

# Development of a mathematical apparatus for state estimation of an integrated safety system for an artificial traffic installation

Alexander Ginzburg<sup>1\*</sup> and Sergey Kozminykh<sup>2</sup>

1 Moscow State University of Civil Engineering, 129337, Moscow, Russia

2 University of Finance at Government of the Russian Federation, 125993, Moscow, Russia

**Abstract.** The article is devoted to the development of a mathematical apparatus for the state estimation of the integrated security system (ISS) artificial traffic installation (ATI). The article analyzes the state characteristics of ISS ATI. It is noted that the estimates of the mathematical expectation show the general direction of changes in ISS average functioning. The article provides a selection of statistical analysis criteria for ISS ATI. The features of state analysis for ISS ATI are given. It is noted that checking the normal distribution should be a preliminary stage in the ISS ATI state analysis and a significant deviation from the normal distribution law can serve as a diagnostic sign of the of the system non-stationarity. Therefore, such a comparison is a sufficient tool for identifying the differences, and it should be used in the first stage of the analysis of the ISS ATI state, making possible to draw fairly reliable conclusions about the system state. The article presents the developed methodological principles of ISS ATI diagnostics. For stationarity ISS ATI analysis a methodology for diagnostic analysis, which consists in comparing the statistical current (dysfunctional) characteristics and reference implementations of the system functioning, has been developed. In conclusion, it was concluded that the use of this mathematical apparatus in the ATI security systems development makes it possible to significantly reduce the cost of their implementation and significantly increase ISS efficiency when implemented on ATI.

## 1 Introduction

Ensuring comprehensive safety of an artificial traffic installation (ATI) in emergency situations is a complex process that largely depends on various factors of a probabilistic nature. At the same time, the methodological apparatus for ensuring the integrated security of objects in the credit and financial sphere, a complex of models, methods and algorithms, makes it possible to solve the problem.

The developed methodology is based on the principle of comparing the state of the security system under study on a specific ATI with some ideal reference state. In this case, it is assumed that ISS functioning in real conditions, the implementation of potential threats is a stochastic process.

---

\* Corresponding author: [ginav@mgsu.ru](mailto:ginav@mgsu.ru)

## 2 Materials and methods

### 2.1 Condition characteristics of ISS ATI

Let the concrete implementation of the functioning ISS ATI be described by a selective function  $X_i(t)$ . Then the set of sample functions  $\{X_i(t)\}$  describes an ensemble of such implementations. Thus, ISS functioning is described by a vector of time functions:

$$\bar{X}(t) = (X_1(t), X_2(t), \dots, X_m(t)), \tag{1}$$

where  $X_m(t)$  is a random time function.

Let us consider an arbitrary random function  $X(t)$  and suppose that as a result of  $n$  experiments,  $n$  its implementation  $x_1(t), x_2(t), \dots, x_n(t)$ . Then at a fixed moment in time  $t=t_k$  the function  $x_1(t), x_2(t), \dots, x_n(t)$  meanings form the so-called random function section.

Considering the functioning ISS in general, we can talk about a single observation in the form of a cross section at  $t = t_k$ , represented by the matrix

$$X = \|x_{ij}(t_k)\|, i=1, m; j=1, n. \tag{2}$$

Considering a random system in the form  $m$  random functions  $X(t)$ , we get to a coordinate-wise analysis of the vector  $X(t)$ . In this case, the main numerical characteristics  $X(t_k)$  in the section are:

- mathematical expectation estimate:

$$M_x[X(t_k)] = \bar{X}(t_k) = \sum_{j=1}^n (x_j(t_k))/n \tag{3}$$

- variance estimate:

$$D_x[X(t_k)] = \sigma_{x_2}(t_k) = \left( \sum_{j=1}^n [x_j(t_k) - \bar{X}(t_k)]^2 \right) / (n-1) \tag{4}$$

where  $\sigma_x(t_k)$  is the standard deviation estimate.

To determine the statistical connection degree of a single random realization at different points in time  $t_k$  and  $t_l$  the most suitable is the correlation function ([1]), the estimate of which has the following form:

$$K[X(t_k), X(t_l)] = K_x(t_k, t_l) = \sum_{j=1}^n [x_j(t_k) - \bar{X}(t_k)][x_j(t_l) - \bar{X}(t_l)] / (n-1). \tag{5}$$

It can be shown that  $K_x(t_k, t_l) = K_x(t_l, t_k)$  и  $K_x(t_k, t_k) = \sigma_{x_2}(t_k)$ .

The estimate of the normalized correlation function is taken as:

$$r_x(t_k, t_l) = \frac{K_x(t_k, t_l)}{\sigma_x(t_k, t_l)} \tag{6}$$

Considering that  $\sigma_x(t_k) = \sqrt{K_x(t_k, t_l)}$ , we get:

$$r_x(t_k, t_l) = \frac{K_x(t_k, t_l)}{\sqrt{K_x(t_k, t_k)K_x(t_l, t_l)}} \tag{7}$$

Hence, knowing the correlation function estimate, it is possible to find the estimate of the normalized correlation function.

It follows from the definition  $r_x(t_k, t_l)$  that  $|r_x(t_k, t_l)| \leq 1$ , and the estimate of the normalized correlation function has the same probabilistic meaning as the estimate of the correlation coefficient: the closer  $|r_x(t_k, t_l)|$  to unity, the closer the linear statistical relationship between the cross sections.

For multidimensional random systems ( $m > 1$ ) the statistical properties of any two random functions' connection from different sections are proposed to be described by the mutual correlation function ([2]). Let two random functions be given as  $X(t)$  and  $Y(t)$ , then the estimate of the cross-correlation function will have the form:

$$K[X(t_k), Y(t_l)] = K_{xy}(t_k, t_l) = \sum_{j=1}^n [x_j(t_k) - \bar{X}(t_k)][y_j(t_l) - \bar{Y}(t_l)] / (n-1). \tag{8}$$

It should be noted that in the general case  $K_{xy}(t_k, t_l) \neq K_{yx}(t_k, t_l)$  is the statistical relationship of the cross sections  $X(t_k)$  and  $Y(t_l)$  does not have to be the same as the statistical relationship of the cross sections  $X(t_l)$  and  $Y(t_k)$ . At the same time, with the simultaneous permutation of indices  $x, y$  and the arguments  $t_k$  and  $t_l$  we have:

$$K_{xy}(t_k, t_l) = K_{yx}(t_k, t_l). \tag{9}$$

Similar to the correlation function  $K_{xy}(t_k, t_l)$  the normalization of the cross-correlation function is also admitted:

$$r_{xy}(t_k, t_l) = \frac{K_{xy}(t_k, t_l)}{\sqrt{K_x(t_k, t_k)K_y(t_l, t_l)}} = \frac{K_{xy}(t_k, t_l)}{\sigma_x(t_k)\sigma_y(t_l)} \tag{10}$$

moreover  $|r_{xy}(t_k, t_l)| \leq 1$  and  $r_{xy}(t_k, t_l)$  has the same probabilistic meaning as the correlation coefficient estimate.

Let  $t \in [0, T]$ . Let us divide the time span  $[0, T]$  into intervals by dots  $t_1=0, t_2, \dots, t_s=T$ , and for each section, using the above-mentioned formulas, we construct a time series that describe the dependence of the corresponding numerical characteristics on time:

$$\begin{aligned} M_x(t) &: M_x(t_1), \dots, M_x(t_s)\}, \\ D_x(t) &: \sigma^2_x(t_1), \dots, \sigma^2_x(t_s)\}, \\ s_x(t) &: \sigma_x(t_1), \dots, \sigma_x(t_s)\}, \\ r_x(t) &: r_x(t_1, t_2), \dots, r_x(t_{s-1}, t_s)\}, \\ r_{xy}(t) &: r_{xy}(t_1, t_2), \dots, r_{xy}(t_{s-1}, t_s)\}, \end{aligned} \tag{11}$$

Here the mathematical expectation estimates  $M_x(t)$  show the general direction of change in ISS average functioning. If their values increase, then this is evidenced that over time there is a certain tendency for the average value of the trait to increase  $X$ . Variance estimates  $D_x(t)$  and standard deviations  $\sigma_x(t)$  characterize the ISS functioning in the sense of changing the spread of values  $X$  relatively  $M_x(t)$ . According to the estimates' values of the normalized correlation function  $r_x(t)$  and the normalized cross-correlation function  $r_{xy}(t)$  it is possible to trace how the strength of the bond between the sections changes for one  $x(t)$  and any two  $x(t)$

and  $y(t)$  respectively. It should be noted that, the functions  $r_x(t)$  and  $r_{xy}(t)$  for all possible pairs of values from each section may be of interest in the latter case.

ISS functioning is described by the random functions that vary approximately uniformly and have random fluctuations around some constant mean value ([3]). Moreover, the dispersion of these fluctuations does not show significant changes over time. In this case, we can talk about a stationary random system.

Stationarity of a single random realization  $X(t)$  is determined by the conditions:

$$M_x(t)=const; D_x(t)=const \text{ or } \sigma_x(t)=const. \tag{12}$$

It can be shown that for stationary ISS  $r_x(t_k, t_l)$  is a function of one argument  $\tau=t_l - t_k$ , i.e:

$$r_x(t_k, t_l) = r_x(t_k + \tau, t_l) = r_x(\tau). \tag{13}$$

Correlation coefficient for any pair of cross-sections  $t=t_k$  and  $t=t_k + \tau$  for any  $k$  and  $\tau=const$  for a stationary system remains constant.

The same can be said for  $r_{xy}(t_k, t_l)$ : for stationary connected systems  $X(t)$  and  $Y(t)$  the cross-correlation function depends only on the difference of the arguments  $\tau = t_l - t_k$ , and, hence:

$$r_{xy}(t_k, t_l) = r_{xy}(t_k, t_k + \tau) = r_{xy}(\tau). \tag{14}$$

Statistical tests can be applied to check ISS stationarity. Let us set some statistical measure, which is determined by the confidence probability, and we will call ISS stationary in the probabilistic sense (quasi-stationary), if the discrepancies in the estimates for the numerical characteristics in the sections at  $t=t_k$  and  $t=t_l$  do not go beyond the selected measure. Otherwise, the system will be called non-stationary ([3]).

## 2.2 ISS ATI statistical analysis criteria selection

Critical statistical characteristics for ISS ATI safety analysis are: average  $M_x$ , potential security threat; dispersion  $D_x$ , characterizing the accuracy of its reproduction; correlation  $r_{xy}$ , characterizing the structure of connections of all potential threats.

Stationarity ISS is checked at the stage of diagnostic analysis and includes comparing the statistical characteristics of the current and reference implementations of the system's functioning.

Analysis of existing methods for establishing a significant difference between the two averages (estimates of mathematical expectations) showed that the most appropriate is  $t$ -criterion (Student's distribution).

If the variances in the sections are equal, then the statistics are calculated:

$$t = \frac{[\bar{x}(t_k) - \bar{x}(t_l)](n_1 + n_2 - 2)^{1/2}}{(1/n_1 + 1/n_2)[(n_1 - 1)\sigma^2_x(t_k) + (n_2 - 1)\sigma^2_x(t_l)]^{1/2}} \tag{15}$$

with the number of degrees of freedom  $\nu = n_1 + n_2 - 2$ , where  $n_1, n_2$  are the sample volumes, respectively, in the sections at  $t=t_k$  and  $t=t_l$ .

If the variances are not equal, then the statistics can be used:

$$t = \frac{\bar{x}(t_k) - \bar{x}(t_l)}{(\sigma^2_x(t_k)/n_1 + \sigma^2_x(t_l)/n_2)^{1/2}} \tag{16}$$

In this case, the number of degrees of freedom is the largest number not exceeding:

$$\frac{(n1 - 1)(n2 - 1)[(\sigma^2_{x(tk)} / n1 + \sigma^2_{x(tl)} / n2)]^2}{(n1 - 1)\sigma^2_{x(tk)} / n1 + (n2 - 1)\sigma^2_{x(tl)} / n2}$$

If in both cases  $|t| < t_{cr} = t_{tab}(\alpha, \nu)$  ( $\alpha$ - level of significance), then the difference in estimates is considered statistically insignificant, and the system is quasi-stationary relative to the average. In the case of a significant difference in the estimates of the average, the value is taken as a measure of the non-stationarity degree:

$$\delta x = |[\bar{X}(tk) - \bar{X}(tl)] / \max\{\bar{X}(tk), \bar{X}(tl)\}|$$

To compare the variance estimates, it is proposed to use  $F$  - criterion (Fisher):

$$F = \sigma^2_{x(tk)} / \sigma^2_{x(tl)}$$

with  $\nu_1 = n1 - 1$  and  $\nu_2 = n2 - 1$  degrees of freedom.

If  $F < F_{cr} = F_{tab}(\alpha, \nu_1, \nu_2)$ , then the variances are considered equal, and the system is recognized as quasi-stationary with respect to the variance.

Just as for the mean, the following quantity may be used as a measure of the non-stationarity degree in this case:

$$\delta \sigma^2_x = |[\sigma^2_{x(tk)} - \sigma^2_{x(tl)}] / \{\max\{\sigma^2_{x(tk)}, \sigma^2_{x(tl)}\}\}|$$

Analysis of the known criteria for comparing the correlation characteristics showed that a statistically significant difference between two estimates of the normalized correlation function or the normalized cross correlation function (coefficients  $r_x(\tau)$  and  $r_{xy}(\tau)$ ) given  $\tau$  is most effectively verified with  $z$  - Fisher transformations ([4]). For this, the statistics are calculated:

$$T = (z_1 - z_2) / S, \tag{17}$$

where:

$$\begin{aligned} z_1 &= 1/2 * \ln\{(1+r_{xy}(tk, tk+\tau)) / (1-r_{xy}(tk, tk+\tau))\}, \\ z_2 &= 1/2 * \ln\{(1+r_{xy}(tl, tl+\tau)) / (1-r_{xy}(tl, tl+\tau))\}, \\ S^2 &= (n1-3)^{-1} + (n2-3)^{-1}, \end{aligned}$$

$k, l$  ( $k \neq l$ ) are the section numbers for  $t = tk$  and  $t = tl$ .

In the case when  $|T| < T_{kp} = T_{tab}(\alpha)$ , where  $T_{tab}(\alpha)$  is searched from the table of the standard normal distribution depending on the significance level  $\alpha$ , the difference in the correlation coefficients estimates is considered insignificant. In the opposite case  $r_{xy}(tk, tk+\tau) \neq r_{xy}(tl, tl+\tau)$ , with the ISS non-stationarity measure relative to the normalized cross-correlation function, the following quantity can be correct:

$$\delta r_{xy} = |r_{xy}(tk, tk+\tau) - r_{xy}(tl, tl+\tau)| \tag{18}$$

Similar reasoning can be carried out in relation to the normalized correlation function  $r_x(\tau)$ , calculating the statistics for it and choosing the measure of the system non-stationarity.

When calculating numerical characteristics in time  $t \in [0, T]$  in order to study ISS quasi-stationarity with the help of time series, in general, it is advisable to use the statistics, making a pairwise comparison for all cross sections ([5]) at  $t = t_1, t_2, \dots, t_s$ . In general, the most effective for diagnosing the functioning ISS are the following procedures of statistical checks: to check the equality of several means – single factor analysis of variance; to establish the fact of a significant difference in the estimates of several variances – Bartlett’s test for the samples of

different sizes, Cochran’s test for the samples of the same size; to check the equality of several correlation coefficients – z Fisher transformation with criterion statistics  $x^2$ .

Let us turn now to the multidimensional system  $X(t)$ . Within the framework of statistical analysis, an exhaustive understanding of it is given by the task of a set of mathematical expectations estimates:

$$\bar{M}_x(t) = \{ \bar{x}_1(t), \bar{x}_2(t), \dots, \bar{x}_m(t) \}$$

and the estimates of correlation matrices:

$$K(t) = \|K_{ij}\|, i=1,m; j=1,m,$$

the elements of which, in the general case, are mutual correlation functions  $K_{ij} = K_{ij}(t_k, t_l)$ . It is obvious that  $K_{ij}(t_k, t_l) = K_{ji}(t_l, t_k)$ , and at  $k = l$   $K_{ij}(t_k, t_l) = K_{ij}$  are the estimates of the correlation moments of random variables  $X_i$  and  $X_j$  in the section  $t = t_k$ .

From the point of view of establishing the strength of connections between random functions in different sections, a representation by the matrices of estimates of the correlation coefficients is given ([6]):

$$r = \|r_{ij}\|, i=1,m; j=1,m,$$

where in general  $r_{ij} = r_{ij}(t_k, t_l)$ ;  $r_{ij}(t_k, t_l) = r_{ji}(t_l, t_k)$  and at  $k=l$   $r_{ij}(t_k, t_l) = r_{ij}$  is the correlation coefficient estimation between random variables  $X_i$  and  $X_j$ .

Summarizing the procedures for checking statistical analysis of ISS multivariate for quasi-stationarity, we can talk about checking a significant discrepancy between the estimates of the vectors of means and estimates of the covariance matrices.

If the quasi-stationarity analysis with respect to the vectors of means is carried out, provided that the covariance matrices of the investigated sections are equal, but unknown, then from the whole set of criteria,  $T^2$  – Hotelling’s criterion, since it uses generalized estimates of the links structure. The statistics are calculated

$$T^2 = \frac{n_1 * n_2}{n_1 + n_2} (\bar{M}_x(t_k) - \bar{M}_x(t_l)) S^{-1} (\bar{M}_x(t_k) - \bar{M}_x(t_l))^T, \tag{19}$$

where:  $S = \frac{1}{n_1 + n_2 - 2} (X_k X_k^T + X_l X_l^T)$  is a generalized covariance matrix;

$\underline{X}_k = \|x_{ij}(t_k) - \bar{x}(t_k)\|, i=1,m$ ; is an observation matrix for  $j=1, n, t = t_k$  in the form of dimensional deviations  $m * n_j$ ;

$\underline{X}_l = \|x_{ij}(t_l) - \bar{x}(t_l)\|, i=1,m$ ; is an observation matrix for  $j=1, n, t = t_l$  in the form of dimensional deviations  $m * n_2$ ;

$n_1, n_2$  are the volumes of samples, respectively, in the sections at  $t=t_k$  and  $t=t_l$ .

The hypothesis about the quasi-stationarity of the multidimensional system with respect to the vector of means (the vector of mathematical expectations) is accepted if:

$$T^2 < F(\alpha) \frac{(n_1 + n_2 - 2)m}{n_1 + n_2 - m - 1}, \tag{20}$$

where  $F(\alpha)$  is searched by the Fisher distribution table ( $F$ -distribution) with  $V_1 = m$  and  $V_2 = (n_1 + n_2 - m - 1)$  degrees of freedom and level of significance  $\alpha$ .

Criterion  $T^2$  is also advisable to use in the cases where it is assumed that the correlations in the investigated sections differ insignificantly ( $K(t_k) = K(t_l)$ ). In the case when the

differences in the correlation matrices are significant and  $n_1=n_2=n$ , then the hypothesis on the mean vectors' equality can also be tested using  $T^2$ -criterion.

If there is a reason to consider the differences between the correlation matrices and  $n_1 \neq n_2$ , then the comparison of the means should be carried out using the Behrens - Fisher test ([7]).

Let  $n_1 < n_2$  be for determination. Then the criterion formula has the form:

$$T^2 = n_1(\overline{M}_x(t_k) - \overline{M}_x(t_i))S^{-1}(\overline{M}_x(t_k) - \overline{M}_x(t_i))^T, \tag{21}$$

where:  $(n_1-1)S = U_k U_k^T$ ;

$U_k = \|u_{ij} - \overline{U}_{ij}\|, i=1,m; j=1,n$ , defines a dimensionality deviation matrix  $m * n_1$ ;

$$\overline{U}_{ij} = \left( \sum_{j=1}^n u_{ij} \right) / n_1, i=1,m;$$

$$u_{ij} = x_{ij}(t_k) - \sqrt{n_1/n_2} x_{ij}(t_i), i=1,m; j=1,n.$$

Random value  $T^2$  has  $F$ - distribution (Fisher) with  $V_1=m$  and  $V_2=n-1$  degrees of freedom. The hypothesis about the vectors' equality of the means is accepted if:

$$T^2 < F(\alpha) \frac{(n_1 - 1)m}{n_1 - m} \tag{22}$$

The hypothesis about the vectors' equality of the means is accepted if the value  $W$  is less than the table value  $x^2$  with the number of degrees of freedom  $V = m$  and the level of significance  $\alpha$ .

In order to check the quasi-stationarity of a multidimensional random system with respect to covariance matrices, it is necessary to use the Kullback criterion ([8]), since in it, as well as in the criteria proposed above, the generalized estimates of the structure of connections are used:

$$J = (n_1-1) \lg(|S|/|K(t_k)|) + (n_2-1) \lg(|S|/|K(t_i)|), \tag{23}$$

where  $|S|$  is a determinant of generalized covariance matrix.

A significant discrepancy is estimated using the distribution  $x^2$  with the number of degrees of freedom  $\nu = m(m+1)/2$  and the level of significance  $\alpha$ . The hypothesis of covariance matrices' equality is accepted if  $J < x^2_{tab}(\alpha, \nu)$ .

For an express analysis of ISS functioning modified the criterion for simultaneous testing of the hypothesis about the vectors' equality of the means of two samples and covariance matrices for the synthesis of the numerical scheme of the corresponding IT.

For these purposes, a random variable is used:

$$V = \frac{|A_1|^{(m_1-1)/2} |A_2|^{(n_2-1)/2}}{\left| A_1 + A_2 + mn_2 / (n_1 + n_2) (M_x(t_k) - M_x(t_i))(M_x(t_k) - M_x(t_i))^T \right|^{(m_1+n_2-2)/2}}, \tag{24}$$

where:  $A_1 = \|X_k X_k\|, A_2 = \|X_i X_i\|$ .

It can be shown that for large sample sizes  $n_1$  and  $n_2$  magnitude  $(-2 \rho \ln W)$  is approximated by the distribution  $x^2$  with the number of degrees of freedom  $V = m(m+3)/2$ , where:

$$W = \frac{(n_1 + n_2 - 2)^{m(n_1+n_2-2)/2}}{(n_1 - 1)^{m(n_1-1)/2} (n_2 - 1)^{m(n_2-1)/2}} V \tag{25}$$

and

$$\rho = 1 - \left[ \frac{1}{n_1 - 1} + \frac{1}{n_2 - 1} + \frac{1}{n_1 + n_2 - 2} \right] \frac{2m^2 + 3m - 1}{6(m + 3)} - \frac{m}{(n_1 + n_2 - 2)(m + 3)}$$

Then if  $(-2 \ln \rho W) < \chi^2_{tab}(\alpha, \nu)$ , then the hypothesis is accepted, the random system is recognized as quasi-stationary simultaneously with respect to the vectors of mean and covariance matrices. Otherwise, the multidimensional system should be considered nonstationary, and as a measure of the discrepancy with respect to the means, we can propose:

$$\delta M_x = \sqrt{\sum_{j=1}^m \omega_j [\bar{x}_j(tk) - \bar{x}_j(tl)]^2} \tag{26}$$

where  $\omega_j$  is a normalizing factor.

### 3 Results and discussion

#### 3.1 ISS ATI state analysis features

Thus, the analysis of the ISS ATI state characteristics and selection of statistical analysis criteria ISS ATI allowed to determine the features of the ISS ATI state analysis which will be discussed below.

As already noted, a single observation of the ATI integrated safety system state is a random vector  $x=(x_1, \dots, x_m)$ . Due to the presence of a large number of independent small influences on the formation of parameters and in accordance with the De Moivre-Laplace limit theorem  $x_i, i = 1, m$  basically have a normal distribution law. Along with this, other types of distributions which are reduced to normal are possible, only preliminary functional transformations to the normal form are necessary for them. In addition, most statistical methods are highly robust to small deviations from the normal law, so they are applicable to almost any unimodal and symmetric distribution ([9]).

Checking the normality of the distribution should be a preliminary stage of the state analysis. ISS artificial traffic installation and a significant deviation from the normal distribution law can serve as a diagnostic sign of the system non-stationarity. Since the vector  $\bar{x}$  distribution parameters are unknown, then it is necessary to have parallel data samples of sufficient volume for the statistical analysis. Then a single observation of the system state is the matrix:

$$X = \|x_{ij}\|, i=1, m, j=1, n, \tag{27}$$

where:  $m$  is a number of parameters,  $n$  is a number of observations.

Then the procedures for estimating ISS ATI stationarity are built on the basis of a comparison of the current  $X$  with the previous one, either with an average for a sufficiently large number of measurements, or, even more efficiently, with some reference  $X^0$  values.

The simplest way of statistical estimation is coordinate-wise comparison  $X$  and  $X^0$ , in which the corresponding columns of the matrices  $X$  and  $X^0$  are individually compared, as the

samples of scalar random variables. Scalar statistics are much better developed than vector statistics and provide more reliable results. Therefore, such a comparison is a sufficient tool for identifying the differences, and it should be used in the first stage of the analysis of the ISS ATI state, allowing to make fairly reliable conclusions about the system state.

The advantages of the coordinate-wise comparison are as follows ([10]):

- great theoretical validity, elaboration, branching of methods;
- obtaining higher confidence probabilities with lower volumes;
- computational simplicity;
- availability of standard applications.

The disadvantages include:

- abundance of non-collapsed information and, as a result, difficulty in interpretation for decision-making;
- lack of interaction consideration of the parameters.

The coordinate-wise comparison of the means is used as an auxiliary diagnostic technique, which makes it possible to concretize the features of non-stationarity ISS ATI. Coordinate comparison of variances shows how much the degree of random influences on each individual parameter has changed.

The greatest diagnostic value in the ISS ATI study has a comparison of the mean over the population, since it:

- does not require additional information folding;
- evaluates the cumulative effect of individual parameters;
- can be unambiguously used for decision making;
- more resistant to random fluctuations, that is, less likely to give a "false alarm".

With a simultaneous comparison of means and variances, the following cases are possible:

- 1) simultaneous equality of grades  $\bar{x}_{(tk)} = \bar{x}_{(tl)}$ ,  $\sigma^2_{x(tk)} = \sigma^2_{x(tl)}$  means a quasi-stationary system;
- 2) inequality of variances with equal mean values characterizes an increase in the influence of random factors on the system under study, a change in its reproducibility – non-stationarity with respect to variances;
- 3) inequality of means with equal variances indicates a change in the ISS adjustment level while maintaining reproducibility – non-stationarity relative to the average;
- 4) inequality of means and variances  $\bar{x}_{(tk)} - \bar{x}_{(tl)}$ ,  $\sigma^2_{x(tk)} \neq \sigma^2_{x(tl)}$  characterizes a non-stationary system in relation to the level of adjustment and playback accuracy.

Along with the average values, the ISS ATI state is essentially characterized by the relationships between its individual characteristics. The most accessible and commonly used for the study of statistical relationships are various correlations (paired, partial and multiple correlations). The advantages of these methods are ([11]):

- easy accessibility;
- easy computability;
- sufficient visibility - an experienced technologist, using only the correlation coefficients, can determine some failures in the system and determine the direction of their elimination.

Disadvantages:

- with complex parameters' interactions, the phenomenon of false correlation arises (will be discussed below);
- lack of a causal orientation of the established relationships of parameters;
- a large effect of relationships linearity, since by its theoretical nature, the pair correlation coefficient is best suited for identifying the linear component of a statistical relationship, therefore, in the presence of large nonlinear effects, its information value decreases, but the situation can be corrected by introducing the correlation relations.

### **3.2 Methodological principles of ISS ATI diagnostics**

As a result of studying the features of ISS ATI state analysis it might be necessary to proceed to the development of methodological principles of ISS ATI diagnostics.

For ISS ATI stationarity analysis a methodology for diagnostic analysis has been developed, which consists in comparing the statistical characteristics of the current (dysfunctional) and reference implementations of the system functioning. All diagnostic procedures form a hierarchical structure consisting of three levels: primary diagnosis, condition diagnosis, causes diagnosis.

The methods consisting in the statistical analysis of each parameter separately are at the lowest level. As a rule, this diagnosis does not even require comparison with the standard. Not only the average value, but also the variance of the indicator, the correlation between pairs of indicators, etc. can have a diagnostic value ([12]).

The second level is formed by various statistical methods for comparing the next implementation with the standard. They must answer two main questions: are the differences between a given implementation and the reference significant, and what implementation features are they related to? Multidimensional statistical procedures are used here, requiring complex programs and rather voluminous calculations. These include: comparison of vectors of means; comparison of covariance matrices, tracing relationships between the parameters, etc.

The third level - diagnostics of the causes - occurs after the system malfunctions detection at the second level and is intended to determine the characteristics of the system, the control of which is able to return it to the reference state. The methods used here can be internal (for example, factor analysis, which establishes the grouping of parameters according to their relationship and influence on the system; informational, regression and other types of analysis), as well as external, i.e., related to the study of data on modes and input parameters. The third level procedures are selectively linked to the second level procedures. So, variance analysis is performed when a breakdown in variances is detected, factor analysis is performed when there is a breakdown in relationships, regression analysis is performed when the mean is deviated, etc. Third-level procedures directly serve to support decision-making by a technologist or other authorized employee.

In contrast to a fairly clear distribution of diagnostic procedures over three levels, which is valid for any systems, the hierarchy of procedures within the levels either does not exist, or is associated with the specific features of each ISS ATI. At the same time, the establishment of such a hierarchy can be very useful, first of all, for the second-level procedures as the most time-consuming and expensive. It should also be taken into account that the procedures of the first and third levels solve, as a rule, different tasks for each procedure, and the order in which they are carried out is most often determined practically. But the second-level procedures solve the same problem of comparing the current implementation with the reference, and the choice of preferences without additional information is impossible here ([13]).

Let us consider the practical value of an objectively established hierarchy of second-level procedures. First, it will make it possible to establish the necessary frequency of monitoring and diagnostics for various criteria. Secondly, it is convenient for diagnosing the system "to the first failure", i.e., to the first criterion that found deviations in the parameters of the current implementation from the standard (similar to Bayesian troubleshooting methods). Thirdly, it will accelerate the corrective effect on the system during express diagnostics with a step-by-step identification of the reasons.

Based on the main goal of diagnosing the condition, it is natural to put the principle of greater sensitivity to violations of ISS ATI stationarity at the basis of the second-level procedures hierarchy. As a result, it becomes possible to establish the frequency of a particular diagnostic procedure taken separately; if the procedures are carried out completely,

then the hierarchy will permit setting the order of their implementation - from more sensitive to less sensitive.

Thus, the final hierarchical structure of the second-level procedures should have the form of a diagnostic tree, each branch of which corresponds to the positive or negative response of the previous diagnostic procedure ([14]).

The main type of second-level diagnostic procedures (condition diagnostics) is testing the statistical null hypothesis associated with a particular statistic calculated from the current and reference matrices  $X$  and  $X_0$ .

The rules for comparing and ranking two procedures of this type have been developed. In this case, it is considered that the quality of the final product is assessed by one numerical indicator  $k$  (for the reference  $k_0$ ), for example, the batch average yield factor. We will also assume that the type of probabilistic distribution of matrix elements is known due to purely random factors (or the absence of failures); most often, this distribution is close to normal, which can be verified by special fitting criteria.

So, let  $S(X)$  be some statistics with a known theoretical distribution; the corresponding general statistics will be denoted:  $\tilde{S}(X)$ . Let us denote null hypothesis probability  $S(X)=\tilde{S}(X_0)$ , estimated by sample values  $S(X)$  and  $S(X_0)$  by  $P_S$ . In other words,  $P_S$  denotes a possibility that the difference between  $X$  and  $X_0$  from the point of view of statistics is purely random, which means that we can expect the final equality of quality indicators  $k=k_0$ . The less probability of  $P_S$  exists, the more doubt about the achievement of such an outcome, the more attention should be paid to the indicators related to statistics  $S$ . Therefore, it is proposed to take the value  $J_S = -\ln P_S$ , justifying the use of the logarithm, just like in information theory.

The values  $J_S$  according to different statistics  $S$  are used to rank the statistics in terms of their diagnostic value, based on the data from one current implementation  $X$ . However, such a ranking will largely depend on the random features of a given implementation, which is undesirable. Drawing on the same material on other implementations, it is possible to get different rankings, for averaging of which additional rules are needed. The most significant in this case seems to be the connection  $J_S$  with a final quality indicator  $k$ , more precisely, with the difference  $\Delta_{k=k_0-k}$ . High value  $J_S$  at small  $\Delta_k$  is a "false alarm" and should make a reduced contribution to the overall quality estimation of statistics  $S$ . In other words, this overall estimate should be a weighted average using  $\Delta_k$  as a weighing factor:

$$\tilde{J}_{S=} = \frac{\sum J_S \Delta k}{\sum \Delta k}, \tag{28}$$

where the summation is over all considered realizations  $X$ .

Comparison of ratings  $\tilde{J}_S, \tilde{J}_T$ , for all targeted statistics  $S, T$ , will make it possible to select the best statistics for the first diagnostic procedure, place it at the base of the diagnostic tree. To continue this tree, it might be necessary to set the level of significance  $P$ , where the null hypothesis  $\tilde{S}(X)=\tilde{S}(X_0)$  rejected (for example,  $P=0,05$ ). After that, all implementations  $X$  split into two sets  $G_1$  and  $G_2$  by rule  $X \in G_2$ , if the null hypothesis is rejected (the answer is "yes" to the question about the presence of failures), and  $X \in G_1$ , if the null hypothesis is accepted (answer "no" to the same question). For each of the sets, the estimates of all statistics are again calculated, except for the one that was previously worked out as the best. Comparing these estimates separately for each set, we obtain the best statistics for diagnostics of the second order on each branch outgoing from the base. Continuations of the diagnostic tree for all subsequent diagnostics are constructed in a similar way.

Deeper relationships between the statistics  $S, T, \dots$  with the indicators of the quality of finished products can be established using the methods of variance analysis. First of all, they

are able to quite strictly establish (or reject) the very fact of such a relationship, and then assess the strength of this relationship using standard methods. Finally, variance analysis allows to obtain not only individual, but also mixed estimates of different statistics, their interaction in influencing the final indicators, which is essential when constructing a diagnostic tree ([15], [16]).

Let  $G_S$  be the class of individual (for each implementation) estimates  $J_S$  for statistics  $S$ . We arrange these estimates in ascending order and divide them into  $n$  intervals ( $n=3 \div 5$ )  $G_{S1}, G_{S2}, \dots, G_{Sn}$ . Each statistic will be considered as a factor, and the constructed intervals - as the levels of this factor. Calculating grades  $J_S, J_T \dots$  for each implementation  $X$  can be viewed as a passive element for the simultaneous implementation of the factors  $S, T, \dots$  with the levels that will be obtained as a result of these calculations. The result of this experiment will be considered the same difference  $\Delta_{k=k_0-k}$ . With such an interpretation, we form all the prerequisites for the use of multivariate variance analysis.

Since the experiment is passive, the repeatability of certain combinations of factor levels varies does not depend on LPR. This repeatability can be partially controlled by the initial choice of intervals.  $G_{S1}, \dots, G_{Sn}$ , as well as discarding some redundant data. The results of variance analysis are known to have the character of variances of all operating factors, their interactions, as well as the factor of randomness  $\sigma^2$ . Leaving only second-order interactions, we get a set of numbers:

$$\sigma^2, \sigma_s^2, \sigma_T^2, \sigma_{ST}^2.$$

First of all, comparing all variances with  $\sigma^2$  using Fisher's criterion, we leave only those that are significantly larger; statistics with other variances in this situation should be recognized as unsuitable for diagnostic purposes. We choose the largest of the remaining variances - it will determine the leading statistics. To select the statistics of the second stage, it is necessary to use the interactions variances. Statistics  $T$ , corresponding to the maximum  $\sigma_{ST}^2$  for all  $T$  (here  $S$  - the leading statistic already selected), will determine the second-order diagnosis, provided that the first gave a positive result; minimum statistics  $\sigma_{ST}^2$  will determine the second order diagnosis in case of a negative result of the first diagnosis.

The use of ANOVA to rank statistics in terms of their diagnostic sensitivity makes it possible to give such ranking a deeper explanation (not to mention filtering out unpromising statistics) ([17]). However, for sufficient reliability of the conclusions, a very large amount of initial data is required.

Nonparametric statistics can also be used. Let, for example, the matrices  $X$  and  $X_0$  be scanned in some order (for example, by columns), as in the inversion criterion, and each element exceeding  $x_{ij}$  over  $x^0_{ij}$  be marked with a plus sign, while the opposite excess is marked with a minus sign. The number of sign changes in the resulting sequence from  $m*n$  elements are denoted by  $N$ . For purely random discrepancies between the matrices  $N$  there is an almost normal distribution over the mean  $(m*n-1)/2$  and variance  $(m*n-1)/4$ . Based on this data, the probabilities  $P_N$  and the summary score  $J_N$  are calculated.

There are not so many significantly different statistics (like the ones mentioned above), even with known distributions. However, the methods we propose can also be used to compare the same statistics applied to different sets of ISS ATI functioning parameters, which will allow identifying the parameters most indicative for the stability of the process, simplifying and reducing the diagnostics cost.

To develop algorithms and programs for ranking diagnostic procedures to control ISS ATI stationarity, three most important statistical characteristics of multivariate populations were selected: mean, variance, covariance. Due to the multidimensionality, we will deal with the vectors of means and variances:

$\bar{x} = \{x_1, x_2, \dots, x_m\}$ ,  $\sigma^2 = \{\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2\}$  and the covariance matrix  $Q = \|q_{ij}\|$ ,  $i=1,m, j=1,m$ .

Here  $m$  is a total number of parameters,  $x_i$  and  $\sigma_i^2$  denote mean and variance of the  $i$ -th parameter parallel observations,  $q_{ij}$  is a sample covariance for the  $i$ <sup>th</sup> and  $j$ <sup>th</sup> parameters. For similar characteristics calculated from the reference implementation, we will use the notation  $\bar{X}_0$ ,  $\bar{\sigma}_0^2$  and  $Q_0$ . Null hypotheses mean the coincidence of the general mean variances or covariance matrices of the current and reference implementations.

The choice of criteria for assessing the discrepancy between the vectors of means and covariance matrices was justified above.

Unlike the two previous statistics, variances are considered and compared only for one-dimensional random variables. Therefore, a procedure for combining not the variances themselves, but the probabilistic estimates of their comparison, using the probability multiplication theorem has been developed. It is convenient to use this theorem for independent events and provided that there are not very many of them. In our case, neither one nor the other condition, as a rule, is fulfilled, therefore, a preliminary grouping of parameters, their enlargement, is necessary.

There are various methods for combining and averaging random variables. For our purposes, factor analysis is most acceptable, pursuing precisely the goals of consolidation and independence of factors. For any set of random vectors, it is possible to construct a matrix of factor loadings  $R$ , multiplying which by the matrix  $X$ , to get a new matrix  $W=RX$ , with a small number of weakly dependent columns. Matrix  $R$  is built on the matrix  $X$ , in this connection, the number and nature of factors will vary from one ISS ATI implementation to another, which is unsuitable for our problem of comparing variances.

Therefore, it is proposed to calculate the matrix  $R$  for only  $X_0$  reference, applying it in the future to any  $X$ . This will give a single grouping of parameters into factors and a single number of columns in all matrices  $W$ . Increasing dependency between columns  $W$  in this case turns out to be an additional diagnostic moment in the ISS stationarity estimation.

Let us calculate the sample variances  $\sigma_i^2$  and  $\sigma_{i0}^2$ ,  $i=1,k$  for each column of matrices  $W$  and  $W_0$ , where  $k$  is the number of columns. Using Fisher's criterion, we estimate the probability of a random discrepancy between  $\sigma_i^2$  and  $\sigma_{i0}^2$  for each column. Considering the standard is always the best in relation to any current implementation, we will try to avoid accidental exceeding  $\sigma_{i0}^2$  over  $\sigma_i^2$ , assuming, that

$$F_i = \max \{1, \sigma_i^2 / \sigma_{i0}^2\}.$$

We denote the probability of exceeding the number  $F_i$  in the Fisher distribution by  $P_i$ . Ignoring the residual relationship between the matrix columns  $W$ , we find the combined probability:

$$P = 1 - \prod_{i=1}^k (1-P_i), \tag{29}$$

taking the magnitude  $J = - \ln P$  as the efficiency of statistics  $S^2$  for the following calculations.

Let us consider an algorithm for constructing a diagnostic tree for the statistics presented above, in which we denote:

- $S$  – comparison of averages;
- $K$  – mean covariances;
- $D$  – variance comparison.

The reference matrix is taken as a basis  $X_0$  and many matrices of current implementations  $X_j, j=1,N$ . We represent the algorithm in the form of an enlarged procedure blocks sequence.

1. Primary statistical processing of the matrix  $X_0$ . Here, the vector of column averages  $x_0$  is calculated, the vector of sample variances  $\sigma^2$  and covariance matrices  $Q_0$ , memorizing the number of parallel observations (matrix rows)  $n_0$ .

2. Primary statistical processing of matrices  $X_j$  for  $j = 1, N$ . Here the vectors of means  $X_j$  are calculated, variances  $\sigma_j^2$ , covariance matrices  $Q_j$ , number of parallel observations  $n_j$ . Additional inspection and quality improvement activities can be included in this procedure  $X_j$  (identification and filtering of gross deviations, smoothing and restoration of missing data, etc.).

3. Computing summary covariance matrices:

$$Q_{1j} = \frac{1}{n_j + n_0 - 2} ((n_j - 1)Q_j + (n_0 - 1)Q_0), j=1, N \quad (30)$$

4. Matrix inversion  $Q_{1j}$ .

5. Calculating Probabilities by Hotteling's Criterion using the formulas:

$$T_j^2 = \frac{n_j n_0}{n_j + n_0} (\bar{x}_0 - \bar{x}_j) Q_{1j}^{-1} (\bar{x}_0 - \bar{x}_j)^m, \quad (31)$$

$$F_j = \frac{n_j + n_0 - m - 1}{(n_j + n_0 - 2)m} T_j^2, \text{ for all } j = 1, N \quad (32)$$

6. Calculation of the yield factors suitable for the standard  $k_0$  and for each implementation  $k_j$ , vector drawing:

$$\bar{k} = (k_1, k_2, \dots, k_N),$$

$$k_j = \max \{ 0, k_0 - k_j \}.$$

7. Calculation of the final indicator of the  $S$  statistics effectiveness:

$$J_S = - \sum \ln P_j \Delta k_j / \sum \Delta k_j. \quad (33)$$

8. Calculating the determinants of a matrix  $Q_0$ , matrices  $Q_j$  and  $Q_{1j}$  for all  $j=1, N$ .

9. Calculation of probabilities by  $\chi^2$  - criterion using the formulas:

$$\chi^2 = (n_j + n_0) \ln |Q_{1j}| - n_j \ln |Q_j| - n_0 \ln |Q_0|, \quad (34)$$

For all  $j = 1, N$ .

10. Calculation of the final indicator of the  $K$  statistics effectiveness:

$$J_K = \frac{\sum \ln P_j \Delta k_j}{\sum \Delta k_j}. \quad (35)$$

11. Determination of the factor loadings matrix  $R$  using the factor analysis of the standard, memorizing the number of factors  $k$ , matrix computation  $W_0 = RX_0$ .

12. Matrix transformation  $X_j$  in the matrix  $W_j$  by multiplying (left) by  $R$ :

$$W_j = RX_j, j = 1, N.$$

13. Calculation of variances over the matrix columns  $W_0, W_j$ :

$$\sigma^2 = \{ \sigma_{10}^2, \dots, \sigma_{k0}^2, \sigma_j^2 = \sigma_{1j}^2, \dots, \sigma_{kj}^2 \}$$

and estimation of their adequacy according to Fisher’s criterion:

$$F_{ij} = \max \{ 1, \sigma_{ij}^2 / \sigma_{10}^2 \}, \quad i = 1, k,$$

$$P_j = 1 - \prod_{i=1}^k (1 - P_{ij})$$

For all  $j=1, N$ .

14. Calculation of the final indicator of the  $D$  statistics effectiveness:

$$J_D = - \frac{\sum \ln P_j \Delta k_j}{\sum \Delta k_j} \tag{36}$$

15. Comparison of indicators  $J_S, J_K, J_D$  and the selection of statistics corresponding to the largest indicator. This will be the statistics of the first stage, the base of the diagnostic tree. ([18]).

16. Based on the chosen statistics of the first stage and some level of significance  $D$  (for example,  $P=0,05$ ), the entire original set of matrices  $X_j$  is divided into two groups: the first (it corresponds to a set of numbers  $G_1$ ) for the matrices of which the probabilities  $P_j$ , calculated according to the statistics of the first stage, are more than  $P$ , and the second (it corresponds to a set of numbers  $G_2$ ), for the matrices of which  $P_j < P$ .

17. For clearness, the remaining two statistics will be denoted as A and B. For them, for each of the groups built above, new performance indicators are calculated as follows:

$$J_A^1 = - \frac{\sum \ln P_j \Delta k_j}{\sum \Delta k_j}, \quad j \in G_1, \tag{37}$$

$$J_A^2 = - \frac{\sum \ln P_j \Delta k_j}{\sum \Delta k_j}, \quad j \in G_2 \tag{38}$$

and similarly,  $J_B^1, J_B^2$ .

18. Evaluations  $J_A^1, J_B^1$  are compared with each other, and the one that is larger determines the statistics of the second stage for the case when the first stage corresponds to the answer “no differences”. Similarly comparing  $J_A^2, J_B^2$ , we find the statistics of the second stage for the case when the first stage corresponds to the answer "there are differences".

After choosing the statistics of the second stage on each branch of the diagnostic tree, the statistics of the third stage can be determined unvaliantly ([19]).

## 4 Conclusion

Thus, we can conclude that the developed mathematical apparatus designed for diagnostic analysis of the ISS ATI state, based on the multivariate statistics methods, makes it possible to estimate the values of ISS parameters in the process of its functioning (in dynamics), and also to assess the interdependence of these parameters. Selection of ISS ATI statistical analysis criteria allowed to go to the analysis of the ISS ATI state. Checking the normality of the distribution should be a preliminary stage of the ISS ATI state analysis and a significant deviation from the normal distribution law can serve as a diagnostic sign of the system non-

stationarity. This comparison is a sufficient tool for identifying differences and should be used as a first step in the ISS ATI status analysis, allowing to make fairly reliable conclusions about the system state. The developed methodology of diagnostic analysis, which consists in comparing the statistical characteristics of the current and reference realizations of functioning ISS makes it possible to effectively assess the quality of developed and implemented ISS for ATI.

Application of this mathematical apparatus in the ISS ATI development allows to significantly reduce the cost of their implementation and significantly increase ISS efficiency when implemented on ATI.

## References

1. E.V. Sandrakova, E.V. Sumin, *Differential forms on smooth manifolds* (2nd edition, Yurayt Publishing House, Moscow, 2020).
2. A. Ginzburg, S. Kachanov, S. Kozminykh, *Maintenance of comprehensive safety of tunnel-type road interchanges FORM-2020*, IOP Conf. Series: Materials Science and Engineering **869**, 052030 (2020). IOP Publishing doi:10.1088/1757-899X/869/5/052030. This content was downloaded from IP address 109.252.64.245 on 09/07/2020.
3. A.V. Ginzburg, L.A. Shilova, L.A. Shilov, *The main trends of computer-aided design systems development*, VI International Scientific Conference “Integration, Partnership and Innovation in Construction Science and Education”, IPICSE-2018, Moscow; Russia, MATEC Web of Conferences **251**, 05001 (2018). doi: <https://doi.org/10.1051/matecconf/201825105001>.
4. A. Ginzburg, A. Bolotova, A. Dolganov, I. Vedyakov, M. Vaynshteyn, *Organizational-technological reliability of monolithic construction*, Proceedings of the 19th International Scientific Conference “Energy Management of Municipal Transportation Facilities and Transport EMMFT 2017”. Khabarovsk, Russia 10–13 April 2017. *Advances in Intelligent Systems and Computing* **692**, (2017). Springer, Cham. [https://doi.org/10.1007/978-3-319-70987-1\\_124](https://doi.org/10.1007/978-3-319-70987-1_124).
5. A.V. Ginzburg, International Scientific Conference Environmental Science for Construction Industry, ESCI 2018; Ho Chi Minh City; Viet Nam: MATEC Web of Conferences **193**, 05030 (2018).
6. S.I. Kozminykh, V.B. Gisin, S.V. Dvoryankin, V.I. Korolev [et.al.]; edited by S.I. Kozminykh, *Modeling of systems and processes of information protection at the facilities of the credit and financial sector* (Monograph, Information security of financial and credit organizations in the context of digital transformation of the economy: Prometey, M., 2020).
7. A.P. Potapov, *Mathematical analysis. Differential calculus of FRR, equations and series* (Yurayt Publishing House, Moscow 2020).
8. S.I. Kozminykh, S.A. Kachanov, *Technologies of Civil Security* **1**, 31-36 (2020).
9. O.D. Maksimova, *Foundations of Mathematical Analysis: Inequalities and Estimates* (Yurayt Publishing House, Moscow, 2020).
10. S.I. Kozminykh, *Cybersecurity Issues* **4**, 14-28 (2020).
11. A.D. Polyanin, A.V. Manzhirov, *Integral equations in 2 parts. Part 1: a reference book for universities*, 2nd edition, revised and added (Yurayt Publishing House, Moscow, 2020).

12. V.F. Zaitsev, A.D. Polyanin, Ordinary differential equations in 2 parts. Part 1: a reference book for universities, 2nd edition, revised and added (Yurayt Publishing House, Moscow, 2020).
13. S.I. Kozminykh, R.B. Raschevskiy, Methods for identifying threats to information security by analyzing network interactions Information security in the banking and financial sector: collection of articles; edited by S. I. Kozminykh (Russcience, Moscow, 2020).
14. S.I. Kozminykh, Scientific and practical journal Finance: theory and practice **5**, 105-121 (2018).
15. S.I. Kozminykh, Issues of Cyber Security **1(25)**, 54-63 (2018).
16. Yu.M. Beketnova, Models and methods for solving analytical problems of financial monitoring (Monograph. Publisher: Limited Liability Company "Publishing House Prometey", Moscow, 2018).
17. S.A. Ayvazyan, *Application of multivariate statistical analysis in economics and quality estimation*, Proceedings of the XI-th International Conference Moscow, August 21-23, 2018 Central Economics and Mathematics Institute RAS, Moscow, 171 (2018).
18. V.B. Gisin, E.S. Volkova, Fuzzy sets and soft computing in economics and finance. Tutorial. Financial University under the Government of the Russian Federation (KnoRus Company, Moscow, 2019).
19. A.D. Polyanin, Equations and problems of mathematical physics in 2 hours. Part 1 reference book for universities, 2nd ed., Rev. and add. (Yurayt Publishing House, Moscow, 2020).