Fig. 2. Threshold field dependence on γ and ρ

If there is no nonmagnetic interlayer ($d_s = 0$), and the ferromagnetic system consists of two equal layers ($\gamma = 1$), the equation (15) will depict the known expression:

$$h^{th} = \left(\frac{\pi}{2}\right)^2 \frac{\alpha}{4d^2}.$$

In other extreme cases ($\rho \to \infty$) the second equation of (14) gives:

$$\cos\frac{\pi}{4}\sqrt{\frac{h^{th}}{h_c}} = 0,$$

$$h^{th} = \left(\frac{\pi}{2}\right)^2 \frac{\alpha_1}{d^2}.$$

This result corresponds to the ferromagnetic layer the thickness of which equals d as expected.

It's also necessary to investigate the behavior of magnetization curves. Let's examine a particular case where two identical ferromagnetic layers are separated by a nonmagnetic interlayer. Only variable ρ varies. The average value of the magnetization vector projection is presented by the following expression:

$$\overline{m}_x = \int_0^1 m_x dz = -\int_0^1 \cos\varphi dz.$$

The integration gives:

$$\overline{m}_x = 2\left[1 - \frac{1}{u_1}E(\mathrm{am}\,u_1) - \frac{1}{u_1}E(\mathrm{am}(u_1 + F_2)) + {} + E - [\mathrm{am}\,F_2, k_2]\right], \quad (16)$$

where am(u, k) is the Jacobi amplitude (fig. 3):

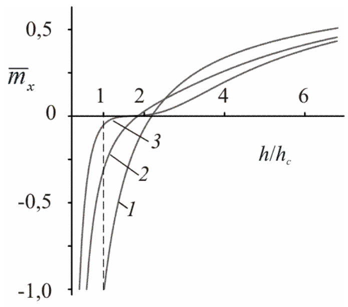


Fig. 3. The magnetization x-projection dependence on the applied field:
1 – $\rho = 0$; *2* – $\rho = 1{,}4$; *3* – $\rho = 2{,}8$

Parabola *1* corresponds to the magnetic system consisting of one ferromagnetic layer with a thickness of $2d$ because ρ and, consequently d_s equals, zero in this case. Parabolas *2* and *3* correspond to the magnetic systems involving a non-magnetic interlayer and the curve shifted to less values of the applied magnetic field at the beginning of magnetization; corresponds to a greater thickness of the interlayer.

Currently, a research on interlayer influence on threshold fields in systems of ferromagnetic-interlayer-ferromagnetic-antiferromagnetic types has been conducted. Also, parabolas of magnetization with different values of effective parameter ρ, characterizing the interlayer exchange interaction have been plotted.

References

1. Aharoni A., Frei E. H., Strinkman S. J. // Appl. Phys. 1959. Vol. 30, № 12. P. 1956–1961.

2. Zakharov Yu. V. Static and dynamic instabilities caused by switching of ferromagnetic layer // DAN. Vol. 344, № 3. P. 328–332.

3. Zakharov Yu. V., Okhotkin K. G. Nonlinear bending of thin elastic rod // PMTF. 2002. Vol. 43, № 5. P. 124–131.

4. Zakharov Yu. V., Isakova V. V., Okhotkin K. G. Analogy of magnetization reversal of a magnetic bilayer system and nonlinear bending of an elastic rod with compression // Vestnik SibSAU. 2009. Vol. 2 (23). P. 122–125.

5. Magnetization thresholds of soft-magnetic layer in exchange coupled structure depends on surface exchange / U. V. Zakharov [et al.] // New magnetic materials of microelectronics : Collected papers of XX international seminar-school. M. : Physfac of MSU, 2006. P. 201–203.

6. Magnetization reversal of the multilayer magnetic film / Yu. V. Zakharov, K. G. Okhotkin, A. D. Skorobogatov, V. V. Isakova // Magnetism on a Nanoscale (EASTMAG-2007) : Brief outline reports of Euro-Asian symposium Kasan : KSU, 2007. P. 184.

7. Vlasov A. Yu., Okhotkin K. G. Non-linear bend of stiffness-variable compound rod with longitudinal and transversal stress // Physical-mathematical science. 2006. № 1. P. 26–28.

8. Akhiezer A. I., Bariakhtar V. G., Pelemitskii S. V. Spin waves. M. : Nauka, 1967.

V. V. Zolotarev, E. A. Danilova
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

ON APPLICATION OF FACTORIAL ANALYSIS IN PROBLEMS OF SECURITY ESTIMATION OF AUTOMATED SYSTEMS ELEMENTS

The possibility of factorial analysis application in theestimation of the state of information systems security is considered. The procedure of selection and classification of factors as well as calculation of factors influence on the resultant indicator size are described.

Keywords: risk management, information risk, factorial analysis.

Factorial analysis is one of the possible methods of automated systems security analysis. This method of analysis allows both to establish cause-and-effect relations between negative events and to characterize them quantitatively.

Let's consider the application peculiarities of a security estimation factorial model (further a factorial model) in the problem of security estimation of electronic document management system (EDMS). At the same time we will introduce universality elements into the offered model which will allow to use it for the estimation of various elements of both EDMS and other automated systems. We will especially note the applicability of the offered model and the solutions found on its basis for a human factor estimation.

The model description. Let there be an information system *IS* which consists of *N* numbers of *E* elements, each of which in its turn consists of *K* components. A component of each element in a certain period of time can accept *x* of states *s* with probability *r*:

$$IS = \{E_l\}, \qquad E_i = \begin{Bmatrix} s_1^1 & s_2^1 & \dots & s_{x1}^1 \\ s_1^2 & s_2^2 & \dots & s_{x2}^2 \\ \dots & \dots & \dots & \dots \\ s_1^k & s_2^k & \dots & s_{xk}^k \end{Bmatrix},$$

where

$$l \in [1 \dots N]. \tag{1}$$

Let x be the quantity of degrees of a component freedom. It is obvious that in using a similar model of the system it is possible to use the method of a morphological box of Zwicky [1, p. 196] in various variants. At the same time it is possible to calculate the quantity of cause-and-effect relations between the states of information system elements if we calculate them as a number of placings with repetitions:

$$k_{\text{сост}} = \left(\sum_{i=1}^{N} \sum_{j=1}^{K} x_{s_{ij}} \right)^m, \tag{2}$$

where m is the quantity of interlinks between system elements components. During such calculation a number of assumptions was made which is necessary to mention as these assumptions restrict the model application range:

– it is necessary to reduce the quantity of freedom degrees to some uniform value which assumes a standard set of states of system elements components;

– it is necessary to provide the completeness of an initial set of freedom degrees of each system element component which assumes a certain approach to the choice of indicators defining freedom degrees;

– the private function of utility should be calculated for each system element separately, thus resorting to simplification of calculations;

– it is necessary to possess the information about internal connections of analyzed system elements. Without updating such approach is inapplicable for a system with incomplete information about internal connections.

Let's consider basic elements of an applied factorial model:

– the private function of element *E* utility for performing the main task of *IS* system (further – private function of utility):

$$u^*(t) = f(E),$$

where

$$E = \begin{pmatrix} r_1^{(1)}(t) & \dots & r_1^{(x)}(t) \\ r_2^{(1)}(t) & \dots & r_2^{(x)}(t) \\ \dots & \dots & \dots \\ r_k^{(1)}(t) & \dots & r_k^{(x)}(t) \end{pmatrix}, \tag{3}$$

where $r_i^{(j)}(t) = p_i^{(j)}(t) * w_i^{(j)}(t)$, for i-th component in j state according to [2], $i \in [1…K], j \in [1…x]$.

– here $p_i^{(j)}(t)$ characterizes the probability of unfitness of an element component in a certain degree of freedom in a certain period of time for performance of the set function; $w_i^{(j)}(t)$ is the probability of influence of the established security facilities on suitability of an element component in a certain degree of freedom in a certain period of time; as a whole $u^*(t)$, hence, characterizes the ability of a concrete element to resist the influence. The function $f(E)$ essence is the calculation of an average factor allowing to use value u^* for the security estimation of a system element as a whole, instead of the security estimation of its components characterized by values $r_i^{(j)}(t)$. At the same time it is possible to additionally prove the decomposition of a system element, considering it as a set of interconnected components, at the same time keeping the possibility of an inverse operation of decomposition into components, as it will be shown further in an example of use of the offered model;

– directed graph $G(V, E)$, representing the system model. Nodes of the graph–system elements – are characterized by pair $u^*(t)$ and B_e (average cost of security facilities of an element), and tree edges are characterized by value Δu^*, which shows a utility correlation of the related elements;

– the integrated function of utility of a system as a whole:

$$U(t) = \underset{i=1}{\overset{N}{H}} u^*_i(t), \qquad (4)$$

where N is the quantity of investigated components; H = {*,+, max} is a set of operations of interconnection. The choice of a concrete operation is defined by a kind of interconnection (or absence of that) of investigated system elements [2].

The following correlation can be used as a criterion function of risk:

$$R(t) = U(t)\sum_{i=1}^{N} B_i(t). \qquad (5)$$

Thereby, the utility of work in a certain period of time [a; b] will be equal to (according to Neumann–Morgenshtern function [3]):

$$D = \int_a^b R(t)dt. \qquad (6)$$

Taking into account (5) for an optimum configuration:

$$\sum_{\tau \subset T} D_\tau \to \min, \qquad (7)$$

where, T is a set of all analyzed periods of time.

Obviously, the offered approach can be supplemented with the criterion of elements connectivity for the choice of the system optimum structure [4] taking into account a prior indicator of elements pairs contents. The considered factorial model together with the offered function of utility allows to solve the problem of optimum distribution of resources.

Then we will consider separate aspects of the offered approach.

Selection and classification of factors. The functioning of any information system occurs in the conditions of complicated interaction of a complex of internal and external factors. A factor is a reason, motive power of any process or phenomenon, defining its character or one of the basic characteristics.

There are various principles of factorial analysis [5]. Deterministic multiple analysis is used to perform the set task.

To solve the problem of the account of all set of factors influencing the information, circulating in EDMS, we suggest breaking it into the basic components. First of all, for each element we will define possible states which it can accept, then we will consider combinations of these states, creating a basic model of all possible private functions of utility and modelling a risk function in the dynamics of system development on the basis of a minimum set of initial data.

Usable model of an element. Let's start with distinguishing the system elements components. Taking into account the fact that the object of this research is to estimate the security of EDMS whose structure can be rather easily described organizationally, technically and by means of functional-logic models, the description of element components becomes the primary goal which defines the efficiency of integrated estimation. Taking into account (2), we will specify the problem as the development of a morphological box of Zwicky satisfying the chosen conditions. So, in view of the definition: an automated system (AS) is a system realizing information technology of performance of the established functions [6] and consisting of personnel and a complex of means automating its activity.

On the basis of the definition, we can distinguish four basic components of each AS element, namely: hardware (HW (fig. 1 and table, 1–ТС)), the software (SW (fig. 1 and table, 1–ПО)), the personnel (P(fig. 1 and table, 1–Ч)) and organizational measures of information processing and protection (various kinds of instructions, regulations and orders (the OM)).

Taking into account model restrictions let the ability of AS element to carry out the set functions be described by a static set of states:

1. Up state (U(fig. 1 and table, 1–Р)) – a state of AS element component when values of all parameters characterizing the ability to carry out set functions correspond to requirements of the specifications and technical documentation and (or) design (project) documentation [7];

2. Down state – a state of AS element component when the value of at least one parameter characterizing the ability to carry out set functions does not correspond to requirements of specifications and technical documentation and (or) design (project) documentation. Let's divide down states of AS element components into three categories:

– failure (F (fig. 1 and table, 1–Отк)) – an event consisting in malfunction of an upstate condition of AS element component after which a component of AS element stops functioning and demands extraneous intervention for restoration of normal work;

– malfunction (M (fig. 1 and table, 1–C)) – a transient fault or a single fault eliminated by slight intervention of an operator [7];

– error (E (fig. 1 and table, 1–O)) – incorrect or incomplete performance of separate tasks of a component of AS element without loss of its functionality.

Let's set a way of a morphological box formation as investigation of combinations of the listed set of AS element components states. To estimate security it is necessary to define combinations which can negatively influence the change of a productive indicator, i. e. the change of an information system security state. At the same time to exclude the combinations which are of no interest for further research and don't influence the change of an integrated indicator of risk or element utility private function we will introduce some conditions. We will exclude:

– interrelations between the states of one AS element component as we accept that AS element component can be only in one state in some concrete period of time;

– interrelation between up states of various groups of elements as it does not influence the element utility private function negatively.

Let's take into account the dependence on AS element generating state which defines technically impossible combinations of a morphological box states:

– change of a state of organizational measures (OM) can be caused only by personnel's (P(fig. 1, Ч and table)) actions;

– "Failure" (fig. 1 and table, 1–Отк) state of hardware can become a cause of refusal in personnel and/or software, but not simultaneously;

– if hardware is in the state of "Failure" (fig. 1 and table, 1–Отк) the software cannot be in "Upstate" (fig. 1, 1–P) condition;

– "Failure" (fig. 1 and table, 1–Отк) in personnel's work cannot become a reason of change in the states of other AS elements.

On the basis of introduced restrictions it is possible to define all combinations of AS elements states whose change can lead to the change of AS security state (fig. 1). The total number of such connections can be presented if we exclude some of them according to the conditions set above.

So, we receive 128 combinations with which it is possible to describe all set of influences on AS element, using minimum initial data, that is, failure rate of AS element separate components. Completeness and reliability of the revealed connections between states of AS elements influencing or able to influence the information is reached by way of considering a set of states of all AS element components and, as a consequence, of all factors influencing all AS elements at all stages of information processing [2].

Calculation of the probability of states combination. If the probabilities of occurrence of each state for all components of AS element $P_{sx}(E_i)$ are known to us, then it is possible to represent the probability of occurrence of connected event P_s, as the mutual one, thus applying the following expression, considering (1):

$$P_s = r_{si} r_{sj}\text{, here } i, j \in [1 \ldots x]. \qquad (8)$$

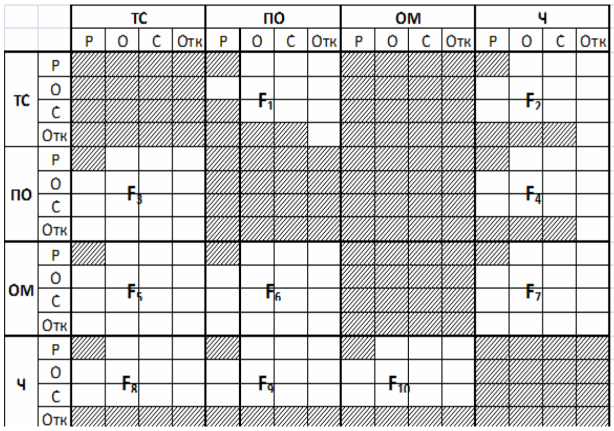


Fig. 1. The morphological box of AS element components states

By means of (8) it is possible to calculate probabilities of occurrence of all 128 connections between AS element components states, presented in fig. 1.

Definition of dependences between AS element components states. Connections between AS element components states can be presented in the form of directed graph *GE*, whose nodes are states of AS element s_{Ei} components, edges of the graph are dependences between them (fig. 2):

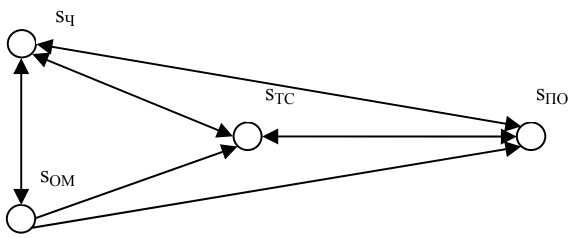


Fig. 2. The graph of dependences between element components of an information system

Having investigated the dependences of the chosen states on each other (fig. 2), we form a private function of an element utility on the basis on factors (average probabilities of sets of the states generated by a concrete component of AS element):

$$u^*_{E_i} = F_1 \cap F_2 \cup F_3 \cap F_4 \cup F_5 \cap \\ \cap F_6 \cap F_7 \cup F_8 \cap F_9 \cap F_{10}, \qquad (9)$$

where u^*_{Ei} is a private function of utility of AS element; F_1, F_2, ..., F_{10} are factors influencing the change of AS state; $\cap$ is a symbol of intersection if there is dependence between factors; $\cup$ is a symbol of association if there is no obvious dependence between factors.

Using methods of mathematical logic, it is possible to present (9) as:

$$u^*_{E_i} = F_1F_2 + F_3F_4 + F_5F_6F_7 + F_8F_9F_{10}\,. \qquad (10)$$

Modelling of investigated AS element security. To carry out the information risks factorial analysis it is necessary to present the object of research – an automated system – as a set of constituent elements. The better the model of research object is studied, the more exactly an integrated function of utility can be defined. Some standard methods of reliability theory, for example, an exhaustive method of states [8] can be used in this process. However it is necessary to remember that modelling the research object should be made from the position of integrity and systematicness, excluding redundant parameters which do not give any useful information for realization of the analysis purposes. The object of research should be considered from the point of view of logic, technical and structural schemes of information processing in an organization. In factorial analysis one can use both each scheme separately and all of them together, which can give fuller information for calculation of a resultant indicator, i. e. the risk level of each element of research object. Thus the risk level of each element of research object is represented from the point of view of factorial indicators (with possible decomposition to separate states of AS element components). On the basis on these data it is possible to draw a conclusion not only about most vulnerable elements, but also to point the concrete reason reducing the level of information safety.

Numerical modelling of security change. Using the offered approach, it is possible to carry out factorial modelling and estimation of security of both separate AS elements and the integrated indicator of security and risk for all research object. Taking (6), (7) and the consequences of the specified correlations into account, it is possible to operate information security of an object, using minimum initial data.

For example, we will consider numerical modelling of security level of the centre of collective access for organization ESDM (that is an AS element in terms of the model) (table), in brackets there is a final state after reaction of an element to introduction of the offered protection measures, nearby there is an initial state of an indicator.

The failure rate was considered for a certain period of time (one operational month, that is 30 days) which was agreed in the course of carrying out of the analysis; intensity calculation was done in terms of 1 day of operational time of the centre of collective access on the basis of reports of technical support services in the organization. Obviously, methods of information protection of ESDM element, applied in this case, have not influenced ESDM as a whole significantly though they were preliminary estimated by an organization management (judging by their influence on concrete components of system elements) as effective ones. You can see from the example that the offered approach allows to specify the complex problems of ESDM element, but does not allow to predict sharp changes and individual security infringements of research object. At the same time the value of the offered approach is obvious both for the solution of operative, short-term problems of information security management and for modelling the systems of protection without taking into account the influence of the information security infringer (preliminary calculation and choice of security facilities and configurations which are optimal form the point of view of cost).

The application of the offered approach takes into account cause-effect relations of processes of information processing which influence the level of information resources security. The use of factorial analysis is a step towards reception of objective quantitative results in information security management.

An example of numerical modelling of AS element security

S_x	P_{sx}	F_1	F_2		u^*
Р(ТС)	0.84(0.94)	0.0328 (0.302)	0.0567 (0.512)		0.0038 (0.0030)
О(ТС)	0.06(0.03)				
С(ТС)	0.1 (0)				
Отк(ТС)	0 (0.03)				
		F_3	F_4		
Р(ПО)	0.68(0.79)	0.0286 (0.0224)	0.0595 (0.0560)		
О(ПО)	0.23(0.09)				
С(ПО)	0.06(0.09)				
Отк(ПО)	0.03(0.03)				
		F_5	F_6	F_7	
Р(ОМ)	0.91(0.97)	0.0157 (0.0123)	0.0254 (0.0227)	0.0436 (0.0421)	
О(ОМ)	0.06(0.03)				
С(ОМ)	0 (0)				
Отк(ОМ)	0.03 (0)				
		F_8	F_9	F_{10}	
Р(Ч)	0.38(0.51)	0.0564 (0.0520)	0.0620 (0.0594)	0.0540 (0.0487)	
О(Ч)	0.5 (0.4)				
С(Ч)	0.06(0.09)				
Отк(Ч)	0.06 (0)				

References

1. Волкова В. Н., Денисов А. А. Теория систем и системный анализ : учебник для вузов. М. : Юрайт, 2010.

2. Применение факторного анализа и эволюционного алгоритма оптимизации для решения задачи управления информационными рисками систем электронного документооборота / В. Г. Жуков [и др.] // Системы управления и информационные технологии. 2009. № 3(37). С. 41–50.

3. Беллман Р., Калаба Р. Динамическое программирование и современная теория управления. М. : Наука, 1969.

4. Антамошкин А. Н. Алгоритм расчета прогнозируемого трафика при проектировании распределенных систем обработки и хранения информации // Вестник СибГАУ. 2006. Вып. 1. С. 5–10.

5. Сафонов А. А. Теория экономического анализа : учеб. пособие / под ред. Л. В. Моисеевой. Владивосток : Изд-во Владивост. гос. ун-та эконом. и сервиса.

6. ГОСТ 34.003–90. Информационная технология. Комплекс стандартов на автоматизированные системы. Автоматизированные системы. Термины и определения. М. : Изд-во стандартов, 1990.

7. ГОСТ 27.002–89. Надежность в технике. Основные понятия. Термины и определения. М. : Изд-во стандартов, 1989.

8. Половко А. М., Гуров С. В. Основы теории надежности. 2-е изд.. перераб. и доп. СПб. : БХВ-Петербург. 2006.

N. N. Zagirov, A. A. Kovalyova, E. V. Ivanov
Siberian Federal University, Russia, Krasnoyarsk

TECHNOLOGY OF PRODUCING FIBROUS STRUCTURE WIRE FROM CHIPS OF ALUMINUM–MAGNESIUM–SILICON ALLOYS

A technological scheme for processing the scrap of aluminum–magnesium–silicon alloy in the form of friable chips into rods and wire is introduced. This scheme is based on the powder metallurgy methods. The characteristic structure and the level of mechanical properties of the produced wire are denoted.

Keywords: friable chips, briquetting, combination of rolling and pressing, drawing, fibrous material, structure, mechanical properties.

A complex composite material is implied in cases when wire is not made of compacted metal material. This material has a metal coating consisting of a hard plastic body and a powder core, which is a friable mixture of heterogeneous particles [1]. During mechanical processing, the metal coating is in a complex interaction with the powder core; this causes complex movement of the powder particles and their elastic-plastic interaction under external load.

According to the suggested technology, making wire from friable fine chips (filing) of aluminum alloy АД31 and putting it into the metal coating is not accomplished. The process of making the final product can be divided into two parts:

– the technological scheme of producing an intermediate workpiece for drawing, which includes the preparation of chips for compacting, briquetting, briquette heating for extrusion, and hot extrusion for the rod of a specified diameter;

– the technological process of making wire consisting of multiple repeated operations of drawing the workpiece through dies and other auxiliary operations.

The method of chip processing in which the quantity of secondary raw material is quite high provides a higher yield ratio of metal chips in comparison with molting. Besides, energy consumption and harmful environmental impact are being reduced, which is a key issue for any type of industry.

It is well known that the suitability of metal chips for making press-items and wire depends on the compressibility during briquetting. The traditional scheme of pressing in rigid molds for making lengthened briquettes of cuboids form with height to width ratio 1 and length to width ratio 10 is not effective (such a ratio is determined by specific character of the equipment and the combination of rolling and pressing). Due to comparatively low briquettes density and cohesion of chips particles there is high probability of fillets rupture (breaking) during the pressing-out.

Briquetting of chips *2* is made in molds (fig. 1) consisting of upper *1* and lower *4* plugs, split matrix *3* and chase *5* with sloping contact surfaces. The experiment shows that briquetting pressures for providing integrated briquette density of 70–80 % must not be lower than 80–100 MPa.

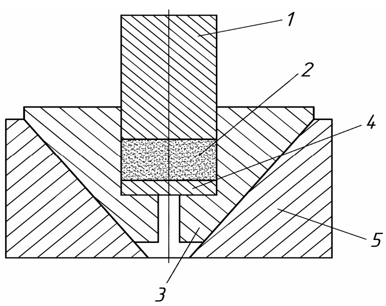


Fig. 1. The scheme of the briquette making mold for combined rolling and pressing

Fig. 2 shows the equipment for combined rolling and pressing to manufacture press-items. The stand consists of two steel frames *1* fastened by tie bolts *2* and mounted on the same base with the engine, gearbox, reducing unit (not shown in the figure). In the bronze tilting pads of sliding bearings *3* there are shafts *4*, on which fastened rollers *5* make the close roll-pass. Adjustments for the roller gap are made by special device of simultaneous and divided rotation of the push screws *6*.

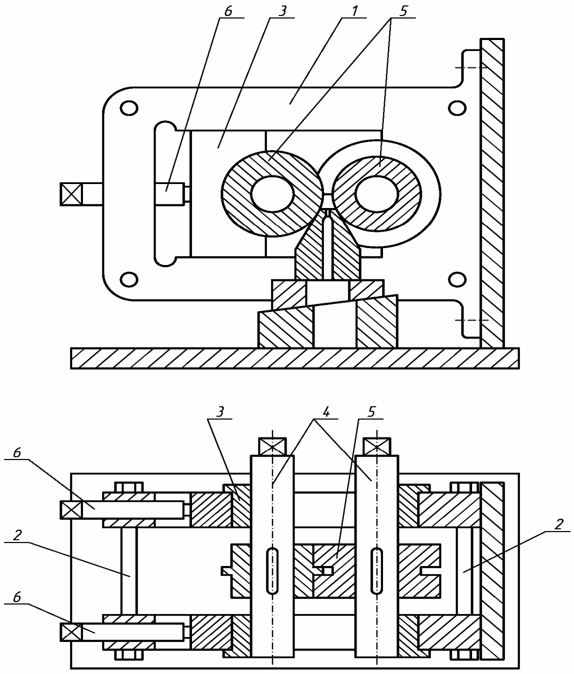


Fig. 2. Laboratory equipment for combined rolling and pressing of aluminum alloys on the basis of rolling mill DUO200

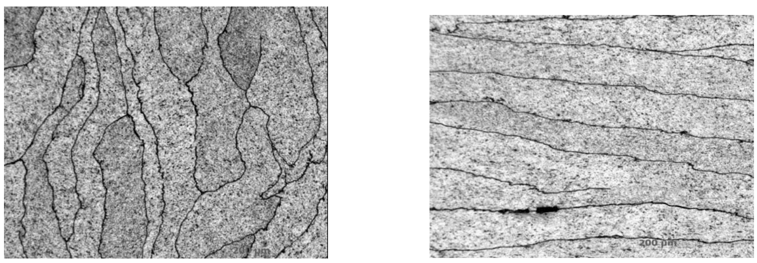
Before being placed into the rollers, the briquettes are heated to a temperature of 500 ± 20 ^{0}C in a furnace. At the same time the rollers are heated to the temperature of 80–100 ^{o}C. Briquettes inputting to the close roll-pass (fig. 3) is accomplished successively with the reduction of pauses up to their minimization.

A 7–8 mm gap between the rollers makes the density of the chips compact to 85–90 %. Therefore, the compacted material now comes to the area of pressing-out and this makes the deformation chips settling crosswise the pass easier.

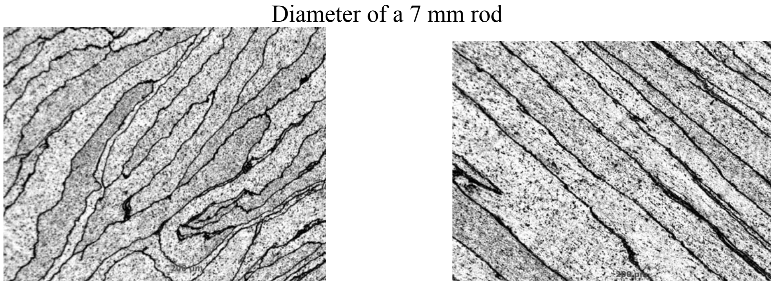
The diameter of the holes for making press-items is 7 and 9 mm and that corresponds to the drawing coefficients 8 and 5.

Fig. 3. The moment of inputting billet on the rollers

Diameter of a 9 mm rod



Diameter of a 7 mm rod



a *b*

Fig. 4. Typical microstructure of the rods (×160) in longitudinal (*a*) and crosscut (*b*) direction

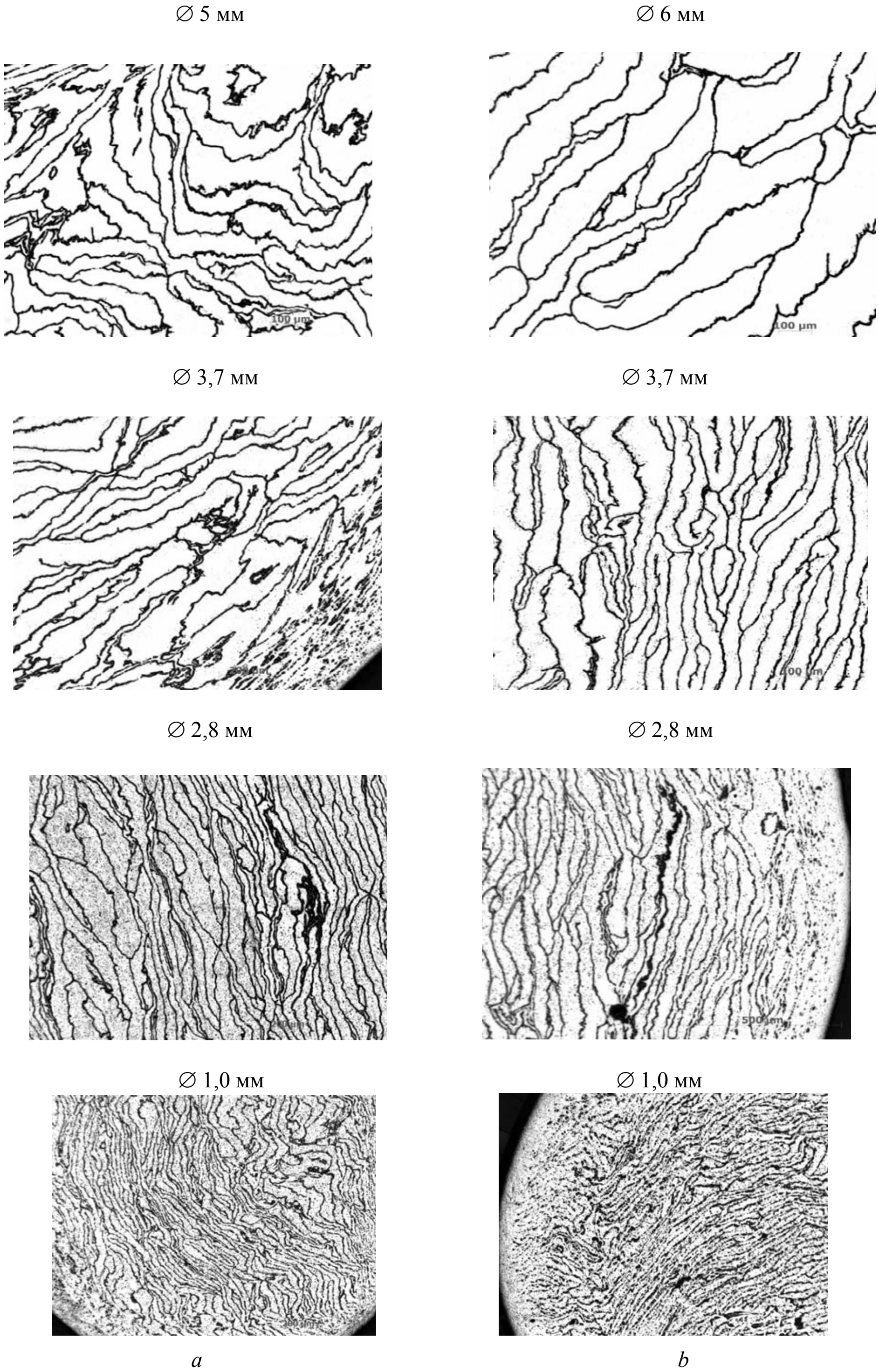


Fig. 5. The cross-sectional microstructure of the wire made in different compression modes from rods with diameters of 7 (*a*) and 9 (*b*) mm

Samples were taken from the middle part of the rods to determine their characteristic structure and mechanical properties. These samples were tested afterwards according to the existing standards.

Fig. 4 shows the results of metallographic research, which shows that the level of deformation during laboratory testing is inefficient for setting the chips particles. The microstructure has clear boundaries between particles in the form of oxide film and uncommon continuity. There is no difference between the sample structure of rods with diameter 7 and 9 mm. Consequently, we may speak about a typical stable structure in chip particles of different thickness, lengthened to the direction of extruding and separated by steady oxide film. There is no setting between chip particles. In other words, the physical contact generally takes place on the chips irregularities with spreading (not breaking) of oxide film over the contact area.

The next technological stage was the cold drawing of the pressed rods and making wire with a final diameter of 1 mm which was carried out on the catenary drawbench applying stress of 50 KH and without intermediate annealing of average compression ε_{cp} = 15–20 %.

The experiment did not have a task of optimizing the compression mode.

Sampling for metallographic testing and mechanical properties assessment was carried out on the intermediate diameters. The deformation degree by that moment was given by:

$$\varepsilon_{\Sigma} = \frac{d_0^2 - d^2}{d_0^2} 100\,\%.$$

Fig. 5 shows the microstructure of wire made in different compression modes from rods of 7 and 9 mm in diameter. The trend in structure changes along with the increase of relative compression is the same in both cases, i.e. the decrease of wire diameter results in the reduction of the structure and its granulated property in the surface layers of the wire. This is caused by an irregularity of deformation in the wire section when the layers contacting the dies are subjected to greater deformation. Fig. 5 shows the dispersion of oxide film in surface layers of the wire which must affect the properties of product. Surface discontinuity in structure is unlikely to affect it considerably, as it has random a character and is not connected to the suggested technological scheme.

The estimate of mechanical properties for the produced wire was carried out by tensile testing with determination of tensile strength σ_B, specific elongation δ and reduction of area ψ. Fig. 6 shows the testing results. The points indicate the averages for the five tested samples.

In result of our research we now have the technology of manufacturing aluminum wire from anisotropic composite material with the properties determined by the orientation of fiber in one direction. In this case it concerns "specified" technological anisotropy which occurs in definite schemes of plastic deformation. The fiber pieces (prolate chips) have different length depending on their thickness. This is why there is a different number of fiber pieces in the rod samples per cross sectional area in both longitudinal and crosscut direction. In addition, the higher the mode of deformation in drawing (the less the wire diameter), the longer are the boundaries between chips in their cross section.

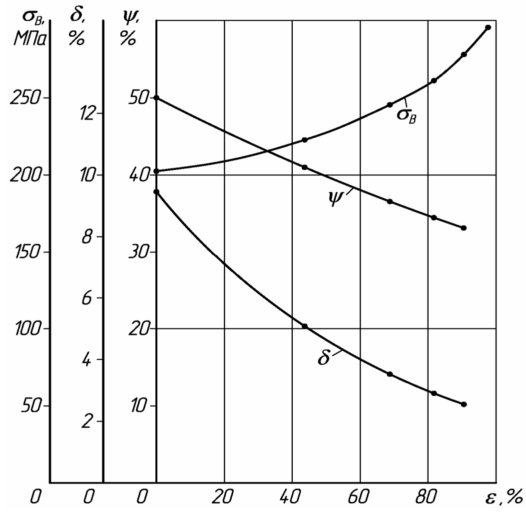


Fig. 6. Changing of mechanical properties of semi-finished items made of alloy АД31 chips after hot pressing (ε = 0 %) and cold drawing

Though it is difficult to indicate the application field of wire from friable fine chips of aluminum alloy АД31, we can suppose that due to its low production costs and mechanical characteristics, it can be used in fuse welding. According to the All-Union State Standard 7871 the tensile strength of such type of wire should not be less than 100 MPa. In addition it can be used for such mass consumption goods as binding wire. Mechanical characteristics for such goods are not specified, however the tensile strength in solid (not annealed) state should not be less than 140–160 MPa, and elongation not less than 2–3 %.

References

1. Patsekin V. P., Rahimov K. Z. Powder wire manufacture. M. : Металлургия, 1979.

Part
2

AVIATION AND SPACE-ROCKET ENGINEERING

O. G. Boyko
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

THE DEPENDENCE ANALYSIS OF DURABILITY, THE CATASTROPHE RISK AND THE TRANSPORT EFFECTIVENESS ON THE AIRPLANE CONSTRUCTION MASS

The analysis of conditions to provide durability, reliability and the transports effectiveness of the civil aircrafts is performed.

Keywords: durability, flying mass, empty mass, payload mass, effectiveness, catastrophe risk.

One of the main perfection characteristics of the civil aviation airplanes is ratio between the construction mass to the quantity of passenger seats, i. e. the airplane construction mass per a passenger seat. The lower this index is, the higher the airplane economic effectiveness will be. From the other hand, decreasing the construction mass is connected with decreasing its durability and increasing the catastrophe risk. In the work the attempt to analyse the conditions and restrictions within the limits of minimizing aircraft deadload is done.

The maximal take off mass of the airplane $M_{\max}$ may consist of: the payload mass $M_{\text{П.Г}}$ (passengers and cargos), the fuel mass M_T and the airplane construction mass M_K. For example, for the airplane Tu-204, the flying mass components are:

$$M_{\max} = 107{,}9 \text{ T};$$
$$M_{\text{П.Г}} = 25{,}2 \text{ T};$$
$$M_T = 32{,}7 \text{ T};$$
$$M_K = 50 \text{ T}.$$

When M_T and $M_{\max}$ are fixed, the change is possible only due to construction mass and the payload mass.

Aircraft construction durability the ability to resist the flight loads without destruction are defined by the section area of its power elements. When the construction materials are non-changeable, the durability is directly proportional to the construction mass. The durability defines the destruction risk or the catastrophe risk and it is exactly connected with payload mass. For Tu-204 airplane the connection is defined as:

$$M_{\text{П.Г}} = M_{\max} - M_T - M_K = 75{,}2 - M_K.$$

The task is to specify the influence of construction mass on its destruction probability, i. e. catastrophe realization.

Tu-204 airplane mass is 50 t. It is defined due to the condition of its ability to resist the double overload *n* without destruction under the maximal mass with the structural load factor *k* = 1.5. And the probability of destruction, due to NFC, must be $\le 1 \cdot 10^{-9}$ [1].

If the airplane construction mass $M_{\max}$ is divided into the overload value *n* = 2 and the safe load factor is *k* = 1.5, then under M_K = 16.6 t the airplane with maximal take off mass will be destructed with probability $Q = 1$.

The viewed conditions define the two points of destruction functional dependence $Q = f(M_k)$ of construction mass. It is absolutely obvious, that when $M_k \succ 50$ t, then the probability of destruction exists, although it becomes less then $1 \cdot 10^{-9}$. We can predict, that the probability of destruction will asymptotically tend to zero while the construction mass increasing. This prediction may be realized if the exponential dependence equals to $Q = f(M_k)$.

Then, according to the earlier highlighted conditions, we can reveal Tu-204 aircraft construction destruction probability:

$$Q = \exp 0.6138(16.6 - M_k).$$

Due to that formula the calculations of construction mass destruction probabilities and payload masses are calculated. They are shown in the tab. 1.

It is very difficult to build the graphic dependences $Q = f(M_k)$ in its changeable range from 1 to $5.82 \cdot 10^{-15}$. That is why the dependence $LnQ = f(M_{\text{П.Г}})$ is shown in fig. 1.

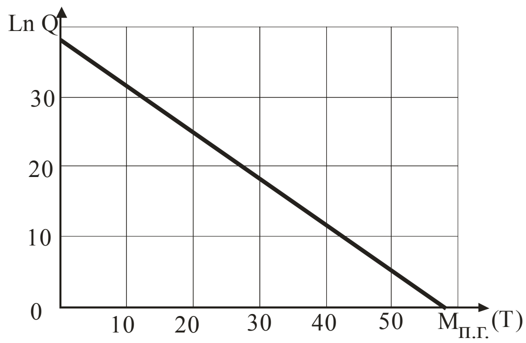


Fig. 1. Dependence of the airplane destruction probability of useful load mass

Tab. 1 and fig. 1 show us that in the range the construction mass is increasing, the probability of construction destruction is continuously decreasing. In order to have the graphic representation of the curve character $Q = f(M_k)$ there is a view of that line at the beginning and ending parts of M_k changing range in fig. 2 and 3.

Table 1

Construction masses, useful load masses and airplane destruction probabilities

M_k, т	16,6	20	30	40	50	60	70
$M_{п.г}$, т	58.6	55.2	45.2	35.2	25.2	15.2	5.2
Q	1	0.124	$2.6 \cdot 10^{-4}$	$5.78 \cdot 10^{-7}$	$1 \cdot 10^{-9}$	$2.7 \cdot 10^{-12}$	$5.82 \cdot 10^{-15}$

The presented pictures demonstrate that the meaning of rated construction destruction probability $1 \cdot 10^{-9}$ is neither peculiar nor characteristic point on the line $Q = f(M_k)$. There is a question how could we make the border meaning of catastrophic situation probability $1 \cdot 10^{-9}$? The answer is in paragraph 2.4 of ICAO continuing airworthiness manual [2]. It is said that: "…2.4. To estimate the design acceptability, it was necessary to specify the well-grounded probability meanings defined due to the next basement:

a) the exploitation experience claims that serious aviation event due to the exploitative and constructive character, may appear approximately one time per one million hours of flight. In 10 % of events it may be connected with breakdown conditions appearing due to aircraft systems breakdown. It was thought that for new constructions the probability of serious aviation events (that were summoned by the system breakdown) must not exceed that index. That is why it is necessary the probability of a serious aviation event happening due to the breakdown conditions does not exceed one event per 10 million hours of flight, so the probability must be less than10^{-7};

b) in order to be sure that index is realised, it is necessary to complete the complex quantitative analysis of the airplane's systems work reliability. The random assumption was made about approximately 100 potential breakdown conditions, which could prevent the continuing safe flight and landing. The predetermined probability of an event, which was 10^{-7} distributed uniformly among those conditions, finally the probability of each breakdown condition not more than $1 \cdot 10^{-9}$ was stipulated. So, the upper limit of probability of one breakdown condition, which will not allow continuing the flight safely and completing the landing, is based on the level of $1 \cdot 10^{-9}$ for each flight hour…"

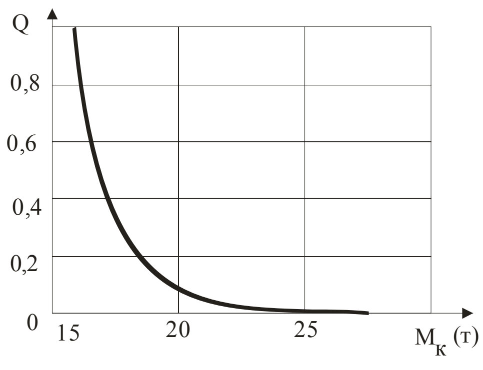


Fig. 2. Dependence of probability during lesser masses

This limit is undoubtedly defined for such breakdown condition, which is stipulated by insufficient reliability and durability of the airplane construction and each of its systems. It is obvious, that there are some essential uncertainties in determination of standard value for probability of catastrophic event $1 \cdot 10^{-9}$.

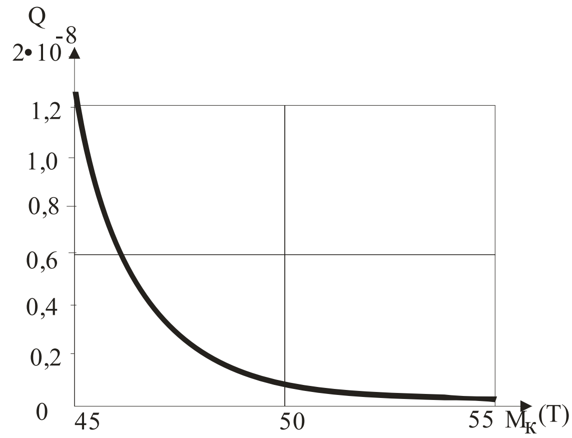


Fig. 3. Dependence of probability of construction destruction in range of the calculated mass

So in paragraph 2.4, *a* definition of serious aviation event frequency is less than $1 \cdot 10^{-7}$. It contains two suppositional estimations changing during the period of exploitation.

In paragraph 2.4, *b* the random assumption about 100 potential breakdown conditions, which guide to a catastrophe, was adopted. Finally, the upper boarder of catastrophic events probability was got caused by the breakdowns of aviation technique, that is $1 \cdot 10^{-9}$ per a flight hour. This estimation is perceived as generalized estimation of airplane constructing experience. It is obvious, careful obtaining of such estimation by the theoretical and experimental materials is impossible.

Airplane developers have to prove the inadmissibility of the catastrophic breakdown conditions with probability not more than $1 \cdot 10^{-9}$ in evidential documentation, which confirms that an airplane meets the demands of NFC. It is also a very difficult task.

There are estimations of attitude of the airplane construction mass to the maximal take off mass in the tab. 2 and 3. The average meaning of relative construction mass for short range airplanes is 0,586, for the medium range aircraft – 0.542, and for the long haul – 0.472. That divergence for airplanes of different classes is quite explainable.

The short range airplanes have the nonstop flight duration of 1–2 hours, and the long haul – 10–12 hours of flight. Having the same resource of 60 thousands of flight hours, the short range airplanes experience 5 times more load cycles, defined by take offs and landings, in comparison with long hauls. The diversity of relative masses for the same type of aircrafts is in limits of 11–12 %, it is very essential and hard to explain.

Judging by data of the table 1 and fig. 3 it is clear that, for examined Tu-204 aircraft, the decreasing destruction probability from meaning of $5.78\cdot10^{-7}$ up to $1\cdot10^{-9}$ for 1 hour demands to increase the mass of construction payload from 40 to 50 t, that in practice will guide to decreasing payload mass from 35.2 to 25.2 t. The reliability of airplane within destruction probability meanings of $1\cdot10^{-9}$ for 1 hour is hard to predict and prove, and its increasing is connected with valuable decreasing payload mass and commercial outcome – the competitive ability. The reliability increasing and decreasing the catastrophe damage, in this case, is connected with increasing the transport cost.

According to this 11–12 % of one class airplane construction diversity, under the same level of machine building production may be reckoned among various risk levels of the developers of airplanes, because decreasing the airplane's construction mass by 10 % increases the destruction risk from $1\cdot10^{-9}$ up to $5\cdot10^{-8}$.

References

1. АП-25. Aviation Rules. Airworthiness standards. M. : MAK, 1994.

2. Doc 9642-AN/941. Continuing Airworthiness Manual / International civil aviation organization. 1995.

Table 2

Calculations of correlation of masses for short range and long haul airplanes

Class	Close-rout					Far-rout					
Parameter type	Tu-134A	Jak-42	MD-81	B-737	A-320-100	IL-62M	B-707-320B	B-767-200ER	IL-96-300	A-340-200	MD-11
Start year	1967	1980	1981	1990	1988	1974	1962	1984	1992	1992	1990
M_{max}	47	57	63.5	52.4	66	167	151.5	175.5	216	251	273.3
M_k	29	33.5	35.5	31	38	73.4	67.1	83.8	117	118.6	126.7
M_k/M_{max}	0.617	0.588	0.56	0.59	0.576	0.439	0.443	0.477	0.54	0.472	0.463

Table 3

Calculation of masses for medium range airplanes

Class	Medium-rout						
Parameter type	TU-154M	B-727-200	B-757-200	A-320-200	IL-86	L-1011	A-330-300
Start year	1986	1971	1984	1988	1980	1972	1993
M_{max}	100	95	108.8	73.5	210	195	208
M_k	55	46.7	58.2	39.8	117.4	108.5	117.7
M_k/M_{max}	0.55	0.49	0.535	0.54	0.56	0.556	0.566

E. N. Garin
Siberian Federal University, Russia, Krasnoyarsk

THE INFLUENCE OF CONDITIONS FOR PASSING GLONASS AND GPS SATELLITE RADIO NAVIGATION SIGNALS ON THE ERRORS OF DEFINING RELATIVE COORDINATES

The influence of the conditions for passing GLONASS and GPS satellite radio navigation signals on the errors of defining relative coordinates is studied.

Keywords: signal delay, multipath propagation, error, frequency, navigation spacecraft.

The retransmission of navigation spacecraft signals of the GLONASS and GPS systems is known as repeated radiation after reception and processing in relaying equipment (further a repeater). Signal processing in a repeater can include intensification, frequency conversion, filtration and additional coding of a navigation spacecraft signal. The relayed signal passes the radio channel and goes on to the retransmitted signal receipt and processing equipment (RSRPE).

The RSRPE processes signals. The signals are transferred by a repeater according to the algorithm provided for the task. It is here that the measurement of radio navigation parameters (parameters of delay, Doppler frequency shift etc.) of navigation spacecraft signals are received by the object and the completion of navigation time task for the object is finished. The repetition of navigation spacecraft signals from a board of an object can be used for different purposes.

1. Remote determination of the coordinates and speed of an object (mobile or fixed) at a basic station (mobile or fixed).

2. Determination of objective coordinates that do not receive navigation spacecraft signals (or receive signals insufficiently to decide navigation-time task) on repeated signal of navigation spacecraft (NSC) radiated by the repeaters that are in favorable conditions for reception of NSC signals.

Navigation parameters of an object (coordinates, speed and direction of movement, parameters of time-and-frequency scale, spatial orientation) are possible to determine on the basis of measurement results of radio navigation parameters (delay, Doppler frequency shift etc.) of NSC repeated signals.

Such an approach can be used as an alternative to the system of remote navigation parameters definition of an object based on the set placed on the object from traditional radio navigation equipment completing the navigation time task and equipment transmitting information about the results of the navigation-time task completion. When using retransmission, the information transmission equipment is replaced by repeating equipment; this permits considerable decrease in navigation equipment weight, size, energy input, and cost by eliminating the block of digital processing of signals (BDPS) and the computing block (CB) from its structure.

An additional advantage the of navigation-time task on the relayed signals completion is the possibility to realize through the RSRPE a special condition for determining objective navigation parameters in real time. To organize conditions for transmission of such a sort from one object to another, no additional information is required.

It should be noted that as the navigation time task is performed with the RSRPE with the given object there is no information about its navigational parameters. This situation is normal for the object monitoring problems. If the object is required both to complete navigation tasks and/or to solve problems of control (guidance), there is a need in an additional channel for transmitting information to the object.

When a signal of NSC is transmitted to the relaying channel, the additional signal delay and Doppler frequency shift take place. The additional signal delay is measured by the distance between the repeater and the RSRPE. The additional Doppler frequency shift is caused by relative displacement of the repeater and the ERPRS. These factors distort the measurement results of NSC radio navigation parameters.

An additional factor that distorts the results of HTT completion for the object retransmitted NSC signal is a shift of the RSRPE frequency time scale from a repeater frequency time scale. To accept the propagation path impact of NSV retransmitted signal on NTT completion results in the RSRPE and impact on parameters of the RSRPE frequency time scale, a special pilot signal can be used. The special pilot signal is formed in the repeater and transmitted from the object along with the NSC retransmission signal. Principally, such parameters as signal delays and Doppler frequency shifts (DFS) of the pilot signal to the RSRPE frequency time scale are measured. The results of this measurement allow us to determine an additional NSC signal delay resulting from propagation path and the RSRPE time scale shift from retransmitter time scale and an additional Doppler frequency shift of NSC signals caused by relative motion of the retransmitter and the RSRPE and a frequency deviation of RSRPE reference generator from a frequency of retransmitter reference generator.

It is possible to suggest some methods of pilot signal use.

The first method is pilot-signal synchronization of the RSRPE reference generator. In this variant, the locking of the RSRPE reference generator (analog or digital) is regulated on the same level as a frequency of received pilot signal. The advantages of this method are the lack of additional complication of the RSRPE software in the primary processing and NTT completion. A disadvantage of this method is the limitation of the RSRPE carrying capacity. This way, the RSRPE is capable of simultaneous signal processing of only one retransmitter.

Another variant of the pilot signal use is a measurement of pilot signal frequency relating to the RSRPE reference generator. According to this method synchronization of the RSRPE processes is performed on the base of its own reference generator. Here the RG frequency does not fit the pilot signal frequency. In the RSRPE frequency deviation of a receipted pilot signal is compared with the nominal value and use of this information additional component is excluded from the frequency estimation of NSC retransmitted signal. An additional component results from propagation path and parameters of the RSRPE frequency time scale.

The aforementioned variant is more difficult to implement. Considering this fact the implementation requires processing software for the primary processing of the RSRPE information. The advantage of this technique is the absence of principal restrictions to the RSRPE carrying capacity. The RSRPE is capable to operate several retransmitters simultaneously.

Also it is possible for the RSRPE to operate a similar object. In these conditions the navigation time task is completed with the RSRPE the same way as if it was accomplished immediately on the object. Here the RSRPE represents an extension of the object navigational receiver, for example the block of digital processing of signals (BDPS) or the computing block (CB). Retransmitter equipment, retransmittion path and RSRPE receiving block are inserted into the gap between the receiving antenna, radio path (or its part) and other traditional radio navigation equipment blocks.

The RSRPE operation in the conditions of relative coordinate determination will be the following. In these conditions the RSRPE along with the reception of retransmitted signals from a retransmitter in the frequency range of NSC signal relaying receives in the standard frequency range. Cooperative processing of the measurement results of NSC signal received on the object and retransmitted through the repeater makes it possible to realize high-precision determination of the relative coordinates of the object relating to the RSRPE. In such conditions realization is possible for mobile objects both code and phase mode to determine relative coordinates. Phase mode realization is possible, due to the fact that in the RSRPE there is complete information about the frequency-timeline of an object by using a pilot signal.

I. A. Maximov
JSC "Academician M. F. Reshetnev" Information Satellite Systems", Russia, Zheleznogorsk

PROBLEMS OF SUPPORT OF MODERN SPACE CRAFT RELIABLE FUNCTIONING UNDER THE CONDITIONS OF THE DESTABILIZING INFLUENCE OF SPACE AND ANTHROPOGENIC FACTORS

The statistic analysis results of cases correlation of abnormal functioning of SC with increased levels of geomagnetic activity are represented in the article. The main results of completing work package in JSC "ISS" to research negative influence of space environment on SC, and the main results of the development of the protective means and methods are described in the article.

Keywords: space craft, SC (space craft) operation, space factors, anthropogenic factors and protective means.

The main operating characteristic of any satellite system is durability that is the ability to perform the specific tasks by all included space craft (SC) during the required life time.

The modern SC includes hundreds of radio-electronic blocks, optical devices and operating surfaces, thousands of structure elements and cable assemblies. This technical complex must operate during all life time (up to 15 years) under the conditions of negative space environment influence.

In the process of full – scale operation the SC is subjected to the influence of the wide spectrum of space factors (SF) and anthropogenic factors (AF). The results analysis of domestic and foreign SC operation indicates the presence of SC abnormal operation correlation of significant degree with variations of solar activity, space geomagnetic disturbance and anthropogenic conditions of SC operation.

According to the conclusion of the Federal Space Agency and Space Forces "…One of the main factors affecting on the characteristics stability and reliability of on-board radio – electronic equipment is space ionizing radiation.

Reportedly, there is 30–50 % of on – board radio – electronic equipment failures per the part of these effects" [1], and "the support of required stability is the most important task of SC manufacture with the long life time (10–15 years), contemplated by the Federal space program of Russia" [2].

Researches conducted by NASA and USAF also indicated, that up to 1/3 of failures at foreign SC operation have operational character and stipulated by the geophysical factors.

During researches the readable dependence between the level of geomagnetic activity and the rate of SC failure of different missions was detected.

It is determined the number of failures of SC on – board equipment rises several times due to solar activity increase [3].

The data analysis on anomalies in the operation of the domestic SC conducted by the specialists of 4 RF CSRI of Ministry of Defense, indicated that general number of failures of on – board SC systems, exchange dysfunction of the control and target information during high heliogeophysical activity increases 2–2,5 times, that, in its turn, sharply shortens the mean time of their active functioning. More than 50 % (on some systems up to 90 %) of them occur because of environment external action on SC on – board equipment. More than 80 % of such failure influences somehow on the performance of specific tasks [3].

At present all these problems are becoming more actual due to transition to the non-pressurized SC performance. The transition to such SC structure is caused by necessity of Life time increase up to 15 years and also by the increase of power supply capacity of newly developed SC. But at the same time the new mechanisms of the space environment influence on-board SC systems appear. All these influences without special analysis and necessary protective means can lead to the serious failures of on-board systems of the modern domestic SC.

Considerable effect on SC operability can also be produced by anthropogenic factors. So during the operation of stationary plasma thrusters (SPT) the number of anomalies in the operation of on-board systems was fixed. First of all it was connected to intensive interaction of SPT jet with the surface, proper outer atmosphere and high voltage SC equipment. Besides, SPT operation can generate optical interferences, influence radio signals transceiving, generate interferences in the control and power circuits, cause thermal, force, contaminating, erosive effect on the SC elements, influence the intensity and amplitude of the charge – discharge processes on the SC structure elements and interference situation on SC board.

It is obvious that the range of destabilizing factors influencing SC in the process of orbital operation is very wide. Therefore one of the main problems in the applied cosmophysics at the SC manufacture and reliable operation support is the problem of resistance to the space factors influence and anthropogenic factors.

To solve the problem, JSC "ISS" performed the work package to research SF and AF levels and influence mechanisms on SC on – board systems, influence modeling, followed by development of the protective methods and means and also production of SC resistance control and diagnostics system, used in the process of ground tests and SC full – scale operation.

At present the following main results were obtained:

1. Influence mechanism of SF и AF on SC was revealed, though earlier it had not been taken into account at SC manufacture and full – scale operation:

– quasi-static potential difference, induced between power buses and SC body;

– development of the discharge processes in the result of SPT plasma influence on the high voltage SC equipment.

2. The indicated mechanisms influence on SC were manufactured and on their base there were necessary protective means and methods developed, which can provide the reliable functioning of the modern domestic SC.

3. In bench tests the effectiveness of filters, developed for the protection against potentials quasi – static difference between power buses and SC body, was checked.

4. The complex of experimental researches on determination of appearance and development criteria of the discharge processes in high voltage equipment in the condition of SPT plasma influence was performed.

5. The full – scale experiments were prepared and conducted, they permit:

– to estimate the influence level on SC charge of artificial plasma formations;

– to confirm the effectiveness of the developed protective means from the induction between power buses and SC body of potentials quasi – static difference, occurred in the result of differential charge of SC structure elements;

– to determine the plasma parameters of the nominal SPT in the places of on – board equipment and also to estimate the degree of SPT plasma influence on SC structure charge.

References

1. The problem items of perspective space complexes manufacture, functioning cycle durability and reliability and the development of the corresponding technologies : Resolution № 4 of section of scientific and engineering board joint meeting of the Federal Space Agency and Space Forces (18-th of October, 2007). M., 2007. P. 2–3.

2. Resolution № 3-RK board of № 27-2002 Russian Aviation – Space Agency (20-th of October, 2002). M., 2002. P. 4.

3. Et al the development of the technical reasoning of outer environment control equipment installation on "Glonass-M" SC and "Glonass-K". STR on integral part of Experimental Design Work "Glonass-MK" / V. S. Krymov [et al.]. M., 2005. P. 19–20.

Part
3

TECHNOLOGICAL PROCESSES AND MATERIALS

A. A. Bezrukih, R. A. Meister
Siberian Federal University, Russia, Krasnoyarsk

M. A. Lubnin, S. A. Gotovko
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

INVESTIGATION OF SURFACING TECHNOLOGICAL OPPORTUNITIES BY THE INDIRECT THREE-PHASE ARC WITH CONSUMABLE ELECTRODES

This paper is devoted to the high-effective process of surfacing by the indirect three-phase arc with covered electrodes. The share of the parent metal in the surfaced one does not exceed 10 % which is infeasible for one-arc surfacing.

Keywords: surfacing, indirect three-phase arc, consumable electrodes.

Operating conditions of technical systems, including the space machinery, require manufacturing of units and parts with specific properties of working surfaces, such as: heat resistance and thermal stability, wear resistance, corrosion resistance, etc. Thus, economical use of expensive materials is necessary. The problem is solved by using surfacing processes. However, widespread one-arc method of surfacing with consumable electrodes has grave disadvantages such as low productivity and necessity of coating with four or five layers for required structure of surfaced metal.

Application of alternating current in surfacing processes allows using simple and cheap power supplies that have higher efficiency in comparison with rectifiers. The alternating current arc doesn't have such a disadvantage as magnetic blowout. Therefore, using welding transformers, especially-three-phase ones, is rather perspective in surfacing processes.

Three-phase arc has the following advantages: high efficiency of the process reached due to simultaneous melting of three electrodes at once; considerable decrease of fusion penetration in the absence of current in part being surfaced; intensive stirring of molten metal in the pool providing homogeneous structure of rollers. The distinguishing feature of the method is the following: the product being surfaced is not included in a circuit, the three-phase arc is located between three cores, i. e. three one-phase arcs are simultaneously on fire in one general torch, mutually influencing each other and electrodes. Such interaction considerably increases productivity of surfacing operations, reduces the specific consumption of the electric power and strengthens ionization of arc space. Thus, the fusion penetration of the parent metal considerably decreases in comparison with one-phase surfacing processes. It occurs only due to the thermal influence of three-phase arc and molten electrode metal. As a result of 1,6–3,0 mm diameter electrode manufacturing, it is possible to reduce the weight of electrode holders and current-carrying wires for three-phase arc processes that has solved a problem of the method realization [1].

However, in references there is no practically information about application of surfacing by indirect three-phase arc with three electrodes. As a result, investigations have been carried out on three-phase surfacing with consumable electrodes for the purpose of roller formation with the minimum share of the parent metal in the surfaced one by optimization of process parameters.

For the experiments with three-phase processes, a special plant with TSHS-1000-3 transformer has been assembled. The surfacing scheme is presented in fig. 1. Dropping characteristic was provided with single throttles with ring magnetic conductors. It gives the chance to regulate and equalize current in each phase that provides uniformity of burning-off.

The following demands, besides standard ones, are made of electrode covering for surfacing by three-phase arc: electroconductivity and slag viscosity. Their violation may cause short circuits between cores on the inetrelectrode covering layer, complicating a process schedule. Electrodes with three cores in general coating are not produced in industry. They are assembled in laboratory environment from normal 1.6–3 mm diameter electrodes of MP-3, O3C-12 marks, etc. Bundle manufacturing includes coating of the cores tied up by threads with silicate glue, after its drying they are fastened through equal intervals with a thin wire and are dried at 200 °C [2]. Conglutination prevents a divergence of electrodes at bundle burning. The identical distance *I* between cores (fig. 1) provides them with equal electric parameters of modes; that promotes their uniform melting. Assembled bundle was fixed in a simple electrode holder with three mutually isolated copper plates.

The initiation of the indirect arc is carried out by contact of the product with three electrodes while connecting a neutral main to the product. After initiation of the arc and the part warming up, the neutral main is disconnected from the part. It is impossible to turn off the arcs without power cutting; it is a disadvantage of the process. Unlike one-arc welding it is impossible to change arc voltage during surfacing process in this case. Voltage on the arcs depends on the distance between electrodes and the thickness of their covering.

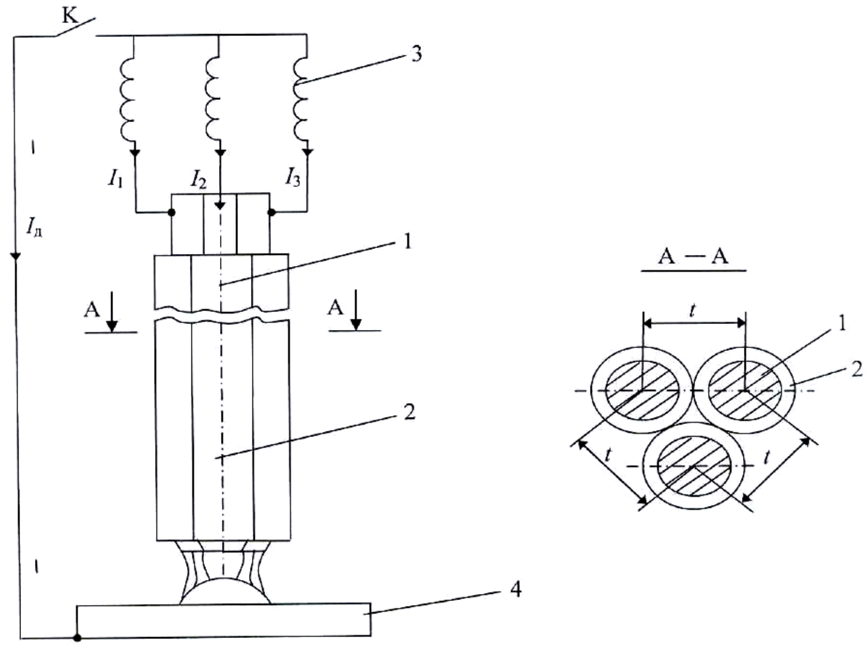


Fig. 1. Scheme of surfacing by indirect three-phase arc:
1 – electrode core; *2* – coating on electrode; *3* – secondary windings of three-phase transformer; *4* – part; I_1, I_2, I_3 – currents in electrodes; $I_д$ – current in neutral main; *K* – contact of neutral main switching off; *I* – distance between electrodes

The most important indices of independent three-phase arc burning stability include simultaneous initiation of three arcs, uniformity of three electrodes melting, burning and spitting losses, stability on fusion penetration depth and roller formation, distance from an independent three-phase arc to the part, absence of three electrodes meltback during breaks in arc burning [3].

Surfacing factor $Q_н$ (g/(A·h)) in experiments was defined as follows:

$$Q_н = G_н \cdot 3600/(3 \cdot I \cdot t),$$

where $Q_н$ – surfaced metal mass, g; t – time of arc burning, s; I – linear current in electrodes, A ($I_1 = I_2 = I_3$).

Percent of burning and spitting losses – ψ was defined according to the formula:

$$\psi = G_{бр}/ G_н \cdot 100 \ \%,$$

where $G_{бр}$ – spark mass, g.

Surfacing was made on scraped bright to metallic luster plates of low-carbon steel of 3–5 mm thickness at following parameters: $I_3 = 55–60$ A; $U_д = 30–32$ B; $U_{xx} = 80$. Metal deposit factor was 6.23–6.37 g/(a·hour), burning and spitting losses – 15–10 %. Thus, efficiency of this process is higher, than at single-phase surfacing as energy is consumed only for electrode bundle melting.

The difference between traditional one-arc process and surfacing by indirect three-phase arc is in considerably greater productivity. It is provided with simultaneous burning of three electrodes at once. Moreover, according to the experiments, on identical current modes three electrodes at surfacing by indirect three-phase arc are melt 20 % faster, than one electrode at usual one-arc surfacing. High efficiency is provided also due to the parent metal fusion penetration reduction. At surfacing with one electrode, the value of parent metal share in surfaced one usually is 30–40 %. As a result, most of the surfacing metal penetrates the parent one, and pure surfaced metal may be provided only in 4th or 5th layer. During surfacing using the method being investigated, samples with share of parent metal in surfaced one less than 10 % have been obtained.

After surfacing samples with rollers were cut across, polished with subsequent etching in the 30 % solution of nitric acid for meltback limit revealing. The share of parent metal in surfaced one was defined by means of digital pictures under the following formula:

$$\lambda = F_{пр}/F_п,$$

where $F_{пр}$ – fusion area of the parent metal; $F_п$ – total cross-section area of the roller (fig. 2).

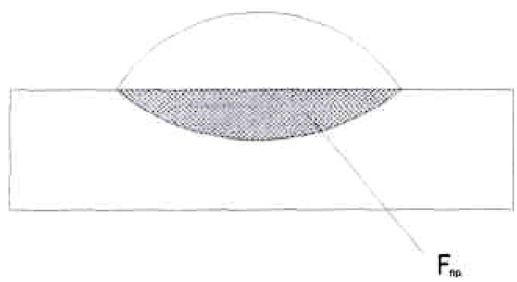


Fig. 2. Cross-section area of the roller

It is established that fusion penetration of the parent metal is minimal; the share of the parent metal in surfaced one has reached 10 %. It is obvious, that the method of three-phase surfacing by indirect arc allows solving a problem of parent metal fusion penetration. Producing of pure surfaced metal without impurities of parent metal is possible from the first or the second layers that is inaccessible with standard one-arc methods of surfacing. It is revealed, that current increasing considerably improves weld forming, initiation of arc, and stability of its burning. However, fusion penetration increases from 2–7 to 8–10 %.

Surfacing on simple steel with electrodes of austenitic class has shown also good results. 2.5 mm diameter electrodes OK 61,30 of ESAB were used for surfacing. Minimal fusion penetration of parent metal has been obtained at the following modes: $I_{эл}$ = 50 A, $U_{л}$ = 32–34 B, U_{xx} = 75 B. The share of the parent metal in the surfaced one does not exceed 5 % (fig. 3).

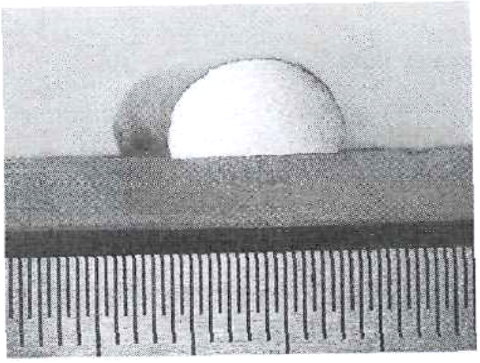


Fig. 3. Outward appearance of the sample surfaced with electrodes of austenitic class in cross-section

Micro section metallographic specimens are made from surfacing samples with electrodes MP-3 for microstructure character revealing (fig. 4). Reagent of the following composition was used for micro section metallographic specimen etching by means of rubbing in: HCl – 60 cm^3, $CuSO_4$ – 12 g, H_2O – 60 cm^3. Examination was carried out using METAM ЛВ-31 microscope with × 50 and × 100 magnification.

The micro section analysis has revealed the structure lamination of the surfaced roller (fig. 5), relating probably to the specific character of the electrode bundle. The dendrite structure typical for melting processes is well visualized. The boundary between the parent metal and the surfaced one is distinguished.

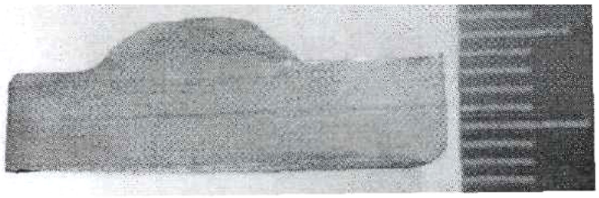


Fig. 4. Cross-section of the roller surfaced with 2 mm diameter MP-3 electrodes

The transitional area (fig. 6), from which a growth of columnar crystallite blocks begins, consists of the parent metal and the surfaced one. Heat-affected zone has a coarse-grained structure, as a result of overheating of the parent metal during surfacing. It is evidence of high efficiency of thermal source. Overcoming coarse-grained structure of the parent metal near meltback boundary is possible by forced cooling of a part back side with running water, by means of surfacing speed increasing, and the subsequent thermal processing (normalization).

Conclusions

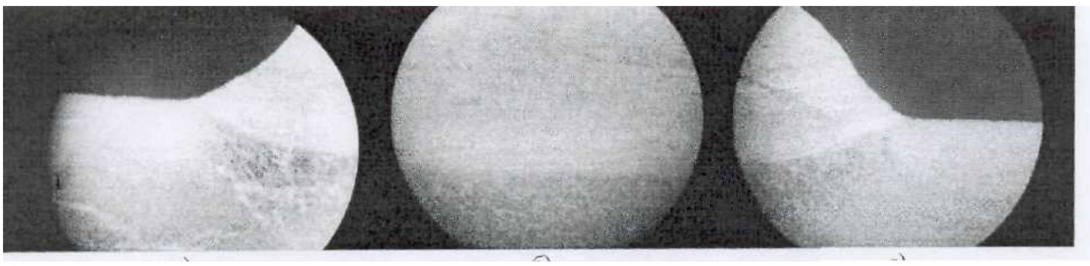
– the minimum current of steady-state combustion of indirect three-phase arc for 2 mm diameter electrodes is 55 A. At reduced values the initiation of the arc becomes difficult;

– the optimal interval of currents is 60–70 A. Exceeding these values causes excessive depth of the parent metal fusion penetration;

– one has obtained samples of surfacing by indirect three-phase arc in which the share of the parent metal in surfaced one is less than 10 %.

High efficiency of surfacing technological process by indirect three-phase arc with covered electrodes is confirmed.

Surfaced roller lamination, related to electrode bundle, and dendrite structure are revealed by metallographic method. The boundary between the parent metal and the surfaced one is distinguished, it relates, probably, to reduced influence of indirect arc on the parent metal.



a *b* *c*

Fig. 5. General view of microstructure in surfaced metal zone (x 50):
a, *c* – roller edges; *b* – the middle

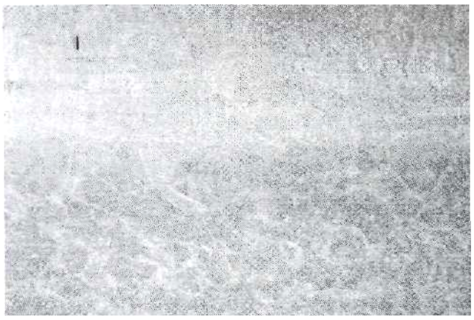


Fig. 6. Transitional area, × 100

References

1. Михайлов Г. П. Сварка трехфазной дугой. М. : Машгиз, 1956.
2. Мейстер Р. А., Падар В. А., Безруких А. А. Наплавка трехфазной дугой плавящимися электродами // Машиностроение: сб. науч. тр. / отв. ред. Е. Г. Синенко. Красноярск : СФУ, 2007. С. 30–34.
3. Мейстер Р. А., Мейстер А. Р., Безруких А. А. Многодуговая наплавка покрытыми электродами // Сварочное производство. 2008. № 8. С. 21–23.

N. V. Morozov, V. P. Karasev
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

STEAM TURBINES WITH A LOW-BOILING WORKING AGENT

The subject of the article is the assembly of a steam-generator plant with a natural working agent. A method of calculation for steam turbines with a low-boiling working agent is offered, which accounts for the correlation between the adiabatic curve indication, pressure and temperature in the overheated vapor area.

Keywords: steam turbine, Freon, adiabatic curve indicator, low-grade heat, coolant.

Nowadays the problems of electric power production are becoming more and more essential. This is the result of the rising cost of energy sources (oil, gas, coal) and consequently their consumer prices are growing.

In the routine work of industrial enterprises (using various heat-carrying agents) a great amount of heat which could be used in some other production cycles is dumped and lost. These heat-carrying agents can vary from slightly warm sewage water with low temperatures to coal coking gases with the temperature up to four hundred degrees (Celsius). Other sources of heat can come from all kinds of plants and systems that dump heat in their working cycle. Some of these sources and their temperatures are listed in the table.

Heat-carrying agents

Heat-carrying agents	Temperature, °C
Sewage water	15–19
Industrial gases flow	250–300
Heating equipment temperature	30–100
Coal coking gas temperature	400–430
TV3-117 engine oil	80–150
VR-14 reductor oil	70–80

Thermal power could be accumulated by means of thermal pumps. But thermal pumps cannot convert heat into other kinds of energy. And it is desirable to get from this heat the power that is easy to be transmitted for long distances, i. e. electric power.

Rather low temperature of most heat sources, their non-gaseous state, such as that of oil in helicopter systems, for example, as well as rather low pressure in gaseous sources do not allow direct application of their thermal energy. The energy is to be extracted by means of low-boiling working agents, such as in thermal pumps. After the heat is extracted from the source, its thermal energy is converted into mechanical energy by a rotodynamic machine-steam turbine.

Taking into consideration the ecological and economic requirements, the working circuit of the plant should be a closed system, i. e. the working agent is to be used many times over. As coolants are usually chemically active and dangerous compounds, this system permits to avoid environmental pollution by the direct ejection of the used coolant from the system. The possible harmful influence on human health is ten times less.

To put this closed system into practice it is necessary to install a component for circulation of the working agent inside the system. The component can be a pump (for the circulation of a liquid working agent) or a compressor (for the circulation of the agent in the form of overheated vapour). Unlike the pump, the compressor system doesn't require conversion of the working agent into liquid phase on its outlet from the turbine.

The installation (fig. 1) consists of the reservoir filled with a working agent, the pump, the evaporator, the turbo generator and the condenser. The coolant Freon R22 (the chlorinedifluoromethane) is considered as a working body. This Freon is the most suitable to the given system under its physical and chemical characteristics and it is widely used in the modern refrigerating equipment.

The system working cycle diagram shows four sections (fig. 2). The first section *1–2* shows the feeding of the working agent to the evaporator, thus increasing the Freon pressure in the system and slightly raising the temperature because of the losses in friction. The second section *2–3* shows the evaporation and overheating of the working agent in the evaporator at a constant pressure and the extraction of heat from the source. The third section *3–4* corresponds to the conversion of the working agent thermal energy into kinetic energy of the turbine shaft, while the temperature and pressure are getting lower. The fourth section *4–1* shows the condensation of the working agent to the initial parameters that were at the pump input, which is necessary for the next cycle.

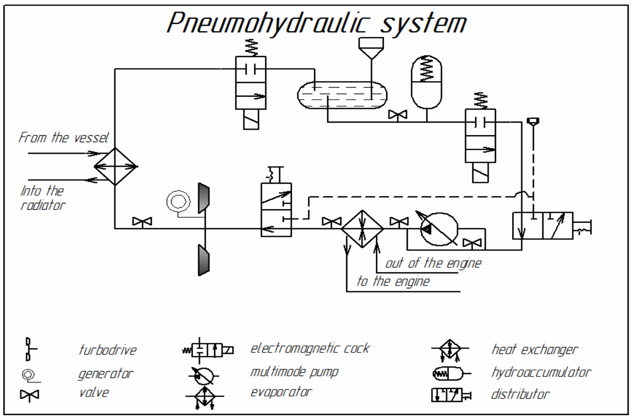


Fig. 1. Pneumohydraulic system using helicopter oil system as a source of heat

The basic points influencing the system power output are points *3* and *4* (fig. 2). Point *4* is the starting point of the working agent condensation and is determined by the type of the condenser and the substance used for heat extraction. Point *4* can be in a low temperature sector if the heat is extracted by a different coolant. But it is more effective to use the environmental media, such as air and, if possible, water. Point *3* is the point of optimal overheat of the working agent. It reflects the optimal overheating of the agent at the given evaporation pressure and is determined by the position of point *4*. Maximum temperature at point *3* is limited by the highest possible temperature of a given coolant. The most suitable heat sources can be determined by evaluation of the working agent parameters at point *3*. When air is used as a condenser (steam parameters at point *4* are 11.8 bar and 36 °C) the system's optimal temperature at point *3* will be 100 °C at 30 bar pressure. For that variant of the system all sources with the temperature of 110 °C and more are suitable. To make the system work more effectively a part of the working agent can be converted into liquid so that there is more used energy to be transferred at heating of the working agent. Modern steam turbines can work with 20 % of liquid agent, but there are still some problems to be solved, as there are no reliable methods of calculation of the percentage of phase correlation for the flow of the axial turbine.

For pipelines with a small amount of working agent and large pressure break the optimal efficiency can be achieved when using an axial-flow turbine with partial working agent feeding.

The capacity at the turbine and compressor shaft is derived from adiabatic work $L_{\text{ад}}$, working agent consumption, turbine efficiency η_m or the compressor $\eta_к$ [1–3]:

$$N = G \cdot L_{\text{ад}} \cdot \eta_m,$$

where N is the capacity at the turbine shaft, W; $L_{\text{ад}}$ is adiabatic work, kJ/kg; G is working agent consumption, kg/s; η_m is turbine efficiency.

Adiabatic work at the turbine and compressor shaft is derived from working agent parameters

$$L_{\text{ад}.m} = \frac{k}{k-1} \cdot R \cdot T_{\Gamma}^{*} \cdot \left(1 - \frac{1}{\left(P_1 / P_2 \right)^{\frac{k-1}{k}}} \right),$$

$$L_{\text{цбк}.s} = \frac{k}{k-1} \cdot R \cdot T_{\text{вх.к}} \cdot \left(\pi_{\text{цбк}}^{\frac{k-1}{k}} - 1 \right),$$

where $L_{\text{ад}.m}$ is turbine adiabatic work, kJ/kg; R is gas constant, J/kg·K; T^{*}_{Γ} is gas temperature at the turbine input, K; P_1 and P_2 are gas pressure before and after the turbines, Pa; k is isentrope index; $L_{\text{цбк}.s}$ is compressor adiabatic work, kJ/kg; $\pi_{\text{цбк}}$ is degree of pressure increase in the compressor; $T_{\text{вх.к}}$ is gas temperature at the compressor input, K.

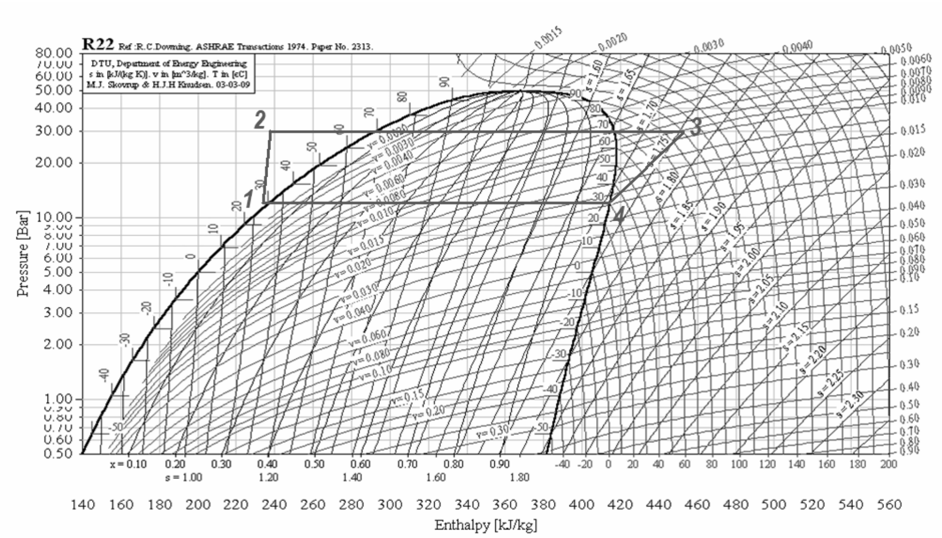


Fig. 2. Plant cycle in *lg P–I* coordinates

Adiabatic work also determines adiabatic speed, sound speed and the temperature at the turbine or compressor outlet:

$$c = \sqrt{2 \cdot L_{\text{ад}}}\ ; \quad a_{\text{кр}} = \sqrt{2 \cdot \frac{k}{k+1} \cdot R \cdot T_{\text{вх}}}\ ;$$

$$T_{\text{вых.т}} = T_{\Gamma} - \frac{L_{\text{ад}.m}}{R\frac{k}{k-1}}\ ; \quad T_{\text{вых.к}} = T_{\text{вх.к}} + \frac{L_{\text{цбк}.s}}{R\frac{k}{k-1}}\ ,$$

where c is adiabatic speed, m/s; $\alpha_{\text{кр}}$ is sound speed in the nozzle, m/s; $T_{\text{вых.к}}$ is gas temperature on its outlet from compressor, K; $T_{\text{вых.т}}$ is gas temperature on its outlet from turbine, K.

These parameters are essential for the kinematic calculation and the design of the directing unit blades, and the compressor driving wheel, and the nozzle assembly, and the turbine driving wheel.

The processes taking place in the compressor and the gas turbine with traditionally used working agents are studied in full and calculation algorithms are valid to obtain the precise parameters for the flow section.

The particular feature of the low-boiling working agents is their variable adiabatic curve indicator depending on the temperature and pressure [4].

There is no analytical dependence of the adiabatic curve indicator on the temperature, there are only tabular values.

Using scientific literature [4; 5] we can draw a surface showing the adiabatic curve indicator changes depending on the temperature and pressure (fig. 3).

The calculation of the adiabatic work at the turbine shaft and the temperature on the outlet, taking into account the variable factor of the adiabatic curve, is also given:

$$L_{\text{ад}.m} = \frac{k'}{k'-1} \cdot R \cdot T_{\Gamma}^{*} \cdot \left(1 - \frac{1}{\left(\frac{P_1}{P_2}\right)^{\frac{k'-1}{k'}}}\right),$$

$$T_{\text{вых}.m} = T_{\Gamma} - \frac{L_{\text{ад}.m}}{R\frac{k'}{k'-1}}$$

where $k' = f(P, T)$.

As the surface has a complex profile with numerous peaks, it cannot be described taking into account only one analytical dependence.

The calculation is based on final differences method, the sector being divided into small pressure values, with the following summing up of the adiabatic work values for each segment.

The calculation results are given in fig. 4 and 5.

The most influencing changes of the adiabatic curve indicator are found in the temperature zone which is close to the working agent decomposition temperature, up to 5 % of adiabatic work and 20 % of the working agent temperature on its outlet. The influence of the adiabatic curve indicator also increases with the growth of pressure at the input.

Similar correlations can be worked out for the compressor. In compressor performance the influence of the variability indicator of the adiabatic curve will be more significant, as the working agent is the state conversion zone.

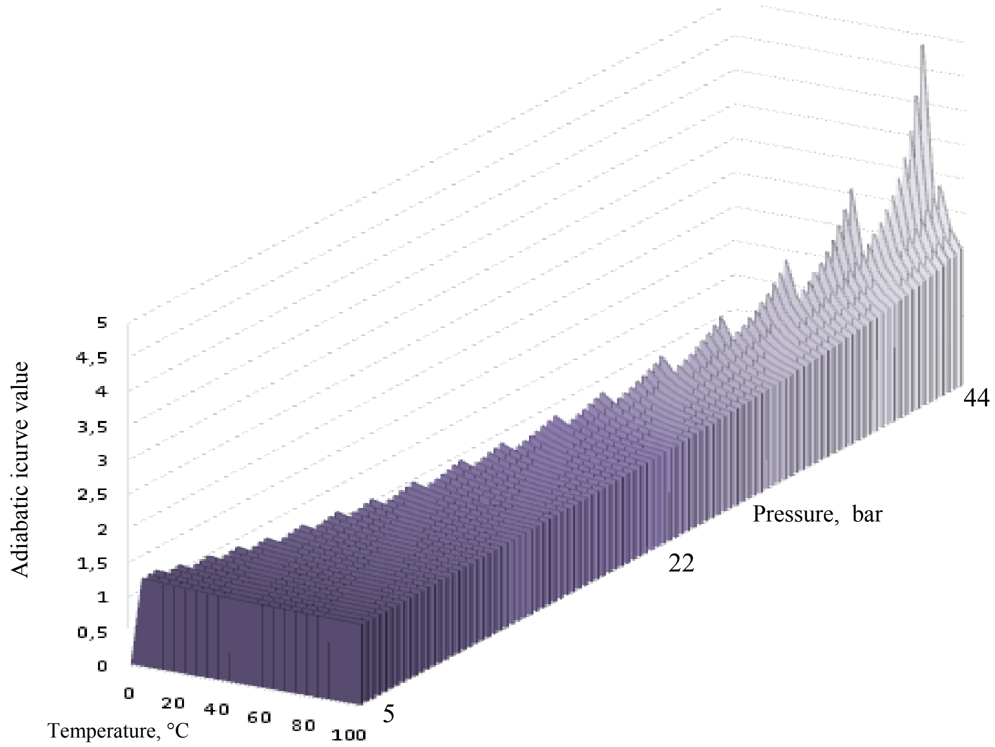


Fig. 3. Adiabatic curve indicator changes in dependence on the pressure and temperatures

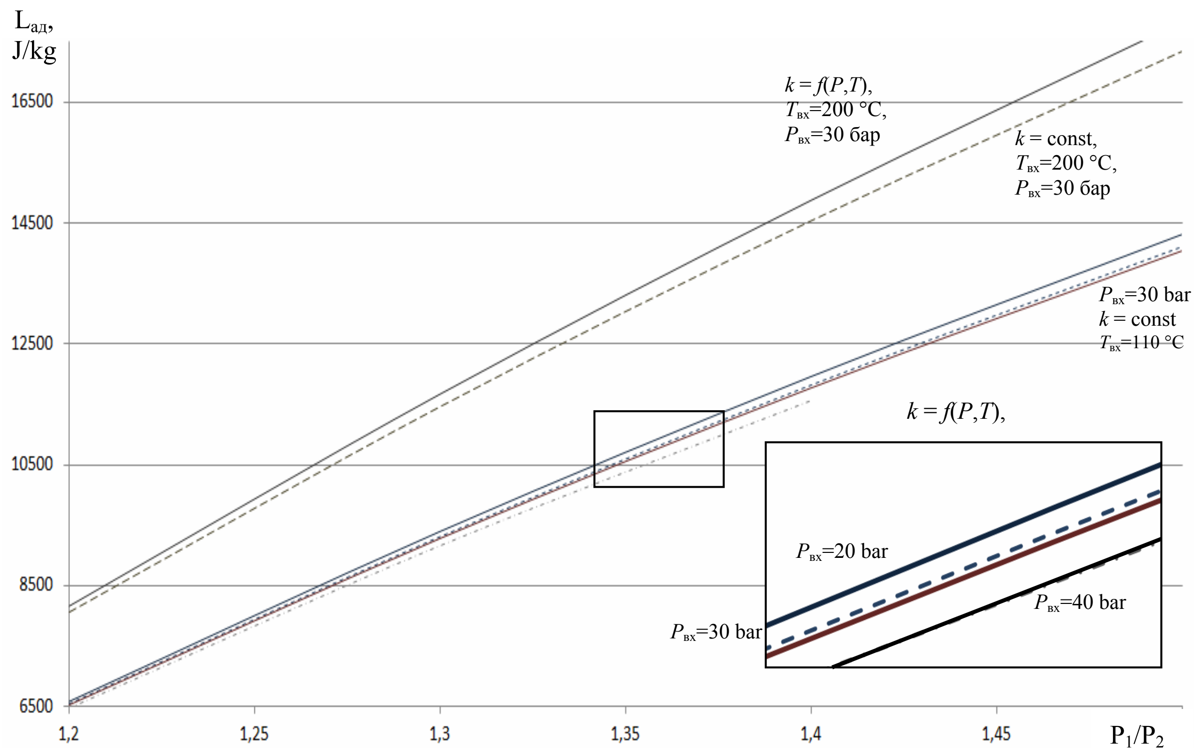


Fig. 4. Adiabatic work value at different parameters:
- - - - – k = const; ——— – $k = f(P, T)$

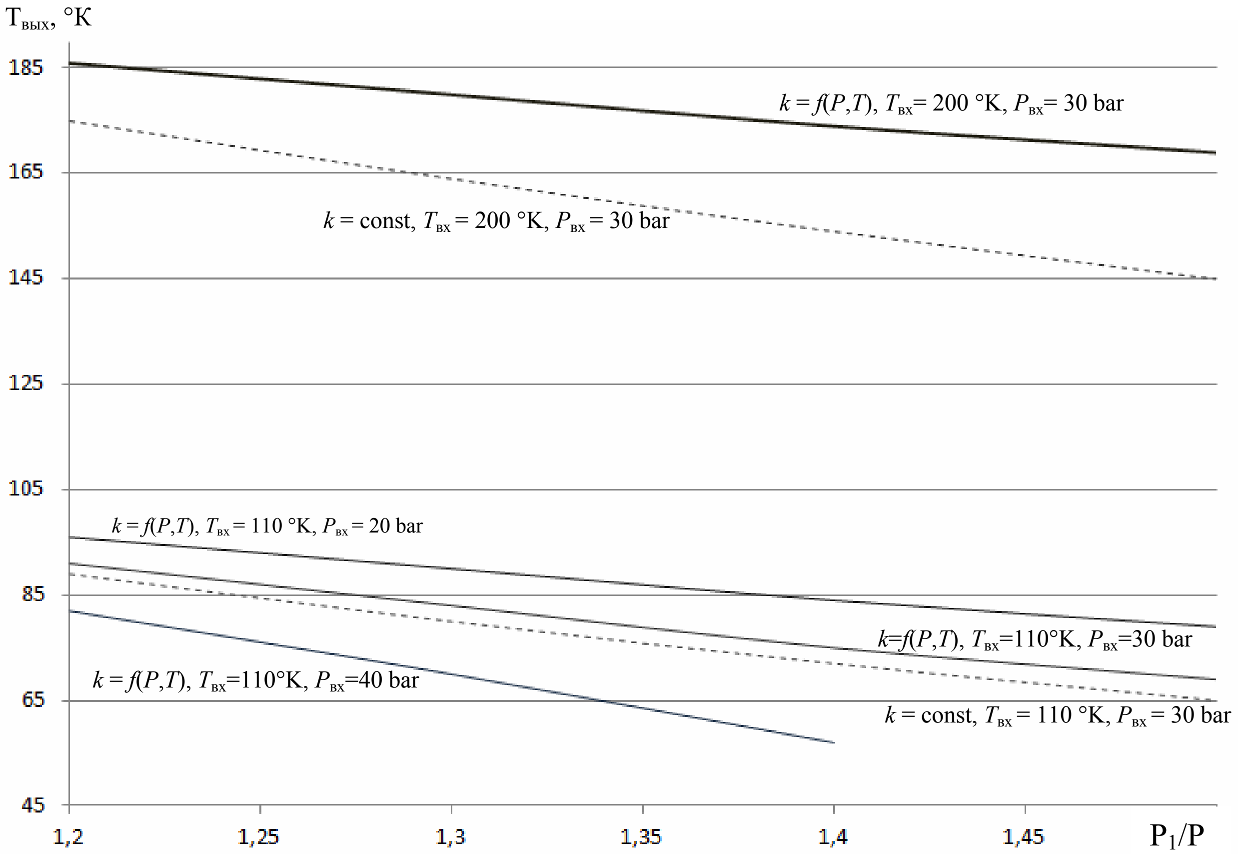


Fig. 5. Flow temperature values at the turbine escape at various adiabatic parameters:
- - - - – k = const; ——— – $k = f(P,T)$

The given calculations show that the registering of the adiabatic curve indications is necessary for more accurate determination of the thermodynamic parameters for the compressor and turbine performance. The calculation of the kinematic parameters of the flow in the system and the design of the flow sector and turbine or compressor blades should also be carried out taking into account the variability of the adiabatic curve indicator.

References

1. Паротурбинные установки с органическими рабочими телами / М. М. Гришутин, А. П. Севастьянов, Л. И. Селезнев, Е. Д. Федорович. Л. : Машиностроение. Ленингр. отд-ние, 1988.

2. Епифанова В. И. Компрессорные и расширительные турбомашины радиального типа : учебник для вузов. 2-е изд., перераб. и доп. М. : Изд-во МГТУ им. Н. Э. Баумана. 1998.

3. Нимич Г. В., Михайлов В. А., Бондарь Е. С. Современные системы вентиляции и кондиционирования воздуха : учеб. пособие. Киев : Изд-во ИВИК, 2003.

4. Теплофизические свойства фреонов. Т. 1. Фреоны метанового ряда : справ. данные / В. В. Алтунин, [и др.] ; под ред. С. Л. Рив-кина ; Госстандарт ; ГСССД. М. : Изд-во стандартов, 1980.

5. Конструкция и проектировочный расчет жидкостных ракетных двигателей : учебник / Г. Г. Гахун, [и др.] ; под общ. ред. Г. Г. Гахунина. М. : Машиностроение, 1989.

Yu. V. Morozov, S. I. Snisarenko
Novosibirsk State Technical University, Russia, Novosibirsk

COMPUTERIZED TEST STAND FOR INVESTIGATION OF STRAIN WAVES IN COMPOSITE CONSTRUCTIONS

The paper presents a test stand for investigation of ramp loading of composite constructions such as beams and plates. This ramp loading was performed using a striker with a piezoelectric force transducer. The velocity of the striker is up to 40 mps. The registration of the measurements results is carried out using a computerized measurement system. The experimental results of contact shock force acquisition are stated with appropriate calculations of non-stationary bending and shearing strain waves in glass-fiber-reinforced plastic beams and plates.

Keywords: composite constructions, shock loading, strain waves, computer measuring system.

The percentage of composite materials content in modern military and civil aircrafts does not exceed 20. Increasing composites percentage up to 50 and more, like in Boeing-787 (introduced in 2007) revealed a number of drawbacks because of the insufficient shock strength of composite materials being used. The matter doesn't concern the damage caused by some extreme situations inside an aircraft but it concerns normal shock loads for constructions made of traditional materials. Thus, at present time it is necessary to create crash-proof construction elements made of fibrous-layered composite materials like glass-fiber and coal-plastic. The application of these materials is not possible without reliable expert evaluations obtained by numerical simulation combined with experiments.

Carrying out shock tests gives more empirical data than carrying out static tests. Moreover, the experiments on shock loading and non-stationary deformation even for simple constructions like beams are complicated by the lack of complete standards for test stands, sensors, instruments and test techniques. So, experimental data are required to be accurate and reliable.

The present paper proposes an automated shock loading test stand with the computerized measuring system based on the LabVIEW virtual instrumentation from the National Instruments Corporation. The developed impact force and dynamical strain acquisition procedure provide sufficiently accurate experimental data to design crash-proof composite constructions.

Automated shock loading test stand. The scheme of the automated test stand for studying beams and plates non-stationary deformation processes under lateral shock loading is shown in fig. 1. It consists of loading device, impact rate acquisition system, and strain gauging system, dynamic force measurement system and computerized measuring system.

The loading device is a light gas ballistic installation for ramp loading of objects with the strikers having mass from 5 to 100 g providing shock loading rate up to 40 mps [1]. The striker rate was measured with drag-type sensors during its flying between magnets N1 and N2 with the inductance coils L1 and L2. Beams and plates non-stationary surface deformations were measured by foil strain gauges KF-4P1-5-100-B-12.

The novelty of the dynamic force measuring system is in using the striker sensor [1] to measure impact contact force including its magnitude, duration and form change in the contact zone between the striker and an object. This sensor is made of piezoelectric tablets that have the diameter 8 mm and are 2.8 mm thick.

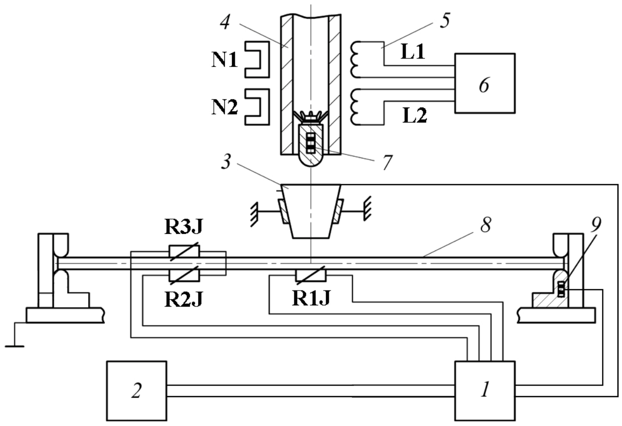


Fig. 1. The scheme of the automated test stand to study beams and plates non-stationary deformation processes under lateral shock loading:
1 – signal conditioning module; *2* – personal computer with built-in data acquisition device; *3* – conic contact surface; *4* – pilot borehole of light gas ballistic installation; *5* – striker rate drag-type sensor; *6* – electronic frequency meter to measure striker rate; *7* – piezoelectric striker-sensor; *8* – specimen under test (beam, plate); 9 – universal support with piezoelectric force sensor

The registration of signals from strain gauges R1J, R2J, R3J and a piezoceramic striker-sensor was acquired by the computerized measurement system shown in fig. 2. The measuring system consists of strain measurement system, force measurement system, hardware and software [2].

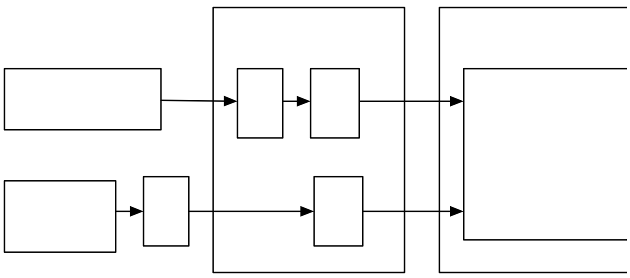


Fig. 2. Computerized measuring system block diagram

Hardware includes the following blocks:

− signal conditioning module SC1 NI SCXI-1520;
− terminal block TB1 NI SCXI-1314;
− signal conditioning module SC2 NI SCXI-1120;
− terminal block TB2 NI TBX-1316;
− chassis NI SCXI-1000;
− data acquisition (DAQ) device NI PCI-6221M.

Software includes the following components:

− graphical application development environment NI LabVIEW;
− drivers NI DAQmx.

SC1 is an 8-channel signal conditioning module converting strain gauge resistance change into electrical voltage. Gauges are connected to SC1 by means of the connector block TB1. Each channel has half-bridge completion resistors. RJ1 is connected by 1/4 bridge circuit and RJ2, RJ3 are connected by 1/2 bridge circuit.

SC2 is an 8-channel high-voltage signal conditioning module. The piezoelectric sensor is connected to SC2 by means of the terminal block TB2 which can accept voltage up to 1000 V. Each channel has an isolated amplifier and 10 kHz low-pass filter to protect the DAQ device and the computer from damage.

SC1, SC2 and TB1 are placed inside the shielded chassis SCXI1000 having power supply and 4 slots for signal conditioning modules and terminal blocks.

Output voltages from SC1 and SC2 are connected to the DAQ device 8-channel analog input subsystem. Two analog input channels are for the strain gauges and one channel is for the piezoelectric sensor. DAQ device also contains a double-channel subsystem of the analog output, an eight-channel subsystem of the digital input/output and two timer/counters.

DAQ device is plugged into a PCI slot inside the computer.

The computer controls DAQ device and signal conditioning modules, analyses data acquired and displays measurement results in digital and graphical formats by means of the graphical application development environment LabVIEW [2]. Interaction between LabVIEW and DAQ device and signal conditioning modules is controlled by DAQmx drivers.

Computerized measuring system error includes DAQ device analog input error and signal conditioning modules errors. Maximal error of analog input for the 16 bit analog-to-digital converter is 0.02 %. Maximal error of SC1 is 0.2 % and maximal error of SC2 is 0.8 %. (Error values are taken from the National Instruments specifications). Application of electronic scopes in old test stands [1] gave the strain measurement error 5.4% and the force measurement error 6.1 %. The computerized measuring system allows reducing the error to 2 % replacing electronic scopes by NI virtual instruments.

Impact force pulse shape identification. Fig. 3 shows the contact force $P(t)$ oscillogram for transverse central impact by the striker-sensor on a hinged beam made of symmetrically arranged glass-fiber $\left[0^0_8 / \pm 45^0_8 / 90^0_8\right]$. The dimensions of the beam are (100.0 × 10.0 × 10.2) mm. The impact rate is 5 mps and the striker mass is 51 g. A glass fiber monolayer has the following characteristics: $\rho = 1800$ kg/m^3, $h = 0.3$ mm, $E_{11} = 270 \cdot 10^8$ N/m^2, $E_{22} = 70 \cdot 10^8$ N/m^2, $G_{12} = 46 \cdot 10^8$ N/m^2, $\nu_{12} = 0.26$. The inclusion volume fraction of the bonding polymer in the composite package is 35 %.

It is known that a single impact takes place if the beam-to-striker mass ratio *m/M* is less than one [3]. If this ratio is more than one, impacts are repeated (fig. 3). In the context of the elementary theory of impacts we can explain this phenomenon in the following way: when the striker hits, the beam rate exceeds the rate of the striker, the beam advances the striker, the contact density decreases and at the moment $t \approx 3$ μs the first relative maximum appears. Then beam rate decreases and it behaves as a compressed spring and the striker does not change rate, the contact density increases and at the moment $t \approx 19$ μs the second maximum appears. After all the striker moves backward and at the time moment $t \approx 86$ μs rebound takes place.

The function $P(t)$ should be approximated to perform numeric calculations. If plates shock loading rate is less than 5 mps, contact force can be approximated rather accurately by a sine half wave with the maximal amplitude P_{max} and duration t_{max} [4, p. 8–46]. The presented experimental results show that if the impact rate is from 5 to 40 mps, other approximations are required.

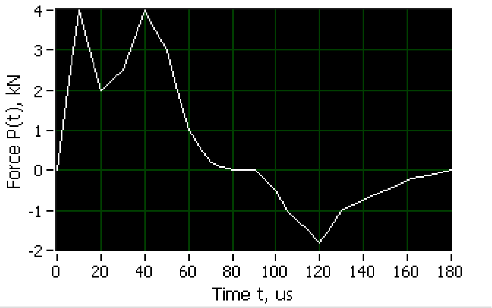


Fig. 3. Contact force time function $P(t)$ at transverse central impact on the glass fiber hinged beam

Fig. 4, *a* shows the contact force time function measured at transverse central impact on firmly fixed glass fiber plate with fiber placing angles $\left[0_8^0/\pm45_8^0/90_8^0\right]$ and geometrical dimensions (100.0 × 100.0 × 7.2) mm. The impact rate is 18 mps and the striker mass is 51 g.

The repeated impacts pulses envelope has a triangular form with rising and falling parts (fig. 4, *b*). The number of repeated impacts increases with the growth of shock loading rate. Such an envelope shape corresponds to impact rates from 5 to 40 mps. The initial part 2, where the load is increasing, is significantly less than the part 1 where the load is decreasing. Moreover, the higher the impact rate is, the less the initial part is.

Non-stationary deformations numeric simulation. Non-stationary deformations under impacts described by the function $P(t)$ have been calculated for composite beams by means of the finite elements method. The displacement equation has the form:

$$[M]\{\ddot{\delta}\}+[C]\{\dot{\delta}\}+[K]\{\delta\}=P(t)\{\delta_j\}, \qquad (1)$$

where $[M]$, $[C]$, $[K]$ are the mass, viscosity and rigidity matrices; $\{\delta\}$ is the generalized displacement vector; $P(t)$ is the contact impact force function; $\{\delta_j\}$ is the Kronecker symbol. This equation is integrated by the Runge–Kutta method.

The finite elements consider transversal shift with respect to the Timoshenko theory [5]. The 3rd order finite element with two nodes and four generalized displacements in a node is selected for a fiber multi-layered beam.

Mechanical behavior of composite materials is described by the linear differential equation named Foight model [6]:

$$\{\sigma\}=[D]\{\varepsilon\}+[W]\{\dot{\varepsilon}\},$$

where σ is stress; ε is strain, [*D*], [*W*] are rigidity and viscosity reduced matrices.

Fig. 5 states calculated maximal bay bending strains ε_b and shear strains ε_s of the hinged beam with fiber arranging angles $\left[0_8^0/\pm45_8^0/90_8^0\right]$ and geometrical dimensions (250 × 10 × 10) mm. The beam was loaded by the transversal central sine pulse with maximal impact level $P_{max} = 2\cdot10^3$ N and duration $t_{max} = 0.2$ ms caused by the striker having 51 g mass and 5 mps rate. The points correspond to maximal experimental values of bending strains. The difference between calculated and experimental values is not more than 10 %.

Non-stationary bending and shear strains in beams made of fiber-layered materials such as glass fibers and coal plastics having low shear rigidity are featured to be of the same order. As it follows from the calculations (fig. 5), the maximal non-stationary bending strain often taking place in the beam center is $\varepsilon_b = 0.565$ % and maximal shear strain is $\varepsilon_s = 0.432$ %.

Therefore, if the shock loading rate is less than 5 mps and the height to length ratio h/L is from 0.04 to 0.10, shear and bending strains are of the same order ($\varepsilon_s/\varepsilon_b \approx 0.8$–0.9). If the beam is short and the ratio h/L is more than 0.1 shear strains can cause damage as they are more than bending ones ($\varepsilon_s/\varepsilon_b \approx 1.06$). If the shock load rate is more than 5 mps, the impact force function $P(t)$ in (1) should be approximated by a triangle pulse (fig. 4). As it follows from the calculations, if the contact impact force pulse duration decreases, shear strains can damage longer beams with the ratio $h/L < 0.04$.

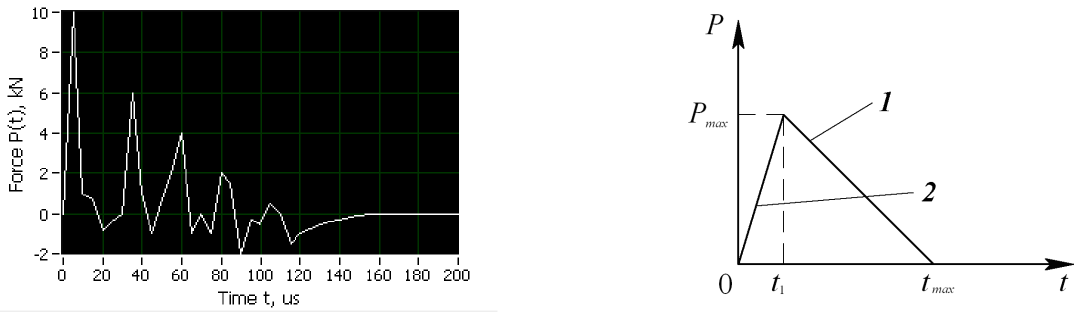


Fig. 4. The contact force time function $P(t)$ (*a*) and pulse envelope (*b*) at transversal central impact on the firmly fixed glass fiber plate: *1* – $P = P_{max}(1 - t/t_{max})$; *2* – $P = P_{max}\, t/t_1$

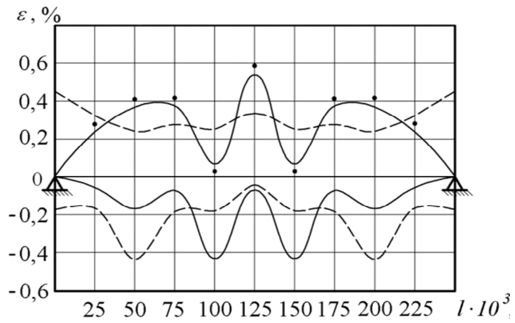


Fig. 5. Maximal bay bending ε_b (solid) and shear ε_s (dashed) strains of the hinged beam loaded by the transversal central sine pulse: lines stand for calculation data; points stand for experimental data

The application of the computerized measuring system in the automated test stand for performing impact tests allows decreasing experimental data acquisition error two and more times. It is possible essentially by means of graphical application development environment LabVIEW supporting virtual instrumentation.

Transversal impact force pulse shapes are established to depend on the impact rate to be approximated by certain functions:

– if the impact rate is less than 5 mps, the pulse shape is close to sine with maximal amplitude $P_{\max}$ and duration $t_{\max}$;

– if the impact rate is from 5 to 40 mps, the pulse shape can be approximated by a triangle pulse. The rising part corresponds to the increasing load and the falling part corresponds to the decreasing load;

– if the impact rate is more than 40 mps, the pulse shape should be approximated by a sawtooth pulse corresponding only to the decreasing load.

References

1. Снисаренко С. И. Универсальный испытательный стенд и методы исследования композитных балок, пластин и оболочек при ударном нагружении // Науч. вест. НГТУ. 2009. № 3 (36). С. 121–129.

2. Евдокимов Ю. В., Линдваль В. Р., Щербаков Г. И. LabVIEW 8 для радиоинженера. М. : ДМК-Пресс, 2007.

3. Голоскоков Е. Г., Филиппов А. П. Нестационарные колебания. Киев : Наукова думка, 1977.

4. Динамика удара / пер. с англ. ; под ред. С. С. Григоряна. М. : Мир, 1985.

5. Рикардс Р. Б. Метод конечных элементов в теории оболочек и пластин. Рига : Зинатне, 1988.

6. Расторгуев Г. И., Снисаренко С. И. Физические соотношения для задач ударного нагружения и нестационарного деформирования композитных конструкций // Прикладная механика и техническая физика. 2009. Т. 50, № 1. С. 187–196.

Part 4

ECONOMICS

S. A. Belyakov
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

TOOLS OF BIOTECHNOLOGICAL SECTOR FORMATION IN THE REGION

At the present time in the world economy a constant increase of importance and influence of hi-tech branches takes place. As a result, appears the natural need in research of various scientific-technological development mechanisms, which correspond to modern views on scientific-technical progress.

Keywords: biotechnology, tools, state support, cluster.

The increase of importance and influence of hi-tech industries in national economies is connected with their considerable influence on all spheres of human activities, and one of such industries is biotechnology that penetrates into all industry branches of national economy: power, medicine, agriculture, building etc. Biotechnology is one of the directions of the fifth and sixth technological levels and the degree of its development defines a country's technological level.

Biotechnology has good prospects of development in Russian practice and will allow to solve many social and economic problems in the development of Russian economy: creation of new biologically active substances, medical products; creation of new technologies in deep processing of agricultural, industrial and household waste; using the country's energy potential more effectively, etc. The leading countries in the field of introduction of biotechnologies are the USA, EU, China, Brazil, and Japan.

In Russian practice the development of biotechnologies still remains at a low level. Their share in the export of high technology products is several tenths of a percent. At the same time it is noted in the program documents that with the development of biotechnological manufacture it is possible to achieve a serious technological breakthrough in many manufacturing industries and in other branches of national economy. The development of biotechnologies requires creation of regional programs considering specificity of natural resources potential of a region and providing the system with supportive measures to create high technology biotechnological manufactures.

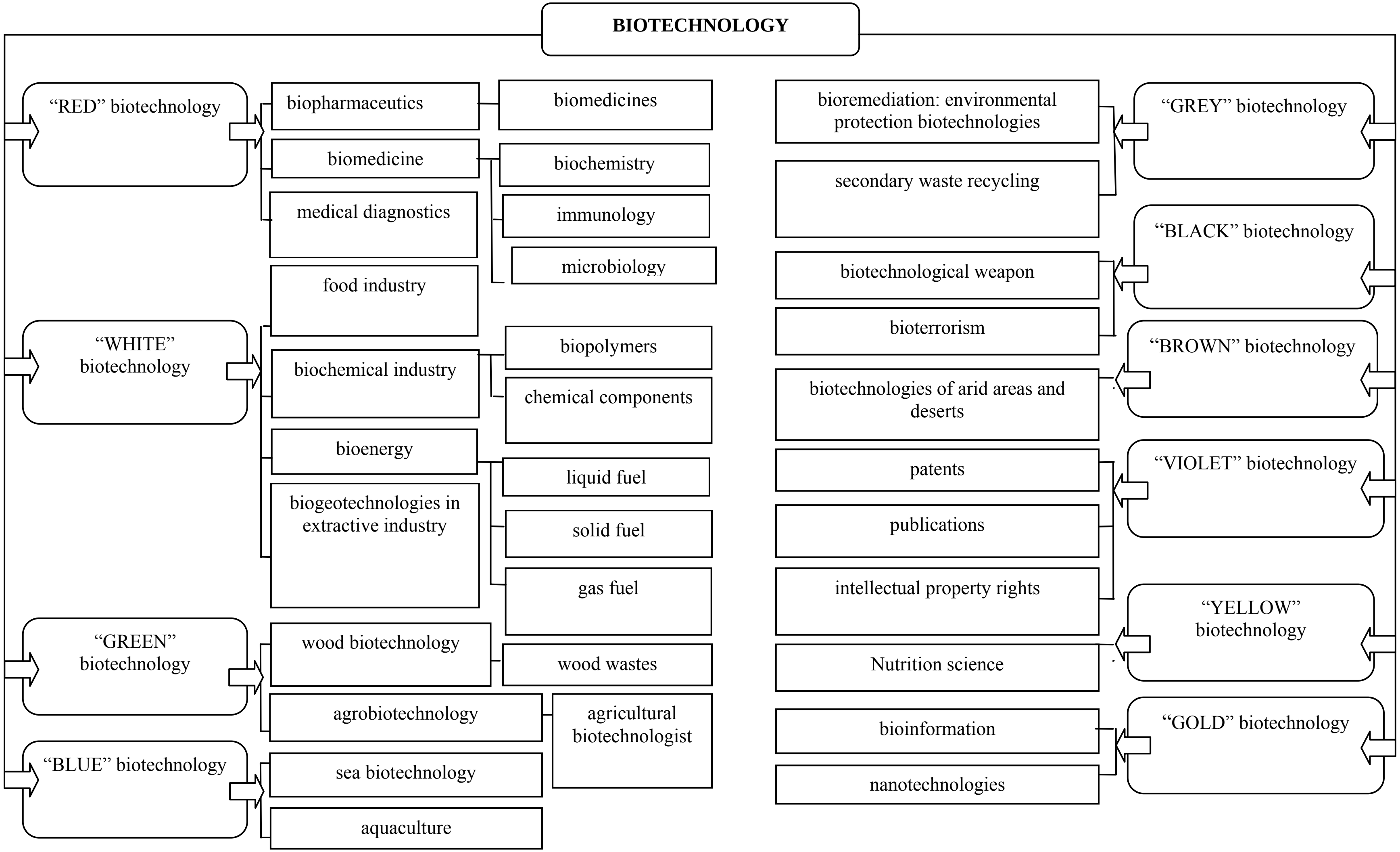
A wide range of biotechnologies application requires their classification based on a defined classification criteria. To such classification criteria belongs a criterion that reflects the actual and potential raw-material base of biotechnologies development, for example, wood biotechnology, agricultural biotechnology, sea, secondary waste processing, etc. The second criterion includes the spheres of biotechnologies application: biomedicine, food-processing industry, biogeotechnologies, etc. The identification of such criteria allows to take into account not only the specificity of technological decisions, but also practical biotechnologies application depending on the quality and availability of raw-material base with respect to regional bio-resources, expediency to use them.

Such organizations as the Society of Biotechnologists of Russia named after Yu. A. Ovchinnikov, the Union of enterprises in biotechnological branch, Institutes of the Russian Academy of Sciences are engaged in scientific and technological studies in the area of biotechnologies. The society of biotechnologists of Russia named after Yu. A. Ovchinnikov has developed a Strategy of biotechnological industry development up until year of 2020 (project). They have identified the directions of biotechnologies development and carried out a foresight analysis. Biotechnology develops rapidly in all spheres of human activities and in the world, as they think. The "color" classification of biotechnology, which is presented on figure [1; 2], has been applied:

The classification derived by the specified sources allows to manage the formation of biotechnological development regional programs more purposefully, to define a choice of areas that correspond to population and biotechnological product market demands.

The presented classification of biotechnologies covers almost all spheres of human activities – from the development of medicine based on essentially new technologies to waste utilization, environment preservation and protection against possible negative use of the research results (biotechnological weapon, bioterrorism) [3]. It is necessary to develop biotechnology on a state level in all areas (from 35 technologies that are recognized as critical, 16 belong to the technologies where biotechnology is a key element), on a regional level those technologies are being chosen that have a sufficient bioresource, effective application and promote social and economic development of a region.

The basis of a program design for biotechnologies development as a target problem of the conceptual approach is a program-target method when a program is presented in the form of target subprograms which reflect application areas of separate groups of resources and includes subprograms: ecological, municipal, power, development of machine-building complex, social, power safety etc. The time framework is defined by the social and economic development Strategy of a region and biotechnological branch development Strategy. Program arrangements depend on the presence of investment projects, their economic efficiency.



Biotechnology classification

A modern form of organizing biotechnologies development in a municipal institution can be a cluster with specific features connected with a raw-material base of manufacture, the potential of manufacture capacities development and the importance of manufactured products, as for the municipal institution itself, as well as for the whole region. The purpose: to create highly profitable small and average business enterprises based on complex use of wood industry waste and industrial development of peat based on modern science intensive technologies with diversified manufactures.

The formation of a biotechnological cluster should be carried out considering several measures, such, as:

– considering innovation level of the offered projects, it is necessary to create a corresponding social infrastructure, including housing, objects of leisure and preschool and school establishments;

– it is necessary to co-ordinate the efforts and resources of investors, state and regional authorities regarding the development of engineer and transport infrastructure;

– to promote the development of domestic manufacturers of modern equipment through the system of state orders and creation of advantages for domestic suppliers.

The basis of cluster structure formation should include the principles defining the system of cluster participants' interaction and forming a potential of its sustainable development in a long-term period.

Based on the study of cluster approach theories of modern management systems, the following principles of cluster formation of the potential product manufacturers are identified on the basis of the complex use of the available raw materials (tab. 1).

The successful formation of the mechanism of scientific-technological development is the result of economic and social conditions that are formed in a society. To create a new mechanism of scientifically intensive manufactures and to promote the formation of new innovative directions in Russian economy, it is necessary to use the existing experience in developed and even developing countries. The special attention represents the experiences of European countries where measures of the state support are actively used, including:

– creating an infrastructure favorable for innovative activity;

– direct state financing of scientific researches through allocation of target grants, providing credits, subsidies;

– providing tax stimulus and other forms.

The state support is provided to both manufacturers of high technologies and manufactures and consumers of high technology products under the condition of importance of creating the new areas of innovative economy (for example, renewed energy resources). That way Swedish house owners receive state grants for refusal from mineral oil heating, in Netherlands and Germany the installation of wind turbines, etc., is subsidized. A special attention is given to biofuel manufacture, as to the major direction to decrease energy dependency of the European countries: the usage of biofuel gives a right to state grants (Belgium, France, Sweden, Italy); VAT discounts (Austria, Netherlands) and other forms promoting its application. Tab. 2 systemizes the tools of high technology manufactures support in foreign practice and identifies the supportive tools of biotechnologies as a new area of high technologies creation.

The basic tools are: granting tax privileges to private business; target management of corporations innovative policy through a stimulating amortization policy and a system of tax privileges; creating a legislative base to stimulate and support the areas of development that are necessary for the state; including innovative support into the system of infrastructural institutes privileges.

The target state support of biotechnologies development is presented in destimulating tools when using not renewed energy resources; setting quotas and "green" certificates; various grants from special funds, created from tax payments for electric power from traditional sources processing. A special attention is given to the development of agricultural production based on "clean" technologies and manufacture of bioproducts. The presented in tab. 2 tools allow to draw a conclusion that in the Russian practice it is necessary to develop target programs on the development of biotechnologies with creation of the corresponding legislative base.

The biotechnology development should become a priority direction of a state policy and include a system of target tools to support scientifically intensive branches of economy including biotechnologies.

The study of domestic and foreign experiences of high technology branches and technologies support, as well as the experience of biotechnologies introduction and the available offers in this sphere has allowed to identify the tools of biotechnological sector industries support (tab. 3).

Such instruments include: tax exemption for a certain period of time for bioenergy manufacturers, for some bioproducts; grants from a special fund; using a system of subsidized loans; grants for carrying out scientific works in this sphere; target support of ecologically clean products manufacturers, etc.

One of the major directions is to include regional programs in federal level programs and to define regional priorities in development stimulation. As the foreign practice shows, for the time biotechnologies require various measures of support due to the formation of a new direction of scientific and technical progress and their high social importance.

The formation of a new direction of innovative economy requires also the support of manufacturers of the equipment for biotechnologies and first of all machine-building branch. At the present stage, the energy saving products as the target problem, in its decision requires the corresponding support of projects and the programs connected with decrease of power consumption by several branches: agriculture, wood industry, housing and communal services etc. A special attention should be given to the usage of industrial and household wastes where support should also be provided and stimulation measures developed.

Table 1

Basic principles of regional biotechnological cluster formation

№ п/п	Principles	Content
1	Orientation on local raw materials	The offered biotechnological cluster is formed on the basis of the present peat and wood industry wastes availability in a municipal body. Peat reserves are sufficient for their long-term use
2	Complex approach to raw materials development and creation of processing manufactures chain	Such approach reflects the specificity of cluster creation with an industrial orientation and allows to considerably decrease production costs
3	Cluster approach to extraction and processing manufactures formation	Thus the system of the enterprises of different forms of ownership and size is specially projected: small, average. Preplanning of the cluster approach assumes the targeted management of cluster formation, reflecting its specificity – biotechnological cluster
4	Creation of a system that guarantees the realization of cluster approach advantages	To the elements of such a system we can relate synergy as an additional effect due to interrelations and interactions, general infrastructure: industrial, social, etc., creating a chain of the added value defining the economic interest of the businesses in a cluster
5	Targeting at innovative, hi-tech manufactures	This principle should underlie all new enterprises created in a region. It is necessary to draw a list of basic indicators, the fulfillment of which is obligatory for the projected enterprises. The enterprises-clusters should possess key competences, based on raw materials features and the innovative level of technologies
6	Step-by-step introduction of new enterprises and capacities	It is connected with the specificity of the used raw materials. Nowadays we have technologies of deep peat and wood wastes processing, however the creation of new hi-tech manufactures requires additional preparation, justification and search of investors and equipment
7	Orientation on interrelations with other branches of national economy	This principle reflects the features of biotechnologies realization which extends to many spheres of economy. The interrelation and projects justification should include agriculture enterprises, building, housing and communal services, mechanical engineering, etc.
8	Including of different level programming materials into the system	This principle allows to create a system of program activities realization due to the inclusion of a biotechnological cluster into the strategies of social and economic municipal bodies development, a region as a whole and programs of biotechnologies development at different levels of management
9	Creating a single control body for a biotechnological cluster	As experience shows, it is most effective to create an operating company at the municipal level. The available advantages of cluster approach can only be realized having a targeted management that is conducted using a single management body.
10	Advancing personnel development at biotechnological clusters enterprises	A cluster has a human resources specificity, the preparation of which should be carried out in the process of involvement of new manufactures into the system of cluster structures

Table 2

Instruments of supporting scientifically intensive manufactures

<table>
<tr><th rowspan="2">Country</th><th colspan="2">Instruments</th></tr>
<tr><th>Scientifically intensive manufactures</th><th>Biotechnologies</th></tr>
<tr><td>Australia, Germany, Greece, Denmark, Spain, Luxembourg, Portugal, Finland, France, Sweden</td><td rowspan="5">1. Granting tax privileges to private business.
2. Changing amortization policy to allow organizations to increase amortization funds as a source of investments and innovations.
3. Tax privileges to encourage those directions of corporate activity that are desirable from the state point of view.
4. Tax deductions for equipment acquisition and installation.
5. The majority of EU member states legislatively allow to reduce taxable profit for the sum of expenses on research and experimental work made during the accounting period.
6. Measures of the state support:
– creating an infrastructure favorable for innovative activity;
– direct state financing of scientific research by allocation of grants, credits, subsidies;
– granting state guarantees;
– support and development of the system of risk (venture) financing;
– Granting tax stimulus</td><td>Extra tariff charges for the energy received from traditional sources
No taxes for “clean” energy manufacturers;
Quotas and “green” certificates
Grants for R&D in the field of nonconventional power
State grants for those who stop using heating on mineral oil
Fiscal measures guaranteeing tax deductions (10 %) when investing in manufacture</td></tr>
<tr><td>The Netherlands, Slovakia, Czech Republic</td><td>Grants from a special fund created from tax payments for electric power from traditional resources processing.
Clearing of manufacturers of “clean” energy from power taxes;
Quotas and “green” certificates</td></tr>
<tr><td>Austria, Bulgaria, the Great Britain, Hungary</td><td>Grants from a special fund created from tax payments for electric power from traditional resources processing
Grants for R&D in the field of nonconventional power</td></tr>
<tr><td>Brazil</td><td>State target support of an industry:
– granting of subsidies;
– tax privileges;
– credits</td></tr>
<tr><td>Italy</td><td>Tax exemption (volume of fuel production up to 300 thousand tones per year)</td></tr>
</table>

Table 3

Instruments to support biotechnological sector of industry

№	Area	Instruments
1	Legislative support	Introducing the program of biotechnological industry sector support in a region. Creating target programs taking into account the priorities in stimulation development
2	Development of biotechnological industry	Compensating a part of the expenses directed at the application of innovative technologies connected with waste recycling of the agrarian complex and intended for independent agricultural production power supply. Compensating a part of the expenses directed at preservation (improvement) of fertility soils. Compensating a part of the expenses directed at the extraction of organic-mineral fertilizers from local raw materials
3	Tax stimulation	VAT decrease or exemption for socially significant types of products. Granting property, profit taxes privileges. Introduction of tax vacations for the period of new technology installation
4	Ensuring a stable growth of domestic products competitiveness	Supporting domestic manufacturers of biotechnologies equipment. Development and support of the agricultural machinery-building complex program regarding the organization of peat extraction machinery manufacture
5	Customs legislation improvement	Decreasing export customs duties: for organic-mineral fertilizers from local raw materials; for liquid and gas biofuel from local raw materials sources; for peat products; for various types of solid biofuel (pellets, briquettes etc.)
6	Legislative support improvement	The priority to approve the Federal Law "About the renewable sources of energy" The priority to approve the Law "About peat". The development of legislative support in intellectual property evaluation and protection
7	Development of the energy saving technologies	Supporting projects and programs connected with the decrease of power consumption by agricultural productions; manufacture in the wood complex. Development and improvement of the program concerning the advancement in wood mechanical engineering, including the organization of manufacture of the power equipment intended for wastes recycling
8	Support of integration communications	Creating an incorporated leasing company to supply peat enterprises with peat extraction and processing equipment. Formation of "technological corridors". Support of enterprise networks formation in clusters. Support of the steady cooperation communications
9	Human resource management	Training and consultation of managers and specialists
10	Solving environmental problems	Compensating a part of the expenses aimed at innovative technologies application, connected with wastes recycling of the wood complex and intended for independent power supply of the housing-and-municipal complex

The offered instruments of biotechnological sector formation support cover the whole cycle of biotechnological products creation and stimulate the usage of bioproducts in separate spheres of national economy, for example, in agriculture for soil restoration, manufacture of ecologically clean products, export development. Stimulation of integrated communications formation, and creation of biotechnological clusters in particular as the most effective organizational form allowing to provide complex usage of bioresources and the production of competitive products.

Thereby, biotechnology development should be recognized as the state priority policy: adequate forms of organizational, financial and information support, both on federal and regional levels, legislative support, business stimulation and private-state partnership. That is, the question concerned is a national project where all key structures of the state and society should be integrated [4].

References

1. Biotechnologies [Electronic resource]. URL: http://www.fos.ru/biology/7178.html.

2. Work materials for the Strategy of biotechnological industry development up until 2020. M., 2009.

3. Kalinina N. Bio weapon and bio security: "under the microscope analysis" // Worl economy and international relations. 2009. № 6. C. 117.

4. State support of innovative activities [Electronic resource]. URL: http://www.erudition.ru/referat/printref/id.29569_1.html.

G. Ya. Belyakova, A. I. Frolova
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

FORECASTING OF TNE LEADING INDEXES OF SOCIAL-EKONOMIC FACTORS ON YENISEISK DISTRICT

In this article presents the forecast of social and economic factors influencing Yeniseisk district over 2009–2012. The main supposition of the forecast is following the existing linear trend of the main social-economic factors' development with permanence of external factors. This results in the positive track record of the growing investment in the fixed capital that is stipulated by involving large-scale investment projects.

Keywords: forecast, social and economic factors, time row, confidence interval, exponential curve, investments.

Nowadays for being included in the world economic system every region must develop and form strategies of social and economic development, which would allow to find forms and ways of realization of the given regional advantages with the purpose to increase the standard of population living and also entering the markets of commodities, services, labour and capital of other regions. For this purpose it is necessary to expose steady tendencies and conformities to the law of social and economic processes, and also influencing factors and reasons predict their impact on the state and development of regional economy and vice versa.

Prognostication of social and economic development is the starting point of work to handle regional economy. The purpose of social and economic prognostication of region development is the adjustment of policy events and priorities in the development of regionall and economic complex.

Yeniseisk district is a municipal district of Krasnoyarsk region which is located in the middle flow of the Yenisei river. Since 1998 the quantity of population here has gone down by 7 017 persons and makes 26 324. The average age of the population is 42.4 years. On the one hand, it is the most active age to work. At this age employees achieve the peak of their professional level. On the other hand, in 5–10 years their labour productivity decreases but in future young employees are expected to relieve and that does not imply a mere substitution.

In this research the social and economic indexes of Yeniseisk district were taken over 2000–2008 as the basis for the local prognoses (tab. 1).

During 2000–2008 the number of registered unemployed persons fluctuates and makes about 1295 persons in 2001, then peaks in 2004 at the point of 2286. Also the analysis to provide the number of organizations in Yeniseisk district was proceeded. Over the period under research there is reduction of enterprises by 18 % in average that negatively effects the economy of district with growing investment in the fixed assets with every period by approximately 8.5 % [1]. Next on the basis of the obtained data the prognosis of the explored indexes was made for 2009–2012 by means of the analysis of basic tendencies in dynamic rows.

The basic pre-condition of the prognosis is saving in the future existent linear progress of basic socio-economic indexes trend at invariability of external factors.

Prognostication for 2009–2011 it is determined the quantity of the registered unemployed persons on equalization:

$$y = -40.16t^2 + 474.7t + 688.7$$

where y is the quantity of the registered unemployed persons; t is the period; R_2 is the index of determinates, in this prognosis it makes 0.901, that testifies to high exactness of trend model.

Table 1

Dynamics of basic indexes of Yeniseisk district over 2000–2008

Index	Period, year								
	2000	2001	2002	2003	2004	2005	2006	2007	2008
Quantity of the registered unemployed persons	1 313	1 295	1 507	1 993	2 286	2 208	2 030	1 667	1 815
Number of organizations	402	370	382	460	347	366	372	379	281
Investments in the fixed assets (million RUR)	26.4	6.7	5.4	5	8.2	13.5	4.4	7.1	9.4

Prognosis of the number of the registered unemployed persons 2011 in the Yeniseisk district is presented on fig. 1.

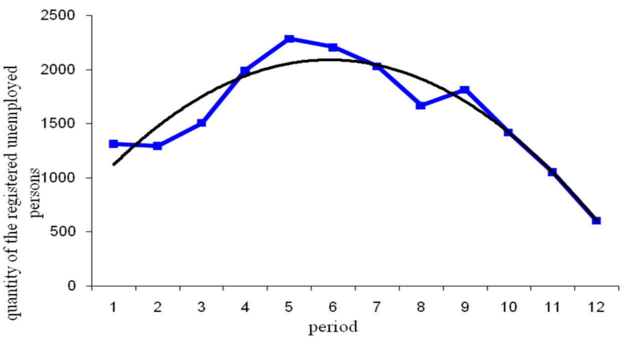


Fig. 1. The prognosis of the number of the registered unemployed persons in the Yeniseisk district by 2011

The actual number of unemployed persons in 2008 was 1667 people and the forecast number of the unemployed persons by 2009 was 1815 persons.

The prognosis for 2010 and 2011 is conducted similarly.

We determine the intervals of confidences for this prognosis for the exponential curve of growth. [2]

The confident intervals of prognosis are calculated with the formula:

$$U_y = \hat{y}_{n+L} \pm t_\alpha S_{\hat{y}} \times$$

$$\times \sqrt{1+\frac{1}{n}+\frac{t_L^2}{\sum t^2}+\frac{\sum t^4 - 2t_L^2\sum t^2 + nt_l^4}{n\sum t^4 - \left(\sum t^2\right)^2}}, \qquad (1)$$

where $\hat{y}_{n+L}$ is the point prognosis on the model for the $(n + 1)$th moment of time; t_α is the tabular value of St'yudent criterion for the level of meaningfulness and number of degrees of freedom; n is the amount of supervisions in a temporal row; $S_{\hat{y}}$ – average quadratic error of estimation of the forecast index; t is the sequence level of row number; t_L is time which a prognosis is done for; L is a period of forestalling.

Results are presented in tab. 2.

Analysing findings from tab. 2 (confidence interval) showed that the model on the basis of which the prognosis was carried out is acknowledged to be adequate with confiding probability 95 %. It is possible to assert that at saving of the folded conformities to the law of development the forecast size gets within the interval formed by lower and upper limits.

Table 2

Confident intervals of prognosis for the quantity of the registered unemployed persons in the Yeniseisk district

Time, *t*	Step, *L*	Point prognosis	Confidence interval of prognosis 95 %	
			upper limit	lower limit
9	1	1 420	2 082	758
10	2	1 051	1 713	389
11	3	601	1 264	–60

Similarly the prognosis is made for the amount of commercial organizations in the Yeniseisk district over the researched period. Thus the given forecast index was added to the factual data that allowed to multiply exactness of the given prognosis.

Prognostication of the number of organizations in the Yeniseisk district by 2009 was conducted on the basis of equalization:

$$y = -2.864\, t^2 - 19.12t + 368.2$$

where y is the number of organizations; t is the period; R_2 is the index of determination, in this prognosis it is equal to 0.855 that testifies the high exactness of trend model.

Information of prognosis for the amount of organizations by 2011 in the Yeniseisk district is presented in fig. 2.

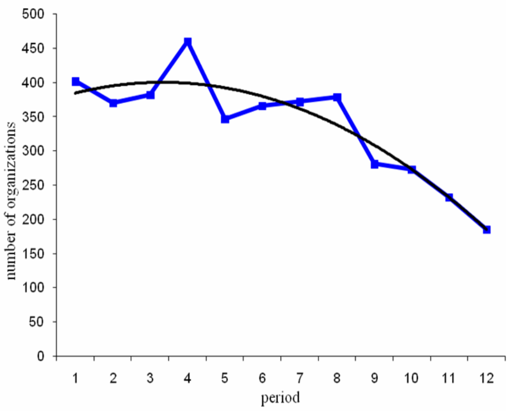


Fig. 2. The prognosis of amount of organizations in the Yeniseisk district by 2011

Actually the number of commercial organizations in the Yeniseisk district by the end of 2008 was 281. The forecast index for 2009 was 273 (the rate of growth went down by 3 %), 2010 it was 232 (rate of growth 85 %) and in 2011 it was 185 (rate of growth 80 %).

Same way as the prognosis of quantity of the registered unemployment was determined, we determine the confident intervals of the prognosis for the amount of organizations in the Yeniseisk district for the exponential curve of growth in formula (1). The results are presented in tab. 3.

Table 3

Confident intervals of prognosis for the amount of organizations in the Yeniseisk district

Time, *t*	Step, *L*	Point prognosis	Confidence interval of prognosis 95 %	
			upper limit	lower limit
9	1	273	399	147
10	2	232	358	106
11	3	185	311	60

Findings (tab. 3) show that the model on the basis of which the prognosis was carried out is acknowledged to be adequate with confiding probability 95 %. Consequently with saving the established tendencies of development the forecast falls within the interval formed by upper and lower limits.

The investment prognosis is similarly made in the fixed assets over the researched period.

The prognostication of the investment in the fixed assets by 2009 was conducted on the basis of the following equalization:

$$y = 0.584\,t^2 - 6.855t + 25.32,$$

where y is the sum of investments in the fixed assets; t is the period; R_2 is the index of determinates. In this prognosis it is equal to 0.715, that testifies the acceptable exactness of the trend model.

Information of prognosis of investments in the fixed assets by 2011 in the Yeniseisk district is presented in fig. 3.

Actually the sum of investments in the fixed assets in the Yeniseisk district in 2008 was 9.4 millions RUR. The forecast index by 2009 was 15 million RUR, by 2010 it was 20.67 million RUR and in 2011 it was 27.27 million RUR.

We determine the confident intervals of prognosis of investments in basic capital in the Yeniseisk district by 2011 for the exponential curve of growth in formula (1). Results are presented in tab, 4.

Findings (tab. 4) show that model acknowledged adequate with confiding probability 95 %.

There is positive dynamics of growth in the given prognosis, such index as investment in the fixed assets is conditioned by realization of large projects of investments. Social efficiency of the project consists of multiplying employment of population in all its implementation phases including multiplying workplaces at enterprises in the district. There is a decline of unemployment in the forecast periods, that will testify the results of realized complex of measures foreseen by the department having the special purpose program "Assistance employment of the Krasnoyarsk region's population".

Table 4

Intervals of confidences of prognosis of investments in basic capital in the Yeniseisk district

Time, *t*	Step, *L*	Point prognosis	Confidence interval of prognosis 95 %	
			upper limit	lower limit
9	1	15.23	33.92	–3.4
10	2	20.66	39.35	1.98
11	3	27.27	45.95	8.58

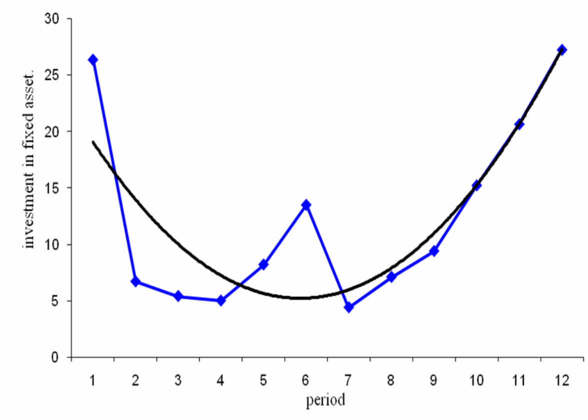


Fig. 3. The prognosis of investments in the fixed assets by 2011 in the Yeniseisk district

The prognosis findings of social and economic indicators can be used in practical activity of municipal education management.

References

1. Social and conomic position of municipal districts of the Krasnoyarskogo region / Krasnoyarsk Statistics. 2009.

2. Vanchikova Y. N., Naletov A. Y. Short-term prognostication of tax profits of the regional budget with the preliminary processing of basic data by the method of main components // Questions of statistics. 2009.

Y. A. Chairan, A. A. Lukyanova
Siberian State Aerospace University named after M. F. Reshetnev,
Russia, Krasnoyarsk

THE FORMATION OF THE ENTERPRISES INVESTMENT APPEAL

The article concerns the problem of investment appeal compounds, approaches to its value. There are groups of actions which can help to raise the investment appeal of the enterprises.

Keywords: investors, investment appeal, investment highlights, approaches to its value, the group of actions which can help to raise the investment appeal of the enterprise.

Nowadays the most important problem is consolidation and effective use of investments for effective functioning of the enterprise in modern conditions.

The encouragement of investment activities, the elaboration of clear investment strategy, the determination its prior ways of development, consolidation of all investment sources is the most important condition of permanent and stable enterprise development.

The most important factor stimulating the investment activities is the state warranty of investors' rights, the defense of the investments and customs and tax incentives for the subjects of investment activities.

There are some main legal acts which regulate relationships in the filed of investments in Russia:

– Federal Law of the 25th of February 1999 year № 39-FL "On investments acts in the Russian Federation carrying out capital investments" (redaction of 24.07.2007);

– The Law of RSFSR of the 26th of June 1991 year № 1488-1 " On investment acts in the RSFSR" (redaction of 30.12.2008);

– Federal Law of the 9th June of 1999 year № 160-FL "On foreign investment in Russian Federation" (redaction of 29.04.2008);

– Federal Law of the 5th of March 1999 year № 46-ФЗ "On protection of rights and legal investors' interests on the securities market" (redaction of 19.07.2009);

– Federal Law of the 29th of October 1998 year № 164-FL "On financial lease" (redaction of 26.07.2006);

– Federal Law of the 22 of April 1996 year № 39-FL "On the securities market " (redaction of 27.12.2009);

– Russian Federation President Decrees, which also define the legal framework of investment financing;

– The Russian Federation government decrees control investments processes.

Unfortunately, the law system of the Russian Federation is not perfect. Different legal norms which regulate the same actions contradict to each other; there is no strict hierarchy of regulatory legal acts. As the result the planning of investment activities has difficulty as the consequences of the particular solution and its expenditures are unpredictable.

Because of the global financial crisis a lot of Russian enterprises have faced several problems which reduce investments into the enterprises for their further development. We can name some deterrent factors which restrain Russian activity:

– high level of the inflation;

– high level of taxes;

– incomplete financial support of the state investment programs;

– low effectiveness of investments;

– lack of own money to renovate the main capital and difficulties to get the commercial credits because of unstable financial situation and high interest rates;

– high investment risk;

– paucity of the first-rate borrowers;

– low quality of offered investment projects;

– unstability of Russian market;

– complicated current situation at enterprises, most of them are in bad financial conditions, inability of top-managers to demonstrate their potential to the investor and show the possibilities to overcome crisis and achieve the revenue growth, or other reasons.

The lack of own funds to finance the investment activity of the enterprise often makes Russian companies to attract foreign investments in current situation.

Nowadays there are several ways of making investments in the enterprise from abroad: investments into the stock capital and granting loans.

1. The investment into the stock capital of the company (direct investment).

There are several main ways to attract investments into the stock capital:

a) Strategic investment – this is the investor's acquisition of the major company stocks. As a rule, the final stage of strategic investment is the takeover of the company or the merger with the investor's company.

Big corporations and enterprise-leaders are the main strategic investors of this branch. The main aim of the strategic investor is to get an access to the new resources and technologies and to raise their own company effectiveness.

b) The investment of financial investors – this is the acquisition by the professional investor or a group of investors the blocking not major share of stocks for further selling this share of stocks in 3–5 years (mainly venture fund and share fund) or placing shares of the company on the securities market for broad spectrum on investors (in this case there may be either an enterprise of any sphere or an individual investor). In this case an investor gets his main income from selling the major share of stock (in fact for withdrawing from the business) [1].

From this point of view, the attraction of financial investments is useful for development of the company; to increase the sales volume, to modernize and expand the production, to raise the effectiveness of work which will result in increasing the company's value and, accordingly, an investor capital.

2. Investment in the form of granting loans.

The main ways – loans (bank or trade), bonded debts and leasing. The main aim an investor is to get the interest from the investment capital with current rate of risk. Therefore this group of investors is interested in further development of the enterprise from the point of view that the company can fulfill the obligations of interest payment and returning the principal debt.

The necessity of attraction investments for further development and forming business competitiveness binds the receiving enterprise to be attractive for investors.

Because of this reason, there exists a concept of the investment attractiveness; this is the main indicator of the enterprise opportunity to attract the potential investor.

The clearest definition of the investment attractiveness is given by L. Valynurova and O. Kazakova. They understand this definition as all subjects, properties, means and possibilities to have the potential demand for investment [2]. This definition is fuller and takes into consideration the views of any member in the investment process.

There are other points of view as L. Gilyrovskaya's, V. Vlasova's and E. Krylatov's and others. They give the definition of investment attractiveness as effective use of own or borrowed capital, solvency and liquidity analysis. The same definition is the structure of own or borrowed capital and its distribution between different kinds of property and also their effective usage [3; 4].

According to Mogzoev's approach the investment attractiveness is determined by the inner and outer state of investment unit (organization). The inner environment is determined by structure, staff, techniques and aims of the enterprise. The inner environment of the investment unit has its own characteristics which determine its effectiveness and investment attractiveness. However it is involved into the outer environment, the main elements determine the current status of investment climate and influence on investor's decision. The distinctive future of these components is its objectiveness so the environment doesn't depend on investor's estimation and it is determined by social and ecological or other characteristics [5].

The most common and abstract definition of the investment attractiveness comes from its meaning – advisability investment into the business or commercial enterprise. It is interesting that investors and organizations put into the meaning criteria which they consider the most interesting and demonstrative.

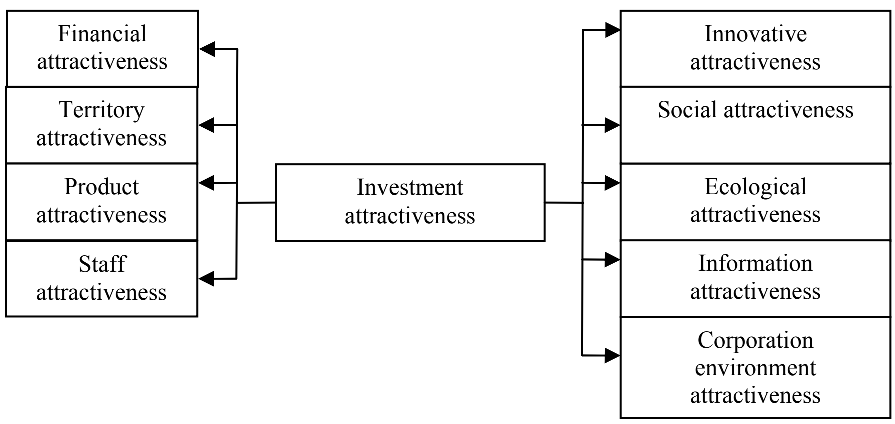
Due to this fact that the enterprises can attract investors of different types, the investment attractiveness is the complex of intercommunication which makes the company attractive for everyone from all sides.

To understand how to create investment attractiveness, let us consider what makes a company attractive to investors and how the investment attractiveness is assessed.

It is important to know that there is no unique approach to estimate the investment attractiveness of the enterprise. Every investor uses his own methods and approaches. It is known that the estimation of the investment attractiveness is the difficult process where the mathematical calculation is only one of the methods.

The features of investment analysis should be completed by classification of some investment attractiveness kinds which are systemized in the figure.

Firstly, the investor is usually interested what is produced at the enterprise and where it is situated also its management and staff. Therefore the initial component of the investment attractiveness is staff, territory and product attractiveness. The main component of the investment attractiveness is financial attractiveness because it reflects the results of its activity (capital productivity, profitability) and financial stability (liquidity index of ample resources). The innovative, social, ecological and informative attractiveness of the enterprise and also attractiveness of corporate culture are viewed as development perspective for investors.



Classification of the investment attractiveness

There are several methods of the investment attractiveness assessment:

1. The market approach is based on the analysis of outer enterprise information. This allows to asses the changes of market share price and also the amount of paid dividends by following methods:

– market value added on share capital (MVA);

– total share revenue on investment into the enterprise (TSR);

– middle value of the capital (WACC);

– capitalization attitude towards capital (MBA).

The method can not be used towards companies which shares are not on the securities market.

2. Accounting method is based on the analysis of inner information and uses the mechanisms of economical activity. The main indexes which are used for this assessment are taken from accounting data of the company:

– net profit;

– cash flow;

– value of assets;

– the proportion of the company's profit for the period to the size of the share capital;

– share added value;

– rate of profit to asset of the company;

– interest coverage on bonds and dividends on share of company profit;

– index of payment ability and liquidity which is accounted due to the payment of the credit;

– liquidity ratio (fast it is relationship of floating funds to short terms commitments and current – the final production, unfinished production and its supply are included in numerator);

– evaluation of business activity;

– evaluation of effective usage of property (index of cost effectiveness, main funds and own funds);

– evaluation of funds statement, estimations and other active assets (bill receivable) is done to reveal the cause of raising the bill receivable, mainly expired or fail, diversion of funds into banking turnover and etc., also to reveal the other resources to improve calculation;

– others.

The method has several disadvantages:

a) financial indexes can differ from current state of the company. For example in data reporting there can be loss or reduce of the revenue it can cause the negative investment attractiveness of the company due to the methods. If the company invests money into new perspective project this fact can attract the potential investors;

b) the method does not take into consideration that using different methods of amortization, the evaluation of property cost, allocations to research and development, exchange transactions, acquisition of assets, real loss can transform into real profit and visa versa;

c) data are shown in the accounting report reflect the financial state of the company only at the definite time but it does not include other periods of time.

3. Compound method can combine inner and outer characteristics of the company. Therefore this method is more correct from the investment attractiveness point of view [6; 7].

The main indexes of stock market and investment managers are:

– ratio of company capitalization to EBITDA (the revenue of the company before taxes, credit interests and amortization);

– ratio of market capitalization to revenue (PSR);

– index comparing the share price to the revenue of it (PER).

The big companies should use other methods for investment attractiveness assessment than smaller ones because they have more economical and political opportunities but at this time they depend a lot on government decisions.

Therefore the following factors become very important for big companies:

– image of the company (either for consumers or counterparts);

– owner's reputation of the corporation;

– interconnections with other companies (formal and informal);

– company activity enhancing its business status and image;

– relations with the government authorities;

– corporate management of the company;

– clarity of data reporting;

– reorganization in the group structure in which companies are involved.

Nowadays there are several rating agencies, the leading ones are Standard & Poor's, Moody's or Fitch which award the rate of investment attractiveness to the companies. The rating method is available for public and the agencies have a good reputation and their opinion is considered to be true and accurate. Service of such agencies is expensive enough (about $50,000) therefore only big companies can afford this service.

To raise the investment attractiveness the company can perform some actions. The main activities can be:

1. The elaboration of long term development strategy. It is important for potential investor as the strategy demonstrates viewing the own long term perspectives and adequate management of the company to the working conditions (inner and outer). It is obvious that the most important thing for investors interested in long term development of the company is clear strategy especially for those who are involved in business. When the company has a long term strategy of development it begins to develop the business planning.

2. The business planning. All aspects of activity are viewed in details in the business planning, the necessary volume of investments, financing scheme and the investment results for the company are proved. The cash flow plan calculated in business plan allows to assess the ability of the company to return borrowed money and pay interests to the investors. The business planning allows to

assess the value of the company, capital and potential development of it for investors.

3. Carrying out of legal examination and release of legal documents in compliance with the legislation. The legal examination allows revealing the drawbacks according to the legislation norms. The correction of these drawbacks is very important as the investor pays attention to the judicial audit. So, for the mortgagee an important stage in the process of negotiations with the company is a confirmation of property rights provided for the collateral property. For direct investors, who buy holding shares of the company, the important point is the shareholders rights and other aspects of corporate management, which have direct impact on their ability to control the direction of expenditures of invested funds.

4. Creation of credit history of the company as it allows judging the experience of getting outer investments and fulfill obligations to investors and owners. It is necessary to follow some steps to create such history. For example the enterprise can carry out the issue and extinguishment of loan on small amount and short period of time. After the extinguishment of the loan the company will enhance its status and get the new level as a creditor who can carry out the commitments. Further the company can attract new loans and investments on advantageous terms.

5. Some ways of reorganization. This point is the most difficult way of raising the investment attractiveness of the company. The full program of reorganization includes the integrated work of the company according to changing terms of the market and its strategy.

The reorganization can develop in several ways:

1. The changing of organized structure and management methods. This direction of reforms aims at improving management processes, which provide the basic functions of effectively functioning enterprise, and organizational structures of the company, which must match the new management processes.

The reorganization of management and structure can include:

– involving some organized chains into the management process;

– carrying out of relevant activities;

– allocation of some business areas in separate legal entities, education holdings, other changing forms of organizational structure;

– improving the information stream in information management;

– finding and elimination of extra chains in management;

2. The reform of the share capital. This direction includes activities aimed at the optimization of capital structure – split, consolidation of shares, all described in the Law on joint-stock companies, forms of reorganization of joint-stock company. The result of such actions is the management rising of the company or its groups.

3. The reform of assets. In the framework of assets restructuring, we can highlight the restructuring of the property, complex restructuring of long-term investments and restructuring of the current assets. This direction assumes any changing of current assets due to the selling extra marginal assets and obtaining necessary assets, optimization of financial investments (shot-term and long-term) assets, bill receivable.

4. The reformation of production. This direction is aimed at improvement of production systems of the enterprises. The purpose in this case is to improve the efficiency of goods production, services; increase of their competitiveness, expansion of the range or conversion.

The reorganization of production can include several measures:

– expansion of product range and sales of profitable products;

– mastering of new commercially promising products or services;

– unprofitable products withdrawal from production, unless there are real for the implementation investment projects to reduce costs, increase of competitiveness of production and etc.;

– other activities.

The complex reorganization of the company includes the combination of measures which are referring to several methods above.

It is necessary to analyze the present situation of the investment attractiveness of the company to determine the necessary activities (diagnostics of the company).

In the diagnostics different directions (aspects) of the company are considered: its sales, production, finances and management. A field of the company activity, which is associated with the greatest risk and has the largest number of weaknesses, is marked; measures to improve the situation on the chosen directions are formed.

The enterprise can form the program of measures for improving the investment attractiveness on the assumption of individual distinctions and capital market. The realization of such program helps to speed up of the investment financial resources and decrease its value. It is worth saying that probable measures do not take material expenses but the result of its realization is the rising of effectiveness besides of investors' interest towards the company.

References

1. Маленко Е., Хазанов В. Инвестиционная привлекательность и ее повышение [Электронный ресурс] // Автоматизация управления компаниями. 2007. URL: http://www.insapov.ru/investment-attraction.

2. Валинурова Л. С., Казакова О. Б. Управление инвестиционной деятельностью : учебник. М. : КноРус, 2005.

3. Гиляровская Л. Т., Лысенко Д. В. Комплексный экономический анализ хозяйственной деятельности : учебник. Ендовицк : Проспект : Велби, 2008.

4. Крылов Э. И., Власова В. М. Анализ финансовых результатов предприятия : учеб. пособие. СПб. : ГУАП, 2006.

5. Могзоев А. О некоторых терминах, используемых в инвестиционных процессах // Инвестиции в России. 2002. № 6.

6. Гайданский А. И. Факторы риска в инвестиционных проектах и оценка эффективности капитальных вложений // Финансовый менеджмент. 2004. № 7. С. 28–34.

7. Воронов К. Основы теории инвестиционого анализа [Электронный ресурс]. URL: http://www.eup.ru/Documents/2007-03-19.

Y. V. Danilchenko
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

A. S. Khozhaev
Private security enterprise “Rodimichi”, Russia, Krasnoyarsk

CUSTOMER RELATIONSHIPS MANAGEMENT AS THE KEY FACTOR FOR PROVISION OF THE BUSINESS SERVICES’ COMPANY COMPETITIVE ABILITY

Development trends of modern business services markets lead to the unification of the range, quality and cost of services. Under these conditions, the customer relationship management becomes the most important factor for a business services’ company competitive ability.

Keywords: competitive ability, management, customer relationships, business service industry.

One of the key trends of modern economic development is the growth of the service industry and its share in the global economy. In the last quarter of the 20th century and the beginning of the 21st century, the service industry’s share grew rapidly. Today the service industry’s share in the economy of some countries is up to 50 % while in most developed countries it is up to 70 % [1]. Since the 1990s these trends were clearly seen in the Russian economy.

The rapid growth of the service industry in the Russian economy is particularly evident in the business service industry where there is a great number of small and medium businesses. This growth is particularly visible due to the relatively small size of seed capital required to enter the market, and the significantly increased needs in these services from big business. In addition, the service sector begins to play a significant role in employment issues. In this regard, there is an increasing need to develop scientific approaches to the study of this area.

The concept of “service industry” has long been used in theory and in practice. It has become a familiar and frequently used term.

Kotler firmly believes: “In the end, all should be limited to the services sector. First, a customer buying a product, actually buys the services that the goods offer. Car is a means of transportation; soap provides its owners the opportunity to wash their hands; tutorial provides knowledge and information” [2].

Chelenkov A. P. defines the service in marketing as a coherent process of interaction between two or more subjects of the market where some agents act on others in order to establish, expand or reproduce the opportunities to receive basic benefits (welfare) [3]. Considering the scope of services, the author points out their relationship with the goods. Under the conditions of constant technological processes, goods and services markets have become almost inseparable from each other (production of goods requires their continued service, etc.) [3].

The services sector is divided into consumer services (targeted to a wide range of individuals) and institutional services.

Institutional services are considered services, the consumers of which do not favor individuals, but companies, organizations, institutions and other economic and managerial structures. The bulk of institutional services are business services.

Business (professional) services are services provided to companies, organizations, institutions and other household and administrative structures, as well as separate individuals which contribute to a successful professional activity or profit.

The proposed interpretation of the above terms may create an impression that there are diffuse notions of business services and consumer services. To some extent this is true, but only partly. The concept of business services is integrated between the concepts of institutional services and consumer services.

Business service is a form of business that provides economic services and management structures and individuals to ensure their professional activities or for profit.

Currently, services as an economic phenomenon is an enormous area of production, financial, legal and social relations, which as it develops, promotes and catalyzes the processes occurring in the areas of production, social (state, municipal) administration, distribution and consumption. Market infrastructure is a specific self-regulating system of divergent industries. In addition, each component of this system is unique.

The consumer is one of the most important elements in the market of a company. There are three main groups of consumers by the nature of consumer preferences: the mass consumer, segmented consumer, and the individual consumer. Taking as a classification feature the type of customer, who mainly focuses on service, business services can be represented as follows (sce table).

The characteristics of the service-oriented system of consumer preferences "mass consumer" includes a standardized range of services, relatively low cost, a large number of separate users, the lack of response to the changing needs of individual consumers, and the availability of information about the company provider. For the mass consumer, undifferentiating tools of mass effect are most effective: advertising, sales promotion methods and PR (public relations), as well as the use of an effective pricing policy and the standardization process of providing basic services.

Most business services are focused on the segmented consumer. The characteristics of the service-oriented system of consumer preferences "segmented consumer" includes an extensive range of services, optimal price and quality of services, availability of groups interacting with each other, and consumer selectivity of the reaction to the changing needs of individual consumers and the company known provider. In the target market segments the company comes out with a specific set of services, with support of their target set of tools specific to the market and product. This strategy ensures a higher volume of sales, and lower level of risk when compared with the strategy of concentrated exposure. The most effective strategy for differential impact for this group of consumers is the use of instruments of direct impact: telemarketing, fax, email, mailing lists, etc. However, this strategy first and foremost requires a high level of management organization. There is a need for permanent, continuous contact and interaction with customers to create positive consumer attitudes to the products of the enterprise. To do this you must use a system of individual actions. The application of modern information technology becomes important then.

The group "individual users" are market participants who have different internal properties, but a combined demand for certain services. Characteristics of the service-oriented system of consumer preferences of individual consumers includes a personalized range of services, flexible pricing policy, a strictly limited number of individual consumers, a company known provider that is responsive to the changing needs of consumers. The most effective strategy of concentrated exposure for this group of consumers is the use of instruments of direct impact, as well as personal sales.

Under conditions of instability, consumer preferences require businesses to diversify services and specialized companies to provide certain types of services. Timely detection of consumer preferences, building services and, increasingly, the choice of appropriate tools to deliver information about new services to consumers can reach the consumer in a timely manner, and thereby significantly improve competitiveness in business services.

Classification of Business Services

Services	Characterization of services by consumer preferences	Consumer	Collaborative tools
– In the area of real estate (real estate, appraisal, etc.); – Printing; – Communication services; – Transportation, etc.	Standardized range of services, relatively low cost, a large number of disparate users, the lack of response to the changing needs of individual consumers, the availability of information about the company-provider	Mass consumer	Tools of mass effect: advertising, sales promotion methods and PR (public relations)
– Services of research and development; – Advertising services; – Civil engineering and architectural services; – Leasing (equipment rental services); – Financial services; – Educational services; – Services of security companies; – Legal services, etc.	Extensive range of services, optimal balance of price and quality of services, availability of groups of interacting users, the electoral response to the changing needs of individual consumers, known company provider	Segmented consumer	Tools of direct impact: telemarketing, fax, electronic mailing lists, etc.
– Urban and landscape planning; – Services, personal security, etc.	Individualized range of services, flexible pricing, a strictly limited number of individual consumers, known enterprise provider, responsive to the changing needs of consumers	Individual consumers	Personal sales

In modern literature there are three main approaches to the interpretation of the concept of "competitive companies": product, economic, managerial (administrative). The essence of the product approach is a focus on the ability of continuous improvement of the product (product line). The essence of the economic approach is to recognize the priority of economic characteristics under companies matching, the most common of which are the profits of a business entity, its costs and, as a target parameter, the ratio between them. The management approach assumes that to achieve the effective market position, the leading role is played by enterprise management and that its improvement is given priority in the competitive environment [4]. The predominant trend in modern conditions is a managerial approach.

For businesses, business services, working in highly competitive markets with established rates, relatively high cost of services, with an assortment side, slightly different in their qualitative characteristics, the most important factor in ensuring the PCB is the level of management. The use of other factors to increase competitiveness is largely limited. In this situation, the introduction of effective enterprise management systems rises to the fore, including systems for customer relationship management.

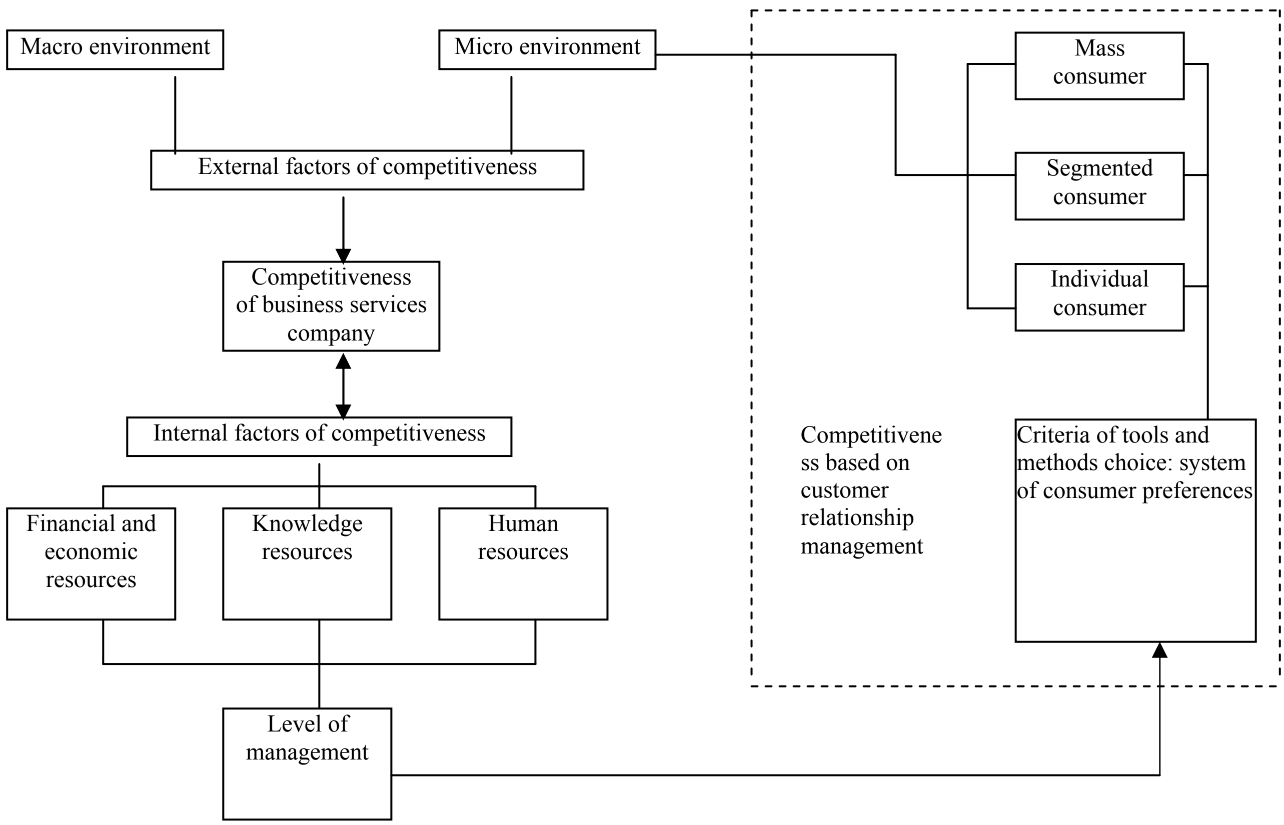
A task-oriented management system of enterprise competitiveness to meet the needs and requirements of clients and the application of principles of interaction between producers and consumers requires the implementation of continuous consultation and coordination tools of promotion.

Changing priorities to ensure maximum efficiency in the short term, focus on long-term cooperation, and building partnerships with customers and suppliers and other market participants can achieve sustainable long-term competitiveness. The effective use of tools and methods of promotion emerges as a leading factor in ensuring the competitiveness of service industries.

Author's model of competitiveness of business services is provided in the figure.

In terms of control factors of competitiveness are divided into two distinct groups: external and internal to business services. Management of a company's activity is an active adaptation of external uncontrollable factors through direct effects on controlled factors in accordance with the stated objectives.

The consumer is one of the external factors of competitiveness that the company may be exposed to through internal factors. In the highly competitive market of business services, the "factor of the consumer" is of the first importance to ensure the competitiveness of the enterprise. As noted above, there are three main groups of consumers: the mass consumer, the segmented consumer, and the individual consumer. In order to ensure its competitiveness, the enterprise chooses a group of consumers (both existing and potential), which is focused mainly on its services. Then tools are selected depending on the system of consumer preferences and methods of interaction with the selected category of consumers.



Model for the competitiveness of the enterprise business services

The system of consumer preferences of the individual user includes an individualized range of services, flexible pricing, depending on the volume of services, the ability to carry out a significant amount of orders, a known enterprise provider, and high quality services responsive to the changing needs of the consumer. The most effective strategy of concentrated exposure for this group of consumers is using instruments of direct impact, as well as personal sales. Using tools of direct marketing, the preferences of enterprise customers are clarified, the offers are tailor-made to meet them, taking into account the specific needs of the consumer. At this stage personnel of the company should be directed not primarily to obtain maximum economic benefits in the short run, but to create a contact with the consumer in order to stimulate further purchases of regular maintenance. Thus, the activities of the company should be directed to the establishment of "partnership" relationship with the largest possible number of consumers. Establishing such a relationship provides the company with orders in the future, and can increase sales specific to the consumer ("grow" it to the largest category of customers), reduce the cost of sales, allow them to raise prices without the fear of losing customers, and as a consequence lead to the long-term competitiveness of the enterprise.

Nowadays the world economy, the services market in particular, grows with the number of enterprises working in it, expanding the range of services and increasing competition. The pressure raises the question of maintaining itself in the market on the survival and development of the company. These trends are forcing companies to look for ways to preserve and expand their market share, improve the competitiveness of their products (services) and the enterprise itself. Under the saturated conditions of the proposal, a limited number of service users, with a preference for variably effective techniques in relationship management with consumers can be an effective tool for ensuring the competitiveness of enterprises in business services in the long term.

References

1. Outsourcing: the creation of highly effective and competitive organizations. M. : INFRA-M, 2003.

2. Kotler F. Marketing services. M. : Business-book, Ima-Cros. Plus. 1995.

3. Chelenkov A. P. Marketing services. M. A: Center for Marketing Research and Management, 2004.

4. Medvedev V. A. The role of services in economic growth consumer services: current issues and prospects. P. 1. Krasnoyarsk, 2004.

O. L. Egoshina
Siberian Federal University, Russia, Krasnoyarsk

V. M. Kleshkov
Chemical plant-branch of JSC "Krasnoyarsk Machine-Building Plant", Russia, Krasnoyarsk

O. E. Semenkina
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

INTERNAL MARKETING AS THE TOOL OF EFFECTIVENESS INCREASE OF INTELLECTUAL EMPLOYEE USE

Under conditions of postindustrial economics intellectual employees as the main manufacturers and knowledge carriers are the source of long-term competitive advantage which is difficult to copy. Therefore the success of business demands the marketing approach not only to consume service, but also to home market that is employees. The article is devoted to the analysis of possibilities, specifying the internal marketing tools and the principles to increase efficiency of intellectual employees' usage in the organization taking into account features of this human resources group.

Keywords: intellectual employees, internal marketing, requirements to the internal environment of the organization, tools of internal marketing.

Activity efficiency of modern organizations substantially depends on business orientation on the client, presence of the unique commodity and service offer, perfection of business processes, but is even more based on that, how much the management copes with the problem of recruiting, managing and holding the intellectual employees, capable to provide competitive advantages of the organization. Cases are known when prospering corporations sharply reduced the industrial indicators and even were absorbed by competitors because could not provide desired conditions of activity for employees, and with their leaving lost the intellectual capital which acted before as the basis of their investment appeal [1]. Complexity of specified problem is connected not only with the rarity of the intellectual employee, as carrier of unique abilities, but also with limited possibility of this group management under its features. Specified factors require review and updating of the administrative tools traditionally applied by organizations, and the offer of additional ones; among the first mentioned can be internal marketing. The article is devoted to the analysis of its application possibility as tool of efficiency increase of intellectual employee using and to working out necessary organizational conditions.

In general the value of intellectual employees in modern organization is defined by that they:

a) are a considerable share of the personnel;

b) create competitiveness of the organization;

c) are the key source of development.

Let's consider the basic characteristics of the intellectual employee:

– handling the information and knowledge in work and ability of their processing in ready products, new knowledge;

– independence of professional activity of the property on means and conditions of production;

– more self-identification with the profession, the field of activity, than with the concrete organization or workplace;

– poor working control because of the results of the intellectual activity sold to the employer and their achievement process is hidden;

– considerable ability to self-organization, therefore their independence does not do harm to the company, and increases the share of such employees in the organization, reduces requirement for the traditional management of the personnel;

– the main motive of activity is possibility of intellectual growth and development, instead of material compensation;

– gradual comprehension by intellectual employees of their leading part in the production process and possibility to play it successfully outside the organization and other;

– responsibility for productivity lays on the intellectual employees who should control themselves, and require independence at work performance;

– the basic indicator of intellectual employee productivity is the quality of solution for the given problem, instead of quantity or volume of made product, as it is peculiar for manual worker [2].

Category of intellectual employees usually includes programmers, lawyers, designers, advisers, experts, analysts, experts in marketing and representatives of other professions capable to create a ready product without physical means of production [3].

The increase of relative density of intellectual employees in modern organizations and competition toughening for them on the labor market increased requirements to the organization from the point of view of creating attractive working conditions, and internal marketing is capable to help the organizations to generate such conditions.

"Internal marketing" concept relates to active development of service business in the developed countries in 70s, and K. Gronroos who offered the concept within functional and tool model of service quality. K. Gronroos also brought in such terms as "internal product) (personnel work)" and "internal consumer" (firm staff) [4].

Today theoretical scientists and experts offer a lot of "internal marketing" concept treatments. Let's consider some approaches to definition of this concept from the point of view of applicability to the intellectual workers (sce table).

In general these interpretations are detailed enough, but concerning intellectual employees are arguable. In particular, D. Lobanov's definition is disputable regarding stimulation, as they cannot be stimulated under inherent features; influence is applicable only through the internal motivation of this employee group. K. Gronroos' and E. N. Golubkova's approaches seem a little limited since the first one is mostly focused on the sphere of services and the process of service, which is carried out by the contact personnel, not considering other categories of intellectual employees, and E. N. Golubkova's approach is focused on external clients and uses employees as means of their gaining. The most adequate, meeting the targets of this research, we consider definition offered by L. Berry and P. Parasuraman as it places emphasis on well-qualified personnel that is intellectual employee.

The internal marketing is focused on home market of the organization, precedes external marketing, as it can be futile to introduce any product into the market if the organization staff cannot provide its high quality. The purpose of internal marketing is to create a true team that is people, whose favour for the business surpasses limits of official duties. They define their work contents from the point of view of external client serving [5].

Approaches to "internal marketing" concept

Author	Definition
K. Gronroos	Marketing assuming work with contact firm staff is intended to create motivational and organizational working conditions promoting to form functional service quality (how consumer service process is carried out) [4]
D. Lobanov	Use of marketing approach with reference to employees, this is forming, stimulation, co-ordination and integration of the staff for effective application of corporate and functional strategy, for the purpose to meet requirements of the consumer through the process of interaction with motivated personnel oriented on clients [6]
E. N. Golubkova	Marketing carried out within the organization and directed to effective training and motivation of employee work, carrying out just contacts with clients and providing these contacts support, and also creating conditions when employees work as united command, providing the most complete satisfaction of clients [5]
L. Berry, P. Parasuraman	Recruiting, training, motivation and retraining of qualified personnel creating working conditions which would satisfy personnel needs [7]

One of the main concepts of internal marketing is the concept of home market as the point where staff of the organization outs values with external clients in exchange for material and moral and social reward given to them by the organization management. As R. Townsend considers, all organization staff should be engaged in marketing, from the owner and the director to the office-cleaner, instead of separated structure [8].

Thus, internal marketing is based, on the one hand, that the personnel is the resource of the organization necessary for its purpose achievement, on the other hand, the personnel is one of the major organization client groups whose wishes are to be satisfied.

The concept of internal marketing in the majority of sources is considered according to analogy with traditional marketing complex (“4P” – product, price, place, promotion). Let us analyze elements of internal marketing complex, offered in [5]:

1. The product is the work offered to the employee by the organization. The organization offers a special product – a post in the organization with its specific problems, rights and duties. The employee “buys” this product, “paying” it with his or her labour. Personnel satisfaction with internal product (work) depends on that, as far as consumer properties of this product (work contents, its volume, creative kind of work, its performance deadline, independence level of decision-making, etc.) correspond to personnel expectations.

To be satisfied with work the intellectual employee needs in creative character of this work (to wide extent), rights and official powers for qualitative performance of the duties, participation in acceptance strategic and operative decisions, instead of their implicit performance, and other characteristics of the work allowing to realize employee potential. Need for conformity of work and personnel expectations speak about expediency as input research of employee expectations in relation to work contents and process at employee admission to the organization, and periodic research in order to track changes of the situation and its timely updating.

2. The internal product price is the cost of the material, social and other goods that employees have for the work. In other words, the internal product price is defined by stimulation degree of employee work.

Application of direct economic influence in relation to intellectual employees is limited by weak material interest of the latter. So these employees, as a rule, have no considerable problems to satisfy material requirements (they have already apartment, car, certain savings and so forth), therefore they can take the liberty of selecting the organization providing not the maximum wages, but possibilities for professional and personal advance, interesting work purposes of which the employees understand and share, where friendly relations in collective are established and so on. However, it does not mean that the organization can save on intellectual employee remuneration of labour. On the contrary, such careerists are informed about their cost on the labour market, but the salary level is not the main motivation for them. One must also pay attention to the size of pecuniary recompense that directors of Russian enterprises usually offer, have no stimulating effect on intellectual employees. On the contrary, size of recompense, capable to stimulate such employees, much more exceeds the sums the directors are ready to spend on these purposes. Therefore, direct material stimulation does not render considerable motivational influence on intellectual employees. Application of indirect economic influence by means of cost accounting introduction (self-recoupment) and profit sharing with organization stock granting can serve as recognition of importance and responsibility increase of this employee category.

3. Place (way of internal product finishing) is the proper allocation of employees according to the charged work within the organization. This component is considered, first, from the point of view of the organizational effectiveness and problem, rights and responsibility distribution.

Organizational influence on the intellectual employee, assuming use of organizational structure, staff list, job description and so forth is limited on the score of performing their production targets the employees are not tied to the concrete workplace and time, organization and means of production and are capable to carry out the duties out of them. Moreover, intellectual employees, as a rule, carry out the creative problems that demand considerable set-up labor, particular mood and situation, inefficiency of administrative and disciplinary influences on intellectual employees is demonstrated by the following example. So, the designer works out the idea during a month, during one more month he finishes it, but in spite of the fact that the result appears in two months, it does not point to the fact that the designer did not carry out the chief charged task, also as well as does not speak about the opposite: the result could be achieved in half an hour, and all two months the designer created visibility of work, sabotaging the chief task, that is not the authority for him, not understanding and, hence, not sharing the organization purposes. In other words, if the employee is interested for any reasons to stay in the organization which purpose he does not share and the chief of which is not the authority for him, owing to mentality he will find possibility, not clashing avowedly, to sabotage administrative and legal influence.

One of decisions for specified difficulties can be simplifying the organizational structure. Realization of existing market possibilities of the enterprise is reflected in its organizational structure. The main tendency of organizational management structure development is to plane sequentially and to increase adaptation. Experts define the following requirements for effective organizational structure formation [9]:

– to reduce division sizes and staff them with more qualified personnel;

– to reduce management level number;

– to plane organizational structure, to decentralize, to expand economic independence;

– cooperation and partnership, group work organizing;

– to delegate authority and to prompt the employees to take responsibility upon themselves;

– to create organizational conditions for rapid reaction to changes;

– to orient the current work including schedules and procedures according to customer’s needs and so forth.

Features of the organic organizations, that is to stimulate creative approach to work, to increase awareness level, to change authorities and responsibility of each employee according to solved problem character, etc, lead to increase in importance and participation of each employee in decision-making, disclosing of organization personnel potential, activation of innovation thinking that is highly typical for intellectual employees.

4. Internal product promotion assumes to create system of effective marketing communications between home market participants. Along with marketing communications, it is necessary to consider common organizational communications as well. Internal marketing problem is the analysis of formal and informal internal communications purposely to reveal efficiency and expediency of communications for personnel quality work.

To use methods of marketing communications in internal marketing means to apply the following tools:

– sale promotion;
– personal sale;
– advertising;
– public relations (PR).

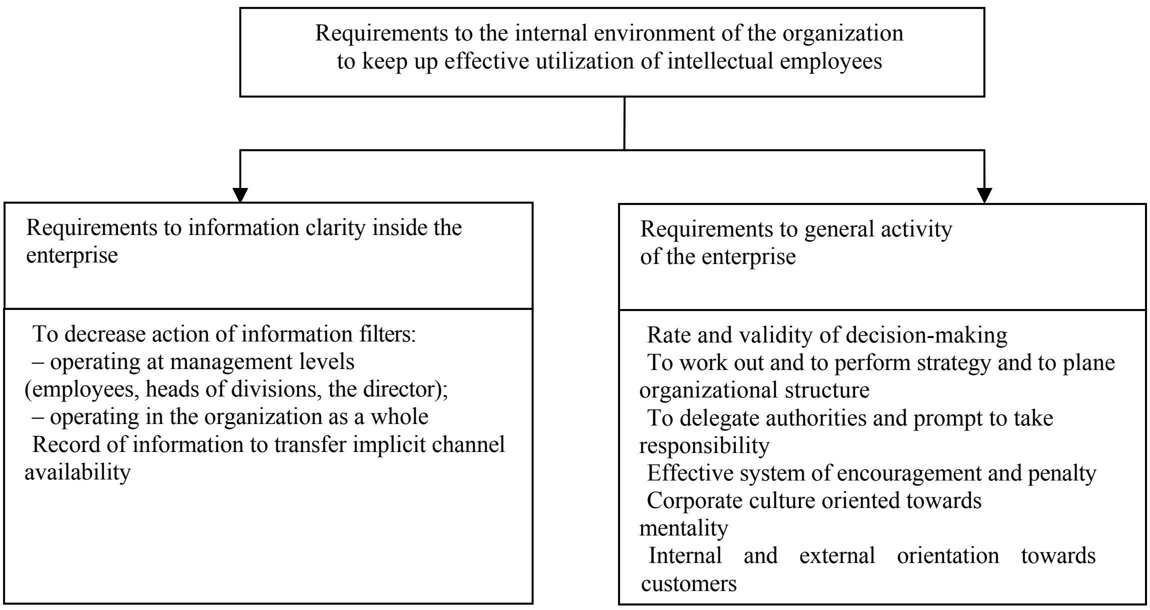
Sale promotion in internal marketing is understood as the short-term incentive measures promoting work "sale" to employees of the organization, for example, granting additional payment or other privileges for performance of off-schedule, important work. The intellectual employee, as well as other categories of the personnel, can be induced with similar tool, more rare with financial incentive, more often by means of authority accrual, participation in the interesting project, etc., and well understanding expediency of increase in the consumption of his (her) labor.

Personal sale is understood as oral presentation by the chief in conversation with one or several potential internal buyers for the purpose of its sale. It is the effective tool for work-product promotion at certain stages of its sale when officials and employees oral communications can be more effective methods of motivation, rather than methods of sale promotion. Moreover, personal sale methods are used during interview at recruiting. Personal character of such meeting allows understanding more precisely employee inquiries, to explain them organization positions and possibilities, to display sold works to advantage, to eliminate misunderstanding, to establish long-term confidential relations. Such promotion tool may be significant enough for the intellectual employee, as his work status requires often direct contact with the officials, and a work personal sale, as internal product of the organization, is often more effective, as allows to minimize procedural encumbrances and misunderstanding of future (or current) participants of home market.

Advertising in internal marketing is, basically, internal, presented by various printed internal information materials coinciding with public relations materials. External advertising should attract new employees to the organization. The targets of this advertising are in many respects identical with organization image advertising as a whole. The tool is applied in relation to intellectual employees under condition of information clarity in and out of the organization.

Internal marketing public relations, first of all, are targeted at the inside of the organization (internal function of public relations is used). The public in this case is understood as various employees of organization divisions, including the contact personnel. Internal function is directed to create and to keep up corporate social responsibility inside the organization, in particular, to meet requirements of external customers. It is a matter of organization's high reputation among its personnel, friendly climate inside the organization, responsibility and interest in administration affairs.

External function of public relations is directed to attract highly skilled personnel for a work in the organization may be used also in internal marketing. Such tool can be used in relation to intellectual employees only under the following conditions (sce figure).



Shows requirements to the internal environment of the organization to keep up effective utilization of intellectual employees (authors' working out)

As a whole, the analysis shows, that organization management problem is to use internal marketing elements to increase application of intellectual workers' abilities that provides enterprise development. According to the above mentioned requirements towards intellectual employee management, the authors offer the following principles of internal marketing use in relation to intellectual employees of the enterprises:

– the principle of information clarity inside the organization: the information is a paramount resource of intellectual employee activity; therefore, the purposes and problems, strategy and tactics of the organization should be opened and clear for the personnel to understand what, how and why must be done. It provides implication feeling in production, improves possibilities for personnel motivation and team creation;

– the principle of personnel partnership in decision-making: joint decision-making on the important questions eliminates contingent opposition of intellectual employees in relation to coercive methods of decision-making and creates feeling of participation, strengthens spirit of "cause";

– the principle of incentive administrative influence preference to punishing and forbidding ones in relation to intellectual employees, underuse limitation of the latter;

– the principle of "triune" identity: what the chief thinks about the intellectual employee, what the employee thinks about the chief and what the chief and the employee think about the customer – this all should not contradict each other, and as the result form the identity of corresponding information which is transferred to the customer and creates that what the customer thinks about this organization. As the result of the identical information environment providing internal and external customer orientation to this organization devotion is created;

– thus, we have revealed, that internal marketing is one of the tools to create friendly organizational environment, where the effective application of the intellectual employees is possible. We have specified elements of internal marketing that is caused by features of intellectual employees and have offered requirements to the internal environment of the organization to provide effective application of intellectual employees and principles of internal marketing use in relation to intellectual employees of the enterprises. These tools provide, in turn, the enterprise success within actually mobile market conditions.

References

1. Inozemtsev V. L. Prospects of postindustrial theory in the varying world // New postindustrial wave in the West. M., 1999.

2. Drucker P. Management problems in XXI century : text-book. M. : Williams, 2007.

3. Yurtaikin E. Features of intellectual employee level // Human potential management. 2005. № 2.

4. Novatorov E. V. International models of service marketing // Marketing in Russia and abroad. 2000. № 3.

5. Golubkova E. N., Mikhailov, O. Z. Internal marketing and personnel management – tools of management efficiency increase // Marketing in Russia and abroad. 2000. № 3.

6. Lobanov D. Internal marketing of the personnel in the Russian companies [Electronic resource]. URL: http://www.hr-portal.ru/article/vnutrennii-marketirm-personala-v-rossiiskikh-kompanivakh.

7. Berry L. L., Parasuraman P. Marketing services. N. Y. : The Free Press, 1991.

8. Townsend R. Secrets of management or how to keep the company from suppression of people initiative and profit decrease. M. : HKK "Intercontact", 1991.

9. Highnish S. V., Tokareva N. U. Organization structure: from the reality to virtually there is one step. Principles of organizational and structural providing of innovative processes at the enterprise. M. : MNIIPU, 1999.

E. A. Kalashnikova
Krasnoyarsk State Agrarian University, Russia, Krasnoyarsk

COMPETITIVENESS OF THE HIGHER EDUCATIONAL INSTITUTION IN THE MARKET OF EDUCATIONAL SERVICES IN KRASNOYARSK CITY AND KRASNOYARSK REGION

The competitiveness of higher educational institutions in the market of educational services of Krasnoyarsk region is considered in the article, the concept "the higher vocational training" is also given in the present research. The problem urgency is proved, the general approaches and conditions of research of competitiveness of higher educational institutions are shown, major factors for a competitiveness estimation are allocated.

Keywords: competitiveness, a higher educational institution, formation, services, competitiveness factors.

In the conditions of modern lines of development of the Russian higher education, and also in accordance with the world tendency of globalization some questions of the competitiveness of higher education and the competitiveness of the higher educational institutions become very important. They have been defining value for successful development of Russia. Marketing in higher educational sphere gradually gets a very independent value. It requires the further study of scientifically-categorical mechanisms, revealing the peculiar features and tendencies of marketing activity inherent in a higher educational institution as the subject of the mark of higher educational services and a labor market. Questions of high school adaptation to market conditions of managing become very important. Its competitiveness and formation of effective marketing of educational services increase either. Ways to increase the competitiveness of higher educational institutions and methodology of its estimation are developed insufficiently because the general aspects devoted to the estimation of the competitiveness of the organizations are considered in various scientific works. More over it is considered without specific features inherent in the higher school.

The state higher educational institutions are the object of the present research. We also analyze the competitive environment of their functioning in Krasnoyarsk region.

To understand the essence and specificity of the market of educational services better we give some the general definitions of formation, educational sphere itself and educational services.

According to the definition education means the process and results of personal skills, abilities and behavior development, when a person reaches the maturity and the individual growth. This definition shows that they do not differ from each other. It was confirmed by the twentieth session of General conference of UNESCO in 1978.

According to the Law of the Russian Federation "About education" from 10.07.1992 year № 3266-1, education is understood as purposeful process of upbringing and training in person's interests, society's, state's interests. It allows to ascertain citizen's achievements in all educational levels established by the state (educational qualifications). At the same time, the sphere of educational services is defined as complete, social and cultural complex which has the specificity and features of functioning and interaction with all public structures.

Educational service is a great complex of actions of educational and training features of educational, research, organizational and educational areas. It is directed on satisfaction of requirements of an individual. During realization of this process the individual gets the development of his professional skills.

Any educational institution, including high school, presents educational services of the certain kind. Pupils and students are consumers of these services. Also any educational institution simultaneously represents the results of its activity on a labor market. This duality brings considerable mess in definition of this activity results. Target commodity markets and groups of consumers. Operating simultaneously in two connected and interdependent markets, the market of educational results and services and a labor market. Nevertheless, the high school has one result with which it enters on both markets. All researchers converge that the basic direction of activity of high school is granting of educational services, but the do not give unequivocal definition what exactly "the educational service" is, whether it is, for example, a lecture of the teacher, a complete course of preparation of specialists? Is it an organization of a students' scientific society or organization a conference with branch representatives. More exact definition education results is extremely important for understanding a concrete service in general activity of high school.

From the point of view of general activity of high school and the classical theory of marketing the result of the market of educational services is educational program which is developed to satisfy the requirement for education, vocational training, or retraining. In other words, it is developed to achieve a certain social effect (change of educational or professional level). Educational institution enters the market educational institution just with this result not separate educational services in the form of lectures, seminars, etc. Thus a high school product it can be defined as an educational program, that is the complex of educational services aimed to change the educational level or vocational training of the consumer. It is provided with corresponding resources of the educational organization. Depending on the possibilities and requirements of clients high schools offer various assortment of such programs.

Speaking about a role of the market of educational services in a society, it is necessary to notice that the higher education realizes not only tool function in order to train highly skilled professionals, but also functions of social mobility, escalating of a mental potential of society, distribution of the most socially significant cultural norms. In other words, the higher education solves various problems: social, political and individually-personal.

The higher vocational education is the most dynamically developing level of educational system of Russia. Russian and foreign experts notice that economically developed countries preparing highly skilled experts, who represent a mental potential of the country will have an advantage in XIX.

Traditionally the higher education is divided into two stages: a bachelor degree and a magistracy/postgraduate study, but before joining to Bologna process in Russia such a division concerned only postgraduate studying. Instead of bachelors and masters the Soviet high schools prepared specialists. The basic establishments of higher education are universities and colleges. Graduates of high schools usually receive the diploma, and post-graduate students get a scientific degree after the defense of thesis. During the last educational level students and post-graduate students are obliged not only to get education in the usual meaning of this word, but also to take an active part in scientific researches. Getting a diploma or a scientific degree reception depends on the results of scientific work.

The education system represents a difficult organizational structure, its constant updating is caused by shifts in a social structure of the country. It is also caused by changing interests of various social classes and population groups. It is also caused by the general tendencies of developing processes of modernization all over the world.

Researching the condition of the educational services market of Krasnoyarsk region is the increasing in connection with the increasing of the image of the higher vocational training. It is also connected with the reform of the Russian educational model to the western one the Bologna process.

The quantity of students who get higher educational institutions which increases from year to year shows the growth of good image of higher education in Krasnoyarsk region. It is caused by that fact that nowadays experts with higher education are claimed on a labor market. In turn, the growth of number of the students wishing to graduate led to the expansion of branches of Krasnoyarsk state high schools (in the cities of Norilsk, Lesosibirsk, Achinsk, Kansk, Minusinsk, Dudinka and Tura), it is also led to the fact that a lot of branches of high schools from other cities of Russia (Moscow, Omsk, Novosibirsk, St.-Petersburg, Irkutsk) was opened in Krasnoyarsk region [1].

It is necessary to notice that the expansion of the supply in the market of the higher vocational education in Krasnoyarsk and Krasnoyarsk region occurs not only at the expense of growth of quantity of high schools, but also substantially thanks to occurrence and competition toughening in the higher education market. Striving to hold competitive positions in the market of the educational services, almost all high schools of a city open new specialties and specializations. For example, the demand which increased in the late 1990s for economic and legal specialties led to the fact that nowadays each high school of Krasnoyarsk, without taking into consideration its specificity and a pattern of ownership offers these specialties.

Nowadays every high school has the aim to expand the offer, but this process has no such "spontaneous" form any more. Before to enter the market of educational services with the new offer, the high school management tries to understand, the consumer requires and interests. They do a research if this or that specialty is claimed or not. For example in 2005–2008 following innovations have been spent: the Krasnoyarsk state agrarian university opened three new specialties: "Protection of plants" – at the agronomical faculty, "the Finance and the credit" – on the economical faculty and "Management of the personnel" – on international faculty. It was done because of the questioning of school graduates and employers. This questioning made it possible to reveal that graduates of these specialties will be in demand. In Krasnoyarsk state university at the philological faculty there were new directions such as: Regional country studying (Polish language and the literature); philological maintenance of document circulation (highly skilled reviewers will be prepared there).

The volume and the supply structure from the point of view of modes of study extend gradually. Nowadays in higher educational institutions of region there are full-time, correspondence and evening modes of studying. Full-time and correspondence modes of studying are very popular, while the demand for such mode of studying as evening courses is lower, and caused practically a total absence of this offer.

In a number of high schools (within the limits of particular faculties or specialties) there is a new mode of studying called externat. Its occurrence was caused by the necessity of getting higher education for short terms. Officially this form of education is licensed in all high schools, however factual within the externat is carried out only by eight high schools of Krasnoyarsk (Krasnoyarsk State Pedagogical University, Siberian State Aerospace University, Krasnoyarsk State University, Siberian State Technological University, etc). Besides there is a certain alternative to this form of higher education – education according to so-called "Individual plan" or "Individual schedule". High schools management staff say that the system of individual plans has the future. They say that it has a great advantage – considerable economy of means of students [1].

The remote method of studying is becoming very popular. Its main point is that students are taught and trained without visiting the institute, credits and examinations are taken in written and electronic form. In this case the presence of a tutor or another specialist is necessary in order to supervise the process. Today the remote method of studying is used by not all high schools of Krasnoyarsk so as externat. It is entered only at separate faculties or specialties.

Meeting the requirements of the Bologna agreement full-time courses in many high schools, are divided into a bachelor degree and a magistracy. It is necessary to notice once again that despite the existence of the federal program of reforming the system of higher education, the management staff of each high school independently solves this question for itself. Nowadays we have a paradoxical situation: by 2010 all high schools of Russia should carry out preparation of students according to clauses of the Bologna agreement, but at present not many high schools of Krasnoyarsk try to enter this system in educational process (only partially within the limits of certain faculties and specialties).

High schools of Krasnoyarsk and Krasnoyarsk region expand the offer in the field of educational services in reply to raising demand. Occurrence of new specialties, specializations, training methods is the increasing of competitiveness of high schools in the market of educational services. It is another level of its development.

Almost in all leading high schools of Krasnoyarsk and Krasnoyarsk region there are specialties of an economic profile. On the basis of market research of educational services of Krasnoyarsk and Krasnoyarsk region a specialty "the Economics and management" is very claimed, number of students trained on the specialty "the Economics and management" reached 32 157 persons, in 2009–7079 persons were accepted to this specialty, 5 059 persons graduated from this specialty in 2008. The offer in the field of the higher vocational education on the specialty mentioned above is narrow specialization. Almost each high school offers the specialization «Enterprise economics and management» in a certain branch of a national economics (trade, agriculture, power, etc.).

According to data fro 2009 thee are 11 high schools and 26 branches of higher educational institutions in

Krasnoyarsk region. The total number of students in 2009 was 118 822 persons, and 117 906 persons were in 2008. This fact confirms that the number of students of higher educational institutions is growing. Of this number there were 63 219 full-time students, 2 207 persons were full-time and correspondent students, 44 875 persons were correspondent students, and 8 521 students studied on the externat. On the basis of these data it is possible to draw a conclusion that for today the full-time course of studying has a priority, though it is necessary to notice that the externat is rapidly developing. The data for 2008 shows that the externat had 4 458 persons, and 8 521 were trained in 2009. This fact means that the quantity of externat students is promptly growing [2].

Thus it is possible to present the certain competitive environment of higher educational institutions of Krasnoyarsk region and Krasnoyarsk. Undoubtedly each high school has its own specificity, its own potential, various reputation and its own position in the market. However possibilities used at the wrong time and unnoticed threats can lead to serious consequences. Considering the increased competition in the market of the higher vocational education, high educational establishments of Krasnoyarsk aspire to constantly trace the changes occurring in the external and internal environment.

It is necessary to notice that for the majority of large state high schools of Krasnoyarsk the main advantages are the highly skilled teaching-staff, good reputation, the provided material base etc. For branches in other regions the main advantages are low cost of training, the possibility of a correspondence mode of studying, a remote method of training etc. The main weak points for the majority of high schools are high cost of training, the impossibility to organize a good practice for students and a narrow orientation within the limits of one specialty. Branches also have some weak points. These are a low reputation, mistrust of the population, and also the absence of full-time courses of studying.

Economic and political transformations occurring in Russian society affect the educational system in Krasnoyarsk region and in the country as a whole. The huge potential of the Russian higher education endures one of the heaviest stages in its development today. An economic crisis in the country, liberalization of the prices, inflationary processes, deficiency of the state budget and crisis of budgetary financing led to that fact that educational establishments of Krasnoyarsk region have the lack of financial and material resources. State financing does not cover all requirements of high schools. As the result scientific researches are being reduced, scientific schools are being closed, the social differentiation in access to a quality education is increasing, level of social support of pupils and workers of the education system is decreasing.

Formation of market relations in the education sphere and the development of the infrastructure of the market of educational services in Krasnoyarsk region are very contradictory. On one hand, as a result of active adaptable activity of many educational institutions, the educational system develops and has positive tendencies. The new legislative base is formed, there is a redistribution of administrative functions, there are elements of multichannel financing, the maintenance of educational programs, etc. On the other hand we have a number of problems. First of all, it is the impracticality of the major high schools to work in the conditions of an accruing competition.

As the result there is a problem of increasing competitiveness of high schools in the market of educational services of Krasnoyarsk region. Therefore research of theoretical and methodological bases of increasing competitiveness of high schools and development of the practical recommendations considering regional features of their functioning becomes very actual.

The educational institution is a part of the education sphere generated in the region. It is also a part of spheres of a national economy. Development of market relations in Russia has radically changed economic operating conditions of the state high schools. Thereupon high schools of Krasnoyarsk have urgent problems of the maintenance of viability, the maintenance of a financial condition, and searching of sources of sustainable development. The basic characteristic reflecting the ability of high school to offer the educational services in the market of formation successfully is its competitiveness. The key moment of enterprise activity of the high school is the organization of works to increase its own competitiveness [3].

Considering that the high school is the organization with a wide spectrum of various kinds of activities, the estimation of its competitiveness can be only the main criteria.

In the conditions of the competition any disadvantages and advantages of high schools have its competitive lacks and competitive advantages. These become the key estimated criteria to define the degree of high school competitiveness. The creation of competitive advantages and overcoming competitive lacks is t of the competitiveness in the market of educational services. In turn the presence of competitive advantages provides the participants of this market with a recognition and interest from outside environments. The presence of competitive lacks weakens their positions in the market of educational services. It does them unattractive for society.

These advantages and lacks come are estimated and analyzed during the process of high schools comparison. Their values often have relative character that can lead to risks when using the received data. Therefore there is a necessity to analyze and study the problem of competitiveness of higher educational institutions and to work out new methods and models of an estimation of competitiveness of high schools. It is necessary to work out the actions for to increase the competitiveness of high schools [4; 5].

For to solve a problem of the estimation of high schools competitiveness the author allocated the factors with the help of which it is possible to lead a qualitative and quantitative estimation:

– guality;
– the price;
– a market share;
– advertising and information service.

Having researched after research the factors of competitiveness of the organizations, the author decided to stop on four factors mentioned above as the majority of methods of competitiveness estimation of organizations and enterprises are based on one or two factors and as it was told above, the competitiveness estimation of organizations can be only many criteria if the given estimation aspire to objectivity.

During the research the author proved a choice of four factors of competitiveness mentioned above.

Quality of educational services is the most significant factor of competitiveness for high school. Choosing a suitable educational establishment in order to increase the educational level according to existing educational qualification, consumers of educational services carry out their choice focusing on the ability of suitable educational institution to satisfy their requirements.

In State Standard ISO 9000–2001, there is a following definition of the term "quality": "Quality is a degree of conformity of characteristics inherent in object to the established requirements".

The price of educational services. Nowadays at the time of market relations development, the important factor at a choice of a higher educational institution is the level of payment for the course. Most of students pay for their course. The quantity of budgetary places, as well as budgetary financing of the state higher educational institutions is constantly reduced. When a consumer do not have sufficient financial resources or possibility to pay for educational course, the price is the defining factor when choosing a higher educational institution.

The market share. This factor is important for maintenance of competitiveness of high school because the more students study in a concrete institute, the more its competitiveness is, it points how many students gave their preference to a concrete higher educational institution. If throughout several years it is possible to trace dynamics of increasing a market share of a concrete high school, it is possible to draw a conclusion whether it is competitive or not.

Advertising and information service is the important factor to maintain the competitiveness of high school because if the high school is accredited, its teaching structure is completed by qualified personnel staff and the high school has all necessary resources for granting qualitative educational services, it is necessary to provide its potential consumers with the information about the educational services of a concrete high school. It is necessary for high schools to inform the consumers of all services, about their quality, the prices and so on, it will promote the growth of a market share and it will promote the competition development in the market of educational services.

It is necessary to notice that the quality estimation as the separate factor of competitiveness was carried out on the basis of several indicators:

– number of the integrated groups of specialties;
– percent of the faculty personnel staff with scientific degree and academic status;
– percent of doctors of science and (or) professors;
– percent of the faculty working in high school as a permanent personnel staff;
– number of branches of scientific specialties for postgraduate students;
– number of branches of a science where scientific researches are carried out.

For the estimation of each factor of competitiveness of a higher educational institution the author developed a special estimated scale, then the author applied some mathematical methods and finally the mathematical model of an estimation of competitiveness of higher educational institutions was developed.

Complex research of competitiveness of higher educational institutions in the market of educational services of Krasnoyarsk and Krasnoyarsk region. We also analyzed which factors of competitiveness can be applied by working out the actions for increasing the competitiveness of high schools in the region.

The basic theoretical conclusions of the present research can be used as teaching materials for the following subjects: Marketing, Economics of education, the Basis of business and others.

References

1. Tereshchenko N. N., Bondarenko I. V. Research of the market of educational services of the higher school : monograph / under the general ed. N. N. Tereshchenko ; Krasnoyarsk State University. Krasnoyarsk, 2005.

2. Educational institutions, average special and higher educational institutions in Krasnoyarsk region on the beginning 2008/09 academic years : the Statistical bull. № 5-116. Krasnoyarsk, 2009.

3. Azoev G. L. A competition: the analysis, strategy and practice. M. : The economy and marketing Center, 2006.

4. Fathytdinov R. Orientation of training on competitiveness // Higher education in Russia. 2007. № 9. P. 38–44.

5. Fathytdinov R. Management of competitiveness of high school // Higher education in Russia. 2006. № 9.

L. A. Mochalova
Altai State Technical University named after I. I. Polzunov, Russia, Barnaul

THE MODEL OF THE CORPORATION RISK-ORIENTED FINANCIAL STRATEGY UNDER THE CRISIS CONDITIONS

The decomposition model of the corporation financial strategies development and selection of the risk-oriented strategy in respect to the financial risks criteria is represented. The definition concept of the risk-oriented financial strategy is given.

Keywords: model, financial strategy, financial risks, criteria.

Market environment, increasing competition, the financial crisis force each enterprise to search for new, more effective approaches to choose a financial strategy of the enterprises, which should guarantee financial stability at the assigned level of risk.

Strategies of contemporary corporations are directed mostly for the economic increase. The scenario of possible unfavorable changes in the ambient conditions is not considered there. The optimism, which a priori prevents the crisis trajectories of economic development under the conditions of the contemporary crisis, has led many corporations to the loss of financial stability and solvency or to bankruptcy.

As the author considers, for successful formation and implementation of the financial strategy it is necessary to find the approach, which ensures reaching and maintaining the financial stability of the corporation on the basis of the balance of the sales volume dynamics, assets and the sources of finance structure in respect to financial risks.

At present the concept of the financial strategy in financial management is given in different aspects by different authors: as an element of strategic or financial management as a system of actions to reach the objective, or as a mechanism of making and implementing administrative decisions in financing, etc. The analysis of this definition interpretation allows to provide a generalized definition of the financial strategy.

The financial strategy is one of the main types of the functional strategy, a system of actions to reach long-term objectives of financial activity, on the basis of which the policy of attraction and use of the corporation financial resources is studied, which includes mechanisms of the required volume formation and effective investment of these resources into the enterprise assets considering the environment changes.

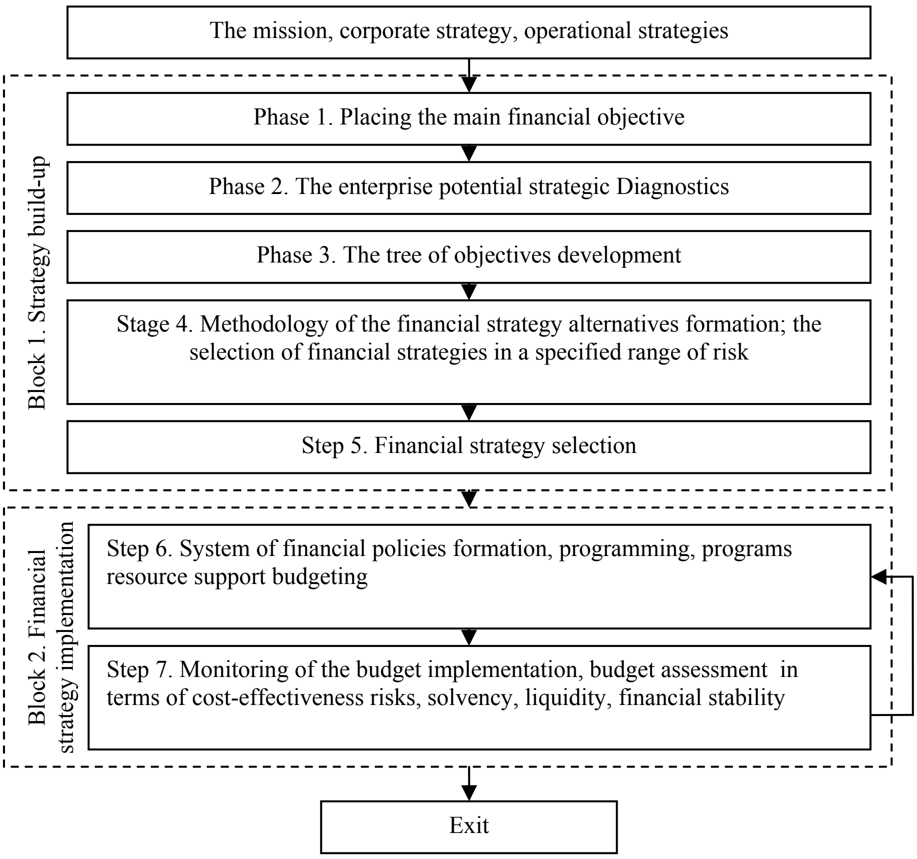


Fig. 1

The formation of the financial strategy should be aimed primarily at maintaining a stable financial situation of the organization. Therefore, we offer the selection of growth financial strategies according to the given criteria of financial risks. It needs the notion of the risk-oriented financial strategy (ROFS).

ROFS is a system of managerial decisions to meet the prospective goals of financial resources formation, distribution and use, ensuring financial stability of the organization through balancing sales volume dynamics, assets and sources of finance structure in respect to financial risks, taking into account the environment changes.

To develop of the risk-oriented financial strategy it is appropriate to use a set of indicators for assessing financial risks as criteria: financial activity effectiveness, liquidity and solvency, financial stability, dividend policy.

This approach will allow managers to make managerial decisions to reach and maintain stable financial situation during the implementation of the corporate strategy.

The main idea behind the proposed methodical approach is the choice of the financial strategy aimed at achieving and maintaining the corporation financial stability by identifying the maximum increase in sales, limited by acceleration of profit rate in a specified range of financial risks.

Fig. 1 shows the block diagram of the financial strategy formation and implementation which includes 7 consecutive stages.

The main element of the methodical approach is the methodology of the financial strategy alternatives formation and the methodology of ROFS choice. The procedure of ROFS formation (stage 4 of the systematic approach) is the sequential passage of the following steps (fig. 2).

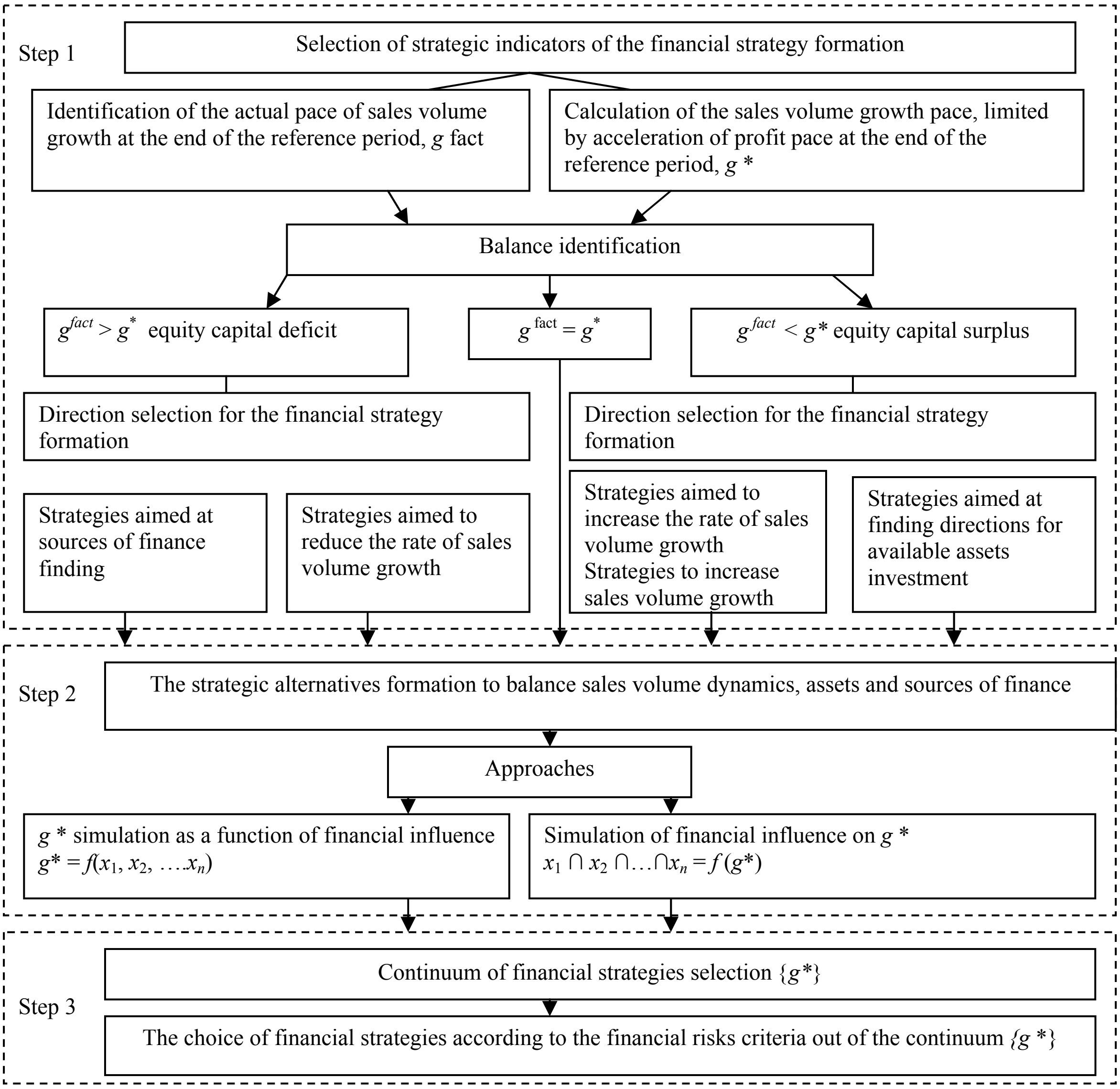


Fig. 2

Step 1. The balanced correlation of sales volume growth, the assets structure and capital is determined. The Model of Optimal Growth Strategy by *R*. Higgins is used:

$$g^{*} = \frac{\Delta E}{E_{by}} = \frac{RR \times NP}{E_{by}}, \quad (1)$$

where g^* – sales volume growth rate, limited by equity capital growth (profit), %; E_{by} – equity capital at the beginning of the reporting year, monetary units; ΔE – equity capital fluctuation for the reporting year, monetary units; *RR* – reinvestment factor (capitalization); *NP* – net profit for the reporting year, monetary units.

The development of the risk-oriented financial strategy includes the following procedure:

1) determination of the actual rate of sales volume growth at the end of the reporting period, g factual;

2) determination of the sales volume growth rate limited by the equity capital growth rate at the end of the reporting period, g^*;

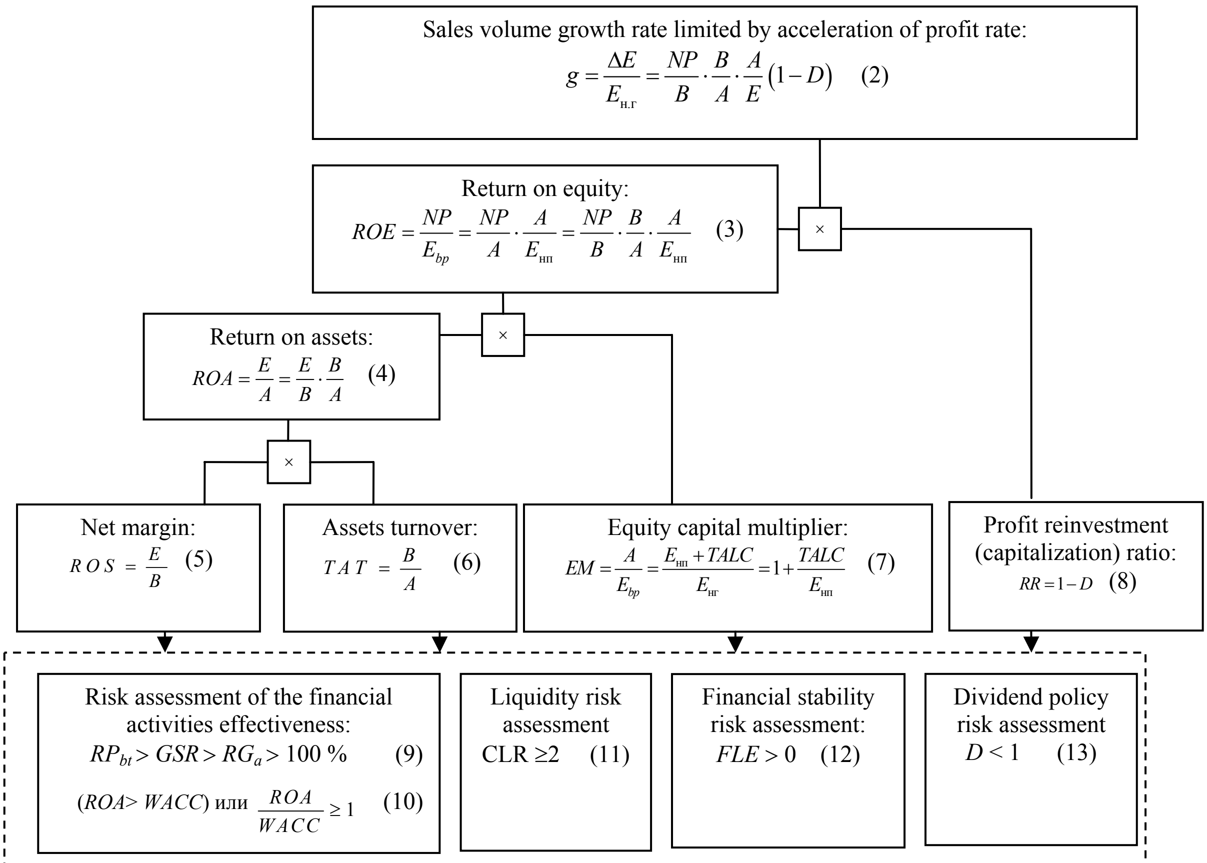
3) identification of the factual growth rate and the sales volume growth rate, limited by the equity capital growth for balance.

4) a managerial decision about the balance of the factual growth rate and the growth limited by the equity capital growth is made on the basis of the following criteria:

– if $g^{fact} > g^*$, then the organization experiences lack of money and must take measures to achieve the of balanced growth level;

– if $g^{fact} < g^*$, then the organization has a surplus of money and needs to decide what to do with profit, which exceeds the organization needs.

Step 2. To calculate the alternatives of the risk-oriented financial strategy we offer a decomposition model that includes a combination of Higgins' model, DuPont's model and the criteria for financial risk assessing: financial activities effectiveness, liquidity risk, financial stability, dividend policy risk (fig. 3).



Convention:

NP – net profit,

E_{bp} – equity capital at the beginning of the reporting period

TALC – the total amount of loan capital, monetary units

A – gross assets

B – the total proceeds

RP_{bt} – the rate of before-tax proceeds, %

WACC – weighted average cost of capital, %

CLR – current liquidity ratio;

FLE – financial leverage effect, %

D – dividend payment ratio

RG_a – Assets growth rate, %

GSR – Sales revenue growth, %

Fig. 3

The decomposition model reflects the possibility of the financial strategy development under the financial crisis conditions, when not only sales volume growth, but also financial stability maintenance is given priority. The proposed decomposition model has the following features: firstly, an important element of the model (2) is the allocation of profits factor (8), which sets compromise between sales volume growth and dividend payout growth, that generally affects the corporate valuation;

– secondly, the factor model (3) allows to determine the relationship between the three main indicators. The organization stability depends on their balance. They are sales net profitability (5), assets turnover (6) and sources of finance structure (7);

– thirdly, risk assessment (9)–(13) allows to prevent the threat of financial resources and financial stability losses.

The proposed model includes seven variables: rates of sales volume balanced growth, equity capital profitability, capitalization of earnings factor, assets profitability, sales net profitability, assets turnover, the correlation factor between equity capital and loan capital. This allows to set the relationship between the listed many variables (x_i) and sales volume growth rate (g).

Based on the decomposition model there is a simulation model of balance indicators: rates of sales volume balanced growth, capitalization of earnings factor, sales net profitability, assets turnover, the correlation factor between equity capital and loan capital. As a result, effective alternatives for the financial risks criterion have been identified.

The model works in the following way:

– actual values of three out of five listed factors are recorded;

– a target (independent variable) (either g^* or x_i) is chosen;

– a grid of the target values with particular spacing is specified;

– the resulting factor (dependent variable) value is calculated: $x_1 \cap x_2 \cap \ldots \cap x_n = f(g^*)$ or $g^* = f(x_1, x_2, \ldots. x_n)$;

– the values limit (passage) of financial risks indicators is specified.

Step 3. ROFS choice is made out of many alternatives according to the criteria: “the sales volume growth rate limited by the acceleration of profit rate” and “acceptable level of financial risks”. To assess risks admissibility threshold limit values of efficiency indices and change limits for these indicators are specified.

The choice of the financial development strategy is based on the financial strategies matrix presented graphically (fig. 4).

In the matrix there are four main types of possible financial strategies defined by two factors: the ability of the business to finance its sales growth and the attitude to risk.

The matrix horizontal axis shows the attitude to risk (from risk avoidance to risk orientation); the vertical axis shows the business’s possibility to finance its sales growth (from money deficit to their excess).

Strategies of “*A*” and “*B*” sectors are typical for slowly developing organizations whose sales are growing more slowly than the ability of the business to finance this growth, thereby cash surplus is generated. This situation allows to raise the bar of the acceptable risk level.

Financial strategies of “*A*” sector aimed at preventing, localization and avoiding risk are chosen by organizations dominating in the market having stable enough profitability, advanced technology and skilled personnel.

Financial strategies of “*B*” sector are focused on the risk typical for those enterprises which have chosen reorganization and restructuring, technical re-equipment and reconstruction of their industrial and technological base, organizational and managerial innovations.

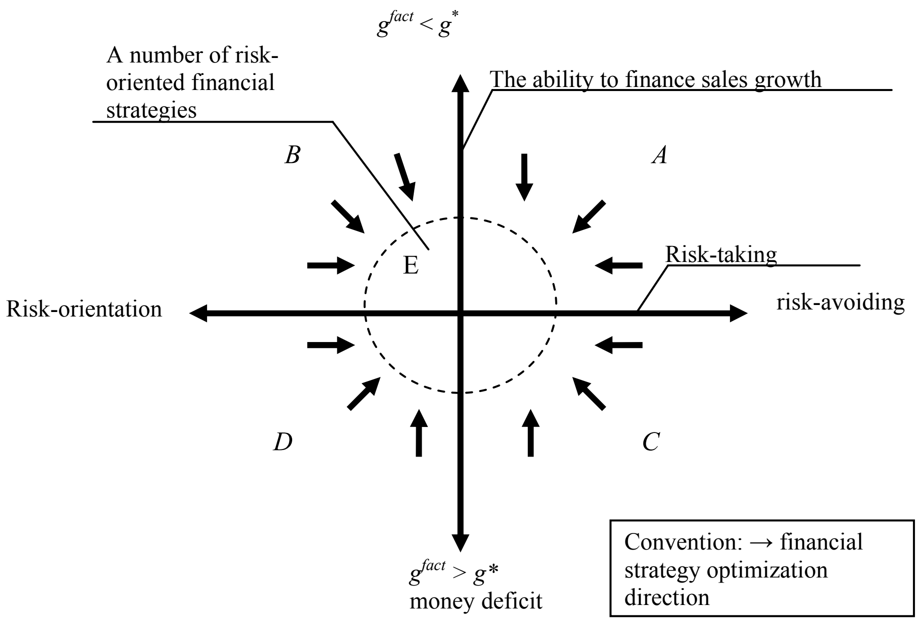


Fig. 4

Strategies of “*C*” and “*D*” sectors are suitable for organizations having money deficit to finance fast sales volume increase. High rate of sales volume increase requires additional investment, which the company often lacks and the company management has to borrow money. If there is no special control and management measures high growth rate inevitably lead to bankruptcy.

Strategies of “*C*” sector achieve financial recovery and restore solvency, they are aimed at achieving optimal criteria of financial risks.

Strategies of “*D*” sector are risk-oriented. “*D*” sector strategies which are closer to the matrix center are targeted at the risk giving the “possibility of breakthrough”, this requires changes of the basic economic characteristics and the business capacity as well as great investment. The risk here is very grave, but the income must be very significant (otherwise there is no sense to risk). Such risk-oriented strategy may at some stage replace “*C*“ sector strategies and activate a new spiral of the organization life cycle. Sector “*D*” strategies that are further from the matrix center work by the principle “everything or nothing”. Today Russia’s experience shows that participation in risky projects in a difficult financial situation, when all the property is at stake, is not a single case in manufacturing business. As a result, the organization goes bankrupt.

The vector of financial strategies improving is directed towards the matrix center. If E means a number of ROFS, the challenge is to select strategies located in E space.

Thus implementing decomposition and simulation ROFS model according to the methodology described in the article will allow the corporation to maximize its sales volume when the values of financial risks are given.

T. N. Odintsova
Saratov State Technical University, Russia, Saratov

PROCESS OF SERVICE PRODUCTS CREATION: LOGISTICAL APPROACH

The logistical approach to solving a complex problem, in particular, creation of new service products is considered. The logistical approach is based on increase of service technologies flexibility and allows to adapt them for current changes within demand for rendered services. Application of the present approach requires development of various variants of effective business processes of service in order to avoid eventual “failures” during the process of service product consumption by the consumer. Urgency of the article is caused by increasing demand of the society for qualitative services including tourist services.

Keywords: logistics, service product, tourism services, business process.

The XXI century markets are characterized by impetuous growth of goods and services, shorter product life cycles and growing rates of new product development. Business gets more and more complicated and conditions of its development become more and more uncertain. Usual fast response at the right time is not enough to satisfy needs of such markets. Logistic covers all parts of an enterprise activity and it is really nessesary. The mission of modern logistic is to provide conditions for the necessary products which satisfies the ultimate consumer’s certain needs to be delivered to the right place at the right time. Such problem means, that the ultimate consumer is of the principal importance as there is nothing more important for the entrepreneur than the consumer of his goods or services. The concepts of logistics help to achieve the ultimate goal of any business, namely getting maximum profit due to the client’s needs satisfaction.

The purpose of the article is to consider the process of service products creation from the point of view of complex applying of logistical management.

Nowadays importance of service products logistical management is continuously growing, that is caused by the service sphere development, with the increasing number of companies concentrating there, aiming their activity at the ultimate consumer, developing the concept of service quality total management. It has an effect on the activity of service sphere organizations which have to change their product strategy periodically, to create new kinds of offers to their consumers and also to develop technologies of products planning and creating. The process depends on market factors as firms constantly search for new opportunities to satisfy their consumers of target segments’ needs and try to differentiate characteristics of their offers in the competitive environment.

Due to occurrence of innovational service products (for example, using the Internet for service rendering) firms develop new business processes to render existing services that results in the effective change of service rendering and allows to create new competitive advantages. Transition of tourism enterprises to the system of on-line service using the Internet (booking, selling and communications) can be given as an example. The Internet is used practically in all basic business-processes which take place inside the travel agency, from search and attraction of clients as the communication and marketing tool to the tourist product formation. A more radical form of a service product innovations consists in applying technological achievements to satisfy clients' both potential and latent (unconscious, unexpressed) needs.

Tchernyshev B. defined "a service product" as an independent service or system combining both a material product and accompanying services and having certain utility effect, that means the quality estimated by the consumer [1]. A distinctive feature of a service product, for example, a tourist product, due to its intangibility and ephemeral nature (the client's certain dream or expectation is bought) is the process of service rendering in the form of certain accumulated experience. Physical objects which the client has a right to possess, for example, dinner at the hotel restaurant, are obviously involved into the service rendering. A significant part of the price which the person pays for the service is the value of service elements, including work, experience and use of specialized equipment.

Another important feature of a service product is the simultaneity of its manufacturing, rendering and consumption. The consequence of the service specific feature is that the quality of service products is connected with the character and level of interaction between the service company personnel and its clients, and also of various employees groups among themselves; of other people actively involved into the service process or observing it passively; the personnel abilities at fixed time and sometimes immediately to react their clients' needs while servicing them, and if it is necessary to amend this process.

Service products cannot be stored, however, if demand fluctuations are great, while capacities and technologies of service rendering have the limited network capacity, service companies face serious problems (for example, queuing).

Logistical approaches based on increase of service technologies flexibility and allowing to adapt them to changes occurring in demand for given services help to solve these problems which directly influence competitiveness of service companies. These approaches demand development of various variants of effective business processes of servicing directed to avoid probable "failures" during the process of the service product consumption by the consumer.

The business process is performance of definite work in the companies, for example, the client's order performance (beginning from a phone call up to getting payment and signing the agreement). Every work has some algorithm which should be described, be optimum and really be followed. Business processes are a basis for process-guided management, a complex but effective approach to the company management. It is the ideal administrative tool, which is not only lowering unproductive expenses, but also raising product quality. The tool allows to have full information on the current state of the business process and to accept duly and strategically right decisions. The process approach has made up the basis of quality management standard ISO 9000:2000 where process is understood as "a set of interconnected and cooperating activities which transform inputs into outputs" (item 3.4.1).

Business processes delimitation can be carried out on the basis of the value creation chain described by M. Porter where basic (initial) business processes providing creation of the product consumer value; supporting (auxiliary) business processes providing business functioning and accompanying the product creation along the whole length of its life cycle are distinguished. Solving the problem of processes borders M. Porter assumed, that link borders of a chain and, hence, of business processes are there where addition of the product consumer value is performed.

Examining the nature of service products, it is important to understand how they can be presented evidently to consumers, how to create them. Paying attention to the basic components of marketing mix, a complex of "4P" elements (product / service, promotion (stimulation of sales), price, place and a way of selling), we shall consider the basic components of a service product and the ways of its value and efficiency increase with the help of the logistical approach, that is defining and modelling business processes of the service product creation and offerring.

Logistical and marketing strategies are directed to smoothing and updating of demand fluctuations, their realization saves clients from the necessity to stand in queues, and also allows the firm to balance its capacities work, guaranteeing servicerendering at the appointed time.

Let's determine a place of a logistical component in effective marketing strategy which is realized with use of resources of "4P" elements and puts before itself the purpose of the maximal satisfaction of need of the client. Logistical strategy providing consumers' demand concerning time and the product place and accompanying services is directed to the maximum satisfaction with the service process, i. e. customer servicing.

Service offers usually consist of the basic product surrounded with a set of additional elements of the service. The basic product is responsible for the client's basic need satisfaction, for example, for transportation of the passenger to the certain destination, providing hotel accommodation. Additional services help to achieve the client's maximum satisfaction and raise efficiency of the basic product use. Information and consulting services, documentation providing, services for problems solving and hospitality services refer to this group.

For example, considering the expanded service product with a high degree of contact directed on the person – rendering a hotel accommodation with all additional services – we shall determine its basic components which are necessary to take into account while modelling business processes of creation and offer.

The first component is the basic product which should provide the base advantages connected to problem solving for which sake the consumer addresses the service company. In our case it is providing a hotel accomodation. The second component takes the following elements into account: stages of the process of the basic service rendering, the way and the schedule of the basic product delivery to the consumer, the role and the degree of the consumer's participation in this process, duration and the level of the process efficiency, the style of the given service rendering.

The third component is represented by the group of additional services which accompany the basic product, promoting and facilitating its use, raising its utility and appeal to consumers. Additional services in hotel can be: booking in advance, parking service, check-in and check-out technology, the porter service, meals, use of paid telechannels, room service, etc. Each of these additional elements in its turn demands definition and modelling of business processes variable system of rendering to the consumers and beforehand stated level of service.

The fourth structural component is sequence and duration of the logistical chain operations of rendering services in time and space (parking, check-in, the porter services, room using, room service, meals, etc.). In fact neither the basic product, nor additional services are rendered to the consumers during all the process of servicing.

It is necessary to note, that the account during modelling the basic logistical parameters – time and place – is obligatory and important not only for scheduling the service, but also for definition of the consumers' potential costs connected with the rendered service, and also for correct resources distribution for the company offering services. It is very important to determine optimum time which the consumer should spend to get each element of the service. For example, placing the order, checking-in at the hotel, paying, etc., consumers, as a rule, aspire to minimize the time expenses or they do not want to spend their time for the kinds of activity they consider unproductive at all.

On the basis of the above-stated analysis of a service product structure, it is possible to draw a conclusion, that logistic, being responsible for time and place in all specified structural components, should provide presence of the product or service then and there, when and where it is necessary for consumers. Undoubtedly, that giving time and place utility to products demands significant efforts and expenses. The companies which basic force is the produced innovational service product differ from their competitors by high quality of service. Having extremely flexible logistical chain of deliveries, they are capable to cope with changeable demand.

Creation of new service products is a complex problem, which performance demands careful analysis of many aspects of its activity from the company: processes, personnel and operations, results and advantages. Processes can be displayed with the help of special structural plans where employees' tasks and operation sequence are specified, and consumers' experience at each stage of service rendering is traced.

In most cases while modelling business processes developers consider each process as a discrete unit of a uniform commodity distribution mechanism. As a result there is a problem of compatibility of separately developed logistical business processes and their integration into the general system of the company processes.

Definition of the company logistical business processes and their connections, both among themselves and with other business processes of the company is the first and basic stage of formation of the logistical system of the service company. Carrying out the given problem it is necessary to have enough profound knowledge of the process and system approaches allowing initially, at the development stage to consider each business process as an element of the uniform mechanism.

Modelling of creation and offer business processes of a service product is a complex task. Success of performance appreciably depends on use of the schemes known as logistical chains due to which all processes are displayed with the help of flows, sequences, interrelations and dependences. Logistical chains of the service process allow to reveal the character of interaction between the consumer and employees of the service company, and also show in what way this interaction is supported by "invisible" additional operations and systems. These chains display interrelation between roles which are played by the company employees, operational processes, information technologies and contacts with consumers, with their help it is possible to considerably raise the degree of integration (marketing management, operational management and personnel management) of any service firm.

For example, the first link of a traditional tourist logistical chain is the tour operator and insurance firm (a bank or another credit organization), carrying out financial maintenance of the responsibility of tour operators. This responsibility provides the tour operator's certain guarantees (a bank guarantee for the certain sum of deposit money or another security, insurance of professional responsibility, etc.). In case of any misunderstanding the tour operator is responsible to the tourist, and the insurer provides monetary payments in accordance with "the contract of civil liability insurance for non-performance or inadequate performance of obligations under a contract of a tourist product realization". As a rule, the tour operator does not sell a product directly to the tourist but carries out activities on its formation (makes contracts with hotels, transport companies, etc.) and realization to tourist agencies. Tourist agencies are the next part of the logistical tourist chain between the tour operator and the client.

The tourist agent is the major link in a logistical chain of the process of a tourist product sale which by virtue of its natural properties and the consumer's character should be delivered to each potential consumer. Relations between travel agencies and tour companies are built on the basis of agency contracts, contracts of agency (of commission) and also mixed contracts including elements of contracts of agency (of commission) and paid services. Such documents act as the basic civil-law tool of protection of consumers' rights, i. e. tourists when they travel. That causes necessity of its detailed studying. Summing up, it is a sequence of some actions, fixation of all possible variants of succession of events. These are high-grade models of logistical business processes. If the client is pleased with the conditions offered the procedure of making an agreement is considered completed.

According to managers of travel agencies' experience, the contract of the tourist product realization is attentively read only by every second client or even more seldom. Patrons do not open it at all and sign it actually automatically. Tourists' ignorance of their rights and travel agency's duties results in growing number of conflicts concerning quality and volume of tourist service (claims concern hotels, flight delays, etc.). As lawyers say the great bulk of tourists' claims is caused by discrepancy between the expected and real service product.

According to the law of tourism activity, tourists should be informed by managers of travel agencies about the country entrance rules, local traditions, features and customs, and other features which can affect the voyage quality. The fact of granting the necessary information about the service product is supplying the tourist with the commemorative booklet, he states in the contract that he/she is acquainted with rules of stay in the country. At this stage of the service product creation the client can express a wish to add some points concerning behaviour in another country which are not reflected in the booklet but seem to be significant for him. It has important value for the tour operators forming the tourist product and realizing the strategy of the client's integration into the process of service products creation.

Some travel agencies use imperfection of the legislative base in their own interests, fulfilling their duties improperly. Judiciary practice shows that quite often not the real party in fault is responsible for mistakes, that is the service product supplier, but the most unprotected participant of the logistical business process, that is the travel agency. Development of various models of logistical business processes will help to secure its participants against similar conflicts at the stage of the tourist product creation. It is necessary to take into account, that only contracts specially developed for a certain company, instead of "typical" ones, are reliable protection both for the client and the travel agency, taking the most of opportunities of the current legislation.

Within Russian market the following properties are considered to be prominent features of a tourist product: high cost (a week of economical holidays costs the client his/her monthly salary at least, not including additional charges), limitation of use (the planned trip is impossible to transfer or decline, having returned all the money back), crediting on the part of the client (a trip to visa countries is paid some weeks and even months beforehand). The risks concerning technical and natural accidents, and lately terrorism threat can be added. Therefore dynamics of travel agencies number change is very mobile, and terms of their existence are insignificant.

To describe and develop logistical business processes of a service product creation and offer it is required to determine the basic kinds of activity connected with service rendering, duration of their performance and to reveal interrelations between them.

While developing models of logistical business processes it is necessary to pay special attention to information flows reflecting all industrial and administrative processes of the service company, and also to ways of processing and the analysing information.

To stimulate ideas development while modelling business processes of service products creation and offer it is possible to use the following principles:

– the client whenever possible should participate in modelling process him/herself (for example, in parameters monitoring). Involving him into the process of a service product creation, the company, thus, gets, maximum information concerning his preferences;

– relations with suppliers should be partner ones and constructed on the principles of "synergy" (for example, development of common service standards);

– creation the various variants of business processes which are taking into account every possible "script" of the service process (for example, a service product formation in view of the client's individual inquiries, etc.);

– concentration of information interchange with decentralization of its participant divisions (a logistical chain: a tour operator – a travel agent – an airline – a hotel – a client, etc.) [2].

Using the logistical approach to modelling business processes of service products creation and offer also enables managers to determine potential "bottlenecks" in service process, i. e. the moments which are fraught with great risk that the process will be a failure and the quality of service will go down. On the basis of this knowledge managers can develop business processes models, allowing to avoid such negative moments. Besides in a logistical chain operations of the service process which, as a rule, are accompanied by queueing are determined. To eliminate delays while servicing there are standards being developed and implemented for each kind of activity of the company employees including time, assigned to perform each operation, consumers′ maximum latency between their performance, and job descriptions regulating interaction between the company employees and consumers.

Starting modelling of business processes of service products creation and offer it is necessary to present their role in the general number of the company business - models. Generally, the problem of the logistical approach use to the structural description of business processes can pursue the following purposes:

– strategic analysis of the company processes organization with the purpose to model interaction of its divisions among themselves and with their contractors;

– organization and optimization of the logistical chains mentioned before. However not many companies presume to have a full description of their activity with such a degree of detailed elaboration which is necessary for management. Therefore, the "key" processes of the company affecting its competitiveness most strongly should be chosen as objects for formal description. It is necessary to start revealing such processes with the help of the typical technique of logistical analysis, namely methods ABC and XYZ, etc.);

– establishment of the quality system also connected with achievement of the enterprise strategic purposes as one of requirements of quality standard ISO-9000. It is recommended to describe the company activity as a set of business processes that allows to achieve its best transparency concerning “a guaranteed quality level of product development, manufacturing and delivery”;

– formal description of business processes is a necessary condition for their computerization. It is necessary to present each of processes – resources, documents, executors, actions, branching conditions, etc. even more precisely.

Thus, detailed and consistent analysis of logistical business processes allows to reveal operations, where expenses can be reduced due to automation, increase of labour productivity, rating, and in some cases due to exception of the given operations or transfering them for outsourcing. However, it is necessary to note, that desire and maturity of the company management is not enough. In this case it is important to have logistical operators in the region, which are capable to suggest a similar sort of service that is the biggest restriction of logistical outsourcing.

After the detailed description of the company business processes including logistical ones, there begins modelling and creating of the integrated logistical system of the whole company where all business processes are interconnected and optimized.

The necessity of such a system is caused by the reason that in modern conditions not only optimum use of available resources and increase of labour productivity are necessary and important, but also a high degree of managing that is expressed in flexibility and speed of reaction to changes in the external business situation being oriented on constant and active interaction with service products consumers (CRM-technology, call-centers and etc.).

References

1. Tchernyshev B. Management in Service Economy: Essence and Content // Problems of the Theory and Practice of Management. 2004. № 1.

2. Ilyin V. V. Modelling of Business- Processes. Experience of the Developer. M. : Williams, 2006.

A. A. Parkhomenko
Krasnoyarsk State Pedagogical University named after V. P. Astafev, Russia, Krasnoyarsk

NECESSITY AND CHELLENGES OF THE RUSSIAN FINANCIAL SECTOR MODERNIZATION

In the article the possibilities of the Russian financial sector modernization for entering the sustainable development trajectory are considered. It touches upon the issues of the division of the Bank of Russia into two functions – a monetary regulator and a body responsible for the creation of a macroprudential supervision system. The questions of the state-owned banks share growth in the Russian banking system, the establishment of the adequate minimum level of capital and the creation of a three-level banking system are examined.

Keywords: Bank of Russia, banking system, macroprudential supervision, state-owned banks, consolidation, the minimum level of capital.

The global financial crisis is coming to an end and the question of choosing directions of the Russian economy development is topical again. It is necessary to admit, as one of the major lessons from the crisis, that the growth of gross economic indicators and qualitative economic growth based on increased competitiveness do not always coincide. If the current financial policy, focused on favorable external economic activities, continues, the solution of the structural problems, accumulated in the system, will be set aside for future where this solution will cost significantly more.

Currently there are all prerequisites for the development and realization of the strategy for entering the sustainable development trajectory both of financial and real economy sectors. Thus the combination of rapid growth and the achievement and preservation of its stability and resilience seems the most rational.

The structural changes in the financial sector, especially in the banking system, where the changes should begin with the Central Bank of the Russian Federation (Bank of Russia), should become the major element of this strategy.

One of the main the Bank of Russia long-term objectives is the maintenance of low, stable and predictable inflation rates, which should be close to the rates of our closest neighbours – the European countries. Currently, however, the activities of the Bank of Russia are under constant pressure of various interested parties. So its policy significantly depends on the vision of general economic situation by the RF Government, which independently sets the targets of the monetary policy which the Bank of Russia is responsible for. Setting these targets the Government ignores the influence of regulated inflation rates. The absence of the necessary hierarchy of the macroeconomic policy objectives, especially the absence of the rate fiscal policy subordination to the priority of the inflation aim, leads to low efficiency of monetary policy. Increasing the actual, rather than nominal, independence of the Bank of Russia can only occur as a result of successive government non-interference in the management of ruble rates and inflation.

Among the causes of excessive control of the Bank of Russia, it is necessary to mention the performance of a large number of functions that are peculiar for an executive body. It arouses the desire of state leaders to tighten control over its operations and, consequently, leads to weakening of its functional autonomy.

It is necessary to note that there is a conflict of interest within the Central Bank, for example, if for the anti-inflation actions, it is necessary to raise interest rates, that, however, could negatively affect the financial conditions of some commercial banks, that the Bank of Russia has a regulative liability to, the Central Bank must either reduce regulations or refuse to suppress the inflation rates.

Therefore, none of the financial regulation participants are interested in the independence of the Bank of Russia. The Government aims to subordinate it, as it performs the functions that are in the executive authority field of responsibility, and the Bank of Russia does not strive for its independence because it is bound by the regulation of the banking market.

This problem can be radically solved by the division of the Bank of Russia into two parts, the first will inherit the functions of monetary regulation and focus on macroeconomic problems.

The second will focus on the functions of regulation and supervision of banking sector, and subsequently on its bases, the creation of the mega-regulation and mega-supervision body of the whole financial market is possible (tab. 1).

The integration process, actively taking place in the financial sector, the consolidation of financial institutions with the formation of financial conglomerates and holding companies on the basis of major commercial banks requires the concentration of supervision functions over financial institutions of various types and in various sectors of the financial market under a single supervisory body. The Bank of Russia supervision system has developed infrastructure and highly qualified staff, so the creation of mega-regulator based on it seems most appropriate. In future other bodies involved in the regulation of financial markets can be joined to it.

The area of financial regulation and supervision also requires changes. The operating system of macroprudential supervision focuses on the solvency and stability of individual banks. However, the global crisis has revealed its weaknesses, having shown the need for systematic work with the financial sector as a whole, including the regulation of markets and operations, which reflect the relationships between banks and between banks and other participants of the financial system.

Macroprudential supervision should concentrate on the system stability of the financial sector, rather than on preventing the insolvency of individual banks (tab. 2) [1]. This approach pays special attention to backbone institutions and relationships in the financial sector, because the risk of system disbalance depends on the collective behavior of financial markets participants, as a result of this, risks in the financial system gain endogenous nature for the regulator.

Nowadays macroprudential supervision denotes the risk assessment for the entire system, which is not reduced to the summation of individual risks. The main tool of macroprudential supervision is the system of assessment of the indicators of bank soundness (sensitivity to risks, financial leverage, liquidity and various characteristics of savings market etc.). Potentially, it should become an effective mechanism for assessing the probability of failures in the financial market and probability of crises.

For the present no operating system of a global systemic risk assessment of has been developed. The assessments for individual countries based on macroeconomic stress tests under the Financial Sector Assessment Program (FSAP2), carried out by the IMF and the World Bank together with the national monetary controllers can be considered the closest analogues. The other directions of work in this field are the creation of advanced systems of crisis indicators (mainly macroeconomic) and the creation of the valuations on the VAR model (value-at-risk) for the financial sector as a whole [2]. Thus, within the European Union, a new oversight body is being created, whose goal will be to monitor systemic risk – European Systemic Risk Council (ESRC).

Table 1

Distribution of the functions of the Bank of Russia

Bank of Russia, as the monetary regulator	Bank of Russia, as the supervisory authority in the banking sector
Development and implementation of unique state monetary policy (in collaboration with the RF Government). Currency issuing and organization of cash circulation. Organization of the system of refinancing of credit institutions as a lender of an ultimate authority. Management of gold and foreign currency reserves. Drawing up of the RF balance of payments. Analyzing and forecasting the state of the Russian economy as a whole and by regions, primarily monetary, currency and financial, and price relationships	Laying down the rules of making payments and conducting banking transactions in the Russian Federation. Decision-making on state registration of lending institutions, and licensing banking operations of lending institutions. Supervision of lending institutions and banking groups. Registration of securities issuance by lending institutions. Currency regulation and currency control. Laying down the rules of accounting and reporting for the banking system of the Russian Federation

Table 2

Comparative analysis of micro-and macroprudential approaches to supervision in a financial sector

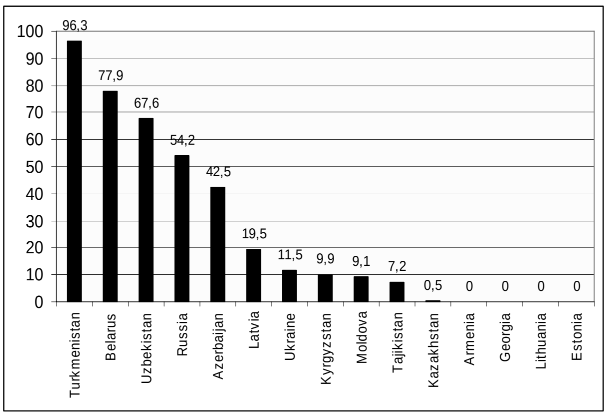
Criteria	Macroapproach	Microapproach
Short-term objective	Limitation of negative processes involving the whole financial system	Limitation of negative processes at the level of individual institutions
Long-term objective	Prevention of a GDP slowdown	Consumer protection (investors (depositors))
Model describing a risk	Endogenous (partly)	Exogenous
Interdependence between institutions	High	No evidence
Choice of prudential control measures	On the basis of negative processes assessment in a financial system (top-down)	On the basis of risks assessment faced by individual institutions (bottom-up)

Thus, we can conclude that the international community focuses on the monitoring systemic risks and the prevention of their transfer between different financial market segments as well as in the real economy. Unfortunately, in Russia the situation is fundamentally different: there are no studies on systemic risks, no monitoring systems, no special bodies dealing with similar problems. This situation calls for urgent changes.

Another significant feature of the Russian banking sector is the dominating state ownership in it. This happens due to several factors: the state-owned banks have an advantage in servicing the financial needs of the state and state-owned corporations, significant amounts of budgetary resources are placed in them, direct public investments are placed in these banks, they are provided with administrative support from the authorities and have already formed an image of stability and reliability.

As a result, over the past decade, the size of state ownership of the banking sector assets has been growing steadily. Thus in 2001 the state and quasi-state (private but state-supervised) banks accounted for about 36.3 % of total bank assets and in 2009 their share increased to 54.2 [3].

Today Russia is included into the small group of countries with defined state ownership in the banking sector. Among the post-Soviet countries state-owned banks dominate only in those countries, where market reforms are not developed properly – Uzbekistan, Belarus and Turkmenistan (sce figure) [4].



State ownership of banking assets in the Former Soviet Union countries, %

In the situation when the share of state sector exceeds 50 %, quantitative changes have turned into qualitative and elements of distributive economy have appeared in the banking market. For example, each of the state-owned banks has a task to increase its lending portfolio for the 2nd quarter of 2009, attempts to control prices are taking place (the upper limit on loans and private deposits) saying nothing of public personal control of the state leaders. At the same time the private sector's share shrinks steadily from year to year, banking market is concentrated on one pole: four banks controlled directly by the state (Sberbank, VTB, Russian Agricultural Bank and the Bank of Moscow) hold 81.7 % of assets among banks of this category and 42.9 % of total bank assets of the country [3].

It should be noted that the nationalization of the banking sector has its advantages. The executive authority can count on banking sector highly resistant to financial shocks; banks with state interest can be forced to fulfill national tasks; expansion of foreign banks is complicated because of multiple preferences for state-owned banks; the Bank of Russia is exposed to minimal credit risks as state have made the third extralegal intermediate level of the banking system, through them primary resources come into the economy and to other market participants.

However, the use of the authorities of the state-owned banks as financial agents raises a number of negative effects. If private banks are concerned about profit maximization or capitalization, the state-owned banks do not have a clearly defined objective function. They carry full activities of commercial organizations that have received preferences from the state and using non-market competitive advantage because they have the opportunity to use both financial and administrative resources of the state. Targets of state-owned banks lead to reduction or even disappearance of competition in some areas and slow the development of financial innovations. In some cases, banks with state interest set prices for financial services that significantly deviate from market prices. This may result in unreasonable prices for their services for customers "linked" to state banks, and in the underpricing to displace competitors from the market.

Alongside the growth of state-owned banks monetary policy becomes less able to affect the real economy and the Bank of Russia loses the ability to affect market conditions by market operations. As a result, the elements of command economy appear in the market.

Despite the obvious negative aspects of state banks activities in the Russian banking sector, a radical change of ownership structure is unlikely.

The main reason is the lack of potential alternative owners. The crises that the banking system has experienced in the recent decade led to private banks market exit because their owners could not restore the negative equity capital of their banks or believed that the banking business has no prospects. Only those state-owned banks remained in the market whose capital is replenished by the State in the required volume. In addition, the long-term banking activities in the Russian market are not sufficiently attractive for investments because the high inflation and regular banking crises lead to the fact that the real profitability of the business is close to zero. Thus, during the years 1998–2009 profits of banks in real terms were less than 3 % of the capital [5]. And one more important factor – there is a shortage of savings on the domestic market so the bulk of investments may be attracted only from abroad. But in this case the most likely scenario is the loss of national control over a significant part of the banking sector.

Thus, in the near future the change of ownership in the largest Russian banks doesn't seem possible, although we can identify a number of measures aiming to its achievement in the long term.

Firstly, it is necessary to consider options for partial privatization of the major state banks to diversify the structure of their property.

Secondly, it is necessary to limit unwarranted expansion of the state-owned banks due to budgetary and administrative resources leading to distortion of competition principles.

Thirdly, it is necessary to optimize the state ownership in the banking sector completely eliminating the state's shares in small banks focusing on the interests in banks that are institutions for development.

The next element that should be reflected in the development strategy of the banking sector is its consolidation. At present Russia takes the third place in the world by the quantity of banks (1 056 as of 02.01.1910 based on data of cbr.ru). The largest quantity of banks is in the U. S. – more than 6 thousand, Germany takes the second position, where there are more than 2 thousand banks, more than a half of them are cooperative. There is no doubt that with the growth of the territory the country needs more banks, while the first two hundred Russian banks account for about 94 % of the total assets of the sector, the share of other players is small and constantly declining.

It should also be noted that the degree of concentration in the Russian banking sector is below average. The market share of five leading banks in Russia reaches 46 %, whereas in Europe it swings around 60 % (from 22 % in Germany to 96 % in Estonia) [6]. Trends in world markets show that the increase in concentration reflects the natural development of the market and the largest Russian banks will become more powerful and bigger. The main element of the stimulation of this process is the toughening of the requirements to banks capital.

It should be noted that the toughening of the licensing terms has both positive and negative effects. The positive effects are:

– Reduction of lending rates – the bigger is the bank, the lower the average interest expense on its liabilities and the lower the interest rates for borrowers from the real sector are;

– Increase of money supply by the growth of credit multiplier – small banks have to maintain high liquidity because of a non-diversified customer base so bank resources are used not very effectively; consolidation of the sector will ensure an increase in money supply as a result of the expansion of lending;

– Consolidation of banks – today the average Russian bank is a modest-sized organization, unable to meet the needs of large enterprises, therefore consolidation stimulation will allow to create banks required by the economy;

– Increasing of the efficiency of financial resources allocation – the informational advantage of the major players, capable to monitor a large part of the market, is reflected in lower information asymmetry and more efficient allocation of resources.

Meanwhile the following negative effects of banking capital consolidation should be noted:

– Reduction of market competition – reduction of the number of players in the market can lead to influence growth of large banks that will allow them to set low interest rates for deposits and overstated rates for loans;

– Reduction in lending supply due to rationing of credit growth – the desire to minimize risks at large banks could lead to cutting off more risky borrowers from lending, that in the long run will lead to a decrease in innovative activities and in economic growth rates;

– Destabilization of the regional economic systems in times of crisis – in periods of crises local branches of Moscow based banks transfer assets in their head offices for accumulation of liquidity, repayment of foreign debts, and other purposes and the regional economic systems are vulnerable to potential shocks because of absence of local banks.

Another argument against excessive consolidation of the banking market is the risk of systemic instability. Large banks are more profitable, their assets are well diversified, and it is easier to supervise them. In theory this should improve the resistance of major banks to shocks and make a financial system more stable. However practice shows that the state considers major banks as "too big to fail" and supports them in times of instability. In turn, the large banks, relying on state aid, become more risky so more unstable. The experience of the recent banking crisis has shown that the threats of systemic instability originate mainly from large banks and the higher the number of players in the sector is, the lower the costs of its reorganization and restructuring are.

The problem of an adequate amount of minimum capital of banks in Russia is actively discussed in the banking community. For example, in November 2009 the Vice-president of the government – Finance Minister Alexei Kudrin said he was ready to initiate a bill to increase minimum capital of Russian banks for five years to 1 billion rubles. Indeed, one of the major contradictions between the financial and real sectors in Russia is a small size of Russian banks. Their weak financial capacity explains why the real sector turned to foreign funding and the level of debts to non-residents is currently comparable to domestic corporate lending.

However, a large group of local banks numbering several hundreds of organizations will not be able to augment equity capital neither to a level corresponding to the financial needs of the leading Russian companies nor to the level designated by the Government of the RF. At the same time this group of banks could help to reduce the impact of the above-mentioned negative effects of consolidation. The best way to save these banks is their isolation in a separate class that will be under special prudential supervision depending on the risk profile.

In most countries where the banking sector has hundreds of players, there are several categories of banks to which different requirements, depending on specificity of activities, are applied. Taking into consideration the available international experience, it is necessary not to force small banks to close or consolidate and let them choose their niche and continue to work in it.

As a result of this offer, multilevel banking system will be created in Russia. The Bank of Russia will be the first level of the national banking system, the second level will be federal banks with general license and a large capital of their own (e. g. from 100 million Euros). They will carry out the whole range of banking operations, operate throughout the country, and have access to foreign financial markets. The third level of the system will be represented by separate groups of banks working at the level of federal districts, federal subjects and cities. Their licenses will include restrictions on the minimum equity capital, the territory of operation (on which the bank may open branches) and the list of banking operations.

In conclusion, it should be noted that, without structural reforms, it is difficult or almost impossible for the Russian financial sector to assist sustained development of the real economy and to resist external shocks. Only the appropriate modernization of the banking system of Russia is able to assist a more sustainable economic growth and strengthening of the competitive position of Russia in the global economy.

References

1. Sarkisyants A. Problems of Banking Supervision // Accounting and Banks. 2008. № 6 [Electronic resource]. URL: www.bankir.ru/publication/regulation/1379862.

2. Kreyndel V. Who will measure the systemic risk? // Institute for Financial Studies. 31.07.2009. [Electronic resource]. URL: www.ifs.ru/upload/systemicrisk_31july.pdf.

3. Vernikov A. Share of state interest in the banking system of Russia // Cash and credit. 2009. № 11.

4. The transition process and indicators of CIS countries and Mongolia in 2009. The transition process in the crisis : Report of the Economic Department of the EBRD under the general direction of E. Berglof. [Electronic resource]. URL: www.ebrd.com.

5. Moiseyev S. Prospects of privatization in the banking sector // Analytical Banking Magazine. 2010. № 1 (176).

6. State and banks: risks cycling // Help to business. 23.06.09. [Electronic resource]. URL: www.bishelp.ru / gde_dengi / detail.php? ID = 80372.

D. V. Parshukov, G. V. Mironov
Krasnoyarsk State Agrarian University, Russia, Krasnoyarsk

A MODEL TO ASSESS THE RISK OF BANKRUPTCY FOR AGRICULTURAL FIRMS IN KRASNOYARSK REGION

In this paper we report on the algorithm of development of a bankruptcy risk assessment model to be applied to agricultural firms of Krasnoyarsk region, which involves factorial and discriminant analysis of relevant data.

Keywords: factors, discriminant functions, tree-like hierarchy, aggregation, membership functions.

The global financial crisis and as a consequence the instability in financial markets have caused a drastic increase in the number of firms going out of business on the background of the overall economic downturn. In this context, an early recognition of pending problems is important for ensuring continuity of one's business. In connection to this there is a necessity to work out an effective model to assess the risk of bankruptcy, which would allow to predict potential distress situations in Russian companies. The purpose of the present work is to construct such a model of bankruptcy risk assessment for agricultural firms of Krasnoyarsk region.

The structure of the model consists of a number of consecutive steps:

Step 1. To select a set of significant financial ratios for further analysis, to define classes of financial condition, put together linguistic characteristics.

Step 2. To reduce the dimensionality of the selected set of factors by applying the method of principal component analysis and to construct factors hierarchy.

Step 3. To derive discriminant functions for the principal components having been identified in the second step mentioned above.

Step 4. To produce an aggregate matrix for level recognizing on a standard 01-qualifier.

Step 5. To perform hierarchy nodes convolution and assign the firm to one of the classes defined in Step 1 mentioned above.

During the first step, we define three groups of firms as follows: financially sound (Class 3), financially unstable (Class 2), and financially distressed (Class 1) firms. Next we create a set of data containing twenty three various financial indicators, such as profitability ratios, solvency indicators, and business activity indicators, which describe different aspects of financial standing of the agricultural firms [1]. The model was constructed according to data for 2006, 2007. Model check was made according to data for 2008.

Let us define the so-called linguistic variable, "Factor level" [2], with the term set of L to have the form:

$$L = \{\text{Low level } (L), \text{ Average level } (A), \text{ High level } (H)\}, \qquad (1)$$

Next we introduce F_0 as a criterion of financial soundness and solvency. The linguistic variable provides a qualitative description of the firm's condition related to F_0. We set the following meaning for the linguistic characteristics:

– low F_0 means the firm is in a critical financial condition and falls into Class 1 enterprises;

– average F_0 means the firm is financially unstable and fits into Class 2 enterprises;

– high F_0 means the enterprise is financially sound and qualifies for Class 1 enterprises.

If the company relates to financially distressed enterprises it indicates a high risk for bankruptcy. Likewise. Financially unstable firms have an average risk for default and financially sound enterprises indicate a low risk for bankruptcy.

The second step is to identify the factors based on factor analysis (after a preliminary analysis for multicollinearity). These factors give the largest contribution into dispersion of the resultant indicator F_0, which describes a probability of bankruptcy through linguistic characteristics. The algorithm of factor analysis can be found in [2].

The following financial ratios are incorporated in the factor analysis:

Factor 1 – $k1$ – inventory coverage ratio; $k2$ – circulating assets in fill rate; $k3$ – economic efficiency of operating assets; $k4$ – a part of working capital in circulating assets; $k5$ – a part of fill rate in operating assets.

The first factor includes indicators showing how effectively the firm manages its assets and inventories. Current assets coverage indicator can be adduced as a short characteristic; $k7$ – financial dependence index; $k8$ – equity flexibility ratio; $k9$ – payback term for equity of the investment.

Factor 2 includes indicators which characterize the use of the firm equity capital.

Factor 3 – $k10$ – equity capital ratio; $k11$ – loan capital ratio, $k12$ – return of assets pricing; $k13$ – common production profitability.

The third factor measures profitability of the firm derived from equity and loan capital indicators. It denotes a profitability factor.

Factor 4 – $k14$ – cash ratio; $k15$ – quick ratio; $k16$ – current ratio; and $k17$ – return on production assets.

The factor incorporates indicators of liquidity and the earning capacity of the production, therefore it can be called a solvency indicator.

To define the significance of the factors, let us consider tab. 1.

Table 1

Total Dispersion

Factors	Initial Eigenvalues		
	Total	Variance (П), %	Cumulative, %
1	6.676	29.026	29.026
2	4.006	17.417	46.444
3	3.178	13.818	60.261
4	1.833	7.967	68.229

Let the sign «≈» denotes the indifference between two factors, i.e. the two factors are equally significant, and the sign «$\succ$» denotes the preference of one factor to another, i.e. one of the factors is more significant than the other for the root element of the hierarchy (F_0). The set of symbols and factors forms a system of relative preference. In this case this system looks as follows:

$$\Phi = \{F_1 \succ F_2 \approx F_3 \succ F_4\}. \qquad (2)$$

It is based on the results of factor analysis. Relative contribution of individual factors to the total dispersion of characteristics [3] is compared:

$F_i \succ F_{i+1}$ if $\Pi_i > \Pi_{i+1}$ is more than on 10 %
and $F_i \approx F_{i+1}$ if $\Pi_i > \Pi_{i+1}$ is less than 10 %,

where Π_i is the percentage of dispersion due to a particular individual factor (i. e. the contribution of that factor to the total dispersion), the subscript i denoting the factor number (thus Π_1 is the percentage of dispersion due to the first factor).

Let us build a tree-like hierarchy of the given factors (fig. 1).

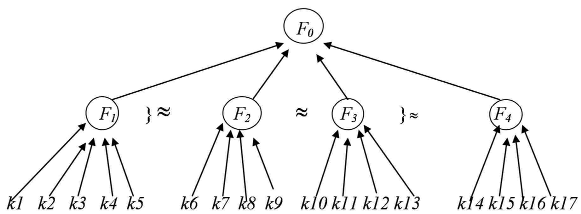


Fig. 1. Tree-like hierarchy of factors

In our case being based on system (2) and the Fishburn method [3] we have the following weights $(\frac{3}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8})$ for factors F_1, F_2, F_3, F_4, accordingly.

Now we have to categorize the firm belonging to classes mentioned above using each factor

(F_1, F_2, F_3, F_4). We attempt to assign the firm to one of the three groups of financial viability by using each factor, based on the indicators selected as described above. This will be the third step in developing the model. Using the indicators determining each factor we reciprocate the functions with the best predictive ability. In this work it is a linear discriminant function. To reciprocate this function (3)–(6), we perform a discriminant analysis (tab. 2). In accordance with its results we have the following functions:

$$F_1 = -0.14 \cdot k1 - 1.055 \cdot k2 + 0.441 \cdot k3 + \\ + 1.534 \cdot k4 - 1.667 \cdot k5 + 2.462, \quad (3)$$

$$F_2 = -0.713 \cdot k6 + 0.738 \cdot k7 - \\ - 0.88 \cdot k8 + 1.658 \cdot k9 - 0.08, \quad (4)$$

$$F_3 = -0.063 \cdot k10 - 0.139 \cdot k11 + \\ + 0.912 \cdot k12 + 2.044 \cdot k13 + 0.802, \quad (5)$$

$$F_4 = 0.071 \cdot k14 - 0.008 \cdot k15 + \\ + 0.462 \cdot k16 + 3.339 \cdot k17 - 1.014. \quad (6)$$

According to the correlation index (> 0.5 for all factors) we can that correlation is satisfactory. Significance p is less than 0.001 for all functions, which implies that the mean values of each of the functions are significantly different for various classes. High eigenvalues (more than one) indicate a good (appropriate) choice of discriminant functions (tab. 3).

Let us calculate the average value between the centroids for the functions of factors and we can establish the intervals of the firms belonging to each class (tab. 4).

Table 2

Statistical calculations for discriminant functions

Function	Eigen value	Variance, %	Cumulative, %	Canonical Correlation	Wilks' Lambda	Chi-square	Sig. p
F_1	1.431	94.8	94.8	0.767	0.382	32.758	.000
F_2	3.316	95.6	95.6	0.877	0.201	55.403	.000
F_3	2.022	89.9	89.9	0.818	0.269	45.249	.000
F_4	3.463	96.7	96.7	0.881	0.200	55.467	.000

Table 3

Group Centroids Functions

Group	Function in Group Centroids			
	F_1	F_2	F_3	F_4
Class 3	1.409	2.324	1.856	2.641
Class 2	.333	–2.211	.014	–.725
Class 1	–1.286	–.223	–1.402	–1.482

Table 4

Intervals of belonging in accordance with discriminant functions of factors

Value range	Parameter level	Risk of bankruptcy by each of the factors
For F_1:		
$F_1 < -0.4765$	Class 1	High
$-0.4765 < F_1 < 0.871$	Class 2	Average
$0.871 < F_1$	Class 3	Low
For F_2:		
$-1.217 < F_2 < 1.0505$	Class 1	High
$F_2 < -1.217$	Class 2	Average
$1.0505 < F_2$	Class 3	Low
For F_3:		
$F_3 < -0.694$	Class 1	High
$-0.694 < F_3 < 0.935$	Class 2	Average
$0.935 < F_3$	Class 3	Low
For F_4:		
$F_4 < -1.1035$	Class 1	High
$-1,1035 < F_4 < 0,958$	Class 2	Average
$0.958 < F_1$	Class 3	Low

Having established the intervals of firm belonging to each class and having calculated the functions for their quantitative estimation we can make a convolution using the hierarchy stages. For the linguistic variable "Factor level" with the L term-set given by (1) and the hierarchy of factors we use the conventional three-level 01-classifier (SFC) [3] which acts as a group of functions of the firm class belonging, where these functions are trapezoidal triangular numbers (fig. 2):

$$\mu_1(x) = \begin{cases} 1.0 \le x \le 0.2 \\ 5(0.4 - x), 0.2 \le x \le 0.4 \\ 1, 0.4 \le x \le 1 \end{cases} \quad (7)$$

$$\mu_2(x) = \begin{cases} 0,\ 0 \le x \le 0,2 \\ 5(x - 0.2),\ 0.2 \le x \le 0.4 \\ 1, 0.4 \le x \le 0.6 \\ 5(0.8 - x),\ 0.6 \le x \le 0.8 \\ 0,\ 0.8 \le x \le 1 \end{cases} \quad (8)$$

$$\mu_3(x) = \begin{cases} 0,\ 0 \le x \le 0,6 \\ 5(x - 0.6),\ 0.6 \le x \le 0.8 \\ 1,\ 0.8 \le x \le 1 \end{cases} \quad (9)$$

Let it be that $F_0 = x$, and x = a 01-carrier in (6) (the [0,1] segment of real line).

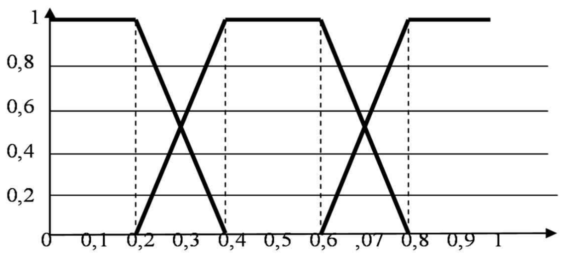


Fig. 2. System of trapezoidal belonging functions on the 01-carrier

The standard classifier projects the fuzzy linguistic variable onto the 01-carrier, and does so in a consistent manner, producing a pattern of symmetrically distributed classification stages (0.1; 0.5; 0.9). In these stages, the value of one particular membership function is equal to 1 (one) while all other functions are zero. The analyst's uncertainty about correctness of classification decreases or increases linearly with the distance from the stages, the sum of membership functions being equal to 1 (one) in all points across the carrier.

We find F_0 by means of matrix convolution:

$$F_0 = PMV = \left(p_1\ p_2 \ldots p_n\right) \cdot \begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ . & . & . \\ \gamma_{1n} & \gamma_{2n} & \gamma_{mn} \end{pmatrix} \cdot \begin{pmatrix} 0.1 \\ 0.5 \\ 0.9 \end{pmatrix}, \quad (10)$$

where P is the coefficient vector of the factors $(\frac{3}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8})$; $\begin{pmatrix} 0.1 \\ 0.5 \\ 0.9 \end{pmatrix}$ is the vector of vertices of trapezoidal numbers SFC; M is the belonging matrix, γ is the belonging of the firm to one of the classes by means of each particular factor (3)–(6). For example if by F_1 (3) the enterprise belongs to Class 1, the element γ_{11} is equal to 1 while the rest of the elements in the first line are zero. The same principle applies to other lines of the matrix.

Let us compare the estimated value with the tabular data and evaluate a probability of bankruptcy in terms of linguistic characteristic (tab. 5).

Table 5

SFC-based classification of the level of F_0

F_0 limit	Parameter level	Degree of evaluation certainty (membership function), magnitude ·100 %
$0 \le F_0 \le 0.2$	Low	1
$0.2 < F_0 < 0.4$	Low	$\mu_1 = 5 \cdot (0.4 - F_0)$
	Average	$1 - \mu_1 = \mu_2$
$0.4 \le F_0 \le 0.6$	Average	1
$0.6 < F_0 < 0.8$	Average	$\mu_2 = 5 \cdot (0.8 - F_0)$
	High	$1 - \mu_2 = \mu_3$
$0.8 \le F_0 \le 1$	High	1

The final result is a linguistic description of the degree of probability of bankruptcy as well as the degree of the analyst's certainty as to the correctness of recognition. Therefore the conclusion about the risk degree not only has a linguistic form but also contains a characteristic of the assertions quality.

All in all we have finished the development of the model for assessing the risk of firm default. Let us illustrate the model by evaluating the factors for the agricultural firm GPKK "Borodinskoye", Rybinsky Region. While this company was included into the group of financially unstable enterprises, its data were not used in the developing of the model. In our calculations we use formula (2) and the data from tab. 6.

According to (3)–(6) we have: F_1 = –0,5 – Class 1; F_2 = –1,6 – Class 2; F_3 = –0,22 – Class 2; F_4 = –0,32 – Class 2;

$$F_0 = PMV = \begin{pmatrix} \frac{3}{8} & \frac{1}{4} & \frac{1}{4} & \frac{1}{8} \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0.1 \\ 0.5 \\ 0.9 \end{pmatrix} = 0.35.$$

According to tab. 5 the company belongs to Class 1 with the probability 5 (0.4–0.35) = 0,25 and to Class 2 with the probability 1–0.25 = 0.75. The company does not have sufficient coverage for floating assets and fill rate (by factor F_1 = –. 05 it falls into Class 1), the company's equity is inefficient, and it exhibits low profitability and solvency indicators (by factors F_1, F_3, and F_4 it belongs to Class 2). The risk of bankruptcy in the nearest perspective is assessed as average.

Table 6

Balance Sheet Data

Item	Thousand roubles	Item	Thousand roubles	Item	Thousand roubles	Item	Thousand roubles
Earnings	16 028	Inventory	12 552	Balance value	36 937	Long-term liabilities	2 023
Cost price	17 686	Long-term receivables	1 967	Charter capital	100	Loans and credits	6298
Profit before tax	–1 658	Short-term receivables	514	Added capital	45 606	Accounts payable	36 110
Fixed assets	20 132	Cash	100	Net profit	–53 200	Current liabilities	42 408
Non-circulating assets	20 132	Floating assets	16 508	Capital and reserves	–7 494	Liabilities	36 937

Based on financial information provided by Agricultural Agency of Krasnoyarsk region, we have calculated the value of F_0 for ten different agricultural firms in each of the classes and compared the results with their initial classification. Full matching (i. e. when probability = 1) of our findings and conclusions with the original classification was 82.5 %. We observed any case when after the analysis a company originally classified as a financially distressed class 1was transferred to Class 3 (financially stable firms). It proves that the model is adequate and appropriate for assessing the risk of bankruptcy.

In addition to conventional methods, the proposed model of bankruptcy risk assessment can be an effective tool in evaluating financial position of a company, that can enable company's management to continuously monitor the financial situation in the company for the risk of default. It is never late to mitigate the risks with the development of a package of measures particular important in the unstable conditions of economic environment [4–6].

References

1. Gilyarovskaya L. T. Economic analysis : Textbook for high schools / ed. by L. T. Gilyarovskaya. M. : UNITI-Dana, 2001.
2. Dubrov A. M., Troshin L. I., Mhitanyan V. S. Multivariate statistic methods : Textbook. M. : Finance and statistics, 2000.
3. Nedosekin A. O. Integrated assessment of the corporate bankruptcy risk the basis of fuzzy descriptions. SPb. : Sesam, 2002 [Electronic resource]. URL: http://sedok.narod.ru/sc_group.html.
4. Davydova G. V, Belikov A. Yu. A quantitave method to assess the bankruptcy risk of enterprises // Risk management. 1999. № 3. P. 13-20
5. Loyko V. I., Yefanova N. V. Quantitative models and techniques of assessment of risks in the agroindustrial integrated systems // The scientific journal of KSU. 2008. № 40 (6) [Electronic resource]. URL: http://ej.kubagro.ru/2008/06/pdf/12.pdf.
6. Fuzzy Logic: An Overview of the Latest Control Methodology. Application Report. SPRA028. Digital Signal Processing Products. 2000.

O. V. Patsuk
Siberian State Aerospace University named after academician M. F. Reshetnev, Russia, Krasnoyarsk

DESIGN OF THE REGIONS' ECONOMIC DEVELOPMET STRATEGY

At present Russian economy is undergoing the similar trends of the modern economic development to those taking place in developed countries, those are globalization, advanced development of the service industries, post-industrial society formation, intellectual component expansion in the outcomes of any industry, informatization of the society, exhaustion of traditional sources of social and economic growth. In such conditions search for new ways and factors of the regional self-development is critical.

Keywords: social and economic development of the territories, regional economy.

In recent years regions in Russia are becoming more independent. They are more responsible now for the results of their economic development. Their social and economic progress is determined by the objective factors (macroeconomic conditions, region's position in the social division of labor, production structure, geographical location, natural resources) and subjective ones, which are in the first place methods of regional management. Economic reforms have shown that regions, using advanced methods of management are less influenced by crisis tendencies.

Strategy of the state regional development is not uniform in different regions. This is caused by their significant differences in natural resources, economic structure and level of economic development. That is why this strategy sets general goals and objectives for a certain period of time and becomes a base for mutually balanced strategies of the region's development design. These strategies are a set of measures aimed at implementation of long-term goals of the country's social and economic development. At the same time it is important to take into account that rational contribution to solution of those problems is made by the regions. Regions' economic development strategy is a general direction for achieving the goals and should be expressed in tactical actions and programs, which will promote regions' development in the right direction.

When analyzing the quality of the regions' development, a conception of growth stages theory is often used. According to that theory economic development goes through three main stages: pre-industrial, industrial and postindustrial. Prevailing branches of the pre-industrial development are extractive industries, agriculture, fishing, timber and mineral resource industries. At the industrial stage manufacturing industries dominate: machine building, chemical, timber and woodworking industry, light and food industries etc. At the postindustrial stage the main branches are those of non-material production – that is science, education, trade, finance, insurance, healthcare etc. The main features of the postindustrial society are relative decrease in goods production and increase in services production, science-intensive production, personnel's advancement.

At present there are two opposite processes in Russia: deindustrialization and growth of the service sector. The first trend is demonstrated by the decline in manufacturing industries and strengthening of the extractive ones. In the first years of the country economic reforms a certain position was obtained by the raw materials sector. The export of oil, gas, ferrous and non-ferrous metals and timber provided the most significant currency earnings and did not allow the whole industrial potential of the country to be destroyed. The trend, which is somewhat forced, but viewed from the country's economic development perspective, is negative as it transfers the society from the industrial stage of development to the pre-industrial one, demonstrating regress. At the same time Russian economy has some features of the postindustrial society, i. e. growth of the service sector, trade and financial institutions. To assess the depth of these directions it is very important to describe not only the production structure, but also the degree of information technology spreading.

Worsening of Russia's regional problems is related to the unjustified inequality of the regions' social and economic conditions. That requires the ecological balance maintenance, improvement of the economic territorial structure, population employment. Therefore new effective methods of influence on the economy should be put in practice, particularly the program and target forecasting which allows to solve single region's development problems in connection with the regional policy of the country.

Regional programs are a kind of complex target programs serving as a regulation and management tool of the regional economic, social and technological development strategy. They are also a method of prioritizing the resources concentration to solve the most urgent problems. Regional programs can be of the following types: interstate, federal, regional, created according to the manufacturing characteristics, and complex. Focused programs are usually designed as target federal programs. Complex programs of the regions' social and economic development usually comprise all the focused programs with the predominance of socially oriented programs. Such programs for a certain region allow to connect all the focused programs according to the needs in different kinds of resources, and prioritize them.

Regional programs are classified in terms of region's location, functional orientation, problems content, significance of the program's objective, etc. Defining the characteristics and corresponding programs types is one of the main provisions for sorting out general features in the process of their design and implementation. Classification characteristics reflect the essence of problems; describe the objects of programs, peculiarities of their management. A regional program may have several characteristics. The main objectives of regional programs include:

– creating an optimal territorial and sectorial structure of the economy;

– equalizing of interregional differences in economic, social and technological development;

– effective use of natural, material and labour resources of the region;

– environment protection;

– overcoming the consequences of natural and anthropogenic disasters;

– spiritual revival of the regions, saving their historical heritage, strengthening of their cultural potential, stabilizing of social, political and legal situation.

The process of selection of the regional problems, requiring program solution, has three main stages:

1. Consideration of the regional problems, requiring program solution, which includes the following steps:

– complex research of the adopted conception and forecasts of the region's economy functioning;

– analysis of the social and economic situation (economic potential, internal and external development trends, market dynamics, main funds movement, level of the needs satisfaction etc.);

– determination of the factors which cause problem situations;

– determination of the problems, which require immediate solution.

2. Justifying the possibility to solve the problem with a program method includes:

– designing of the recommendations for using a target-program method through defining program features, those of certain termination time, determination of a problem's object etc.;

– analysis of advantages and disadvantages of using a target-program method in elimination of the problem situations in relation to the region's conditions.

3. Independent expert assessment of suitability of the problems; solution with a target-program method includes the following procedures:

– preliminary resource assessment of each problem to be solved in terms of different resources;

– determination of the region's resource opportunities and their correlation with the needs;

– determination of economic effectiveness of the problems' solving using a target-program method.

Region's administration, in order to improve the territorial social and economic development, should use different management techniques: regional industrial policy tools, including a wide range of methods of investments mobilization; benchmarking methods. The most effective methods of the region's economic development management are strategic planning and regional marketing. Strategic planning and management are modern management methods. Practice of using strategic planning not only in industry, agriculture, construction, transport but in other spheres of human activity proved to be highly effective. All the procedures of strategic planning care applicable to the regional planning. They can be used not only in designing complex programs of regions' social and economic development but also in implementing anti-crisis measures, in managing large infrastructure projects and other directions of regional development.

Regional marketing is one of the most successful conceptions of the modern management, as it allows to direct al the enterprise's activity to the customer needs. Many marketing methods are successfully designed and developed within regional marketing, implemented by the regional administrative bodies. A region's promotion plan may become its tool. This is a set of measures which create and maintain the region's long-standing competitive advantages.

Region's development strategic plan allows to find a way out of crisis, increase the living standards. The plan's starting point should include assessment of individuals in all aspects of their activity; degree of their compliance with the modern world's development trends and correspondence with the goals set. A region's social development level is determined by its economic state as the living standards are determined by the level and effectiveness of the production. So, economic development should be planned first, and the main objectives of such planning should be the following: production structure renovation, attraction of the new industries to the region, development of the existing ones, development of the infrastructure for people's life and activity support; creation of the new working places. According to the definition, given by A. P. Gaponenko, [1] a cycle of the region's economic development strategic planning can be represented by a set of consequent stages, as shown in fig. 1.

A development plan should be based on a clearly formulated set of development goals, which meet the standard requirements of measurability, attainability, time orientation and consistency. As a rule, there are general and definite goals and objectives. General goals characterize a region's requisite state in certain periods, so they must be clear and relatively concise. They serve as a development criterion and allow to determine suitable measures for their achievement. After that they are transformed into more specific goals, which provide detailed quantitative guidelines and suitable assessment criteria that help to control the level of the region's social and economic development.

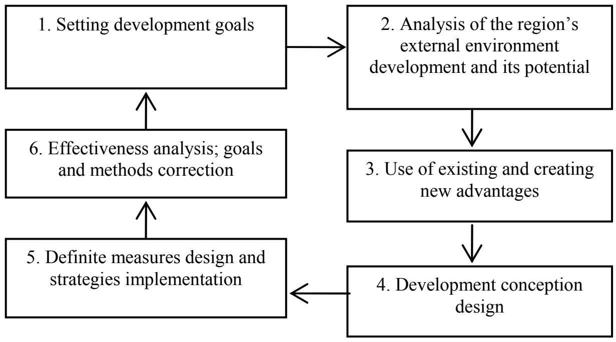


Fig. 1. A cycle of the region's development strategic planning

At the stage of the analysis of the region's external environment development favourable and unfavourable factors of its development are defined, external social and economic opportunities and threats are analyzed. Using qualitative and quantitative analysis of the economic, ecological and institutional external environment, external resources, that can be used in the region's economic development, are evaluated; key social and economic features of the external environment are studied, comparative analysis of the similar regions' economic development is conducted, which leads to the determination of the territorial development external factors. When distinguishing the region's strengths and weaknesses, results of the previous social and economic development programs are assessed, with finding the reasons for success and failure. The region's internal advantages and disadvantages are also analyzed to define the stage of the region's development.

At the conception design stage final conclusions about the goals of the region's social and economic development are made, this development's factors and mechanisms are analyzed, as well as methods of its management. At the same time new trends of the regions' social and economic development are taken into account alongside with the characteristics of the present stage of Russian economic development and local conditions. The conception is based on the design of alternative strategies, their economic assessment, prioritizing the directions of the region's development and defining of the territorial competitive advantages. On the basis of the region's social and economic development conception a plan of specific measures is developed, including the goals, terms of their achievement, responsible persons, expected result, volume and sources of finance, methods of interim control and feedback collection, possible consequences of the plans and programs' implementation.

At the stage of effectiveness of the considered variants analysis or goals and methods correction not only working actions plan is studied but also continuous monitoring of the region's social and economic development is conducted, comparison with the goals and development criteria is made, effectiveness and success of the measures taken are evaluated. The suggested strategy's compliance with the resources is studied, as well as its ability to solve the main problems of the region and residents' support.

One of the characteristics of social and economic systems is that they are goal-oriented. Therefore, the goal is the desired state of the economic system, which has a certain period of existence. Wrong choice of goals when creating a system leads to solving of the non-urgent problems, which can cause more harm than use of an ineffective system for achievement of the goals chosen. When analyzing the frameworks of goal-setting for any studied system it is important to remember that the goal of a particular system can fully correspond with or be a part of a goal of a higher system and strictly comply with the objective needs of the environment. The process of goal-setting is divided into three stages: environment analysis, assessment of the studied system's opportunities, and goals formulation.

In the process of goal-setting economic systems are characterized by conservatism and sluggishness. Very often the goal is defined in abstracto, as a final change state of the studied economic object or process; it is very important to understand what the goal is and what its alternatives are. From the positions of different subjects the goal and alternatives can exchange places, so interaction of goals and alternatives is clearly illustrated with a "goals tree" method. Goals systems and means for their achievements are similar in structure and mutually determined, as the first-level goals are the alternatives for achievement of the main goal, but in relation to the second-level goals they serve as goals [2].

Thus, goal-setting plays an important role in search of a suitable variant, as it helps to make a conclusion about the possibility of the initially set goal achievement. In the process of clear formulation of goals and objectives, goals prioritizing and preferences defining system analysis and different kinds of modeling are used. The analysis is based on studying of the retrospective state of the problem; this is done with use of dynamic indicators, characterizing the levels of needs satisfaction for the program product during the previous years; resources spent on covering of those needs; lagging from the required level; productive forces state; identification of the stable trends in the change of these indicators and characteristics. An analytical picture of the region's state is constructed for comparison with the similar problems in other regions.

The Complex program of the region's social and economic development is a means of the set goals achievement and a form of the economic activity and social territorial organization management. This program contains the volume, pace, proportions and main directions of the region and its structural subsystems' social and economic development. The program also has the substantiation for the system of measures, aimed at rising of production management effectiveness; rising of living standards level is also explained. A single region's development strategy, however, should comply with the country's development strategy and be its integral part.

Sections of the program's project should be designed using a serial-parallel method, that is when design of each next section starts before the end of designing the previous one. Necessary corrections are made in the previous sections for their better compliance with the next following sections. On the whole, there is a logical succession from the goal to the ways and methods of its achievement, then through the goal-achieving activities to the necessary resources, means and sources of their receiving. The Complex program's composition and structure are a list of certain sections, which reflect the contents and functions of the accepted program project, for example:

Section 1. Analysis and assessment of the region's existing social and economic state. Here the territorial social and economic state and its consequences are described; the most critical problems to be solved with a program are defined.

Section 2. Assessment of the natural resources and environmental state. The region's natural resources, their amount and quality, opportunities to use and export them are described, alongside with the existing ecological situation and measures to improve it.

Section 3. Region's demographic conditions and labor market. This section analyzes the territory's economics and sociology of labor and defines the priority problems.

Section 4. Conception of the region's social and economic development. On the basis of the conducted investigation the conclusions on the region's existing economic situation and structure are made; its material and production base is studied and economic disparities are defined. This section also forecasts the opportunities to use natural resources potential of the territory and formulates the conception (forecast) of the region's social and economic development, taking into account the technological progress trends and the region's economic objectives.

The conception should include fundamental substantiation of the most rational use of the large-scale and effective resources and the directions of their use; evaluation of the different branches comparative efficiency and establishment of the ways to correctly form the whole economic complex.

Section 5. Main target subprograms. In this section the main target subprograms for the region are determined. These are based on the production, function and problem characteristics. Such programs have goals, objectives and final results, which will be achieved after the programs implementation. Each program's action is assessed with a number of qualitative and quantitative indicators.

Section 6. Program's implementation scheme. Here a set of mutually related measures, actions and economic tools, that can solve the problem defined by considering and comparing alternative options, is presented. A forecast model of the program is also developed.

Section 7. Program's resource supply. This section contains calculations of the main financial, material, climatic, labour and information expenses. Out of the whole volume of the resources a part from the region's own sources is separated.

Section 8. Program actions coordination. At this stage harmonization of actions of all the organizations, taking part in the program implementation is conducted. The region's cooperation with the neighbouring territories and foreign countries in implementation of the program is substantiated.

Section 9. Assessment of the program implementation efficiency. The program economic efficiency is calculated.

Section 10. Organization, forms and methods of the program management. In this section organizational and functional structure of the complex program management is developed in accordance with the special territorial features.

All the program sections should contain the necessary substantiation and explanations; they are formed for the whole period of the program implementation and are specified for each year of fulfillment. For long-term programs the first layout is made for each year, then – for longer periods.

As a result of the Program's analysis and structuring the defined problems are sorted out, which allows to build a system of goals, consisting of 4 levels and five branches, that are mutually complementary and independent (fig. 2). The carried out analysis is aimed at selecting the most important direction of the region's social and economic development out of the alternative options. At the lowest level of the «goals tree» there are possible resources that can be used for achievement of the main goal. The second-level elements (2.1–2.5) correspond with the fundamentals and structure of the Program.

At the first level Krasnoyarsk region's development variant, which allows to solve the social and economic problems of the territory, create conditions for the economy's gradual transition to the strategy of managed development and maintenance of the stable growth is formed.

At the second level the following courses of actions are analyzed:

2.1. Increase of the region's investments attractiveness.

2.2. Perspective development of the industry.

2.3. Improvement of the agro-industrial complex.

2.4. Increase of the regional population's employment level.

2.5. Provision of the population with the social living standards.

At the third level the problems of the second level are considered in more details:

3.1. Creation and propaganda of the local laws, which would promote investments into the region.

3.2. System work with the potential investors to attract their capital to the region.

3.3. Creation of the conditions for small and medium business development.

3.4. Creation of the conditions for the increase of the fuel and energy complex efficiency.

3.5. Acceleration of the innovation and investments processes at the industrial enterprises.

3.6. Implementation of the main directions of the state policy in the areas of energy saving, industrial and technological cooperation with another countries.

3.7. Participation in the transformations of the region's agro-industrial complex.

3.8. Stabilizing and increasing of the volumes of agricultural, raw materials and food production.

3.9. Improvement of the personnel training and retraining system.

3.10. Formation of the conditions and tools of salary regulation.

3.11. Improvement of the region's living standards.

3.12. Improvement of the population's quality of life.

At the fourth level:

4.1. Creation of the new working places as a result of small and medium business support.

4.2. Maintenance of the stable functioning of the social security institutions, which guarantee the qualitative service to the population.

4.3. Development and implementation of the actions related to the accommodation problems.

4.4. Social and material support of the village residents.

4.5. Development of the social measures to solve demographic problems.

4.6. Coordination of the activities related to family and child care.

4.7. Increase of the cultural and spiritual level of the population.

The goals in the "tree" are mutually complementary, so the method of hierarchy analysis can be applied to the second-level elements to choose the most important goal out of 5 sub-goals. The options are compared; the results are processed on the basis of the matrix analysis and a number of special preferences evaluation procedures.

To calculate the priority for the 2nd-level goals adequate criteria are developed. Among them there are financial support from the region's budget, importance of the achievement in the living standards improvement; compliance with the federal target programs and consequences of non-fulfillment of the certain obligations. The most developed goal is the fifth one, related to the solution of the social problems of the region's population. This goal has specially designed target programs for its achievement. The poorly developed agro-industrial complex is a problem factor, which is caused by unfavourable climatic conditions on the most of the region's territory. Increase of the region's investments attractiveness is connected only with the needed natural resources development and, from the point of view of the future prospects, is not an economically justified condition for the region's development.

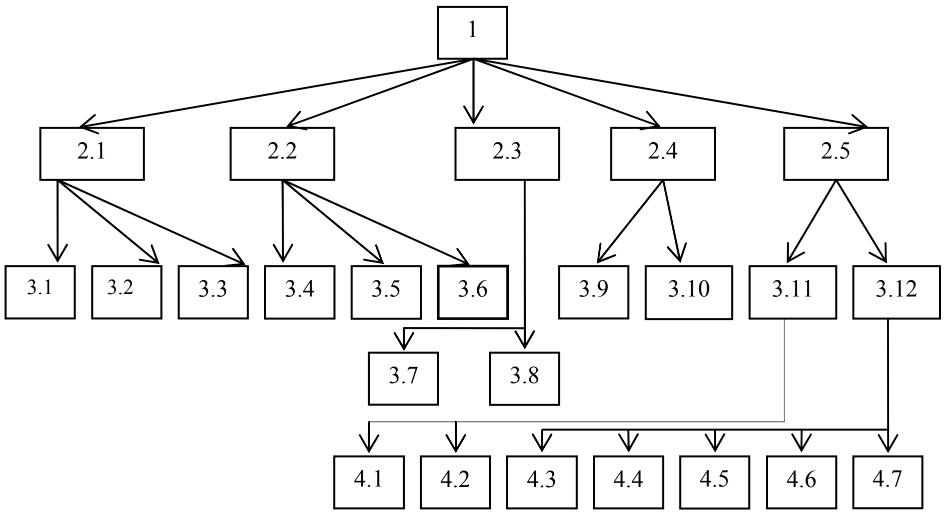


Fig. 2. “Goals tree” of the social and economic development of Krasnoyarsk region

Thus, the main attention is paid to the fuel and raw materials factors whereas machine building is thought to be less important. Meanwhile, machines, equipment, vehicles and other products of machine building industry are not only the most complete kinds of final products but also those very goods, which characterize the scientific and technological potential and industrial development level of the country. That is why perspective development of the region’s industry should gradually become its priority goal, achievement of which will help the region to get a stable economic position in the long-term period, independent of the external or international conditions; become attractive for investments and successfully implement the developed social programs.

References

1. Gaponenko A. L. Design of the Territorial Self-Development // Region: Local Self-Development Resources : Elective Course / scientific ed. Y. N. Alekseev, B. E. Shnilev. M., 1999. Iss. 8.

2. Samofalova E. V., Kuzbozhev E. N., Vertakova Y. V. State Regulation of the National Economy ; ed. by E. N. Kuzbozhev. M. : KNORUS, 2008.

O. V. Patsuk
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

METHODOLOGICAL ASPECTS OF ANALYSIS AND FORECAST OF ECONOMIC SYSTEMS’ DEVELOPMENT

Assessment methodology of the different territories’ economic effectiveness is uniform and independent of a territory’s size and structure. However, certain methodological assessment techniques could differ according to the characteristics mentioned. The complex’s type and special features of the problem being solved are of particular importance when choosing a research method.

Keywords: social and economic development of the territories, forecast of the economic systems’ development.

System analysis of the region’s social and economic development indices is a structure with a number of separate indices. In accordance with the management task it may include criteria that reflect social, economic and other effects of the development scheme. Methodology of the analysis and forecast of the region’s economic system development is based on:

– a combination of formal and informal approaches in a system of decision-making;

– an experimental approach to the implementation of any model;

– mathematical and technological support of a calculating experiment on imitation models.

Designing the models allows to realize the existing problem and facilitate the decision-making process, but can not replace the knowledge and intuition of an expert. Due to this, development of an informal approach to the regional economy management is based on the application of the expert systems and use of artificial intelligence. In such model complexes as regional management systems, in the multicriterion situation the application of planned decisions optimization methods is restricted by the difficulty of their implementation, as it is almost impossible to determine the goals with mathematical exactness and develop an adequate variant, when the information is not complete. That is why when developing a regional economy managing system the main method is imitation modeling, which through previously conducted analytical research mathematically explains the received results. This allows to solve the problems of social and economic development, design actual strategies and choose rational variants of development.

Development of the social and economic sphere becomes possible when the managing bodies are continuously receiving and processing into planning decisions the diagnostic information about its state and the state of the environment; about the deviations in the social and economic sphere's process of reaching its goal. This also allows the managing bodies to change the state of an object according to the goals set. Region-studies research uses a full complex of scientific research methods:

System analysis is based on a principle of consistency and includes setting a goal; determining the objectives, stating a scientific hypothesis, complete investigation of a suitable variant of branches' distribution. This method allows to study the structures of activities, their inner connections and interaction.

Systematization method is connected with the division of the studied phenomena and criteria into sets, which are characterized by certain similarities and distinguishing features (classification, typology, concentration etc.).

Balance method is characterized by establishing different regional balances (natural, financial, labour etc.).

Method of economic and geographical research is divided into three constituent parts: regional method (studying the ways of the regions' forming and development); branch-wise method (studying of the social production's development and distribution according to activity type and geographical factor); local method (studying the ways of establishment and development of a separate town or village's production).

Cartographical method is a visual aid, which illustrates the distribution peculiarities.

Method of economic and mathematical modeling allows to model the territorial proportions of the region's economic development and economic complexes' formation. Using the modern electronic means, it provides the opportunity to process various data much quicker and choose the most suitable variants and models in accordance with the set regional research goals.

Methods of multivariate statistical analysis are connected with the statistical data processing. One of the most widespread methods of a multivariate information analysis is a factor analysis, or cluster-analysis. It researches the influence of different factors (causes) on a result indicator. In today's study of social and economic regions' development the main clusters and ordinal classification methods are used.

One of the first kinds of statistical models used in regional studies was an economic base model, introduced by H. Hoyt (USA) in the '30s. An economic base analysis is an accelerated method of a region's economic growth forecasting, which uses a simplified growth theory and requires minimum information. For creating these models, economic activity indexes (in general, employment) for two periods of time are necessary. It results in forecasting of the basic and servicing sectors' development.

Region-metrics is a scientific branch in regional economy, which uses mathematical methods, i. e. regional modeling. It includes:

Taxoning method is a process of dividing the territory into correlating or hierarchically co-subordinate territorial cells. In fact, zoning at any level is taxoning, as its objects are regions, so it is possible to use the concept of regionalization.

Variant method of the region's productive forces distribution is mostly used in development of the production distribution schemes at the first stages of planning and forecasting. It implies studying of the variants of regions' development levels and territorial economic proportions.

Survey methods include standardized interviews, face-to-face conversations with representatives of different activities and spheres of the region's social and economic complex; analysis of the region's executives, scientists and specialists' speeches etc.

Methods of regional living standards' comparison and regional social infrastructure development's forecasting study living standards on different territories on the basis of synthetic and special indices. The main purpose of the analysis is to determine actual differences in living standards and performance in relation to the equal degree of the population needs' satisfaction in all the regions. These needs are directly connected with the development of the regional social infrastructure.

Thus, study of the regional social and economic systems is based on a wide range of methods and ways which are applied by regions' specialists. Apart from the methods described, there are also conceptions of sinuous development, which are an integral part of the methodology used in long-term forecasting of the dynamic in the national economy's state regulation [1]. At present, there exist several classifications of the sinuous development theories. According to the classification, offered by S. M. Menshikov and L. A. Klimenko, there are:

1. Monetary and credit conceptions, where development fluctuations are caused by monetary factors.

2. Conceptions of intensity changes in producing of capital goods. (Kondratyev, Mandel, Forrester).

3. Theories of fluctuation of the particular production factors, which cause deviations from an economic development trend (Friedman – interchange of excess and lack of labor; Rostow – abundance and lack of food and raw materials; Craig and Watt's theory).

4. Neoschumpeterian conceptions, which study economic system's transition from one equilibrium state

to another on every long wave. Innovations' cluster plays a particularly important role here as the innovations provide the material basis for this transition (conceptions of Mensch, Weidlich, Wordl, Newcamp and others).

5. Institutional conceptions, according to which the long waves are caused by the main economic and political institutions (theories of Peres-Peres, Chandler, Kaletsky, Screpanti).

There is also one more classification of sinuous development conceptions:

1. Investments conceptions suggests that the long wave is determined by periodic accumulation and then devaluation of the durable capital goods (Kondratyev, Forrester, Sterman).

2. Innovations conception implies that the key part in wave-formation belongs to innovations clusters which create a leading sector in the economy, and expansion of this sector causes a corresponding cycle in economic development (Schumpeter, Mensch).

3. Capitalist crisis theory concludes that a tendency of the profit rate to decrease leads to a crisis, that can be overcome due to external factors, which increases profit rate for some time and creates conditions for a new long-run rise of economic development (Mandel, Day).

The main method applied in development of the sinuous theory is determination of the trends using different functions, which describe change of economic indices with time. Using the least squares method, which allows to eliminate unsuitable directions and select suitable ones, they studied different relations, determined by the actual dynamics of the index concerned. Deviations from the trend are processed with the help of the special filter-functions, the most widespread of which are 9-, 21- and 51-year moving averages.

Another widely-used way of eliminating a trend is a transition from the observable indicators to the first differences, which are also statistically processed and subjected to various transformations (squaring, equalizing with help of moving average etc.) Deviations from a trend, considerably different from each other in size and amplitude prove the existence of a wave.

The main difficulty in the sinuous development theory is not statistical study of retrospect calculations, but development of the forecasting models of long waves to study them in future and analyze long-term trends of the economic development. Besides, it is quite hard to design a set of practical methods for anticycle policy in order to avoid lasting economic crises or alleviate their consequences. That is caused by the necessity of taking into consideration irregularity of the economic growth, systematic shifts in the economic system's structure. Today creation of different imitation models with help of the computer is the most effective way to forecast structural shifts.

A variant of the economic systems' cyclic development is Teves' model, which gives a cycle's dynamic interpretation, based on interaction between a multiplier and accelerator in an economic system through an interest rate for capital investments [2]. The model's special feature is that the change in capital investments volume functions as the main source of cyclic fluctuations. Teves' model divides investments demand in two directions: for receiving of or adding to the existing working capital and for creating or maintaining of the economic system's main funds. In the model an interest rate is a time function, and the volume of investments is a function of the interest rate:

$$L_t = L_v\, Y_{t-1} + L_i i_t, \quad (1)$$

where L_t – investments demand (necessary investments); L_v – investments, necessary for operating; L_i – funds-generating capital investments; Y_{t-1} – total yield of the previous period; i_t – current interest rate.

Caldor's model is a dynamic model of an economic cycle, where the savings' volume is a non-linear increasing function of income. Cyclic development can be noticed in the dynamic of savings and investments indicators, taking into account region's various activities. Equilibrium points of the economic systems are crossing points of the investments and savings graphs, and analytically their coordinates are determined when the volumes of those indicators are equal. The main postulate of the model described above is that approaching motion of the investments and savings curves causes change in the equilibrium from stable to unstable. This characterizes the beginning of a new economic cycle.

Economic development is a set of rises and falls in the system, and those changes are connected with the sinuous character of economic dynamics. Samuelson-Heeks model explains the mechanisms of economic dynamics' fluctuations on the basis of acceleration principle and multiplier conception. Acceleration principle is based on the conclusion that investments volume depends on increase or speed of changes in the demand for the final product. Generated by such conditions investments demand is multiple of the demand for final product. The degree of its multiplicity is called an acceleration factor. The model's special feature is that it connects the income of a certain period with the sum of consumption and investments and considers autonomous investments as independent of the income changes. Samuelson-Heeks model can be represented by the following formula:

$$Y_t = C_a + MPCY_{t-1}, \quad (2)$$

where Y_t – total yield; C_a – autonomous consumption; MPC – Maximum propensity for consumption.

In this model dynamics of the total yield fluctuations is determined by the meanings of the maximum propensity for consumption or those of the multiplier and accelerator. The main forms of the total yield fluctuations are the total yield's volume movement, dying fluctuations, explosive fluctuations of the total yield's volume, infinite monotonous growth of an indicator, constant steady fluctuations.

In an extended model investments are divided into 3 constituent parts:

– "extensive" investments, aimed at increasing the capital of the existing technology;

– "intensive" investments, aimed at introducing the capital, which uses new technology. These investments lead to the increase of productivity;

– "innovation" investments, aimed at creating new products and areas of production.

This model is described by a system of linear differential equations and can have only one equilibrium state. The deviation from that state can either decrease with time (stable solution) or increase to infinity (unstable solution). Command qualifiers – labor productivity (reflects intensive character of technological advance) and capital provision (extensive) – influence production through the profit rate or in combination with it.

Forrester's G. system dynamics model is an imitation one. It consists of 6 parts: production, finance, householders, labor resources, government. These parts are responsible for establishing the correlations which determine consumption, investments, employment, prices, governmental policy, interaction among sectors. By distributing investments to different sectors it is possible to model the variants of economic movement, create various scenarios of the region's economic development. For example, the production part contains more than 15 producing sectors, based on Cobb–Douglas' production function with 12 factors as well as the system of balance equations. In non-production parts it is possible to model a large number of different interrelated processes, where there should be distinguished models of financial and credit institutions, state consumption and external relations.

Technique used in determination of cyclic phases of the economic system development is based on a joint analysis of the production increase and its structural changes. These tools of structural and dynamic analysis divide speed of growth into inertial and reconstructive components and studies how the economic system's structure influences its dynamics. Among the main developers of the structural and dynamic approach there is R. Akoff, V. Leontyev, J. von Neiman, L. V. Kantorovich and their today's successors, for example, L. A. Dedov [3]. To solve the mentioned problem a physical volume index is separated into a component related to the production inertia, and a component connected with the structural changes in the production volume. An interim calculation indicator is an index of the production physical volume change λ:

$$\lambda = \frac{\sum y_i q_i}{\sum A_i q_i}, \tag{3}$$

where y_i – actual or reported value of the sector's production volume; A_i – basic volume of the i-sector's production (quantities y_i and A_i are measured in natural merchandising measures); q_i – basic prices, in accordance to which quantities y_i and A_i are transferred into price measurement.

It can be written:

$$\lambda = \sum h_i d_i; \tag{4}$$

$$h_i = y_i : A_i; \quad d_i = (A_i q_i) : \sum A_i q_i. \tag{5}$$

where h_i – speed (index) of growth of the i-component of production; d_i – доля i-sector's share in the basic composition of production; n – the number of sectors; $\sum d_i = 1$.

Apart from basic share characteristics of the production output it is also necessary to operate with the share characteristics of the 'report' composition of P_i, which are evaluated with the formula:

$$P_i = \frac{y_i q_i}{\sum y_i q_i}. \tag{6}$$

Between d_i and P_i there is an interrelation expressed by:

$$P_i = \frac{d_i h_i}{\lambda}. \tag{7}$$

To compare the structures there are several evaluation schemes. The most widespread is evaluation of the common structural shift:

$$m = 0.5 \sum |P_i - d_i|. \tag{8}$$

Economic statistics uses a measure, that is contrary to evaluation of the common structural shift – a quotient of similarity m^*, which shows to what extent starting and factual share structures are similar. Evaluation of the common structural shift characterizes the level of an object's change, i. e. is a characteristic of a reconstructive component in the structure's evolution. Degree of similarity can be interpreted as a quantitative evaluation of persistence:

$$m^* = 1 - m. \tag{9}$$

Separation of the physical production volume index into components is made by a metric approach. There are 3 effects in the dynamics of production output:

1. Effect of the production volume change (as it is) (λ).

2. Effect of eviction demonstrates that if production output has a structural shift, then shares of some product groups will enlarge. Total share of these groups will extend, and they will evict other positions of the output. Measure of the eviction effect is a sum of the corresponding increases:

$$\sum (P_i - d_i) = m. \tag{10}$$

3. Effect of compression shows that shares of some product groups in the total sum of shares are reducing. эффект сжатия выражается в том, что доли некоторых продуктовых групп в общей сумме долей снижаются. In terms of quantity effect of compression is expressed by a sum of the corresponding reductions, i. e.:

$$\sum (P_i - d_i) = -m. \tag{11}$$

Effects of eviction and compression are equal, but opposite in sign; they characterize a common phenomenon – structural shift from different points.

Let us make some transformations like $\lambda = (h_1 d_1 + h_2 d_2 + ... + h_n d_n)$ and determine a share structure of the output:

$$\lambda = \lambda(h_1 d_1 : \lambda + h_2 d_2 : \lambda + ... + h_n d_n : \lambda) = \lambda(P_1 + P_2 + ... + P_n). \tag{12}$$

So:

$$P_i = d_i + I_i, \quad (13)$$

where I_i – change of the *i*-share in transition from a share structure $d = (d_1, d_2, ..., d_n)$ to the share structure $P = (P_1, P_2, ..., P_n)$.

Allowing that $I_i = P_i – d_i$, in the last expression we can distinguish effect of compression:

$$\sum I_i = \sum (P_i - d_i) = -m \quad (14)$$

and effect of eviction:

$$\sum I_i = \sum (P_i - d_i) = m. \quad (15)$$

We will have:

$$\begin{aligned} &\lambda\left(\sum d_i + \sum I_i + \sum I_i\right) = \\ &\lambda\left[\sum d_i + \sum (P_i - d_i) + \sum (P_i - d_i)\right] = \quad (16) \\ &= \lambda(1 - m + m) = \lambda(l - m) + \lambda m = \lambda m^* + \lambda m. \end{aligned}$$

Evaluation λm^* is considered as a measure of similarity of the "starting" and "reported" structures of production output, reduced to the level of growth index λ. Hence, λm determines a reconstructive component of the output. Growth index is decomposed into inertial and reconstructive components.

The next stage is decomposition of the growth rate. Growth index λ can be expressed in the following way:

$$\lambda = y : A = (A + \Delta) : A = 1 + \Delta : A = 1 + N, \quad (17)$$

$$\begin{aligned} &y = \sum y_i q_i; \\ &A = \sum A_i q_i; \\ &\Delta = y - A; \\ &N = \Delta : A. \end{aligned} \quad (18)$$

where N – output growth rate.

It seems logical to move from growth tempos to the growth rate and, consequently, to the decomposition of the latter into reconstructive and inertial.

$$N = \lambda - \lambda_H, \quad (19)$$

where $\lambda_H = 1$ – a starting meaning of the growth indicator, corresponding to the level of the base year.

Logically:

$$\begin{aligned} N = \lambda - \lambda_H = (M_1 - M_{1H}) + (M_2 - M_{2H}) = \\ = M_1 + M_2 - 1, \end{aligned} \quad (20)$$

where $M_1 = \lambda m^*,\ M_2 = \lambda m.$

So, $M_{1H} + M_{2H} = 1$. This unity should be shared between the two components of the starting state: inertial and reconstructive. We will define M_{1H} as $\lambda_H m_H^*$, and M_{2H} as $\lambda_H m_H$. Letter "H" marks the base numbers.

It is already known that $\lambda_H = A : A = 1$. $M_H = 0$ (as it is a base structure shift in relation to itself). So: $M_{2H} = 1 \cdot 0 = 0$. Consequently:

$$M_{1H} = \lambda_H m_H^* = \lambda_H (1 - m_H) = 1(1 - 0) = 1. \quad (21)$$

But then

$$\begin{aligned} n_1 = M_1 - M_{1h} = \lambda m^* - 1 = \lambda(1 - m) - 1; \\ n_2 = M_2 - M_{2H} = M_2 - 0 = M_2 = lm. \end{aligned} \quad (22)$$

As a result:

$$N = n_1 + n_2, \quad (23)$$

where $n_1 = \lambda(1 - m) - 1$ – an inertial component of the growth rate; $n_2 = \lambda m$ – a reconstructive component of the growth rate.

To study the structural cycles, whose analysis is important for monitoring of the region's economic development, a concept of output structural elasticity is used:

$$E = n_1 : n_2, \quad (24)$$

where n_1 – conservative (inertial); n_2 – reconstructive component of the growth rate.

Structural elasticity values correspond to the phases of the structural cycle, i. e.

1. Structural elasticity is positive: growth, based on the traditional output structure is complemented with growth based on structural changes, i. e. $n_1 > 0$ and $n_2 > 0$. Then $E > 0$. A corresponding phase of the structural cycle is called a phase of complementary development.

2. If growth, based on the traditional output structure is decreasing ($n_1 < 0$), but growth based on structural changes is still there, ($n_2 > 0$), is positive and increasing, then $E > -1$. Such situation characterizes a phase of compensating replacement. That is growth due to n_2 – component compensates and replaces decrease based on n_1 – component, which is caused by the purpose of the structural shifts. Growth based on them should expand production capabilities. However, decline of the traditional production capabilities often has a disastrous state; in this case an economic system moves to phase 3.

3. In the situation of cardinal transformations of the traditional output structure decline ceases being compensated by a simultaneous growth based on a reconstructive component, which is related to the late effect of structural changes. We have $N > 0$ and $E < -1$, that is a situation of non-compensating replacement.

4. The fourth phase clearly demonstrates the mentioned factors, i. e. a common (deep) setback of production takes place: $E << -1$ and $N << 0$. Then the setback reduces, and a compensating influence of structural changes becomes significant. An economic system moves to phase 3, and then to phase 2. A new output structure, supported by additional changes, is established and the system returns to phase 1, but with a reconstructive output composition.

Practically, the deviations from the given theoretical scheme are possible (failures and repeat of the phases). It is considered, that a gradual transition of a structural cycle with the regular replacement of its phases is characteristic of a regulated (plan) economic system. The economy in acute instability is characterized by structural and dynamic instability. So, in case of instable correlation between the replacement effect and complement effect a local rise is followed by a local decline. Values of E are changed at random which represents the structural and dynamic instability.

There is a classification of the variants of economic system development, based on decomposition of the output growth (setback) rate:

1. Innovation economic growth takes place when the growth rate is significant, and the components of inertial growth (π_1) and growth caused by the structural changes (π_2), are comparable in volume.

2. Extensive economic growth is characterized by the correlation $\pi_1 >> \pi_2$. An inertial component of growth prevails here, which proves unexpressed efforts to balance production and needs and insufficient use of innovations, so growth rate N can be relatively significant. This takes place when it is defended from external competition and becomes a monopolist on the internal market. As a result, such development turns into stagnation.

3. Stagnation implies extensive management and appears due to absence of innovation, resources deficiency and obsolete ways of organizing economic processes etc. This variant is characterized by the following conditions: $N \approx 0$; $\pi_1 \approx 0$; $\pi_2 \approx 0$. To overcome stagnation is often a serious economic problem.

4. Structural crisis is represented by the following condition: $\pi_1 << 0$ and $N \approx 0$. Elements of the structural crisis include reduction of the traditional production capabilities, complicated changes in technology, assortment and structural unemployment increase. However, it creates conditions for gradual increase of the replacement influence by the growth rate component, based on the structural changes. As a result, such replacement begins to completely compensate the decline of traditional production capabilities, then the growth starts and parameter N becomes $N >> 0$.

Thus, sinuous economic development is continuous fluctuations in business activity, alternation of the extensive and intensive types of the economic growth. The main factor, influencing the length and depth of the cyclic fluctuations is the investments movement. Crisis forms a base for the new mass investments because the main capital devaluation leads to conditions for the production renovation on the new technological base, allowing to reduce expenses, restore the pre-crisis profit rate and then increase it. That is why strategically oriented investments policy allows to influence the cyclic development of an economic system and raise it from the critical states to the growth phase using intensive factors.

References

1. Kouznetsova O. V. Regions' Economic Development: Theory and Practice of State Regulation. M., 2007.

2. Samofalova E. V., Kuzbozhev E. N., Vertakova Y. V. State Regulation of the National Economy / ed. by E. N. Kuzbozhev. M. : KNORUS, 2008.

3. Dedov L. A. Economic Systems' Development: Methods of Evaluation and Analysis. Yekaterinburg, 1998.

T. V. Polozova
Krasnoyarsk State Agrarian University, Russia, Krasnoyarsk

A. A. Kolesnyak
Plekhanov Russian University of Economics, Russia, Moscow

FORMATION OF REGIONAL WHOLESALE FOOD MARKETS SYSTEM

The management structure and the infrastructure of regional wholesale food market, the scheme of agricultural products and foodstuffs movement to the end-consumer are offered. The place of the wholesale market in the system of food supply is defined.

Keywords: food supply, wholesale food market, system.

The organization of physical distribution plays an essential role in the system of population food supply of any region. Population food supply of any region of the country is defined first of all by the condition of food producing and serving branches. The agricultural commodity producers' share in the price of agricultural products decreases every year (15–18 %), a considerable part falls at processors, trade and middlemen. In 1990 commercial structures and black money structures had 0.5 % of costs in the final goods price. Today these figures are 25–28 %. The organizational unit of the food resources promotion system from the sphere of their production to the sphere of their final consumption, that provides increase of commodity producers' share in the final price of agricultural products, is the wholesale food market.

The notion of "the wholesale food market" is treated differently in scientific literature. The wholesale food market is a managed and operated system of goods promotion with a proper infrastructure [1; 2]. We consider that "the wholesale food market" has two meanings: first, it is one of the structures of the mechanism of a large goods consignment concentration and delivering from producers to consumers; second, it is an organizational form that solves certain problems and carries out corresponding functions in the subsystems of food resources formation and distribution. The main advantages of such form of the organization in the physical distribution system are: it takes into account end-consumers' interests, it directs a commodity producer to the market requirements, it promotes funds receipt to the local budget at the expense of shadow trade structures liquidation, and it creates prerequisites for the retail price decrease with simultaneous increase of the commodity producer's profit.

Establishment of wholesale markets allows decreasing retail prices in a trading network by 10–20 % on average [3]. It is possible to influence pricing of particular kinds of foodstuffs and to stimulate their production through regional wholesale markets. A commodity producer can choose different types of sale having wholesale food markets at his disposal: to sell the goods in the wholesale market independently; to deliver the goods to a wholesaler for sale; – to send the goods directly to the shop.

Highly organized wholesale markets lighten the work of Russian Federal Consumer Rights Protection and Human Health Control Service, controlling units that prevent the receipt of poor quality products to a trading network, provide stabile food supply and protect a commodity producer and processor. Mutual settlements of accounts between transactions participants simplify considerably. A bank accreditation in the wholesale market provides payments between purchasers and sellers.

At present there is no organized system of wholesale food markets in Buryatia, though there has been formed a network of wholesale centres and individual businessmen realizing their own interests or the interests of firms which they represent through these centres. Agricultural commodity producers and the system of consumers' co-operation of the republic are out of the established wholesale food market yet. The problem of local foodstuffs sale and creation of wholesale channels of their distribution to consumers has not been solved.

Wholesale food market should take its place in the system of Buryatia population food supply (fig. 1).

The consolidation of economic regions and creation of federal districts open up the possibility of centralized organization and practical implementation of this important and urgent work. Formation of wider markets in terms of territory allows to consider the population needs of the regions, forming a district, predetermines the specification of agricultural production arrangement and specialization, organization of food commodity zones and product formation distribution and trade centers, to increase the food supply level of the districts.

Management of the wholesale food markets system should be carried out by the state. The Ministry of Agriculture and Food, the Republican Service of Consumer Market and License Trade become coordinating authorities of wholesale markets management in Buryatia.

The main tasks of the state authorities in connection with regulation of foodstuffs funds formation and distribution are [4]: formation of wholesale food markets system and coordination of their activity; organization, maintenance and regulation of financial relations mechanism in the wholesale food markets system; development of material and technical basis of products improvement, storage and sale; organization of goods flow movement in the wholesale food markets system; organization of manpower training to work in the wholesale food markets system; rendering assistance in legal, economic and other activities to the participants of the wholesale food markets system, their information servicing.

The possible structure of the republican wholesale food market management offered by the authors of this work is presented in fig. 2.

The work of wholesale food markets is provided by specialized structures that form its infrastructure (fig. 3) [5].

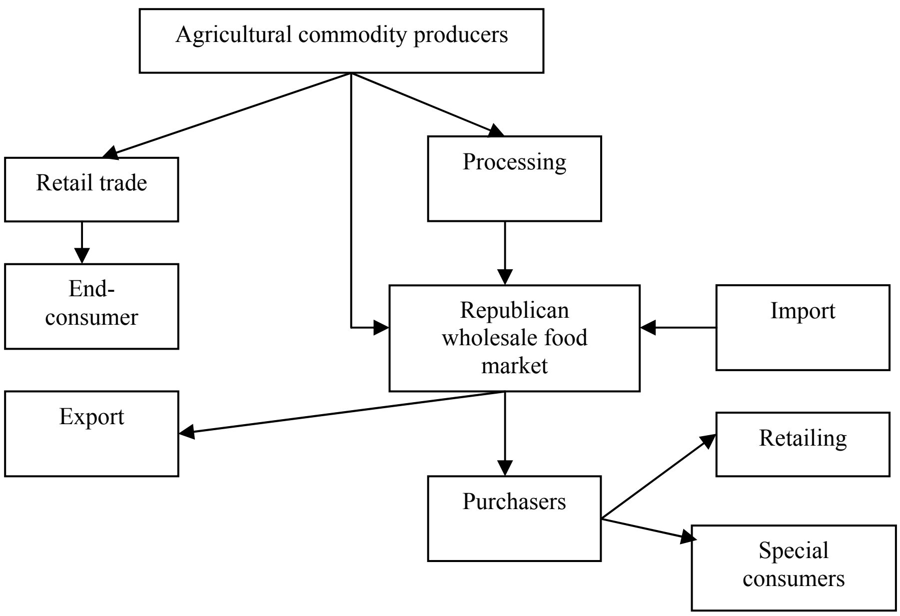


Fig. 1. Republican wholesale food market in the system of food supply

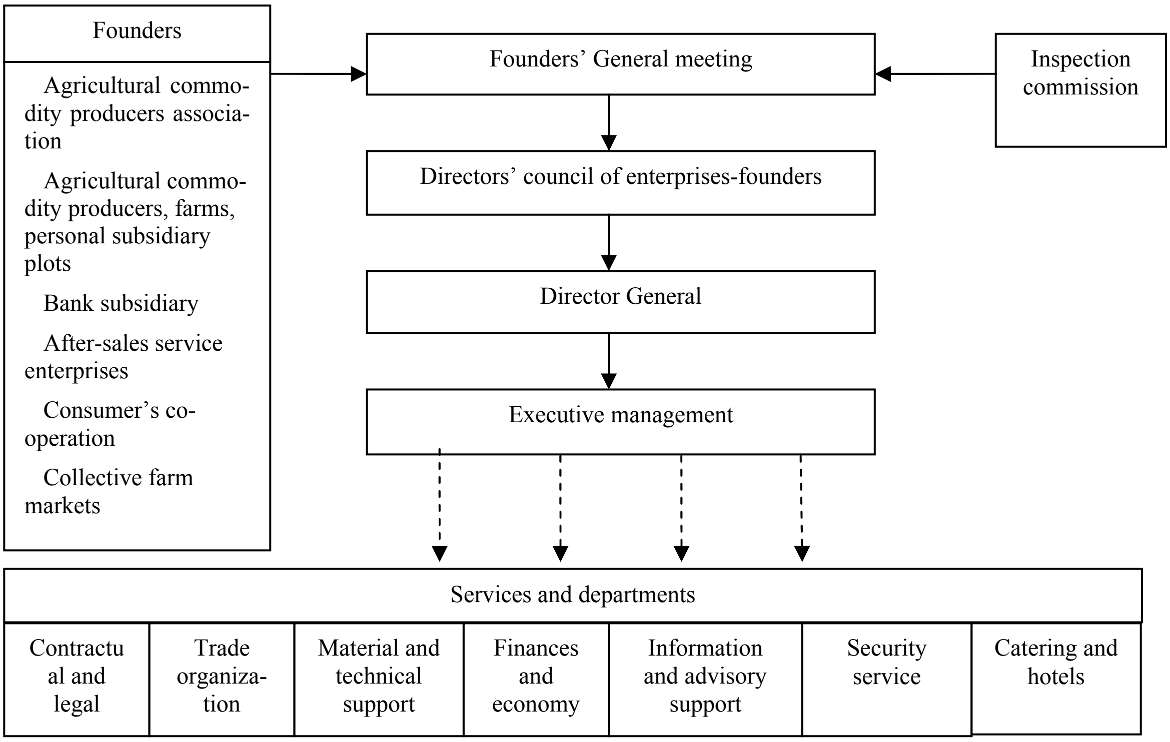


Fig. 2. The structure of the republican wholesale food market management

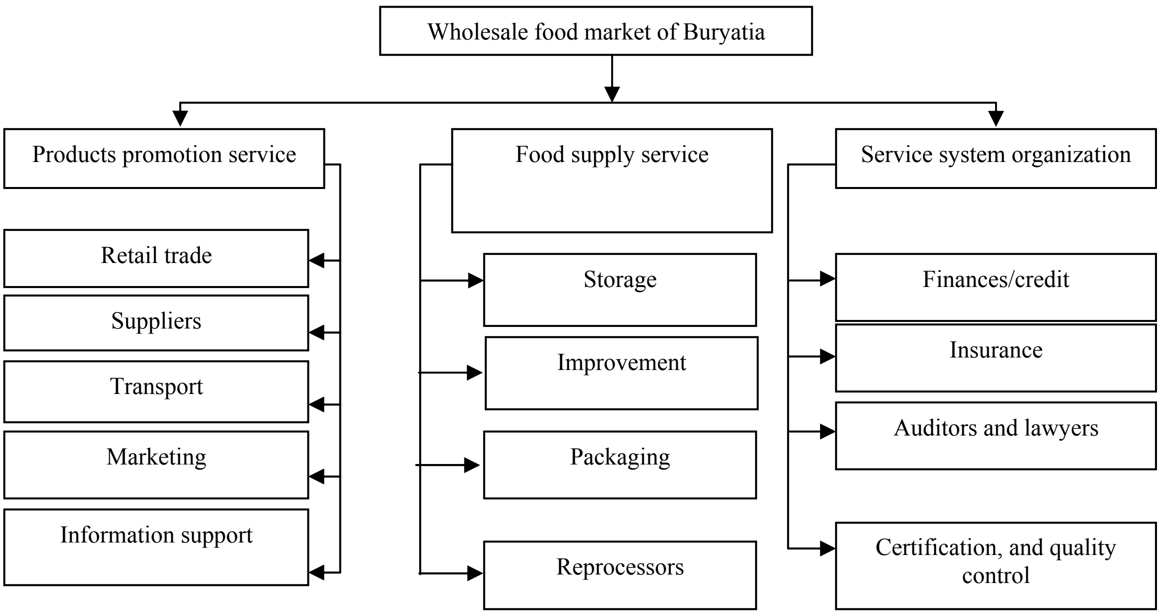


Fig. 3. The infrastructure of the regional wholesale food market in Buryatia

Active work in purchasing, delivery and selling goods of the agroindustrial complex and consumers' co-operation organizations on the basis of interested organizations' funds consolidation with local authorities' support will allow to increase the purchase amount of agricultural products to form the Republic food funds.

The situation improvement in the food market is not necessarily connected with considerable increase of investments, as the main obstacle for the agrarian production stabilization is its unprofitability caused by purchasing and retail prices disparity. Reorganization of the system of economic relations between agricultural commodity producers and processing, storage and sale branches can decrease this disparity.

The scheme of agricultural products and foodstuffs distribution in the system of the main food markets offered by the authors is given in fig. 4.

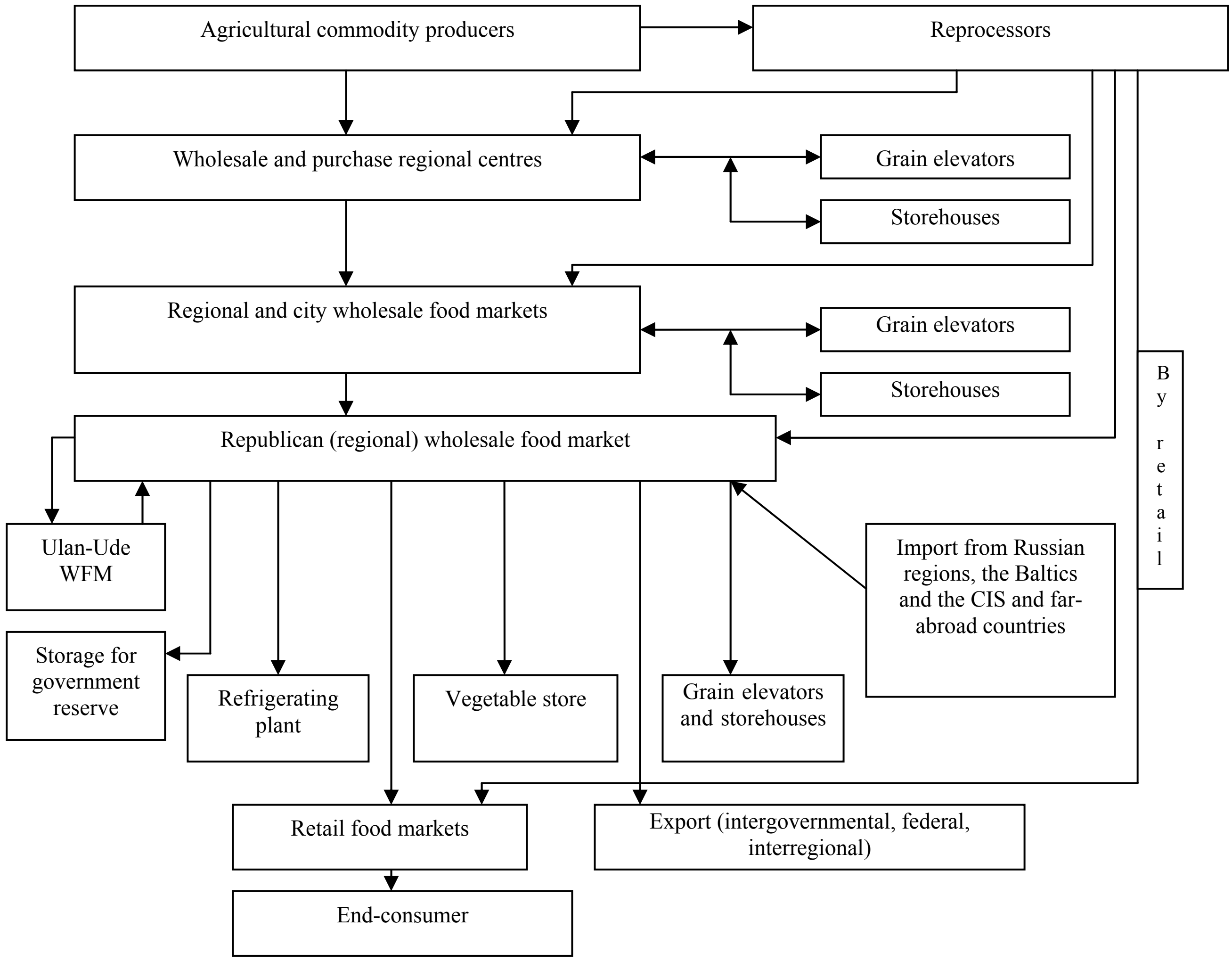


Fig. 4. The scheme of agricultural products and foodstuffs distribution in the system of Buryatia wholesale food markets

The economic basis of the wholesale food market activity is predetermined by its organizational and legal form, which influences both the principles of its initial capital formation and the results of its economical activity. In case of its establishment as a state enterprise the authorized capital stock is formed only at the expense of the state budget and the property of appropriate authorities.

The wholesale market can be organized both on profitable and unprofitable basis. From the point of view of the economic strategy and the wholesale food market purpose, when the market is not the product owner, it is established on the unprofitable basis. In this case its incomes should provide the expenses for the market staff wages, market services, its material and technical basis maintenance and development [6]. All of them should be built into the tariffs.

The wholesale market expenses are provided by the following sources of income: accommodation rent; service rendering payments; the market participants' membership fees; the product quality certification receipts.

Taking into consideration the shortage of means for capital investments wholesale food markets in the republic should be established on the basis of already existing wholesale trade enterprises, fruit and vegetable depots, distribution cold-storage warehouses, which material and technical equipment satisfies the requirements of wholesale markets. They should also attract the territories budgets.

The mechanism of economic relations of wholesale food markets subjects can be different: from the pricing policy coordination to more complicated relations including various forms of futures contracts, selling operations, use of venture capital of foreign investors and great private entrepreneurs [4].

In the first case the wholesale food market provides the commodity producers with the free sale of agricultural products and foodstuffs without making any agreements with the consumers of these products on a security of the wholesale market. In this case there is free pricing on the basis of supply and demand at the moment of the product sale. The market functions come to organization of the service sector (to provide the information about the situation in the market, the product storage, services for the product sale and delivery, certification, etc.).

So, at the given stage the main task of the wholesale food market is to create operating conditions of the competitive market which includes: uncontrollable supply (freedom in production) and uncontrollable price that balances demand and supply, which is possible in all kinds of the Republic wholesale food markets. The price itself does not create the demand and supply balance, but its level warns about unbalancing.

The Republican wholesale market can expand its sphere and forms of relations in the process of its position strengthening; for example, conduct selling operations when agricultural commodity producers transfer their titles in the goods to the market for a certain payment on the basis of a commercial concession agreement due to various reasons (remoteness, employment, incompetence). Selling activity is advantageous for both sides: a producer and a wholesaler. It is a source of investment for the market, and profitable product allocation for the proprietor as the product can be sold more expensively and faster, and if it is necessary it can be improved and packed.

Wholesale food markets can use purchase and commodity interventions to stabilize the market of agricultural products, raw materials and foodstuffs. Problem solving connected with the amount of purchase and commodity interventions, the price level on agricultural products being purchased and sold, raw materials and foodstuffs is the prerogative of Buryatia government.

Executive authorities of Buryatia set activity standards of the republican wholesale market regulation for government agents.

Creation of the republican wholesale food market will provide the commodity producers with organized production distribution, will increase their share in the final price of consumer goods, will promote production volume growth and improve product quality will promote filling the markets with local products, improvement of the social and economic situation in the Republic agro-industrial complex.

The government of Buryatia is working out the issue of a logistic company foundation for the goods delivery to social institutions and formation of the state grain and meat reserves to stimulate local agricultural commodity producers to take part in goods delivery for the state needs. The foundation of such food fund will allow to purchase products for the republican and municipal needs and to regulate prices in the food market.

There is also a possibility to carry out joint competitions to purchase agricultural products for the needs of the republican social sphere.

The Republic of Buryatia is in the region of interests of Asian Pacific Economic Cooperation (APEC) possessing real financial capital, technological capabilities, investment intentions and foodstuffs. Their import is necessary in view of the region geographical position. In its turn Buryatia will have a chance to enter the international food market.

At the same time there occur new possibilities of intergovernmental territories formation with special financial and economic mode. In this case interest of transnational companies in integration of financial, technological and market possibilities of different countries in the local territory raises; the countries-participants' interests in integration for import-substituting and export-oriented policy combination is realized.

It is necessary to develop the system and normative legal acts regulating the relationships of the wholesale food market participants at the level of the Russian Federation and the Republic of Buryatia for the given approach implementation.

Formation and functioning of interstate wholesale food markets is based on the principle of uniform conditions for participating countries, financial and a tax policy, currency transactions, customs regulations, use and conservation of natural resources, an internal infrastructure and management systems. Also it is based on interstate coordination of financial, technological, natural and manpower resources attraction of the participating countries. The state guarantees of the Russian Federation, Buryatiya and other participating countries as well as the system of tax concession and incentives serve as development conditions for domestic and foreign investors.

Transnational companies, city executive boards, large firms of particular countries, governmental bodies, national banks, wholesale markets, associations and corporations can be founders of such wholesale markets.

So, formation of the wholesale food markets system allows to create an effective system of agricultural products and foodstuffs distribution in the region, provides access of agricultural commodity producers to competitive markets and raises their share in the ultimate price of consumer goods up to 60–70 percent, lowers product losses from the field (farm) to the consumer and expenses in the foodstuff distribution system.

References

1. Nuraliev S. U. The formation of the system of wholesale food markets in Russia and the perspectives of their development // Agricultural and processing enterprises economics. 2003. № 7. P. 16–18.

2. Usenko L. N. The theory and practice of food market: Academic and research manual / Rostov state economic academy. Rostov on-Don, 1998.

3. Nuraliev S. U. The organization and development of the system of wholesale food markets in Russia. M. : Rosinformagrotekh, 2002.

4. Kolesnyak A. A., Nekhlanova S. A., Kolesnjak D. A. The concept of development of wholesale food markets system in Buryatiya. Ulan-Ude : Publishing house BGSHA, 2002.

5. Kolesnyak A. A. Food supply: regional aspect. Monograph. M. : Voskhod-A, 2007.

6. Recommendations about the formation of wholesale food markets : Approved by the order № 66 of April, 5, 1994 of the Agriculture Ministry of the Russian Federation.

O. V. Porotova, I. A. Yankina, E. V. Smirnova
Krasnoyarsk State Trade and Economic University, Russia, Krasnoyarsk

EXPRESS-ASSESSMENT OF BUSNESS VALUE FOR MAKING TACTICAL DICISIONS IN MANAGEMENT OF COMPANIES

The management of enterprise value by the method of express-assessment of business value is considerate in this work. The author offers a more accurate formula for calculating the business value by net assets. Also we introduce a system of quadrants, that enable to make management decisions in order to increase business value efficiently.

Keywords: business value, express-method, management of business value, quadrants.

The topicality of the present research is caused by the necessity of using new strategies and methods of controlling companies for accelerating their financial activity effectiveness, steady work in daily and crisis conditions of the market system. At the heart of management concept of the business value is understanding that the increasing of business value strategy has prospects not only for owners, but also for investors, for consumers and society. From such point of view the business value is a resumptive result of enterprise activity. The effectiveness of management clarifies its liquidity, profitability or profit's amount, but and also upwarding business "adjustment", which is the object of financial strategy management.

The aim of the present research is to work out practical recommendations of business value assessment to accelerate enterprise management effectiveness, the offer to use a method of express-assessment of business value as a tool for taking tactical and strategical decisions by managers of companies.

Problems, which have been taken during this research, were rated with reference to medium business class enterprises, which do not have intangible assets on statement of the balance.

Methodical machine of assessment of business value, that exists at the moment, does not allow to define business value fast enough, and also to provide top-management with necessary information about changes in company's value tendencies in process of transformations caused by external and internal factors. Quickness, availability and self-descriptiveness are gaining the more significance in managing of business value.

In a constantly changing internal and external environment of enterprises the efficient definition of business value assumes the character of an objective, focused, continuous process. At the same time, the cost control problem of the company at the stage of stable development as well as in crisis situations stays relevant and insufficiently elaborated. The attention should also be paid to the mechanism of balancing the interests of business and society, the influence of social factors on the dynamics of the value of enterprises.

Company cost management is a modern management strategy targeted at increasing investment attractiveness, competitive advantages and sustainability in the market environment in anticipation of the long term. It is based on a systemic approach of using multiple factors that form the value, taking into account the conflicting interests of different groups of subjects connected with the company. Increasing the market value of the company is the strategic goal of management [1].

To achieve effective governance on the basis of determining the value of business it is necessary to analyze existing methods of valuation. Methods, proposed by foreign researchers, cannot find a proper application in practice for managers of domestic firms at the moment, due to underdevelopment of the equity market and the information base [2]. At the present time there are not so many adapted techniques of value assessment of Russia's business, based on the principle of increasing business value, and each of them has its own nuances that require vast amounts of information. In appraisal practice there are three main approaches to determine the value of the business: profitable, expensive, comparative.

Some appraisers and managers do a prospective analysis of profit on the basis of the retrospective, make adjustments, define by an expert method the proportion of "confidence" to a particular method, and on the base of the three methods by weighting factors we can get so-called integrated index of business value. Immediately the question about the need for a given value emerges. If the owner needs to know the cost of business for management decisions, the method of valuation should posses the following characteristics during the calculation:

– simplicity of calculation;
– self-descriptiveness;
– accessibility for understanding;
– efficiency of definition;
– validity.

Even with very careful check, uncertainty and miscalculation of value of the business will have a place, reflecting a change in the country economics and strategy of the company. The degree of accuracy of assessment will depend on the discount rate (discounted cash flow model), from the stage of the life cycle, from the age of the company. The assessment of the future of the company, based on assumptions about future revenue, will have stochastic nature. Profitable, expensive, or the comparative method of valuation requires the construction of complex multi-factor models based on macroeconomic prognoses, for the calculation by these techniques it requires accessible information about the purchase – sale transaction, the tools of the stock market, which is not always possible in reality. Detailed procedure for determining the "price" of business takes time, effort, financial investment. It also requires the involvement of certified experts-appraisers.

The head of the company has to make decisions that affect the cost of business regularly. Any of the actions of management of organizations, such as: purchase / sale of the property complex, financial investments, sale / purchase of receivables, securities, etc. have huge impact on business value. Modern methods of assessing the value of labor-intensive are complex for perception for an amateur in the evaluation activities. They require a considerable amount of time.

For efficient introduction of these decisions in terms of assessing the business value we suggest to develop a method of express-assessment of the company value. This method is based on accounting data, it allows to take into account the interests of all participants in business processes. The understanding that it is important for shareholders to increase the market value of property is the basis of this concept. It is necessary to adjust its management policies. Investors have to determine the reliability and financial availability of investments. Appraisers have to find the coefficient (rate multiplier) for the industry to use when calculating in the method of analogues. Society has to determine the level of economical development, to take decisions for possible participation in the national IPO.

In our opinion, special attention should be paid to the method of net assets. The authors propose to clarify this method, taking into account the ratio of accounts receivable and accounts payable (to cover the accounts payable for the account receivable) and ratios of profitability of the enterprise.

The method of net assets is one of the methods of the cost approach to value business, the main idea of which is to determine the value of assets with a further subtraction from the amount of assets of all current liabilities of the company.

According to this method the business value is equal to net asset value, which is calculated by the formula:

$$NA = NCA + WA + TFP - LT\&STL, \qquad (1)$$

where NA – net assets of the company; NCA – non-current assets of the company; WA – working assets of the company; TFP – target financing and receipts; $LT\&STL$ – long-term and short-term liabilities [2].

In contrast to existing forms of calculation we suggest a refinement of the net asset formula for determining the value of the business:

$$C = (A - LTL - C\&L - STD / (R/P)) \times \\ \times (1 + Rpr) * (1 + REC), \qquad (2)$$

where C – value business; A – assets of the enterprise; LTL – The long-term liabilities of the enterprise; $C\&L$ – credits and short-term lends; STD – short-term debts; R/P – ratio of receivables and payables; Rpr – profitability, calculated on the profit, REC – profitability of equity capital, defined by profit from sales.

In the revised formula it is suggested to introduce ratio R/P, which plays a significant role in the functioning of the enterprise.

The ratio of receivables and payables, proposed for the calculation of business value, shows how the company is able to pay its short-term debt through the use of receivables, which plays a important role in critical conditions of enterprise activity. It will increase the value of the business if it is more than or equal to 0.5, and reduce the desired rate otherwise.

In addition, it is not the only coefficient, that is suggested by the authors. To clarify it is important to draw attention to the ratios of profitability.

Application of the ratio of profitability of the enterprise plays an important role in calculating the value of the business by the express-method. Profitability of activity shows if the company has positive / negative financial result. This should also adjust the amount of business value, calculated by the proposed express – method. The rate of profitability characterize the position of business in the industry and territorial borders.

The basic idea of calculating the business value is as follows: the cost of business is defined as the product of the difference between the value of assets minus liabilities of the company and the profitability ratio of profit from sales and return on equity capital. Short-term debt in this formula is divided into the ratio of accounts receivable and accounts payable, this action allows you to define how accounts payable may be offset by accounts receivable. Thus, if in the coefficient R / P is more than accounts payable, the value of the business cost is higher, otherwise – the value of cost is less.

The profitability on the equity capital, the profitability of the enterprise, calculated on profit of sales also affect the value of business cost. These coefficients allow to adjust the variable based not only on the value of the assets of the enterprise, but also on financial performance.

Thus, the value of the business in the calculation of the proposed formula (2) will differ from the value calculated by the method of net assets in a big way if the activity of the enterprise is profitable, the receivables will cover the current accounts payable.

To determine the policies of enterprise management based on the definition of business value is suggested to develop a "scenario-model of the enterprise", with the help of which we can define how any decision affects the valuation of the company.

The model is based on two indicators: the liquidity ratio and the ratio of price of the business to net assets of the enterprise.

The ratio of value of the business to net assets (C / NA) allows us to follow the progress trend of enterprise, it shows the effectiveness of the enterprise management. It also reflects the financial attractiveness of the company. According to proposed calculation by the formula (2), the value of the business should exceed the value of net assets in the presence of competent management and implementation of financial and strategic operations, thus, the coefficient of C / NA must be higher than 1 (the limiting value is 1).

To assess the solvency of companies is used the coefficient of current liquidity. The coefficient of current liquidity is defined as the ratio of the factual cost of circulating assets, including stocks, goods, cash, receivables, production, etc. to short-term liabilities.

The coefficient of current liquidity is used to assess the ability of companies to meet their short-term obligations. The liquidity ratio characterizes the ability of the company to pay not only at the present moment, but also in case of emergency.

Standard value for the coefficient of current liquidity is from 1 to 2, the value less than 1 indicates a possible loss of paying capacity, and a value more than 4 is on insufficient using of borrowings and as a consequence, a lower value return on equity capital. It should be noted that the standard value for the coefficient of current liquidity depends on the industry in which the enterprise operates [3].

Figure 1 shows a system of quadrants, where on the *Y* axis the ratio of price of the business to net assets of enterprise is shown, on the *X*-axis we can see the coefficient of current liquidity. In this system imposed boundary values of defined indicators are brought in, the boundary value of coefficient of *C* / *NA* is 1 (optimal value is more than 1), the ratio of current liquidity is 1 to 2.

In figure it is appeared that the boundary values of the indices divide the area into four quadrants.

In the first quadrant the liquidity ratios and the ratio of business value to net assets of the company below the normative values (*KC* / *NA* < 1, *Kliq* < 1), we will call this quadrant "nobody nothing" – "Neither for me nor for the people". Such a state of financial affairs of the company shows the inability of the enterprise to meet its obligations, the fragility of the financial situation, the lack of the strategic value thinking in the company's management.

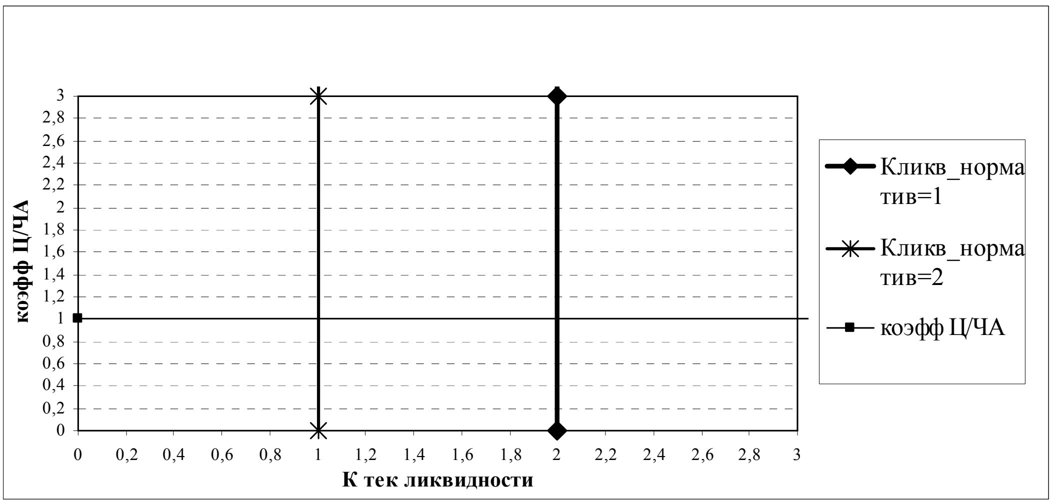
In the second quadrant, the liquidity ratio is below normative values, the ratio of *C* / *NA* is more than 1 (*KC* / *NA* = 1, *Kliq* < 1), let us call it "everybody may be" – "The cost increase at any price"; disparity in the optimal value of the coefficient of liquidity shows the temporary loss of solvency of the enterprise, but the activity of the company is profitable. Management of the organization increases the profits, at the same time paying no attention to the excess growth rate of short-term liabilities over current assets.

In the third quadrant all indicators are favorable at the optimal level. Finding a company in this quadrant indicates competent policy management of the company in terms of the value thinking. Calculations under the third quadrant – *KC* / *NA* = 1, 1 <= *Kliq* <= 2, we will call it "GOOD TOP" – "For people and for me". Estimated location "of the enterprise in these boundary values is advantageously for owners, and investors, and government fiscal organizations, and for society, and it also positively characterizes the company in credit institutions.

In the fourth quadrant the liquidity ratio satisfies the optimal value, *C* / *NA* below 1 (*KC* / *NA* < 1, 1 <= *Kliq* <= 2), so let's call it "No Value" – "No for the values decision". The policy of the leadership in this case allows you to save the current solvency of the enterprise, but does not think about increasing business value.

Thus, using the proposed system of quadrants, leaders can quickly make decisions that will effectively manage the business on the basis of determining the value of the business. The suggested express-method enables to predict the evolution of the price of business, at the same time paying attention to the level of solvency of the enterprise. In a crisis situation, the proposed dependence of the two factors is the ratio of business value and net assets, the liquidity ratio makes it possible to objectively evaluate the process transformed system of functioning of the enterprise, to make decision to mobilize their own resources to restore solvency.

Management strategy based on the approach of increasing the cost of business is one of the most effective, because it allows you to monitor the dynamics of this indicator, which serves as an indicator of financial-economic activity of the enterprise [4].



System of quadrants

Top managers come to the conclusion that the main characteristics of successful and forward-looking company are not only its liquidity, profitability, profit (like it has been considered earlier), but also the increasing of business value (cost of the business). It means that the main task of competent manager is to increase the "price" of the business, which will be the subject for monitoring, strategic management and investment interest of the State and people [5].

So, using the express-method of assessment of business value, management of the enterprise quickly and without considerable expenses can define the "boundary" minimum rate of market cost of the business, and also choose factors, which have an impact on it.

The value of the business, defined by the express-method can be served as minimal cost of the business and used by professionally licensed appraisers for more through and detailed calculation of business price.

For administrative and financial managers it is recommended to pay attention to the business value as controlled indicator, which has a great influence on all results of company's work, such as increase of investment attraction, financial risks managing, forming of funding sources, social part of company's activity.

In conclusion of the present research it is significant to point out at results received by authors:

1. It was suggested to use the express-assessment of business value as the way to increase the effectiveness of company management.
2. It was reasoned that cost management is a process that considers interests of different parties, such as owners, investors, consumers, society, government of territorial entities of Russia.
3. The calculation formula of business value by the method of net assets was modificated, coefficients, which have great impact on value of required indicator, were included.
4. The system of quadrants, which allows to discover management features of business value in different situations, was proposed.
5. It was concluded that the express-method of defining business value has to suit principle of system and guarantee mobilization of all solutions to achieve strategic and tactical goals.
6. It was pointed out, that during crisis situations the express-method of assessing the business value and accepted decisions based on it should be aligned on the preservation of index of value, not on its increase.

The practical consequence of this research is that recommendations about using the express-method to calculate the business value give an opportunity to inform owners and managers of enterprise about business cost dynamics, that can allow to increase the effectiveness of managing companies.

References

1. Valdaicev S. V. Estimate of business and management of company's value. M. : UNITI-Data, 2007.
2. Gryaznova A. G. Estimate of business. M. : Finance and statistic, 2006.
3. Damadaran A. Investment Valuation. Tools and Techniques for Determining the Value of Any Asset. M. : Alpina Business Books, 2008.
4. Efimova O. V. Financial Management. M. : Accounting analytics, 2002.
5. Kouplend T. Business Value, estimation and management. M. : Olimp-Bussnes, 2007.

L. A. Safonova, G. N. Smolovik
Siberian State University of Telecommunications and Information Sciences,
Russia, Novosibirsk

CONCEPTUAL MODEL OF THE INVESTMENT PRIORITIES ESTIMATION IN INFOCOMMUNICATIONS

We offer a new conceptual model for investment priorities estimation. It provides the analysis in a regional profile, industry profile and on a micro level. The model combines a number of economic-mathematical methods of modeling, including cluster and multivariate regression analyses. It is recommended to use theory and systematic tools of the real options concept for the priorities estimation on a micro level.

Keywords: an investment priority, a cluster, a rating, risks, a real option.

Nowadays the problem of investment priorities estimation has become topical. Its urgency is defined by the necessity of socioeconomic development of Russian regions and by the multichoice of capital investment spheres. Shortage of financial resources explains the necessity to stimulate investment mobilization into the infocommunication complex and their efficient use. Lack of reliable methods and approaches to identify investment priorities, appropriate to socioeconomic changes, mechanisms and ways of investment management on a regional and industry level makes development of a reasonable strategy of investment activity for telecoms operators rather complicated.

Russian economic system transformation and its integration into the world information and economic community require studying of the world experience and its use in accordance with special features of our country. In spite of the importance of the problem stated above, issues of investment priorities substantiation and choice are still at the formation stage. There are discussions among analysts concerning criteria, principles and methods of performing appropriate estimations. Taking into consideration that the term of investment priority has a wide interpretation in scientific literature, this work suggests the following definition:

Investment priority is an integral characteristic of a set of factors (social, economic, organizational, legal, political), determining investment expediency in this or that economic system.

This term has an intermediate position in the investment policy between purposes and their implementation tools and plays the role of a guideline in solving investment problems. Substantiation, choice and realization of investment priorities are one of the investment management functions, which implementation is realized in the process of developing, making and realization of management decisions. The systematic mechanism for determining investment priorities is being actively worked out. There are different variants of its classification.

From the practical point of view, the classification according to the level of the problems solved deserves the greatest attention. In accordance with the stated criterion, systematic tools are divided into three groups:

– methods of determining investment priorities on a macro level (national priorities);

– methods of determining investment priorities on a mezo level (regional and industry priorities);

– methods of determining objects for investment on a micro level.

As a rule investment priorities on a macro level are represented by means of rating estimations. The research of the world countries conducted by Harvard business school was one of the first in this direction. The ratings published by international economic journals "Euromoney", "Fortune", "The economist" are well known and are often given in economic literature. To make up practically all ratings expert judgments are used to a certain extent. The stated ratings are used as a criterion for making a decision concerning investment in this or that country.

Rating estimations have become widespread, but it should be noted, that there are some drawbacks in the approach. The rating calculation includes a number of factors which are weakly reasoned and as a rule are restricted by available content of statistical information and by the size of allocated financial resources. Russian and foreign researchers, working in the field of determining investment rating of Russian regions, execute operations with a large number of rates – from 30–70 to 200. It does not provide more profound study of the phenomenon considering multicollinearity of the rates inevitably arises in this case. A complicated interpretation of the results achieved during the estimation produces a certain portion of pessimism. The total integral value does not enable us to trace the cause-and-effect relations and trends of the development of this or that investment complex sufficiently back. Finally, this drawback negatively influences the substantiation of the calculation results. In quite a number of cases the rating calculation procedure is "non-transparent" and does not give the opportunity for the investor to estimate to what extent all the factors taken into consideration when calculating the rating are relevant to the purposes of his investment.

Taking into account high regional differentiation of Russia, the approaches to investment priorities determination on a mezo level deserve special attention. The existing systematic approaches for investment priorities estimation in regional and industry profile borrowed the western traditions to conduct such kinds of research in many respects: indicators system, ways of their quantitative assessment, standardization procedure, total "weighing" and summation. "Expert region" consulting agency has a leading position among Russian rating agents. In compliance with the procedure of this agency, the rating is determined as interrelated estimation of two basic components: risk and potential. The way of using economic mathematical methods is less significant.

Development of the conceptual model and as a result a choice of a computing method must be performed through a number of the estimation objectives. This process influences a list of the factors, criteria and rates of the investment priorities estimation taken into consideration. In the context of the problem considered in this work the investment priorities estimation is performed for an effective strategy development of the investment activity management in the field of infocommunications. It allows to eliminate information inequality both between Russia and the most developed countries of the world community and particular regions of our country. In compliance with the set purpose we propose the conceptual model of investment priorities estimation shown in fig. 1.

According to one of the main paradigms of financial management, while taking a decision an investor estimates an acceptable for him correlation between expected investment profitability and a risk. Depending on reasons of occurrence and possibilities of elimination we should single out two components of investment risks: specific (commercial) and nonspecific (country, region, industry).

Specific, or commercial investment risk, is a variant of entrepreneurial risk connected with the investment activity and caused by peculiarities of commercial (market) activity specific for each investment project or particular enterprise (corporation). Risks spectrum connected with realization of infocommunication projects is wide enough. Let us consider the primary risks.

Risk of a technical development strategy. Any infocommunication network requires further development. In its turn it can be both quantitative and qualitative. Risk of a technical development strategy has two aspects respectively. First of all, it represents danger of advanced growth of the subscriber database in comparison with the new capacities entry rate, the network modernization and optimization. The reason of such a situation can be both a wrong forecast and undue equipment delivery that can lead not only to the failure in the network operation, financial losses but also to the reputation loss. Project partners must be bound by strict contractual obligations which exclude the possibility of undue delivery. Quantitative development implicates numeral increase of network elements which is enough for the incoming load servicing to comply with the stated quality ratings without failure. In addition, there is potential danger of the operator's inability to satisfy users' mounting needs on the basis of the available technology (standard). For example, at this conjuncture it is insufficient simply to provide a subscriber with qualitative and reliable telephony.

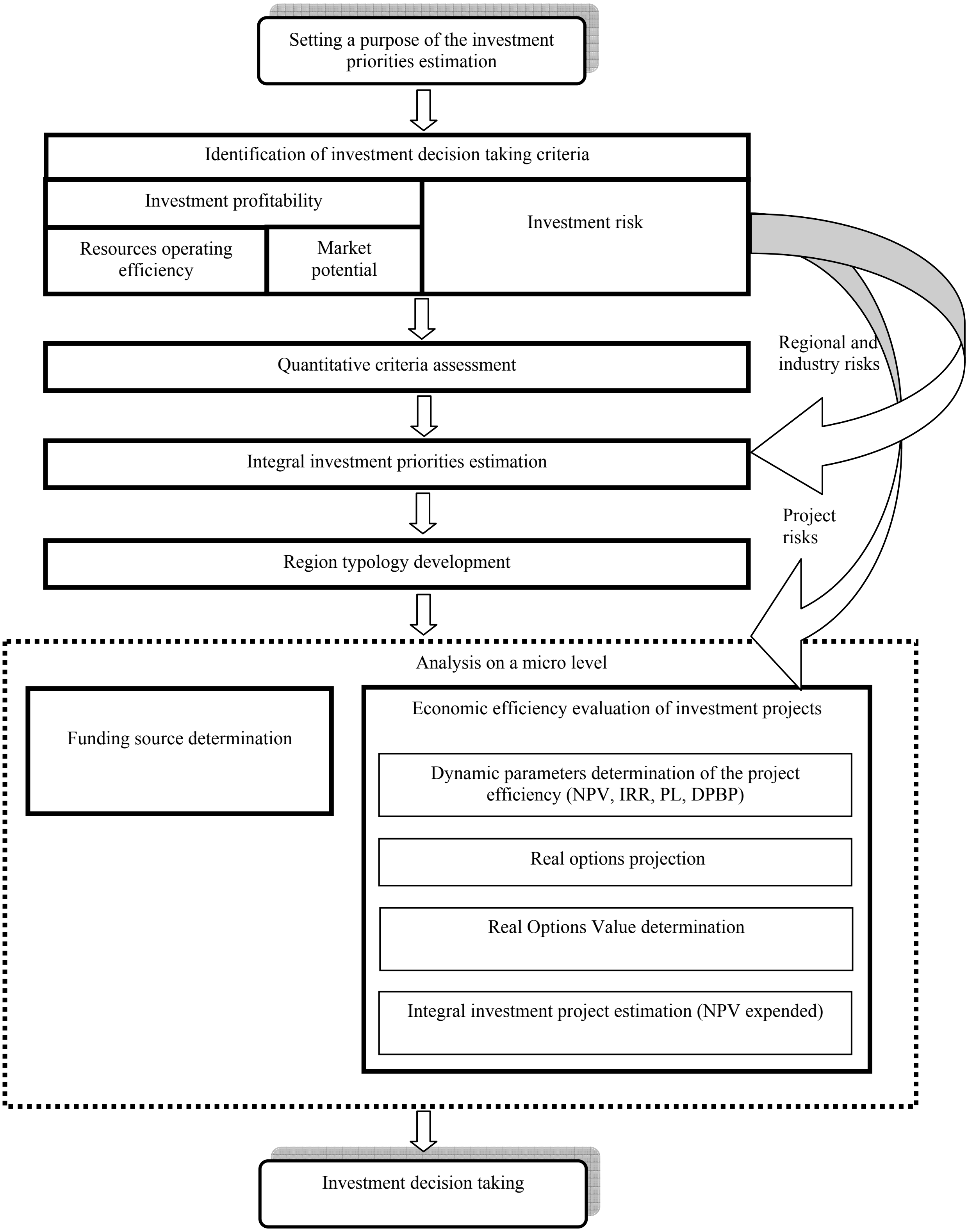


Fig. 1. Conceptual model of investment priorities estimation

New services are getting more and more demanded, for example the multimedia data transferring. In this connection, the problem of the innovation technology implementation on telecommunication networks is highly urgent nowadays.

Market risk is connected with the conditions when an investment project is being realized. It has a possibility of an unfavorable effect of external environment on the company financial performance and its ability to achieve aims. The example is reduction of people's paying capacity level. The further situation of the company in this situation will depend on the due reaction to the changes occurred. To reduce the risk the results of the marketing research are given in the corresponding section of the project business plan, including:

– general characteristic of target markets, their capacity evaluation, market development dynamic and the forecast of its future trends;

– consumers' basic requirements to services;

– competitors' power assessment and their basic services data (tariffs, quality level);

– technological and financial state of competitive organizations and the extent of their influence on this service market.

Marketing policy risk takes place as a result of strategic decisions acceptance or non-acceptance regarding the company marketing policy. The probability of taking inadequate decisions is growing when the competition on a particular market segment is increasing. Operators' growing competition can cause loss of a market share or the financial state deterioration.

Technical risk is a probability of equipment failure. To reduce this risk it is necessary to perform regular monitoring of the operated equipment and improve the service staff's qualification.

Risk of human resources has a few reasons. First of all, it is connected with possible mistakes in the quality assessment of human resources and infringement of conditions of their use. The object understaffing with sufficient quantity of qualified staff can lead to the increase of the required investment volume, because there is necessity to train it, increasing probability of equipment failures and its ineffective operation. The variant of this risk is danger of human resources loss connected with the possibility of new companies setting up which needs advanced and qualified staff. To control this kind of risk it is advisable to undertake measures to improve social policy, career prospect and staff training policy.

Financial risk is connected with the scarcity of financial resources to realize the company's tasks, inability to fulfill obligations and liquidate indebtedness.

Nonspecific or non-commercial risk is caused by external conditions of macroeconomic, regional and industry patterns in relation to the investor.

Macroeconomic (country) and industry investment risks influence equally the probable results of all investment projects implementation in a particular country and in a particular industry respectively. Thus they have no influence on the results of the comparative investment appeal estimation and therefore may be excluded from consideration as the permanent impact factors. Regional risks and risks of a particular project must be taken into account while choosing the procedure of economic efficiency assessment. Specifically, it is suggested to use the optional approach in estimating high-risky and non-standard projects.

Investment profitability is defined, from our point of view, by the work efficiency of industry recourses and potential market capacity. It should be noted, that there is a specific time gap between the moment of the work efficiency of the industry recourses estimation (they can be determined only on the basis of statistic data) and the moment of the investment decision making. The obtained level of the industry recourses work efficiency is determined by its specific character, industry management state and potential market capacity. In its turn, the market capacity (especially infocommunication market capacity) has a constant tendency to its variation. Underestimation of its influence on the work efficiency of industry recourses and, as a result, underestimation of telecom operators' work efficiency may lead to a wrong strategic decision.

Market potential assessment enables us to correct existing work efficiency of the industry recourses taking into account a strategic outlook. In this case, market potential is considered as required capacity of the growing number of access points to the information networks which will give an opportunity to provide economically balanced development of Russian infocommunication complex.

To build the regions typology in this work we used a method of economic mathematical modeling which is based on combining the cluster analysis and the multiple regression. Parameters of infocommunication infrastructure development (stationary telephone density, mobile telephone density and personal computers density) were chosen as factor variables. In general, the model can be presented in the following way:

$$\text{GRP per capita} = a\ (FD)^{b_1} \cdot (MD)^{b_2} \cdot (DC)^{b_3}.$$

Parameters b_1, b_2, b_3 have precise economic interpretation. They are the elasticity coefficients and show how much the result (GRP per capita) will change in average with 1 % change of the corresponding factor when the other two factors are constant. The exponential model making was preceded by the procedure of parameters linearization which was done by taking logarithms. To solve this task the determinant method was used. As a result, we got the following multiple regression equation for the universal set (77 regions):

$$\text{GRP per capita} = 8409\ \cdot FD^{0{,}812} \cdot MD^{(-0{,}241)} \cdot DC^{0{,}211}$$

The analysis of the obtained equation allows us to make a conclusion that the highest gain will be got when investing into development of stationary electrical communication. Taking into consideration high heterogeneity of the regional infocommunication infrastructure development it is advisable to complete the obtained results by making the corresponding multiple regression equations for more homogeneous groups (clusters).

The highest value of entropy was obtained when performing the cluster analysis by Word method. Squared Euclidean distance was used as a measure of convergence. To define the number of clusters objectively existing in the universal set, *E*-criterion calculation was made. We can observe abrupt increase of *E*-criterion at step 71 of integration. Thus the number of clusters is 6. The made analysis of clusterization results showed that out of the regional universal set the Ingush Republic, Samara Region and the Chukchi Autonomous District stand especially out. They form clusters of one object. Moscow and Saint Petersburg are incorporated into a separate cluster. The multiple regression equations for filled clusters are given in table.

Multivariate regression models

Cluster rank	Multiple regression equations
1	GRP per capita = 2,889 $\cdot FD^{1,172} \cdot MD^{1,005} \cdot DC^{0,333}$
2	GRP per capita = 252,03 × × $FD^{1,562} \cdot MD^{(-0,023)} \cdot DC^{0,193}$
3	GRP per capita = 1 550 $\cdot FD^{0,669} \cdot MD^{0,311} \cdot DC^{0,175}$

Analysing the multiple regression equation for the first rank cluster we can make a conclusion that the highest GRP per capita gain will be observed when the funds are invested in the stationary electrical communication development, that is when the stationary telephone density increases. When the funds are invested in the mobile communication development in the regions of the second rank cluster we will observe the GRP per capita decrease. This fact can be given the following explanation: as the regions of a particular cluster are characterized by a high level of mobile communication development, we can claim that further density increase will attract the “little speaking” telephone subscribers that will negatively influence the financial results of the mobile companies activity. In the regions of the third rank cluster, the highest gain will be observed when the funds are invested in stationary electrical communication development, that is when stationary telephone density increases.

Fig. 2 shows elasticity coefficient values in the multiple regression equations according to clusters. It allows us to make an optimal decision about the funds investment in infocommunication development.

Except economic efficiency dynamic indicators calculating (NPV, IRR, PI, DPBP) the investment priorities analysis on a micro level assumes projection and calculation of the Real Options Value. In spite of the fact, that the Real Options Theory is a new direction in the field of investment analysis, there is quite a large number of sources of its theoretical and practical application in the world literature. The works by F. Black and M. Showlz, A. Damodaran, R. Merton, D. Moon, A. Dixit and R. Pindyck, N. Kulatilaka, D. Ingersoll and S. Ross, L. Trigeorgis and the works written by a number of other foreign researchers can be referred to as classical works devoted to the option approach. This problem is less studied in our country. The emergence of the real options concept is explained by the drawbacks of the traditional approach to efficiency evaluation, which supposes passive project management and does not take synergistic effects into account. According to the traditional approach, all that a manager has to do is to follow the project in accordance with the plan developed beforehand. Thus, the managers’ ability to make appropriate decisions in future is excluded from the investment project estimation.

Contrary to the traditional method, the option approach takes the administrative flexibility into account because it considers the investment project to be a system of options. Flexibility is the possibility to change the made decision in the broad sense of the word has its advantage. The more such possibilities are there in the project, the more valuable the project is.

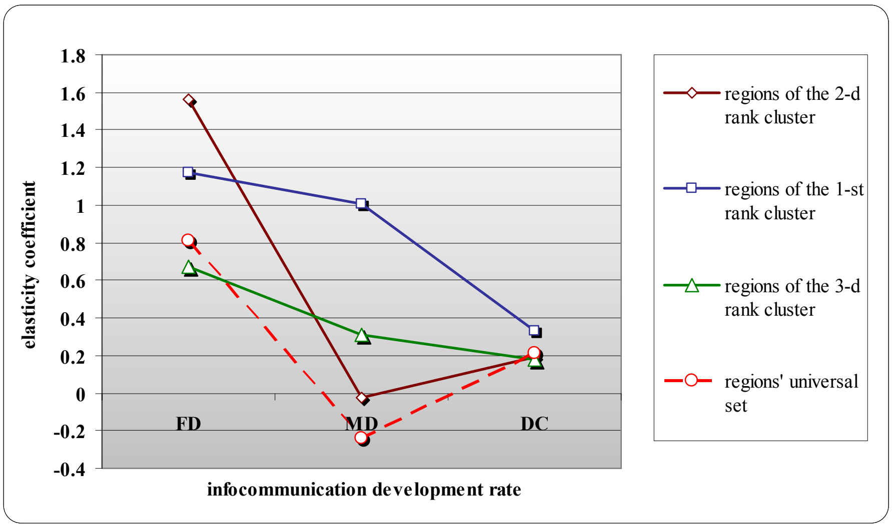


Fig. 2. Elasticity coefficient values by clusters

Conceptually the net present value of IP can be presented as a sum of NPV indicator, calculated according to the traditional procedure and a value of administrative options included in the project. It can be presented as the following formula:

$$NPV_{exp} = NPV_{tr} + ROV,$$

where NPV_{exp} (Expanded NPV) – expanded net present value of IP; NPV_{tr} (Traditional NPV) – net present value, calculated by traditional method; ROV (Real Options Value) – real options value.

There is a large number of methods and models of real options estimation, the most part of which supposes use of a rather complicated mathematical apparatus, in particular stochastic mathematics, which makes their practical use difficult. The most practicable, from our point of view, is the binomial method and Black-Showlz model which were considered in details in [1–3]. The real options concept allows us to estimate the project possibilities quantitatively and thereby include them in the project efficiency estimation. It should be noted, that quantitative estimation plays the key role in the investment decision making.

The conceptual model suggested in this work is assigned to estimate investment priorities and assumes the analysis realization on several levels. It is proposed to study the development mechanism in the regional and industry profile by means of the multiple regression equation. It will give the opportunity to perform the regions positioning. It is recommended to analyze investment priorities on a micro level using the systematic mechanism of the real options concept. All this, from the authors' point of view must provide reasoned investment decision making and the choice of the optimal strategy of the regional infocommunication infrastructure development.

References

1. Safonova L. A., Smolovik G. N. Economic efficiency of investment projects. Methodology and tools of estimation : monography / SibSUTI. Novosibirsk, 2007.

2. Safonova L. A., Smolovik G. N. Use of binominal approach to real options value estimation // Prospects of modern means and telecommunication systems development. Irkutsk, 2006. P. 147–158.

3. Safonova L. A., Smolovik G. N. Use of real options choosing investment project development of telecommunication networks // Infocommunication and computing technology and systems : materials of the II All-Russian conf. with international participation. Ulan-Ude, 2006. C. 117–121.

R. N. Shevelyova
Krasnoyarsk State Agrarian University, Russia, Krasnoyarsk

TECHNIQUE TO ESTIMATE AND FORECAST LIFE QUALITY OF POPULATION

The article presents the expediency to apply life quality indicator of efficiency for activity of authorities and also the essence of the estimation technique developed by the author and forecasting of life quality of the population with regional features taken into account.

Keywords: life quality of population, estimation of life quality, forecasting of life quality.

Since late 80s the theory and sustainable development practice are in the centre of attention of scientists and politicians in our country and abroad. The tendency to design regional (and even municipal) sustainable development programs which began in the mid-nineties in Russia is still in trend. As a rule the goal sets of these programs have regional concretization and are directly focused on use of available preconditions to stabilize and improve the economic and social situation of the corresponding territories. The question of indicators and criteria of regional stability is in fact open.

Considering and analyzing various approaches to an estimation of stability of social and economic systems [1–3], the author comes to the conclusion that all the offered techniques focus attention on the process and development indicators, but do not answer the question "what for?". After all, the given estimation is necessary not only to compare the level of social and economic development of the separate countries and regions and to drawing up their ratings. Today high quality of life of the population should become the overall objective of sustainable development. The importance of life quality problem is increasing in Russia because the human resource in the conditions of progressive ageing and depopulation becomes the most scarce resource. Last version of the long-term demographic forecast of the United Nations shows that in the long term the population of Russia will be reduced, the middle age group will go up and the able-bodied population share will go down [4].

Thus, in the conditions of depopulation and ageing the problem of life quality maintenance is particularly urgent. Achieving and maintaining high quality of life will provide improvement of health and increase of life span of the population, rising of educational level, birth rate growth etc., and all these in their turn will contribute to improvement of manpower quality which is the necessary factor for sustainable development of both separate regional social and economic systems and the country in general.

The reference to problems of life quality estimation is necessary to research economic possibilities of the countries, and also for the analysis of development prospects of the human capital. This estimation is extremely important to define the development level of social sphere as well as to reveal potential possibilities of the country, region. Therefore what is urgent here is formation of a new management paradigm that is quality management of life. In this connection, it is expedient as the main criterion of a sustainable development to use such integrated indicator as life quality of the population as the given indicator

– allows to formulate the purpose of steady regional development accurately;

– is the sensitive indicator of economic, financial, social, ecological, etc. changes in the country and the region;

– gives the chance to strengthen a social orientation of the regional policy and regional development, puts on the first place the social purposes of a society instead of material;

– allows to estimate efficiency of state and municipal management bodies;

– transforms economic growth into the main tool to achieve the social goals of the society.

Considering life quality as an overall objective and the basic indicator of sustainable development of regional social and economic systems, it is necessary not only to define indicators of its estimation but also to develop forecasting tools as forecasting is basis for planning and hence for improvement of life quality.

The analysis of foreign and domestic techniques to estimate life quality [5–10] has allowed to formulate the following conclusions:

– foreign and domestic scientists conduct active work in sphere of designing methods to estimate quality of life;

– every year the world community pays more and more attention to life quality of the population; achievement and quality maintenance of life is the purpose of all developed countries of the world;

– existing techniques considerably differ by quantity and structure of indicators (the quantity of indicators varies from three to several tens, and their structure includes indicators of economic, social and physiological components of quality of a life);

– the majority of the considered techniques is estimated only by objective indicators of life quality and do not consider the subjective ones;

– all the considered techniques allow to estimate only stand-alone qualities of life of the population and cannot apply for universality.

Proceeding from the above-stated, it is possible to say that for today creation of the unified technique to estimate life quality with reference to the Russian regions owing to different natural and climatic, cultural, historical, social and economic conditions of their development is not obviously possible, therefore it is expedient to estimate life quality by the indicators reflecting the most essential factors of ability to live in the given region.

The characteristic economic and geographical position; natural and environmental conditions; natural resource potential; demographic potential and population structure; structures and economy specializations; financial security of Krasnoyarsk region, and also typology of the factors produce positive and negative impact on life quality of the population (tab. 1). All these contributed to the presented allocation.

To estimate life quality it is necessary to choose indicators which to the greatest extend reflect negative factors of the ability to survive. It will help to estimate degree of negative influence, its dynamics that, in turn, will allow authorities to allocate priority directions of the social and economic policy in region.

From tab. 1 it is obvious that the greatest negative influence on life quality of the population is given to Krasnoyarsk from severe environmental conditions, the big extent from the north on the south and from the West on the east, the high industrialization, close to monoprofile character of development of economy. Thus, it is necessary to choose from all the variety of indicators of life quality of the population the ones that most objectively reflect the revealed negative factors.

Taking into account regional features and the requirements, shown by social and economic indicators [11], the author suggests estimating life quality of the population of Krasnoyarsk region in four basic components: population health, availability of social services, standards of living, ecological conditions, with social aspect of the economic indicators presented in tab. 2.

Indicators of life quality presented in tab. 2 are applied to calculation of the general indicators for each component and a complex indicator by methods of the average geometrical and the average arithmetic weighed accordingly.

The estimation of life quality is only one of the management stages, being the basis for standard (target) forecasting. For effective forecasting of life quality of the population the major question is the choice of forecasting method. The optimum method should meet following requirements:

– to provide functional completeness, reliability and accuracy of the forecast;

– to reduce expenses of time and means of forecasting;

– to consider features of the forecasting object, the main feature of life quality as forecasting object depends on a great number of factors (tab. 1).

The analysis of forecasting methods [12–16] allows to draw a conclusion that substantially method DEA (in English the name of this method sounds as Data Envelopment Analysis) meets these requirements.

Method DEA is based on application of methods of linear programming to create a non-parametric linear surface on the basis of certain data. This method has arisen as generalisation of simple factors to analyse activity in the multidimensional case, i. e. activity of complex object is described by a set of entrance parametres (x_1, …, x_m) and a set of target parametres (y_1, …, y_r). For the sake of correctness and pithiness of such statements the set of similar complex objects is considered. Then mathematically such approach will be reduced to the decision of the big family of optimising problems. The founders of the given approach are American scientists A. Charnes and V. Cooper.

Table 1

The factors influencing life quality of the population of Krasnoyarsk region

Economic and geographical position	
Positive influence	The region is located on crossing of the major transport ways that considerably simplifies possibilities of moving not only in the country, but also beyond its boundaries
Negative influence	The big extent from the north to the south and from the West to the east creates a territorial problem of availability for medical, educational and other services, especially for the population of peripheral areas of the region
Natural and environmental conditions	
Positive influence	In the south of the region there is warm summer and moderately severe winter with little snow falls. Dry pure air, an abundance of sunny days in the summer, healing waters of sources and numerous lakes create favorable conditions for building of resorts, sanatoria and bases of rest
Negative influence	The considerable part of territory of the region concerns regions of the Far North, adverse to reside owing to extreme natural and environmental conditions; negative influence on health of people, lack of possibilities for alternative employment and the all-the-year-round transport land message, essential branch ruptures in payment
Resource and raw potential	
Positive influence	Rich in resource, the raw potential is capable to provide throughout the foreseeable future the considerate incomes in the regional budget as well as financing of the expenses directed on improvement of life quality of the population
Demographic potential and population structure	
Positive influence	Age structure of the population is younger than in the average all over the country
Negative influence	The steady tendency of reduction of the population; life expectancy is below the all-Russian indicator; high death rate from the external reasons (a trauma, suicide, alcoholic poisonings); illnesses of blood circulation system; new growths; excess of birth rate of the death one
Structure and economy specialization	
Positive influence	Export branches of the economy in the region provides higher incomes to the population; the modern level of the economy development, realization of investment projects provide high employment of economically active population
Negative influence	The regional economy is close to monoprofile (nonferrous metallurgy defines the industrial profile of the region), it leads to demand for traditionally "man's" specialities, complicating employment of women; high branch and gender differences in payment which is characteristic for the industrial, export-focused regions; concentration of the most productive parts in two cities (Norilsk and Krasnoyarsk); environmental problems, characteristic for the majority of industrial regions; depression of the economy in the northern territories

Table 2

Components and social and economic indexes of life quality

Components of life quality of the population	The social and economic indexes characterising components of life quality
Population health	Pre-supposed life expectancy; Death rate from the unnatural reasons; Death rate from illnesses of blood circulation system; Death rate from new growths
Availability of social services	Coverage by preschool centres; Coverage by educational institutions; Coverage by establishments of primary, secondary and higher vocational training; Coverage by public health services establishments; Coverage by establishments cultural type, libraries, museums, theatres etc.
Standards of living	Level of incomes of the population; Security habitation; Branch differences in payment; Gender differences in payment
Ecological conditions	A condition of water resources; A condition of atmospheric air; A soil condition

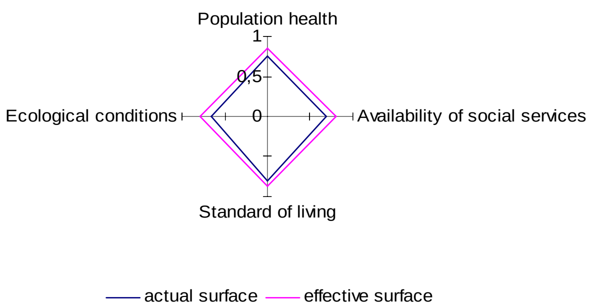
Among the basic advantages of DEA method there is giving the grounds to apply it to forecasting life quality, it is also possible to name the following:

– possibility to estimate efficiency and forecast taking into account a considerable quantity of inputs and exits that allows to avoid necessity to calculate the uniform result indicator or the indicator of resources expenditure;

– possibility to define each forecasting object of optimum volume of inputs or exits, which should reach peak of efficiency;

– no need in the subjective task of the functional form of an effective surface, and also the form of distribution at random error.

In general the main idea of DEA method with four components of life quality (inputs) is presented in figure.



Comparison of actual and effective surfaces

The actual surface is under construction on the basis of indicators of life quality of the investigated region, and the effective surface based on the information of regions, advanced on these indicators. The comparative analysis of actual and effective surfaces gives the chance to carry out target (standard) forecasting, i. e. to define a desirable degree of life quality of the population in the future; ways and terms of achievement of optimum indicators of life quality defined as the purpose.

Schematically offered technique to estimate and forecast life quality of the population taking into account regional features is presented in tab. 3.

The important advantage of the offered technique to estimate and forecast life quality is that applied indicators correspond to the requirement of information availability and enter the nomenclature of the official statistics (or are calculated on values of the last); besides the offered technique:

– considers important regional features; for an estimation of life quality of the population of Krasnoyarsk region the indicators reflect factors producing the most essential impact on population ability to live are selected;

– characterized by simplicity of application and speed of reception of results;

– allows to trace dynamics of life quality of the population of the region, to carry out inter-regional and intraregional comparisons to make ranging of intraregional administrative and territorial units on life quality;

– considers value judgment of the importance of separate components of life quality for the population;

– is flexible and dynamical as at change of the factors influencing life quality of the population in the region, gives a chance to change or add the list of applied individual indicators.

Table 3

Evaluation stages and forecasting of life quality of the population

Sources Information	Stage	Applied methods
The official statistical reporting of the analyzed region	1. Revealing of the factors making the greatest negative impact on life quality of the population	Analytical methods
	2. Definition of components and the indicators reflecting revealed negative factors	
	3. Calculation of the generalised indicators on each component of life quality	Average geometrical
Results of a stage 3, results of expert interrogation	4. Calculation of complex indicator of life quality	Average the arithmetic weighed; a ranging method (for reception of factors of weightiness)
Results of stages 3, 4	5. Construction of an actual surface	Method DEA
The official statistical reporting of regions, advanced on analyzed indicators of life quality	6. Construction of an effective surface	
Results of stages 5, 6	7. The comparative analysis of actual and effective surfaces	
Results of stage 7	8. Definition of the desirable degree of life quality of the population in future; ways and terms to achieve the optimum indicators of life quality defined as the purpose	Analysis and synthesis methods

References

1. Adams R. Performance indicators for sustainable development / Accounting and Business. 1999.
2. The limiting to growth / D. H. Meadows, D. L. Meadows, J. Randers, W. W. Behrens. Potomas. 1972.
3. Bobylev V. Information and methodical basis for calculation of ecological and economic indicators. M. : Moscow State University Publishing house, 2000.
4. Jakovets Y. About the combination of long-term forecasting and strategic planning // Economist. 2008.
5. Zhukov N. V. Indicator of social development as the tool of social programming: foreign experience // Sociology. 1994. № 3–4.
6. Policies of incomes and life quality of the population / under the ed. N. A.Gorelov. SPb. : Peter, 2003.
7. Mstislavsky P. S. Social parameters in comparison to European countries // The Standard of living of the population of regions of Russia. 2003. № 2
8. Osipov G.V. Sociology and social myth. M, 2002.
9. Ajvazjan. S. A. Indicators of life quality of the population: their construction and use in social and economic management and inter-regional comparisons. M., 2000.
10. Kolbasina A. About the technique to estimate life quality of territories (by the example of Krasnoyarsk) [Electronic resource]. Cop. 2010. URL: http//www.ram.ru/activity/comp/bp2003/files/std09.pdf /.
11. Borodkin F. M., Ajvazjan S. A. Social indicators. M. : It juniti, 2006.
12. Trofimov A. M., Demakov A., Mustafin M. R. Forecasting in economic geography. Kazan : Publishing house of the Kazan university, 1990.
13. Egors V. V., Parsadnov G. A. National economy forecasting. M. : Infra-TH, 2001.
14. Forecasting and planning in the market conditions / under the ed. of T. G. Morozova. M. : It juniti, 2001.
15. Krivonozhko V. E., Utkin O. B., Senkov R. V. Parametrical methods in the analysis of efficiency of complex systems // Collection of works of the Russian Academy of Sciences. Non-linear dynamics and management / under the ed. of S. K. Korovin. 1999.
16. Cooper W. W., Seiford L. M., Tone K. Data Envelopment Analysis. Boston : Kluwer Academic Publishers, 2000.

G. P. Tarasova
Federal State Statistics Service in Krasnoyarsk territory, Russia, Krasnoyarsk

T. G. Butova, T. V. Reshetova
Siberian Federal University, Russia, Krasnoyarsk

THE PERSONAL SERVICE MARKET: PECULIARITIES OF THE REGIONAL DEVELOPMENT (BY EXAMPLE OF THE KRASNOYARSK REGION MARKET)

The dynamics of personal services market development in the Krasnoyarsk Region is notable for lagging behind those of all-Russian ones. Besides, there is a lower growth rate of the personal services volume in 2008. Having sufficiently large potential, the market of personal services can be attractive to average and small-scale businesses. The share of personal services in total amount of paid services testifies to it in 2002. The decision of business and state regulation in the branch of personal services should be based on applied scientific research of the market taking into account current trends, in particular the market fragmentation.

Keywords: personal services, the personal services market, fragmentation of the personal services market.

The development level of service can be considered as index of social and economic progress of regions, as well as the most important condition for population's life quality assurance. The basic task of the domestic service for people is comfortable living condition made at the expense of the household rationalization and therefore time save dup for another purposes, e. g. rest, self-education, satisfaction of cultural needs.

The domestic service for people represents the traditional and most volumetric sector of paid services sphere for the population of Russian Federation, which percentage is approximate equal to 10 % based on statistics [1]. The domestic service is sphere for the active work of small businesses. The territorial task program "The development of small and average business subjects in Krasnoyarsk Region" in 2008–2010 contributes to the domestic service organizations' development and their problems solving [2]. This approved program is rather urgent decision in our opinion because the analysis of domestic service market development in Krasnoyarsk Region shows some problems.

Statistical analysis of the personal services market development dynamics has shown the rise of personal services volume in the country in general and in Krasnoyarsk Region in particular. The personal services volume has increased by 3.2 times (from 126.8 to 406.1 billion rbl.) in Russia over a period of 2002–2008 and just by 2.1 times (from 3.2 to 6.7 billion rbl.) in Krasnoyarsk Region (fig. 1, 2).

The contrastive analysis of personal services market growth rates showed its negative dynamics in Krasnoyarsk Region. Thereby the growth rate of the personal services volume was equal to 3.7 % in 2008 in Russian Federation while there is reduction of this index by 1.3 % in Krasnoyarsk Region.

Having estimated the dynamics of personal services volume indexes without taking in account the price influence over a period 2002–2008, we can draw a conclusion that there is a fluctuating market development with a declining tendency in Krasnoyarsk Region in contrast to the all-Russian tendency of steady but slight growth (fig. 3).

The domestic service development depends on population living standards and progress of competition. There is tendency of personal services percentage reduction in the structure of paid services for people in spite of the people income growth, and this tendency has faster pace in Krasnoyarsk Region (fig. 4).

Thereby percentage of the personal services in Krasnoyarsk Region has decreased by 5,3 % and in Russia – by 1,8 % over the investigated period. As the analysis shows, Krasnoyarsk Region has been falling behind Russia indicators since 2006 in terms of the personal services percentage in the total amount of paid services for people.

Meanwhile it is significant to note that the personal services volumes growth in Krasnoyarsk Region has passed ahead of the paid services volumes growth except for the year of 2006, but tendency of lead rates reduction is fixed (fig. 5).

Thus, statistical analysis of personal services regional market in Krasnoyarsk Region allowed detection of its development problem. Meanwhile, in our view, it is possible to determine potential resources of domestic service market on the basis of examination its modern development tendency.

It should be noted that there are changes at the personal services market just like these at the other markets, e.g. market fragmentation to the commodity segments [3].

Fig. 1. The dynamics of personal services market development in Russia over a period of 2002–2008 (made up on the basis of [1])

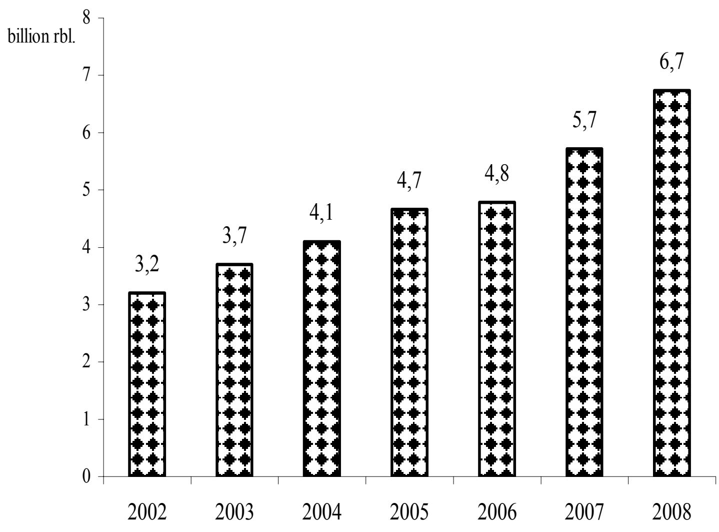


Fig. 2. The dynamics of personal services market development in Krasnoyarsk Region over a period of 2002–2008 (made up on the basis of [4])

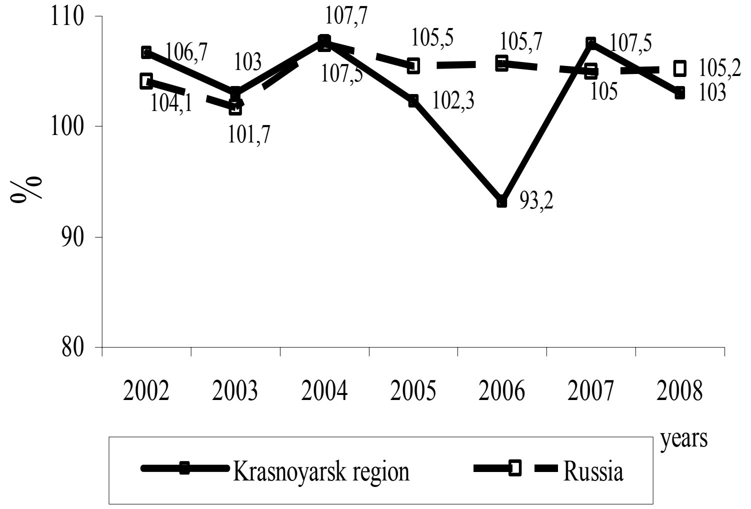


Fig. 3. The dynamics of personal services volume index in Russia and Krasnoyarsk Region over a period of 2002–2008 (Made up on the basis of [1; 4])

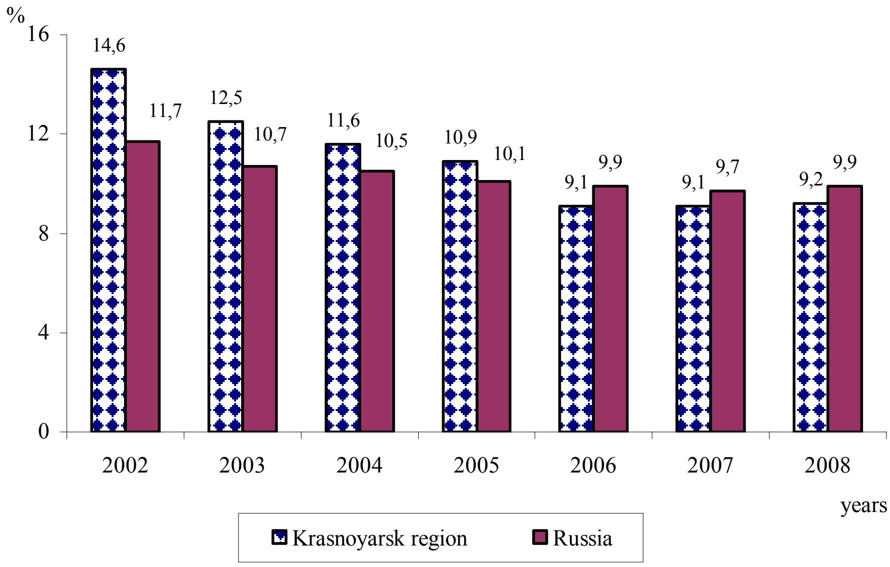


Fig. 4. The specific weight of the personal services in the total amount of paid services for people in Russia and Krasnoyarsk Region over the period of 2002–2008 (made up on the basis of [1; 4])

Fig. 5. The volume indexes of paid and personal services for people in Krasnoyarsk Region over the period of 2002–2008 (made up on the basis of [1; 4])

The dynamics of personal services structure in Russia and Krasnoyarsk Region over the period of 2002–2008

		2002		2003		2004		2005		2006		2007		2008	
		RF	Krasnoyarsk Region	RF	Krasnoyarsk Region	RF	Krasnoyarsk Region	RF	Krasnoyarsk Region	RF	Krasnoyarsk Region	RF	Krasnoyarsk Region	RF	Krasnoyarsk Region
All rendered service		100	100	100	100	100	100	100	100	100	100	100	100	100	100
including:															
I	1.1. Repair and sewing of shoes	2.9	10.3	2.9	10.5	2.8	11.4	2.7	11.1	2.6	11.2	2.4	10.3	2.3	10.1
	1.2. Repair and sewing of clothes, sewing and knitting of knitwear	10.8	3.5	10.2	3.1	9.7	3.7	9.2	3.9	8.3	4.5	7.6	4.3	7.2	4.4
	1.3. Dry-cleaning and dyeing	0.8	0.4	0.8	0.3	0.8	0.3	0.8	0.4	0.7	0.4	0.8	0.4	0.5	0.4
	1.4. Laundry services	1.1	1.5	1.1	0.9	0.9	1	0.8	1.2	0.7	1.4	0.6	1.3	30	1.2
II	Repair and maintenance of domestic electronics, domestic machines and devices, repair of metalware	8	32.9	7.7	32.9	7.4	34.2	6.5	33.1	6.1	27.9	5.7	24.9	5.2	22
III	Manufacturing and repair of furniture	1.9	2.9	1.9	2.1	1.9	1.8	1.8	1.6	3.2	1.2	4.1	1.3	0.8	1.8
IV	Repair and building of dwelling and other structures	25	16.7	25.9	20.3	27	18.7	27.4	20.2	28.3	21	28.6	23.5	26	23.1
V	Maintenance and repair of transport facilities, machines and equipment	21.9	11	21.7	9.8	22	9	24.1	10.6	23.7	12.9	25.4	13	26	14
VI	Services of photographer's studio, photo and film laboratories	3.3	2.4	3.1	1.9	2.7	2.3	2.5	2.4	2.3	2.5	2.1	2.4	2	2.3
VII	Services of bathhouses and shower cubicles	1.9	2.1	2	1.7	1.9	1.9	1.9	2.3	2	3.2	2	3.3	2.2	3.7
VIII	Services of barber's shops and hairdressing salons	6.5	2.2	6.7	2.2	6.7	2.4	6.9	2.1	7.2	2.6	7.1	3	7.8	4.6
IX	Ritual services	6.9	10.5	7.3	9.9	7.7	10.2	8	9.7	8.2	10.6	7.8	11.2	7.2	10.8
X	Rental services	0.3	0.2	0.4	0.1	0.5	0.1	0.5	0.2	0.5	0.2	0.5	0.4	0.7	0.5
	Other domestic services	8.7	3.4	8.3	4.3	7.7	3	6.9	1.2	6.2	0.4	5.3	0.7	4.7	1.1

This leads to the necessity of approaches analysis for commodity bounds of personal services market segments delimitation. The types of personal services are extraordinarily various. This fact brings in the necessity of types classification for the rise of analysis effectiveness. Currently Russia has the classifier of the people services named "The all-Russian classifier of the consumer services" (abbreviation in Russian: "ОКУН"). The aggregative group of this classifier, named "personal services", includes various and not always associated with essence of the domestic service points [5]. The personal services types in compliance with the all-Russian classifier of people service are in the table.

Such eclectic approach doesn't contribute to quality analyzing and developing potential of separate goods markets determining because of modern tendency of personal services market development, i. e. commodity fragmentation and B2B market forming. The imperfection of consumer services classifier led to necessity of relevant personal services classification on the basis of theoretically substantiated features.

The analysis of literature has shown that there are some approaches for the personal services classification problem. Several authors divide services rendered by establishment for homehold services into three groups:

– services associated with new use values making (sewing of clothes, shoes, manufacturing of furniture etc.);

– services associated with renewal of use values created earlier (clothes, shoes, furniture repair, laundry etc.);

– personal services intended straight for people or for their surrounding conditions and non-fixed in the commodities (services of barber's shops and hairdressing salons, rental services etc.) [6].

Other authors include new types of services without their grouping, e.g. services of pawnshop and clock repair, exclude some types, e. g. ritual services, maintenance and repair of transport facilities, repair and building of dwelling and other structures, divide the group «repair and maintenance of domestic electronics, domestic machines and devices, repair and production of metalware» into separate services [7].

The personal service classification, in our opinion, should be done with a glance of its consumer essence determined by necessity which is met by any service.

Therefore prior to describe the proposed approach, we must give a definition of the category "life" essence. In the wide sense this category is understood as extraindustrial social life sphere of people, the walk of everyday life closely associated with material and spiritual needs satisfaction, reproduction of people, ethnos and population of country in general. In the narrow sense the term "life" is used as "domestic life" signified the style of people's everyday life [8]. Relations in this sphere depend on ways of housekeeping. Self-service and hired (i. e. paid professional) labour in the form of personal services can be referred to these ways.

It's possible to distinguish two subgroups in the main group of personal services: "household" and "consumer" services. The first subgroup consists of services which are aimed at the environment change. Effect object in this case is material products which surround the human being in his everyday life. The second subgroup includes services directed towards change of the people appearance.

This approach allows authors to formulate a personal services classification proposal. The division of personal services accordingly to the aggregative groups is presented in table. Taking into consideration the determined essence of a category "life" in this article, we propose to label groups I, II, III, VII, VIII, IX as "household" services.

The comparative analysis of «household» and "consumer" services development in Russia and Krasnoyarsk Region has shown the uneven development of concrete types (sce table). Thus, in Russian Federation in general the basic share is taken by repair and building of dwelling and other structures (28.6 %) as well as the maintenance and repair of transport facilities, machines and equipment (25.4 %). The services of domestic electronics repair and maintenance and services of dwelling and other structures repair and building have the greatest specific share in Krasnoyarsk Region – 24.9 % and 23.5 %, respectively. These services also were prevalent in the domestic service structure for population of Krasnoyarsk Region in 2008; approximately 45 % of the domestic services total volume falls on the share of the given types.

However, in spite of considerable share of domestic electronics, domestic machines and devices repair and maintenance in the total volume of personal services, the demand for this type of services goes down, e. g. their specific weight in 2008 descended by 12.9 % in comparison with 2002, but the share of dwelling building and repair grew by 6.4 %.

The demand for the maintenance and repair of transport facilities, machines and equipment grew by 3 % over the investigated period, the demand for the hairdressing salons and barber's shops increased by 2.4 %, for the services of bathhouses and shower cubicles went up by 1.6 %, for the repair and sewing of clothes services it grew by the 0.9 %. The demand for the following services went down: manufacturing and repair of furniture (–1.1 %), repair and sewing of clothes, sewing and knitting of knitwear (–0.9 %) and other services (–2.3 %). Changes for the other types of services are negligible.

This market analysis of personal services in Krasnoyarsk Region allows drawing a conclusion that, in spite of rendered personal services absolute volume index growth, there are negative tendencies in this sphere too. For example, there is recession of the personal services share in the paid services total volume and of the demand for the specific types of services. Therefore innovative development of the domestic service for population is of great importance at present. The superior organization of public service establishment functioning can make easier the services market growth in Krasnoyarsk Region, provision of population with qualitative services expansion and consequently standard and quality of people's life rising. The personal services market development guarantees the investment activity in this sphere, new vacancies appearance and gross regional product growth.

References

1. The materials of the Federal State Statistics Service [Electronic resource]. Cop. 2010. URL: http//www.gks.ru.

2. The development of small and average business subjects in Krasnoyarsk Region in 2008–2010 : the territorial task program // Krasnoyarsk Region law. 2007.

3. Tarasova G. P., Butova T. G. The Personal services market development in Krasnoyarsk Region // The actual problems of marketing and management. M., 2008.

4. The materials of the Territorial Form of Federal State Statistics Service in Krasnoyarsk Region [Electronic resource]. Cop. 2010. URL: http//www.statis.krs.ru.

5. The All-Russian classifier of the consumer services (OK 002-93) // State Standard of Russia. M., 2008.

6. Zvorykina T. I., Platonova N. A. Technical Regulation : nonmaterial. M. : Alfa-M, 2008.

7. Firsukova V. V., Belobrov M. V. The source of the financial resources attraction to the personal services sphere of region (by example of Moscow Region) // Serviceplus. 2008. № 2. P. 84–91.

8. Great soviet encyclopedia. 3d pub. M. : Soviet encyclopedia, 1971.

O. N. Vladimirova
Siberian State Aerospace University named after academician M. F. Reshetnev,
Russia, Krasnoyarsk

STRATIFICATION OF TERRITORIES OF INNOVATIVE DEVELOPMENT: THEORETICAL APPROACHES

The factors influencing development of regions are studied in the article. The author considers approaches to territories stratification by criteria of innovation, existing in the scientific literature.

Keywords: territories of innovative development, regions typification, innovative development, stratification.

In modern conditions the innovative way of the development is admitted as a priority direction of economy going out from its crisis state and the way to increase its efficiency. Globalization and integration processes indicate the necessity to pay attention to the territories which have high innovative potential. So consideration of existing theoretical approaches to stratification of such formations seems topical.

The significant amount of publications is devoted to the issues of stratification of innovative development territories. The review of their content seems expedient in the following sequence: definition of typification factors of innovative development territories, consideration of stratification groupings and their adaptation in modern conditions analysis. Russian practice is taken as a basis of research.

It is pointed in a number of works by region development experts that spatial organization in Russia has appeared to be inefficient and has led to expenses growth to support infrastructural economy and other negative consequences.

A. G. Granberg, defining specificity of economic, legal and ethno-political space of the Russian Federation, allocates the following forming feature groups:

– considerable differences of natural-climatic conditions that proves the presence of minerals and other resources;

– during the Soviet period the basic principle of productive forces allocation was to form the economy as a "single factory" where the republic and region economic complexes were considered as peculiar "workshops" with the main purpose to satisfy the needs of the country's national economy, and only after that – to satisfy the needs of the country's population;

– high differentiation of social and economic development levels of the Russian Federation subjects, certain indicators show tens times gaps;

– RSFSR formation was based on the national-territorial principle, so it wasn't related to the territories economic potential;

– Russia's national policy (as well as during the Soviet period) is characterized by substitution of its population national interest by the mechanical sum of separate ethnic communities interests;

– qualitative difference of basic values idea and acceptability of various mechanisms of social and economic transformations is also typical for territorial communities;

– presence of various ideas about the most preferable form of the national-state system that involves a various measure of the state and regions responsibility for social and economic problems solution [1].

In our opinion, in modern conditions the list of the specified factors should be supplemented with the ones such as presence of highly developed branches, the state policy orientation and the international environment influence. Interaction of external and internal parameters can influence the formation of innovative development territories today.

Theoretical foundation of stratification issues dates back to the 1940s. The definition of various factors as the basic criterion of stratification (in this article stratification is treated and seen from the position of innovative activity) causes a variety of approaches.

Typification of Russian and the near abroad regions by the degree of innovative potential development and innovative activity scales is conducted in V. L. Baburin's research. On a parity of creative and acceptor components, correlation between innovative productivity of the USSR regions (number of patents per 100 thousand people of urban population) and the level of their innovative consumption (the relation of the introduced patents to the number of the produced ones), or the creativity index, have been calculated. As a result, the following groups of regions have been singled out:

1. Creative regions (using much fewer inventions than they create, having higher density than the Union on average). Moscow capital region and St.-Petersburg were the examples of such regions.

2. Sub-creative (using much fewer inventions than one uses having higher than average innovation density). They included a number of areas of the Volga-Kama interfluve, the Ural-Volga Region, Rostov Region, Novosibirsk-Omsk area.

3. Acceptor-creative (having higher than the Union average innovation generation, but having a large part (exceeding 100 %) of the inventions used). In Russia industrial-agrarian areas belong to them: the Altay and Krasnodar Territories, Belgorod and Voronezh Regions, Chuvashia and Mari-El.

4. Strong acceptors (having rather low innovation generation, but having very high (exceeding 100 %) generation of the inventions used). First of all, these are near-the-capital areas (Leningrad area), and the areas of economic development the Khanty-Mansi Autonomous Area and Yamal-Nenets Autonomous Area, the Far

East southern regions, Astrakhan, Orenburg and other Regions.

5. Weak acceptors (having low innovations generation and having a part of the inventions used higher than the Union average, but not exceeding 100 %). In this group the most developed east areas and also the periphery areas and republics of European Russia are widely presented.

6. Innovative periphery (having the lowest indicators both for inventions density and their use). This group includes mainly the USSR peripheral territories (overwhelming majority of autonomous regions, republics of Siberia, the Far East and the Northern Caucasia), internal areas of Transcaucasia, some "godforsaken places" of European Russia (Tambov, Kostroma, Novgorod Regions) [1].

In N. I. Markova's works regions stratification by innovative activity, conducted both with the help of formal methods of dispersive analysis and by means of expert estimations, allows with a sufficient degree of validity to allocate only one group including eleven subjects of the Federation, which have the innovative potential developed enough and the possibilities of activate its use (Moscow and St.-Petersburg, Moscow, Samara, Nizhni Novgorod, Kaluga, Sverdlovsk, Novosibirsk, Tomsk, Chelyabinsk and Voronezh Regions). The group structure testifies that the innovative potential of these regions is defined first of all by their high scientific potential. These are not only the regions with a high level of social and economic development, but also historically-established scientific centers of the country. Their role in the innovative process is to be innovation donors, first of all in the high technologies sphere [1].

There are two types of territorial innovative-technological formations specified in the scientific literature according to N. V. Beketov. A part of them is the areas with natural (evolutionary) concentration of high technology branches (the so-called technological regions). Absence of the special planning and coordinating foundation at the period of their appearance can be defined as their main distinctive feature. Other formations are the specially created centers of technological development (scientific, technological, research parks, "business incubators", innovative centers, etc.) [2].

E. A. Lurie, generalizing the experience of innovative development territories formation in Russia in the 20 th century, marks out 19 types of them. As the defining features in the given stratification we can see: administrative resource; the society mentality (the level of its susceptibility to innovations, readiness for innovative transformations, adaptation to the special innovative culture); presence of the developed documents regulating innovative activity (the Forecast, the Concept, the Strategy, the Program); involvement of the scientific-educational complex; estimation of the created innovative system influence on the basic indicators of the region development [3].

In the Base report for OECD to review the national innovative system of the Russian Federation "the National innovative system and the state innovative policy of the Russian Federation" 6 groups of the regions are marked out formed on the basis of the newness of the innovation index. On the basis of indirect statistic data the basic components of the innovative process are considered. As a criterion to estimate scientific potential the so-called the newness of innovation index was used. According to its value the regions which are leaders in scientific-technological potential, the regions which are leaders of realization and advancement of scientific work into a final business product, the regions focused on technology loans are assigned. The results of the Russian Federation regions analysis from the point of view of human potential presence to implement innovative activity, to distribute new knowledge, and to launch innovative products to the market are used as components.

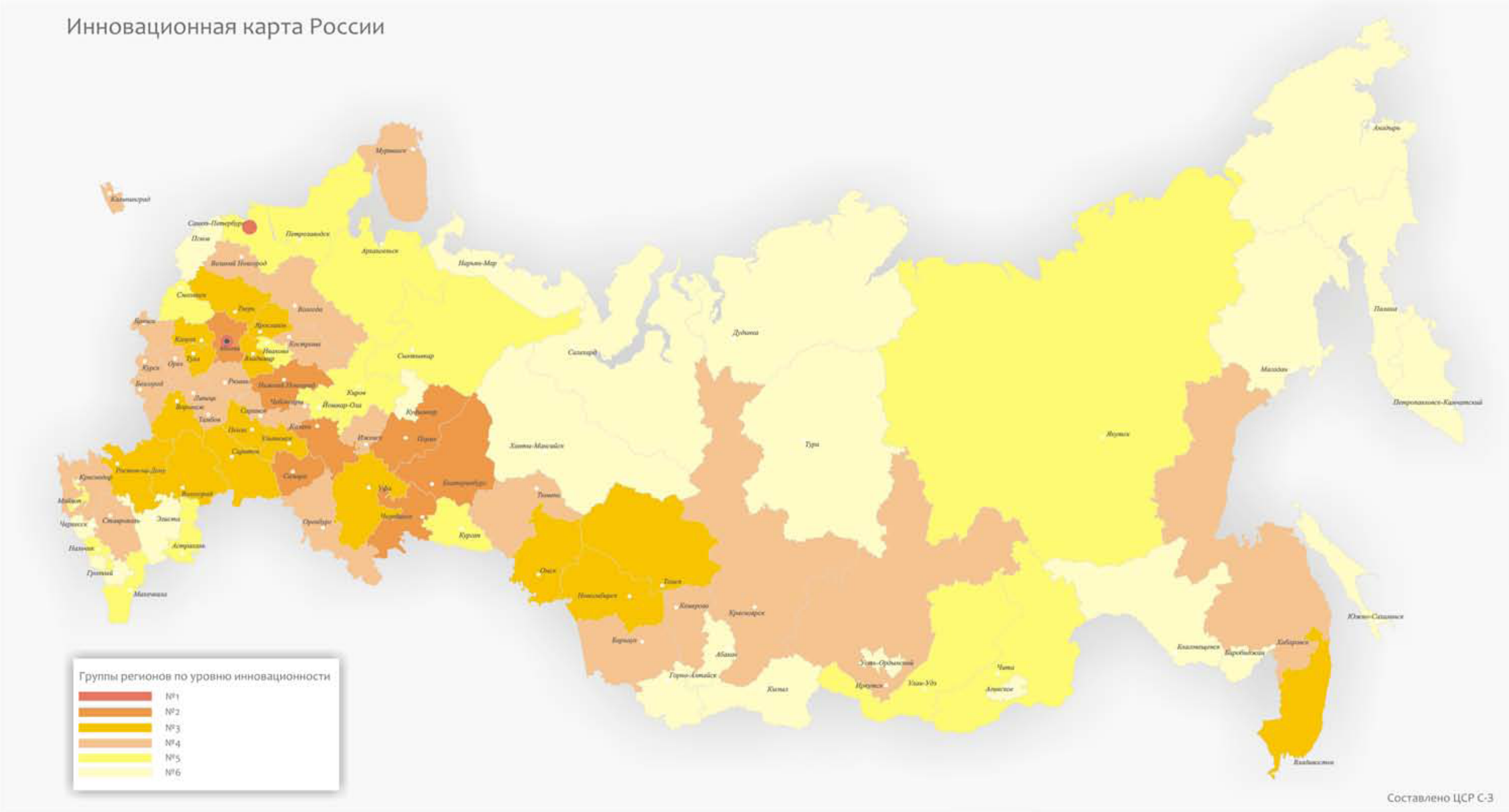
The newness of innovation index under the conditions of Russian innovative system formation rather fixes the regions start position from the point of view of their having some qualities necessary for innovations creation. The index better characterizes regions readiness or ability to innovations, rather than the actual innovative process. It is possible to concern the integration of the indicators used for its calculation as the advantages of the offered index. They embrace (as much as the modern official Russian statistics allows it) the basic stages or elements of the innovative process. A lack of the given tool is that it does not allow to estimate quality and intensity of interaction between the components of the innovative chain.

Proceeding from the considered approach the Russian Federation regions are divided into 6 conventional groups [4]:

Group 1 ("capitals") is the leader by all the indicators. It concentrates highly skilled human resources and implements the market stage of innovations the most successfully. The best representatives are Moscow, St.-Petersburg and the Republic of Tatarstan.

Group 2 is possible to be conventionally named "potential innovative leaders" or "the regions ready to innovations", is the leader by the market component after "capitals" (corresponds to the "capitals" level), lagging behind by characteristics of human potential. The greatest number of high technologies is used and the greatest volume of innovative products is made in the regions of the second group. These are Saratov Region, Nizhni Novgorod Region.

Group 3 includes the regions where sub-indices of human resources carry the greatest weight in the total index. They are slightly inferior to the group of potential innovative leaders by these sub-indices. However "market" sub-indices, especially the "appearance on the market" index, lag behind essentially. It can be explained by inefficient use of qualitative human resources, absence of stable relations between science and manufacture or discrepancy between scientific and industrial bases of the region. This group can be characterized as the regions with unrealized intellectual potential. The subjects of the named group are Tomsk, Novosibirsk, Omsk Regions and the Primorsky Kray.



Innovate map of the Russian Federation

Group 4 is very homogeneous by indices of "market" and "human potential", in the majority of its regions there are big cities, or they are located close to Moscow, so there are sources of human resources for new knowledge creation in these regions. Group 4 is inferior to Group 3 in terms of the «new knowledge creation» index, but they are on the same level by market indices. Thus, the regions of Group № 4 can be characterized as large industrial centers relying on technologies transfer, having an average level of innovative potential as they produce quite a large product volume using high technologies. However there is no due number of specialists for new knowledge creation. This group is represented by the Krasnoyarsk Territory, Kemerovo Region, Orenburg Region.

Group 5 represents the regions which are not among the leaders at present by any of the indicators, and neither their education system nor their industrial base allow them to move onto the next level. The regions – representatives are the Buryat Republic, Arkhangelsk, Kurgan, Chita Regions.

Finally in the last *Group 6* there are regions-outsiders by all the indicators. They are the Altai Republic, the Tuva Republic, the Khakass Republic, the Khanty-Mansi Autonomous Area.

In figure "Innovative map of Russia" made on the basis of the analysis of the regions level of the newness of innovation is presented [4].

A number of legislative acts show attempts of the regions stratification from the point of view of the possibility form conditions for innovative transformations. In particular, in the Program of the Russian Federation Government "Reforms and development of the Russian economy in 1995–1997" the following groups of the regions were allocated: lagging behind, depressive, crisis, and also the regions of the special strategic value.

In the Project of Siberia Social and Economic Development Strategy while considering Siberian Federal district according to its economy structure, specificity and rates of social and economic development, three groups of regions have been distinguished:

– the mastered areas of resources extractive orientation with centered settlement character, rather a high level of industry and its resource branches development, having definite specialization (Kemerovo, Tomsk, Irkutsk Regions, the Krasnoyarsk Territory, the Khakass Republic). Here the basic sources of the Russian Federation budgetary system are formed;

– regions with rather high density of population, quite diversified economy and rather a high level of infrastructure development and the territory cultivation (Novosibirsk and Omsk Regions); basic scientific-educational and agrarian potential, the processing sector of Siberian industry are concentrated here;

– territories with a low level of social and economic development (the Altai Republic, the Tuva Republic, Altai Territory, the Buryat Republic, Chita Region) [5].

The table represents the results of the conducted calculations, in our opinion, reflecting the basic stratification criteria.

Granberg A. G. specifies that correlation of scales, forms and tools of regional policy with specificity and needs of various types of regions, and also with its long-term orientation is possible only with typification of social and economic development levels of the federation subjects. The given typology is not conducted in Russia [1].

A number of scientists consider that studying regions economy manifestations of the so-called of "resource damnation" phenomenon is of great importance, when the countries possessing rich natural resources show lower rates of development than the countries, which do not have these resources. Nevertheless, according to the experts' analysis and estimation, there is no linear dependence between the volume and the kinds of the resources and the economy development rates [6].

The basic indicators of the Siberian Federal District regions development in 2007*

Region	Number, thousand people	GRP per person, thousand rub. per capita, 2006	Investment expenditures, thousand rub. per capita	Average monthly income, thousand rub. per capita	Budgetary security, thousand rub. per capita
The Altai Republic	207	54.39	27.908	6.934	43.76
The Buryat Republic	960	94.168	21.495	8.892	29.01
The Tuva Republic	312	47.967	6.814	5.817	31.85
The Khakass Republic	537	94.949	31.034	7.982	22.27
Altai territory	2508	66.275	16.964	7.438	н/д
The Krasnoyarsk Territory	2890	202.029	40.82	12.654	41.77
Irkutsk Region	2508	128.276	48.769	10.078	27.55
Kemerovo Region	2823	119.124	38.085	11.700	29.98
Novosibirsk Region	2636	108.453	32.306	10.317	25.86
Omsk Region	2018	121.934	32.367	11.318	26.93
Tomsk Region	1035	180.44	61.973	11.890	30.26
Chita Region	1119	77.898	30.365	8.212	н/д

*The source: calculated by the author according to [7].

Consideration of the above-stated approaches allows to make the following conclusions:

– there is no common approach to the criteria choice at present. Every indicator taken as a basis allows to reflect specificity of the formations considered;

– there are quite many parameters applied as stratification indicator criteria;

– imposing of the received stratification results allows to define coincidence of separate groups including, as a rule, the identical list of regions that gives a chance to make an assumption that there is some regularity which appears while grouping by the newness of innovation.

The author's position in this point is based on use of the innovative susceptibility factor as a priority. In our opinion, it fully considers economic, social and motivational components. The approach is expounded in the author's publications n more detail [8].

References

1. Movement of Russian regions to innovative economy / ed. by A. G. Granberg, S. D. Valentey ; Inst. of economy of RAS. M. : Nauka, 2006.

2. Beketov N. V. Methodological problems of formation and development of scientific-innovative systems of regions. M. : Akademia, 1999.

3. Lurie E. A. Territories of innovative development: regions experience // Innovatsii. 2009. № 2. C. 31–44.

4. National innovative system and the state innovative policy of the Russian Federation. The base report to OECD review of national innovative system of the Russian Federation. Moscow, 2009 [Electronic resource]. URL: http//www. mon.gov.ru/press/news/6333.

5. Strategy of social and economic development of Siberia till 2020. [Electronic resource]. URL: http//www.sibfo.ru/…/strdoc.php?action=art&nart=81.

6. Kopein V. V., Filimonova E. A. Structural transformations of the region economy and its financial safety ; ed.-in-chief. V. V. Mikhajlov. Novosibirsk : Pubishers SB RAS, 2008.

7. Russian regions. Social and economic indexes. 2008: Artic. assem. / Rusart. M. : 2008.

8. Baburin V. L. Innovative cycles in Russian economy. M. : Editorial URSS, 2002.

O. B. Voeykova
Siberian State Aerospace University named after M. F. Reshetnev,
Russia, Krasnoyarsk

INTEGRATION BETWEEN HIGHER SCHOOL AND INDUSTRY IN REGION AS THE FACTOR OF IT'S INNOVATIVE DEVELOPMENT

In this article the problems of Integration between Higher School and Industry in innovative development of region are covered. Retrospective analyze of co-evolution between defensive-industrial complex and Higher School in the face of Siberian State Aerospace University is carried out for estimation of dynamic interaction.

Keywords: innovative development, integration, defense industry, higher school, co-evolution.

Increase of national economy competitiveness ability, retention the position of Russia in the row of the world leading countries is possible when the innovative way of development is realized and the growth of intellectual public potential becomes the most important state problem.

The modern Russian government understands the importance of innovations and tries to stimulate everyway the innovative development of the country. So in the article "Message of President of Russian Federation D. Medvedev to Federal Assembly" (November 2009th) the President accentuated the necessity of creation in Russia the big Centre of innovations, the analogous of famous Silicon Valley, where "the attractive conditions for labour of leading scientists, engineers, constructors, IT-specialists, managers and financiers will be formed and new competitive in the world market technologies created" [1].

Recently, at the meeting with the winners of school Olympiads the President said, that the Russian Silicon Valley will be built in the Skolkovo in the suburbs of Moscow [2]. However it's not clear, why the Committee for Creation the Russian Centre of innovations chose this place. As it is known, some Russian territories known as zones of high technologies competed for the opportunity to become such a centre. They are Tomsk and Novosibirsk regions, St.-Petersburg and others. There are famous Universities and scientific centers, also big enterprises for industrial application of new developments tied up by the longtime connections.

Integration between Universities, enterprises and other Institutions doing scientific and research activities is a very important factor for the formation of the Centre of innovations.

Just the integration between Stanford University and the Base of United States Air Force (USAF) in Palo-Alto permitted to create the "Stanford Research Institute", which worked first for defense and then became the biggest Centre of microelectronics in the world [3].

There are some famous Universities of such kind in the USA, for example: "Massachusetts technological Institute" st. Massachusetts, "Texas University" in Ostin, st. Texas, University of Arisona and others, that demonstrate a wide range of different regions in innovative development of the country.

The innovative development orientation of regions is important especially for Russian territories, because they are characterized by extreme regional polarization, which is connected with essential disproportions in allocation of public-valued resources: fertile soils, climate conditions, natural resources , industrial enterprises and others. Innovative activity in Russian regions has to permit to eliminate the inequality of economic development of these territories.

Achievement the purposes of innovative development in a region demands realization of activities considering its competitive advantages. One of the most important competitive advantages in some regions of Russia is the defense industry which always has been a source of advanced scientific, technical, and technological achievements and developments. However, the powerful potential of defense industry is not employed totally under the conditions of activation the innovative development in regions.

Besides, some problems remain without attention, they are:

– estimation the scientific-innovative development influence of the of Higher School (HS) on the condition of research-and-production potential of enterprises in defense industry;

– stimulation of integration between HS and defensive-industrial complex (DIC) of region in innovative-technological direction.

All above-mentioned underlines the importance of researching the evolution of interaction between DIC and HS in region, and finding out the vector of their co-evolution in conditions of region's innovative development.

Krasnoyarsk territory is significant for such researches because it is characterized by the high concentration of defense industry enterprises, which have had tight connections with HS in the face of Siberian State Airspace University for a long time.

Formation of Siberian DIC began with the creation of two big enterprises for needs of space-rockets industry (SRI) in 50–60th. They were "Krasnoyarsk Machine-building plant" ("Krasmashzavod") and construction bureau of M. F. Reshetnev in Zheleznogorsk. In 1960 on thy basis of the "Krasmashzavod" the special secret educational establishment was opened. Later "Research and Production Union of Applied mechanics" named after M. F. Reshetnev and other defensive enterprises of the region entered the number of basic enterprises. The unique educational system "Plant-Institute" was accepted in this Institute, which had to connect the student's studies with the work at the enterprises of this branch [4].

Thus at the first stage of foundation the Siberian defense industry a certain system was formed, that involved the components of two bigger systems: defensive and educational one (fig. 1). Maintenance of this system integrity in the process of evolutional development depends on the degree of interaction stability of its components. The principle of co-evolution means mutual-adaptive variability of the system parts that leads to development acceleration of each of them and the system as a whole. According to this principle every mentioned component of the newly formed system ensured the development of another, and their interdependent development conduced to evolution of the whole system.

Such interpenetration between HS and defense industry resulted in the fact that each of this integrative system components was getting its advantages from fixed connections. The advantage for region defense enterprises from this interpenetration was foremost in permanent renovation of human resources, which had allowed to satisfy the needs of enterprises for modern specialists in concrete directions of engineer's activity (production, designing, researching). High level of specialist's university preparation for defensive-industry was most of all defined by unique educational programs, which were composed for development level of these enterprises, which were equipped enough with modern equipment conforming technological period.

The evolution of basic branch gave the unique opportunities for development of airspace Institute owing to presence of permanently acting integration in educational, research and production sphere. The Institute got especially intensive development in 80th under the exterior influence of DIC (fig. 2).

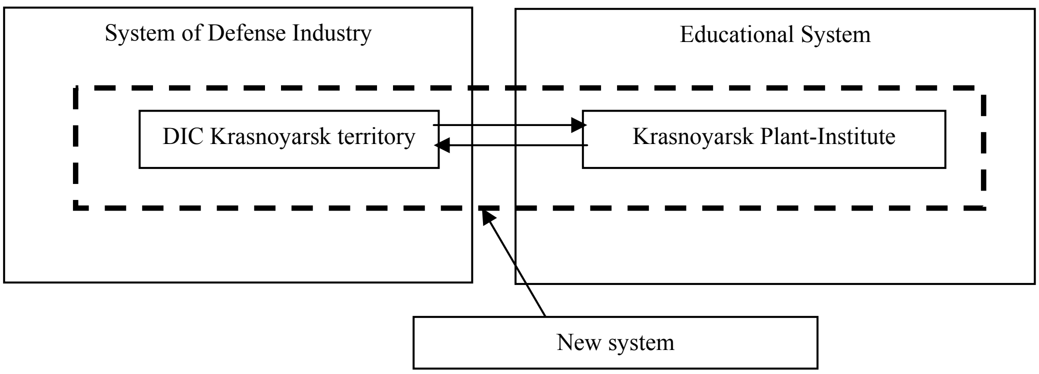


Fig. 1. Formation of the new production-educational system in region

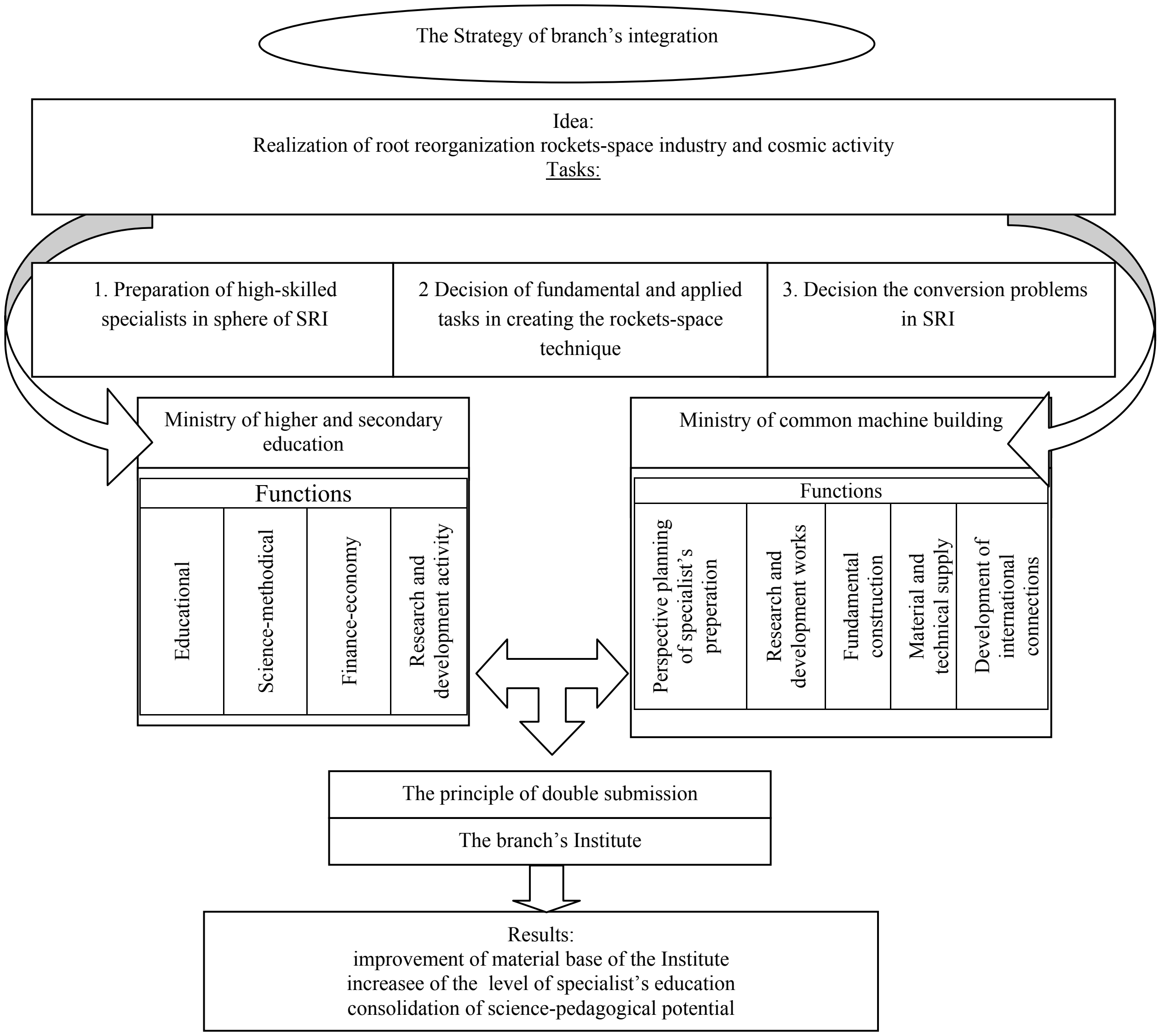


Fig. 2. The Institute as an object of co-management by two branches of economy (the principle of double submission)

Just then the Institute was transferred in double submission by attaching to two departments – the Ministry of higher and secondary education of RSFSR and the Ministry of common machine building of RSFSR, each with certain number of functions. This was the precedent for whole system of higher education in the country and showed the efficiency of strategy of branch's integration for the development of educational establishment, owing to that the multichannel financing of Institute became possible

The change in economic and political conditions in the country in 90th, connected with disintegration of Soviet Union, realization of new liberal market reforms, liquidation of the Ministry of common mashing building in RSFSR put both DIC and HS in a very complicated situation that resulted in considerable deformations of integrative co-evolution connections.

The integrity of Institute-enterprises system DIC was broken. Stagnation of one of the components hampered another. DIC was not interested in keeping up the connections with profile Institute, because it was on the verge of survival itself.

However, in spite of all difficulties of crisis period, the Institute showed its flexibility, mobility, and readiness to transformations developing even faster then the basic branch, whose enterprises were necessary to re-profile the production on the lower technical level.

Searching new ways of basic branch development according to the interests of civil economic sector, transition to market relations gave the impulse to enlargement the spectrum of educational directions among which besides conventional – rocket-space appeared such as: the exploitation of air transport, IT and computers, economics and management, humanitarian sphere. In period since 1990 per 2009 the list of specialties opened in the Institute enlarged essentially.

Economic specialties got intensive development connected with fast increase of the need in economic specialists and managers during the transition to market relations.

The necessity in enlargement of science-educational activity, attracting additional investment determined the Institute entering the international level. Nowadays the Institute is the member of European Convocation Business Education (ECBE) and International Company Engineers Pedagogic (ICIP). Stable connections were contacted with Czech technical University in Prague, High technical School and University of state New York in Oneonta (SUNY), the Rocket-Space Centre and International cosmic camp in Huntsville (USA). The Institute collaborates effectively with different foreign partners, such as Universities of Germany, Holland, Finland and Great Brittan.

Successful international cooperation and dynamic development of economic specialties in the Institute allowed to open new perspective direction of educational, research and development and foreign economic activity by means of formation the Faculty of International Business, now Higher International School of Business. This faculty formation was the important strategic decision of the Institute's leadership, because the space industry in Russia is acquiring the features of market economy due to the increase of services realized on commercial basis. In such a situation the defense enterprises are to start the commercialization of their production, doing business in sphere of space high-tech industry trying to enter and consolidate their position on the world space market.

Actually there is big necessity in engineers, who have got profound economic knowledge, also the knowledge in sphere of international relations, can perform the system analysis of the native and foreign high-tech space markets, and who have high level of speaking foreign languages.

So the High International School of Business in structure of the Airspace University is becoming especially actual in conditions of almost total absence of specialists of such level on the native enterprises in space industry.

On this stage of development the Siberian State Airspace University is characterized by high degree of innovative potential, so it has:

– innovative structure of organization, that includes the educational, production and scientific components;

– high share of innovative specialties (more than 50 %), which are oriented upon the modern requirements of high-tech innovative sector of economy;

– innovative technologies, which are realized in the University;

– a great number of high skilled stuff, professors and tutors, among them there are many representatives from industrial enterprises, branches and academic research and development institutions;

– considerable experience of realization fundamental and applied researches, experimental developments according to science-technical programs at different levels: International, Federal, Industrial, Regional and so on, so the University has got high potential in partial realization of linear model of innovations;

– innovative infrastructure as the combination of interconnected and complementary subdivisions consolidated in functional blocs, which ensure the realization of scientific-research and applied works combined with education;

– the University is a member of Association "National United Airspace University", that is the innovative structure of Higher Airspace Education in Russian Federation and includes 9 Russian Airspace Universities.

Taking into account the fact that the space industry entering the market has to be based on modern conceptual basis, science-methodical support and the highly-qualified stuff that is adapted to the new conditions, it is necessary to emphasize the progressive urgency of consolidation the integrative connections between DIC and Higher Airspace School. At the same time Airspace Universities bares the role of initiator and equal partner of this integration, that permits to inculcate their educational, scientific-research experience and accumulated innovative potential in development of DIC. So the raise of the industry makes the positive dynamic effect on development of airspace higher education as a whole.

References

1. Послание Федеральному Собранию Российской Федерации 12 ноября 2009 г. [Электронный ресурс] // Офиц. сайт Президента России. URL: http://www.kremlin.ru/transcripts/5979.html.

2. Кузьмин В. Президент учил жизни таланты. [Электронный ресурс] // Российская газета. 2010. 19 марта. URL: http://www.rg.ru/2010/03/19/vstrecha-olimp.html.

3. Кокшаров А. Силиконовая Германия // Эксперт. 1999. № 23. С. 20.

4. Аэрокосмический вуз Сибири: к 40-летию Сиб. аэрокосмич. акад. им. акад. М. Ф. Решетнева / Г. П. Беляков [и др.] ; под общ. ред. Г. П. Белякова ; СибГАУ. Красноярск, 2000.

E. A. Zhirnova, A. A. Snezhko
Siberian State Aerospace University named after M. F. Reshetnev, Russia, Krasnoyarsk

Ya. I. Shamlitskiy
JSC Krasnoyarsk Refrigerator Plant "Biryusa", Russia, Krasnoyarsk

THE EVALUATION OF THE QUALITY MANAGEMENT SYSTEM EFFECTIVENESS AT MACHINE-BUILDING ENTERPRISES

In this article we propose a method of evaluating the quality management system effectiveness. The major criteria of the evaluation are formulated, using the Joint Stock Company Krasnoyarsk Refrigerator Plant "Biryusa" as an example.

Keywords: quality management system effectiveness.

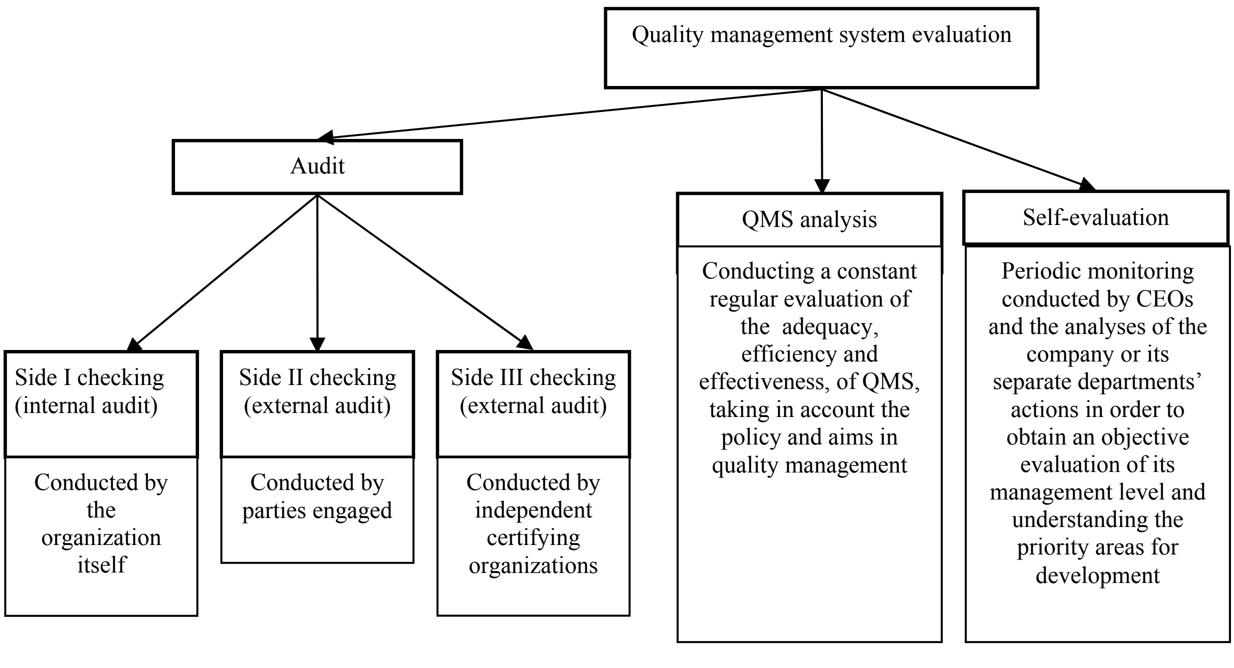
Quality management system (QMS) regulates the work of an enterprise sectors in identifying, providing, and maintaining the production quality at the stage of designing, developing, realizing, and servicing. The purpose of this is to constantly improve the organization effectiveness and efficiency. Evaluation is an essential condition for the functioning and improvement of a quality management system. It can differ by its sector of application and includes such activities as audit (monitoring) and the analysis of the quality management system, together with the organization self-evaluation (sce figure).

The analysis of the quality management system is conducted in order to provide its constant suitability, sufficiency, and effectiveness. It includes the evaluation of improvement capabilities – including policy and goals in the field of quality.

The quality management system analysis and evaluation has been conducted on the basis of product and process monitoring results, internal audits, the evaluation of consumer satisfaction, corrective and preventive actions, as well as suggestions made by workers and process managers.

The most important problem of quality management evaluation is the selection of the most "problematic" processes (from the process quality point of view), ranging them according to the level of "significance" for the organization activity, making decisions about conducting technical-organizational activities and allocation of funds for various purposes. This is why it is necessary to develop a method for evaluating the quality management system effectiveness.

Let's apply the following method to JSC Krasnoyarsk Refrigerator Plant "Biryusa". In order to evaluate the effectiveness three main criteria of goal achievement are highlighted: the improvement of the quality management system effectiveness, its compliance with the GOST R ISO 9001 requirements and internal regulations, the increase of the customers' satisfaction level. To achieve the planned targets each criterion is divided into two parts which include a list of problems, reflecting the accomplishment of planned results (tab. 1).



Types of quality management system evaluations

Table 1

Questions for evaluating quality management system effectiveness

Evaluation sub-criterion	Criterion evaluation questions	Satisfaction evaluation	Percentage of sub-criterion accomplish-ment, %
1.1. Planning quality improvement activities	1. Are the department targets clearly defined?	met	100
	2. Are meetings on quality problems being conducted at all levels?	met	
	3. Are quality indicators of manufactured products being planned?	met	
1.2. Accomplishing planned and correcting activities in quality improvement	1. The accomplishment of required quality indicators for products	not met	50
	2. What actions are taken by workshop managers to prevent the appearance of defects?	met	
	3. Accomplishment of planned activities on time	not met	
	4. The effectiveness of correcting / preventing actions taken	not met	
	5. What are the worker actions if he notices a product defect?	met	
	6. Who supervises corrective activities?	met	
2.1. The quality system level of compliance with GOST R ISO 9001–2008 requirements	1.Compliance of the QMS regulations with GOST R ISO 9001–2008 requirements	met	100
	2. Compliance of department targets to the Policy in the field of quality	met	
2.2. The quality system level of compliance with internal QMS regulations	1. Managing paperwork according to STP SK 222-4005–2002 requirements	not met	50
	2. Managing documentation according to STP SK 222-4016–2002 requirements	not met	
	3. Appropriate qualification of the staff	met	
	4. Research and development according to STP SK 222-4004–2002 requirements	met	
	5. Supervising irrelevant products according to STP SK 222-4013–2002 requirements	met	
3.1. The level of consumer contentment based on conducted surveys	1. Do you regard “Biryusa” refrigerators as reliable?	not met	50
	2. Are you satisfied by the prices of “Biryusa” refrigerators?	met	
	3. Do you suppose that purchasing a “Biryusa” refrigerator is a sign of prestige?	not met	
	4. Are you satisfied with the range of refrigerators the company has to offer?	met	
	5. Are you comfortable with using “Biryusa” refrigerators?	met	
	6. Does the outer design of “Biryusa” refrigerators satisfy you?	not met	
	7. Does the inner design of “Biryusa” refrigerators satisfy you?	not met	
	8. Do the “Biryusa” refrigerators save energy?	met	
	9. Do the temperature regimes of the “Biryusa” refrigerators satisfy you?	met	
	10. Does the after sales service of the “Biryusa” refrigerators satisfy you?	not met	
3.2. The level of consumer contentment based on warranty returns	1. Electronics	not met	50
	2. Compressor defects	not met	
	3. Leakage	met	
	4. Outer shell	met	
	5. Does not fit the doze	met	

It is easier to evaluate the quality management system effectiveness by an amount of points gained for each of the target achievement criteria. A 100 point scale can be used for the evaluation. Each of the criteria is given a quality index. When identifying the quality index it is supposed that the sum of all coefficients is equal to 100 %. The quality index – determined for each criterion is divided into sub-criterion, depending on its importance. The highest quality index is defined by the quality management system effectiveness improvement criterion as well as by the increase of the consumer’s contentment. This is important, first of all, because any economical activity of organizations is bent on supplying consumers with quality products in order to meet their needs.

According to the developed method, problems of the first two criteria are determined by the results of conducted monitoring. The third sub-criterion is determined by the level of consumer contentment. The evaluation of this sub-criteria is conducted by the marketing and service department specialists studying the return and refund statistics.

Evaluation is conducted through three parameters:

– the sub-criterion is achieved – 100 % score;

– it is achieved partially: If more than 50 % of actions to achieve the subcriteria done, then it is 50 %, but if the number is less than 50 %, the score is 0 %;

– if the sub-criterion is not achieved at all – 0 % score.

Then the quality management system effectiveness score is determined.

The common effectiveness score is calculated in the following way:

$$Б = \Sigma В_j \times П_j / 100\ \%,$$

where $В_j$ – is the quality index, $П_j$ – percentage of performed and accomplished subcriteria.

The common score of effectiveness determines the quality management system level (tab. 2).

Table 2

Evaluation of quality management systems

Common grade for effectiveness	The effectiveness level
91–100	Very good
71–90	Good
51–70	Acceptable
Under 50	Poor

According to the developed method an evaluation of quality management system effectiveness at JSC Krasnoyarsk Refrigerator Plant "Biryusa" has shown a number of drawbacks. The results of the quality management system evaluation are shown in tab. 3.

The results of quality management system evaluation at the enterprise according to the proposed method are displayed in tab. 3.

On conducting the effectiveness evaluation a number of drawbacks has been found. For example, the criterion of quality management system effectiveness improvement was 27.5 points. This does not fully satisfy the planned value. At the same time the sub-criteria of scheduled quality improvement activities is 100 % accomplished, because the monitored department targets are defined, regular meetings on the quality management are held and quality indicators of manufactured products planned. The sub-criterion of accomplishing scheduled corrective activities for quality improvement is only 50 % fulfilled. This was discovered during a screening, which revealed that the quality indicators of manufactured goods do not correspond to the ones established. Activities do not meet their deadlines and the corrective and preventive actions applied are not useful.

The effectiveness level of criterion, the process compliance with the GOST R ISO 9001 requirements, and internal regulations is 12.5 points – this does not meet the planned value. The quality management system regulation corresponds to the GOST R ISO 9001–2008 requirements, the targets of the department to the quality policy [1]. Hence the quality management system complies with the GOST R ISO 9001–2008 requirements by 100 %. The qualification of staff working in the department, research and development, as well as control of irrelevant products is done through the use of normative-technical regulations. This cannot be said about paperwork supervision; hence the sub-criterion of quality management system compliance with internal regulations is only 50 %.

The consumer contentment criterion is only 50 % from the planned value – specialists from the marketing department conducted a survey and specialists from the service department analyzed data about return and warranty. All these studies showed that the effectiveness level is only 20 points. The survey showed that the consumer was not satisfied by the inner and outer designs and the after sales service of "Biryusa" refrigerators. They suppose that these refrigerators are unreliable and do not suppose that the mentioned brand is prestige. The after sales service departments carried out the warranty repair caused by defects in compressors and the electric system.

The results of the quality management system evaluation by using the developed method showed that the total effectiveness is 60 points. This means that quality management system functions appropriately but still does not fulfill all of the GOST R ISO 9001–2008 requirements.

Table 3

The evaluation of the quality management system at JSC Krasnoyarsk Refrigerator Plant "Biryusa".

Evaluation criteria	Quality index	Evaluation sub-criterion	Sub-criterion quality index	Percentage of criterion accomplishment, %	Effectiveness score, points
1. Improvement of quality system effectiveness	0.4	Planning quality improvement activities	0.375	100	15
		Accomplishing scheduled and correcting activities in quality improvement	0.625	50	12.5
2. Compliance of the quality system with GOST R ISO 9001 requirements and internal regulations	0.2	The quality system level of compliance with GOST R ISO 9001–2008 requirements	0.25	100	5
		The quality system level of compliance with internal QMS regulations	0.75	50	7.5
3. Increasing the consumer contentment	0.4	The level of consumer contentment based on conducted surveys	0.25	50	5
		The level of consumer contentment based on warranty returns	0.75	50	15
					60

The analysis results showed that the total quality management system effectiveness – in particular, can improve the capability of current activity improvements according to quality system criterions and improve the contentment of the consumer. This value should be increased based on the results of organizational and technical activity evaluation.

Having conducted the evaluation of the quality management system effectiveness applying the developed method, one can identify not only the functioning conditions of the quality management system, but see problematic areas, based on the established evaluation criteria. Using this method of analyzing results of enterprise activity and the functioning of the quality management system, the enterprise management can quickly focus their actions on the planning of improving enterprise activity. The given method of evaluation concentrates on the most important factors influencing the effectiveness of the quality management system provides the development, introduction, and realization of the best solutions. The appliance of the developed method will permit an enterprise in case of well functioning quality management system to reduce significantly the expenditures on quality and the functioning of the company. This will result in a concrete economical effect.

References

1. ГОСТ Р ИСО 9000–2008. Система менеджмента качества требования. М. : Изд-во стандартов. 2005.

CONTENTS

PART 1. MATHEMATICS, MECHANICS, COMPUTER SCIENCE