Evolutionary methods in modelling behaviour of complex system

V.V. Mokshin[1], P.I. Tutubalin[1], L.M. Sharnin[1]

[1]Kazan National Research Technical University named after A.N. Tupolev – KAI, Karl Marx street 10, Kazan, Russia, 420111

Abstract. In this paper, a recursive-regression approach is formulated in the formation of a functioning model of a complex system. The complex system consist of the object of study represented by a set of related signs. The proposed sub-course combines the method of group accounting of arguments with standard regression analysis adapted to fast dynamic processes. As an example, recursive-regression modelling of the functioning of an industrial enterprise was performed.

1. Introduction

There are various approaches aimed at the formation of models of analysis and control of complex systems whose behavior is caused by various external factors. As complex systems can be considered as biological and economic systems, as well as entire manufacturing enterprises, as well as sociological objects (city, district, region, etc.) [1, 2]. Management of complex systems involves the analysis of the corresponding dynamic processes and the selection of the most optimal solutions aimed at achieving a specific end result. A detailed analysis of continuously available information about the "object under study" as a dynamic system [3, 4] creates the basis of operations for the development of new management decisions [5]. One of the main tasks is to find a methodology for the most correct prediction of the particular characteristics (parameters) of the system. For the analysis of complex dynamic systems, methods of linear and nonlinear programming, the method of neural networks, fractal and multifractal analysis, methods of nonlinear dynamics and chaos theory, spectral and correlation analyzes, concepts of self-organizing systems, balance methods, etc. are widely used today. [6, 7, 8]. These methods are successfully used in various tasks related to optimization, prediction and recognition in physics, astrophysics, biology and physiology, geology, meteorology, economics, etc. The most qualitative and reliable results are obtained with the combined use of several methods, when the possibilities of various methods are adjusted and mutually reinforced. In the present work, a recursive approach based on a combination of a multifactorial nonlinear regression model and the method of group accounting of arguments (MGAA) is proposed for analyzing and controlling the object of study [9, 10, 11]. In this case, the method of group accounting of arguments allows one to find the original solution of the problem associated with the choice of a certain set of model functions necessary for constructing a regression model. So, in accordance with the MGAA, on the basis of the criterion of reliability calculated from additional samples or using the sliding control method [12], it is possible to select this finite set of functions.

The recursive-regression approach proposed in this paper is a combination of standard regression analysis with the method of group accounting of documents adapted to dynamic, rapidly changing

processes. The present approach is an iterative procedure in which, based on some "basic" model, as information is updated and statistics accumulate, it is automatically corrected: independent variables are added and / or removed. As a result, the regression model is dynamically updated [13] and at the same time its recursive self-organization occurs. The combination of the aforementioned methods makes it possible to adapt the well-known regression analysis to the study of complex dynamic systems and significantly improve the quality of forecasting. This significantly reduces the time spent on building the regression model [8].

2. Generalized analysis model of the object of study

2.1. Generalized model

As a system of research, let us consider some object and the final combination of attributes (variables) associated with it (see Fig. 1). It is assumed that this set of features can be divided into those that directly affect the object – independent factors (we denote them by x_i, $i = \overline{1, M}$), and those determined by the object itself – dependent factors (we denote them by y_i, $i = \overline{1, K}$). A set of independent factors x_i is also called input factors, and dependent factors y_j are called responses, effective factors, or indicators of the functioning of the system (see Fig. 1).
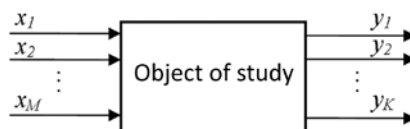
Figure 1. The object of study, which is influenced by factors x_i, and which itself directly determines the factors y_j.

The behavior of such a system, characterized by a combination of features, can be described by a model based on a multifactorial nonlinear regression analysis [14]. In accordance with the regression analysis, after determining all dependent and independent factors, it is necessary to form a model of system functioning. It is assumed that each of the dependent factors y_j can be represented as a function of the input factors x_i:

$$y_j = f_j\left(x_1, x_2, \ldots x_M\right), \; j = \overline{1, K}, \tag{1}$$

where K is the number of dependent (effective) factors of the functioning of the system, M is the number of factors influencing the functioning of the system, $(M + K)$ is the total number of factors. After determining the functioning model (1), it is possible to formulate the optimization problems of the research system functioning.

To construct the model (1), one can use the method of group accounting of arguments — the method of generating and choosing regression models of optimal complexity [10, 15]. For this purpose, an inductive approach is used, according to which models of increasing complexity are sequentially generated until the minimum of a certain quality criterion of the model is found [10]. This means that the subsequent complication of the model does not lead to an improvement in its quality. As a result, achieving the minimum quality criterion will match the desired model.

One of the authors of this method, A.G. Ivakhnenko, writes [5, 10, 13], that at the same time "there is a purposeful search of many models-applicants of varying complexity according to a number of criteria. The result is a model of optimal structure in the form of a single equation or system of equations. The minimum of the criterion determines the mode of optimal structure".

2.2. The method of group accounting of arguments

The method of group accounting of arguments consists in finding the functional relationship between the input factors x_i, and the effective indicators of the functioning of the system y_j, according to their temporal behavior. In the most general form, such a dependence can be represented as a generalized Kolmogorov-Gabor polynomial [10, 13]:

$$y = a_0 + \sum_{i=1}^{M} a_i x_i + \sum_{j=1}^{M} \sum_{i \leq j} a_{ij} x_i x_j + \sum_{i=1}^{M} \sum_{j \leq i} \sum_{k \leq j} a_{ijk} x_i x_j x_k + ... \qquad (2)$$

where the weights a are found using regression analysis and the least squares method for the training set [7]. Polynomial (3) actually contains all possible combinations of input factors $x_i, i = \overline{1,M}$. Note that to improve the quality of models (1), it is also possible to use not only the power dependences of the input factors, but also the elementary functions of these factors.

As an example, consider the following situation. Suppose that there are two independent input factors x_1 and x_2 and one effective factor y. In accordance with the method of group accounting of arguments, various possible combinations of limited complexity S arise, i.e. $y^{(1)}$, $y^{(2)}$, $y^{(3)}$, …. The complexity S here is the number of members of the polynomial (2) being formed [16]. So,

$$
\begin{aligned}
y^{(1)} &= a_{10} + a_{11} x_1, \\
y^{(2)} &= a_{20} + a_{21} x_2, \\
y^{(3)} &= a_{30} + a_{31} x_1 + a_{32} x_2, \\
y^{(4)} &= a_{40} + a_{41} x_1 + a_{42} x_2 + a_{43} x_1 x_2, \\
y^{(5)} &= a_{50} + a_{51} x_1 + a_{52} x_2 + a_{53} x_1 x_2 + a_{54} x_1^2 x_2, \\
y^{(6)} &= a_{60} + a_{61} x_1 + a_{62} x_2 + a_{63} x_1 x_2 + a_{64} x_1 x_2^2, \\
y^{(7)} &= a_{70} + a_{71} x_1 + a_{72} x_2 + a_{73} x_1 x_2 + a_{74} x_1^2 x_2^2.
\end{aligned}
\qquad (3)
$$

The best model is selected based on the minimum value of the quality criterion [12]. Or, the threshold value of the quality criterion is determined and several models are selected whose quality criterion values do not exceed this threshold.

2.3. Quality assessment and model selection

To assess the quality of models (for example, for models of the form (3)) one can use the variance [10, 11, 13]:

$$\overline{\varepsilon^2} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left(y_i - f\left(x_i\right) \right)^2, \qquad (4)$$

where n is the number of considered time intervals of input factors $x_i, i = \overline{1,M}$ and effective indicators of the functioning of the system $y_i, i = \overline{1,K}$.

Then only those values of the weighting factors a of the model polynomial (3) will be acceptable, for which the value $\overline{\varepsilon^2}$ takes the minimum values, i.e.

$$\overline{\varepsilon^2} \to \min \qquad (5)$$

In this case, the principle of multiplicity of models can be realized [12,13], which consists in the following: there are many models on this sample with zero error. To do this, it is enough to increase the degree of the polynomial model. In other words, you can build a whole family of models, each of which, when passing through experimental points y_j, where $j = const$, will give a zero error, and, therefore,

$$\overline{\varepsilon^2} = 0 \qquad (6)$$

Obviously, the variance $\overline{\varepsilon^2}$ depends on the complexity of S. Moreover, with increasing complexity, the variance $\overline{\varepsilon^2}$ will first fall, and then grow. Then you need to choose the optimal degree with a minimum value. In addition, the object of study can be affected by various interferences, which in the present work will mean unaccounted input factors affecting the system, or insufficient complexity of the desired polynomial. Considering the effect of interference, the following situations can be distinguished:

1. At different levels θ of interference, the $\overline{\varepsilon^2}$ dependence on complexity S will change, while maintaining the overall directivity (meaning that with increasing complexity, $\overline{\varepsilon^2}$ will first decrease and then increase).

2. As the level of interference θ increases, the value $\dfrac{\min \overline{\varepsilon^2}}{S}$ will increase.

3. As the level of interference increases, $S^* = \arg\min \overline{\varepsilon^2}$ will decrease, i.e. the optimal difficulty value will shift to the left (see fig. 2). Moreover, $\overline{\varepsilon^2}\left(S^*\right) > 0$ if the noise level θ is nonzero.
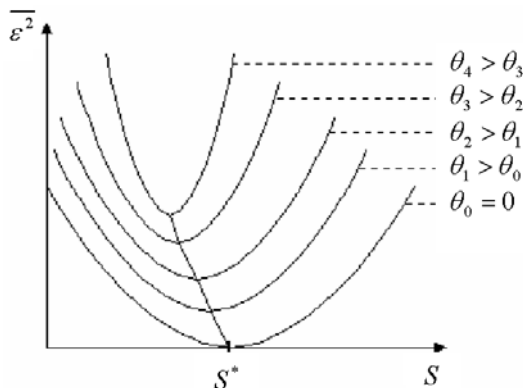
Figure 2. The dependence $\overline{\varepsilon^2}$ of the variance on the complexity of the model S at different levels of interference θ. The minima on the curves correspond to models with optimal complexity S *.

3. Evolutionary approach of formation models for research object functioning

The above-stated basic ideas of the method of the basic accounting of arguments allow determining the interrelations between input and effective factors, which are expressed with the help of functional dependencies of the form (3). Based on this, you can perform (i) analysis of the current state of the object, (ii) forecasting the values of the result factors, (iii) influence the resultant factors through the input factors, thereby controlling the object. However, all this is possible only in the case when the entire system remains unchanged over time. This means the preservation of the number of input factors, their independence from each other and the absence of new external factors. However, this situation is, in fact, idealized. Therefore, there is the task of creating such a model of the functioning of the system, which could adapt to changing external factors. In our case, this means that for each subsequent time step, a new model is created based on the previous model (see Fig. 3).
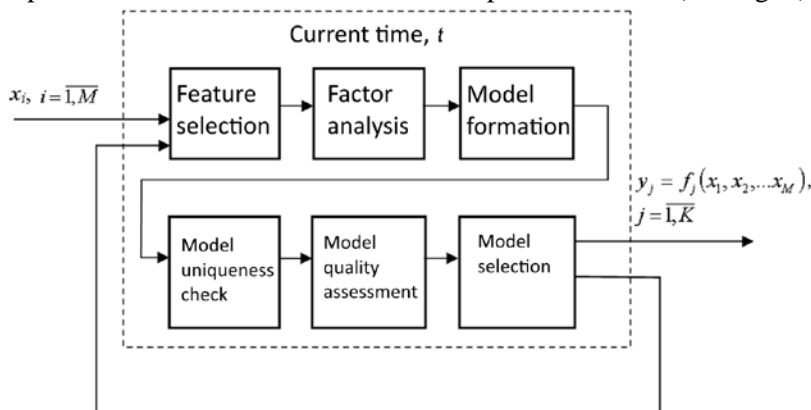
Figure 3. Schematic representation of the formation of the structure of a multidimensional additive series, taking into account changes in the system over time.

The area indicated by the dotted line determines the formation of the functioning model of the object of study at a certain point in time. The whole scheme reflects the possibility of changing the already created model in time, i.e. it becomes possible to add new input factors and re-form the model (1). In fact, this is the formulation of the basic ideas of recursive approaches [14, 15, 16]. Such considerations will be used by us in the dynamic formation of the model of the functioning of the object.

4. Application of the recursive and regressional approach to evaluating the efficiency of managing a manufacturing enterprise

Consider the results of the above approach, in relation to the problem of improving the management efficiency of some industrial enterprises, cooperating with the enterprises of the automotive and petroleum industries. In order to form a model for the functioning of a production enterprise, we will select the values given in Table 1 as input external factors and effective factors y_j. In the study, quarterly values of selected factors were used, and a particular task was to increase the revenue from the sale of goods, products, works, and services. As a result, a functioning model was obtained, including a system of ten regression equations. This number of equations corresponds to the number of selected effective factors y_j. The most significant obtained regression equations are given below:

$$
\begin{aligned}
y_1 = {} & -665 \cdot 10^5 - 0{,}65 x_1 + 6{,}35 \cdot 10^{-2} x_2 x_3 + 3{,}59 \cdot 10^{-4} x_4 x_{15} + 3{,}36 x_5^{0,5} x_{17}^{0,5} - \\
& -2{,}15 \cdot 10^7 x_6^{-1} + 5{,}78 x_7 - 4{,}56 \cdot 10^2 x_8^{0,5} + 1{,}65 \cdot \ln(x_{10} x_{12}) + 84{,}04 x_{11} + \\
& + 22{,}25 x_{13}^{0,5} x_9^{0,5} + 45{,}29 \cdot x_{14} + 4{,}99 x_{16} - 4{,}62 \cdot 10^{-6} e^{x_{17}},
\end{aligned}
\tag{7}
$$

$$
\begin{aligned}
y_2 = {} & 5{,}95 - 1{,}1 \cdot 10^{-8} e^{x_1} - 8{,}18 x_2 + 0{,}6 x_3 - 3{,}73 x_4 x_{15} + 2{,}03 \cdot 10^{-6} e^{x_5} - \\
& -8{,}09 \cdot 10^{-6} x_6 x_7 - 4{,}59 x_8^{0,5} x_{16}^{0,5} - 1{,}86 \cdot 10^4 x_9 + 9{,}53 \cdot 10^4 \ln x_{10} + \\
& + 2{,}53 e^{x_{11}} + 9{,}34 \cdot 10^{-5} x_{12} x_{17} + 7{,}41 \cdot 10^{-3} x_{13} x_{14} + 85{,}44 x_{16}^{0,5},
\end{aligned}
\tag{8}
$$

$$
\begin{aligned}
y_3 = {} & 3{,}96 \cdot 10^7 - 0{,}45 x_1 + 0{,}16 x_2 + 6{,}1 \cdot 10^6 x_3^2 - 9{,}94 \cdot 10^{-6} x_4^2 + 3{,}14 \ln(x_5) - \\
& -5{,}56 \cdot 10^{-5} x_6^2 + 0{,}67 x_7 + 1{,}99 x_8 - 3{,}65 x_9 - +0{,}18 x_{10} + 23{,}2 x_{11} + \\
& + 5{,}63 x_{12} - 50{,}93 x_{13} - 78{,}89 x_{14} - 2{,}77 \cdot 10^5 \ln x_{15} + 11{,}99 x_{16} - 94{,}33 x_{17}.
\end{aligned}
\tag{9}
$$

The whole set of obtained regression equations forms a recursive-regression model, in which the input factors x_i, $i = \overline{1, M}$, have a controlling influence on the resulting factors y_j.

The proposed model allows us to determine how much a change in an input factor x_i, $i = \overline{1, M}$ (or a combination of them) can affect the resulting factor y_i, $i = \overline{1, K}$. To do this, we estimated the degree of influence of factors by their weight and elasticity coefficients. The specific weights of input factors determine, as a percentage, their influence on the resulting factors (see Fig. 4-6). The coefficient of elasticity determines the degree of change of the *j*-th effective coefficient when the value of the *i*-th input coefficient changes by one percent. Positive values of elasticity coefficients indicate the positive influence of the corresponding external factor, and negative values – vice versa (see Fig. 7). Note that the elastic coefficients were calculated under the conditions of the last considered time interval. When calculating them, all factors were assigned values that they took into account in the last time interval, and the coefficient that was taken into account changed over the entire time interval.

From fig. 4 clearly shows that the effective factor y_1 (revenue from the sale of goods, products, works and services) is most influenced by the input factors x_5, x_7, x_{11}, x_{13} and x_{16}, which for the system under study evaluate accounts receivable, currency account , selling expenses, operating expenses and payroll, respectively. From fig. 7 that for the growth of the values of y_1, i.e. increase in revenue from the sale of goods, products, works and services; it is necessary to increase the fixed assets of the enterprise x_1, finished goods and goods for resale x_5, the current account x_6, retained earnings of the reporting year x_8, the average wage per employee x_{17}, and reduce inventory x_2, costs in work in progress x_3, receivables x_5, currency account x_7, business expenses x_{11}, operating expenses x_{13}, non-operating expenses x_{14} and the number of employees with permanent labor contracts x_{15}.

**Table 1.** The list of input and effective factors selected for the study.

| Factor	Factor name
y_1	Proceeds from the sale of goods, products, works, services, thousand rubles
y_2	Cost of goods sold, products, works, services, thousand rubles
y_3	Profit from sales, thousand rubles
y_4	Profit before tax, thousand rubles
y_5	Current income tax, thousand rubles
y_6	Interest receivable, RUB thousand
y_7	Other operating income, thousand rubles
y_8	Non-sale income, thousand rubles
y_9	Income per employee, thousand rubles
y_{10}	Profit per employee, thousand rubles
x_1	Fixed assets, thousand rubles
x_2	Inventories, thousand rubles
x_3	Costs in work in progress (distribution costs), thousand rubles
x_4	Finished products and goods for resale, thousand tons
x_5	Accounts receivable (payments for which are expected within 12 months after the reporting date), thous. Rub.
x_6	Bank account, thousand rubles
x_7	Currency account, thousand rubles
x_8	Retained earnings of the reporting year, thousand rubles
x_9	Borrowed funds, thousand rubles
x_{10}	Accounts payable, thousand rubles
x_{11}	Selling expenses, thousand rubles
x_{12}	Administrative expenses, thousand rubles
x_{13}	Other operating expenses, RUB thousand
x_{14}	Non-operating expenses, thousand rubles
x_{15}	The number of employees with permanent employment contracts (contracts) in people.
x_{16}	Payroll, thousand rubles
x_{17}	Average salary per employee, thousand rubles
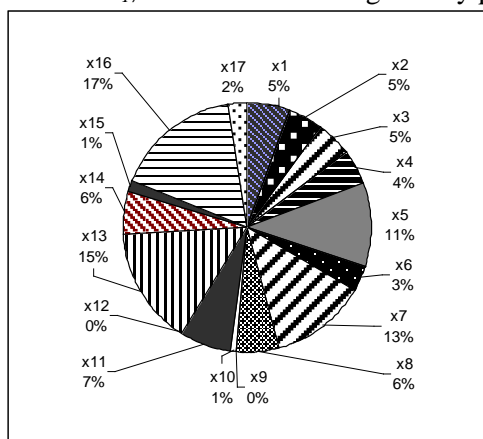
Figure 4. The diagram of the relative influence of input factors on the revenue from the sale of goods, products, works and services, y_1.
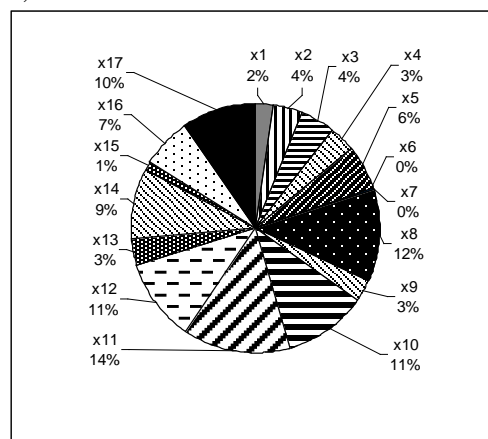
Figure 5. The diagram of the relative influence of input factors on the cost of goods sold, products, works and services, y_2.
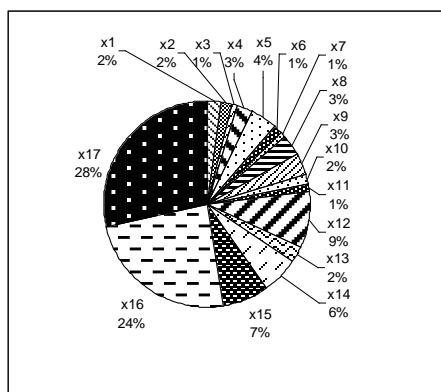
Figure 6. The diagram of the relative influence of input factors on sales profit, y_3.
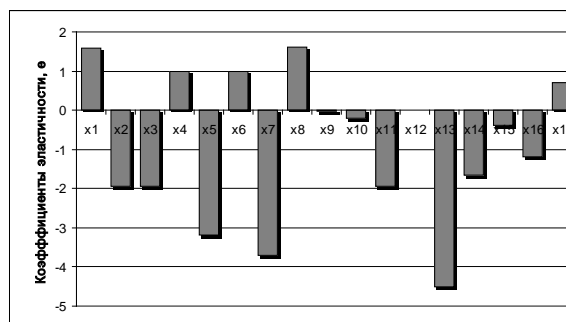
Figure 7. The coefficients of elasticity of revenue from the sale of goods, products, works and services y_1.

Thus, it is obvious that the resulting model can also be used to optimize all the effective factors yj taken into account in the framework of our model. So, for example, if we consider the revenue from the sale of goods, products, works and services y1 as an optimized resultant factor (see Fig.4 and 7), then, in accordance with our results, it is fair to say that increase by 40%.

5. Conclusions

This paper presents a recursive-regression approach to the formation of a complex system functioning model. As such, it can be the object of study (biological organism, enterprise, ecosystem, etc.), whose behavior is determined by a combination of external influencing factors. This approach allows you to differentiate the level of entry of each of the original external factors and, in general, allows you to control the entire system under study.

An important advantage of the proposed approach is that models for effective factors are formed automatically at each time step. A recursive self-organization of models occurs, and the need for their constant adjustment "by hand" disappears. Thus, the proposed approach makes it possible to automate the process of forming the structure of an additive functional series of the form (3), while flexibly adjusting the necessary level of complexity of the series in terms of individual parameters. Moreover, here it also remains possible to go beyond the basis of the classical types of power and harmonic series by increasing their flexibility by introducing polynomial members (3) with fractional multiplicity. The integrated approach presented in the paper provides, in principle, a powerful and flexible tool for analyzing and controlling the object of research, the behavior of which is influenced by many different external factors. It should be noted that there is also the possibility of using the presented approach in conjunction with the methods and algorithms of neural networks.

6. References

[1]   Mantegna, R.N. An Introduction to Econophysics: Correlations and Complexity in Finance / R.N. Mantegna, H.E. Stanley // Cambridge: Cambridge University Press, 1999.
[2]   Chakrabarti, B.K. Econophysics and Sociophysics : Trends and Perspectives / B.K. Chakrabarti, A. Chakraborti, A. Chatterjee. – Berlin: Wiley-VCH, 2006.
[3]   Yulmetyev, R.M. Universal approach to overcoming nonstationary, unsteadiness and non-Markovity of stochastic processes in complex systems / R.M. Yulmetyev, A.V. Mokshin, P. Hanggi // Physica A. – 2005. – Vol. 345. – P. 303-325.
[4]   Mokshin, A.V. Simple Measure of Memory for Dynamical Processes Described by a Generalized Langevin Equation / A.V. Mokshin, R.M. Yulmetyev, P. Hanggi // Physical Review Letters. – 2005. – Vol. 95. – P. 200601(1)-200601(4).
[5]   Ivakhnenko, A.G. Simplified Linear Programming Algorithm as Basic Tool for Open-Loop Control / A.G. Ivakhnenko, G.A. Ivakhnenko // System Analysis Modeling Simulation (SAMS). – 1996. – Vol. 22. – P. 177-184.

[6] Lemke, F. SelfOrganize! - software tool for modelling and prediction of complex systems // SAMS. – 1995. – Vol. 20(1-2). – P. 17-28.

[7] Vapnik, V.N. Algorithms and programs for recovering dependencies // Science, 1984. – 816 p.

[8] Ivakhnenko, A.G. Simulation noise immunity / A.G. Ivakhnenko, V.S. Stepashko. – Naukova Dumka, 1985. – 216 p.

[9] Parallel genetic algorithm of feature selection for complex system analysis / V.V. Mokshin, I.R. Saifudinov, L.M. Sharnin, M.V. Trusfus, P.I. Tutubalin // J. Phys.: Conf. Ser. – 2018. – Vol. 1096. – P. 012089. DOI: 10.1088/1742-6596/1096/1/012089.

[10] Yakimov, I. The comparison of structured modeling and simulation modeling of queueing systems / I. Yakimov, A. Kirpichnikov, V. Mokshin, Z. Yakhina, R. Gainullin // Communications in Computer and Information Science (CCIS). – 2017. – Vol. 800. DOI: 10.1007/978-3-319-68069-9_21.

[11] Yakimov, I.M. AnyLogic, extendsim and simulink overview comparison of structural and simulation modelling systems / I.M. Yakimov, M.V. Trusfus, V.V. Mokshin, A.P. Kirpichnikov // Proceedings of the 3rd Russian-Pacific Conference on Computer Technology and Applications. RPC, 2018. DOI: 10.1109/RPC.2018.8482152.

[12] Saifudinov, I.R. Visible structures highlighting model analysis aimed at object image detection problem / I.R. Saifudinov, V.V. Mokshin, P.I. Tutubalin, L.M. Sharnin, D.G. Hohlov // CEUR Workshop Proceedings. – 2018. – Vol. 2210. – P. 139-148.

[13] Mokshin, A.V. Adaptive genetic algorithms used to analyze behavior of complex system / A.V. Mokshin, V.V. Mokshin, L.M. Sharnin // Communications in Nonlinear Science and Numerical Simulation. – 2019. – Vol. 71. – P. 174-186. DOI: 10.1016/j.cnsns.2018.11.014.

[14] Tutubalin, P.I. The Evaluation of the cryptographic strength of asymmetric encryption algorithms / P.I. Tutubalin, V.V. Mokshin // Second Russia and Pacific Conference on Computer Technology and Applications (RPC), 2017. – P. 180-183. DOI: 10.1109/RPC.2017.8168094.

[15] Lyasheva, S.A. The analysis of image characteristics on the base of energy features of the wavelet transform / S.A. Lyasheva, M.V. Medvedev, M.P. Shleymovich, V.V. Mokshin // CEUR Workshop Proceedings. – 2018. – Vol. 2210. – P. 96-102.

[16] Gabor, D. A universal nonlinear filter, predictor and simulator which optimizes itself by a learning process / D. Gabor, W.R. Wilby, R.A. Woodcock. – 1961. – Vol. 108(40). – P. 85-98.