

57. *Lignite* oxidative desulphurization. Notice 1. Process condition selection / Volodymyr Gunka, Serhiy Pyshyev // International Journal of Coal Science & Technology. – 2014. – Vol. 1. – Issue 1. – P. 62–69.
 58. *Toroptseva A.M.* Laboratory session on chemistry and technology of high-molecular compounds / A.M. Toroptseva, K.M. Belogorodskaya, V.M. Bondarenko. – Chemistry: L., 1972. – P. 416.

DOI: doi.org/10.18372/38228

UDC: 519.23:519.242

1.5 STATISTICAL METHODS IN VIRTUALIZATION

Alexandr Moiseev, Nikolay Grishin, Alexandr Oreshencov

The basis for virtualization is virtualization tools, including families of base functional means and base mathematical models [1]. Functional means usually use for synthesis or simulation of the software of various purpose, including processing of results of qualifying tests [2]. Base models use for testing the software, synthesis of identification models for is modelling – the focused control systems, and also algorithms of processing of results of qualifying tests.

An important component of the virtualization toolkit (see Fig. 1) is a statistical component to solve the following problems: identifications of statistical models by results of qualifying tests, parametrization criteria models for carrying out of forecasting within the limits of virtual qualification, synthesis regression models, the statistical analysis of their adequacy and consecutive optimization during factorial planning. Regression model in structure of the toolkit, calibrated by results of qualifying tests, use for planning qualifying experiments. As a rule, this planning is based on use of the sated plans and gives following opportunities to:

- research the importance and completeness of system of the entered factors;
- carry out consecutive optimization of the parameters corresponding significant factors, during qualifying tests;
- carry out forecasting qualifying tests with use calibrated regression models.

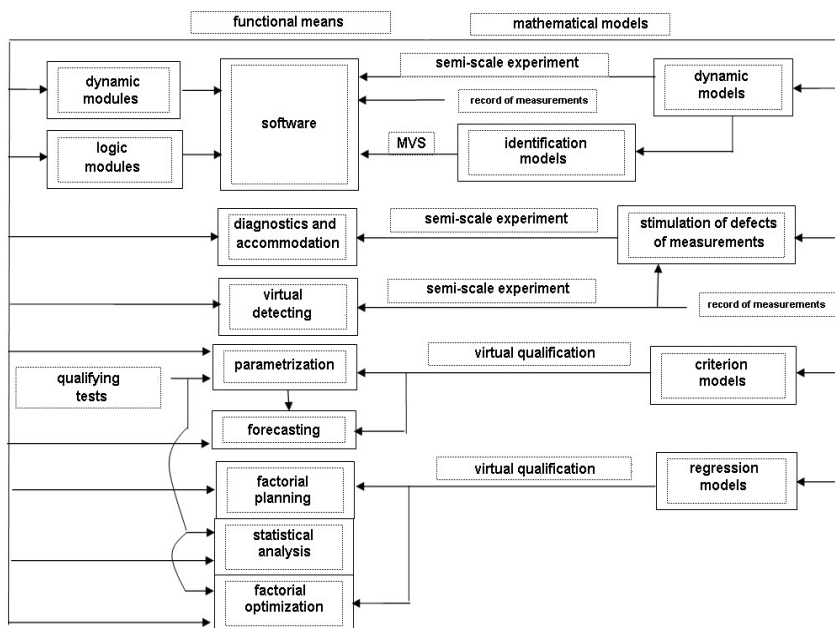


Fig. 1. Virtualization tools

Thus linearity used regression models allows to carry out the control of the importance of factors and completeness of their system over use of traditional statistical methods, and optimization of parameters by means of methods of linear programming.

The statistical analysis of adequacy regression models assumes use of some statistical criteria. In work [3] improvement of methods Cochran and Grubbs methods, providing them robustness and based on preliminary normalization of researched sample by means of corresponding Smirnov's transformation. It is shown, that the developed a process to ensure robustness reduces probability of error of the first sort at decision-making on reproducibility of results of qualifying experiment.

Distribution of Cochran's statistics on the basis of F -Fisher's distribution and Grubbs statistics on the basis of T -Student's distribution are constructed of following reasons. On the normalized samples of Student relations dispersions which then are analyzed by Cochran's criterion [3].

The decisive statistic of the Cochran represents relation of the maximal private dispersion $\max_{i=1...l} S_i^2$ and the sum of private dispersions $S_1^2 + \dots + S_l^2$ of Student relations. Here l is the number of measurement samples by which partial variances for student fractions are calculated. Each of them represents not displaced selective estimation of a kind:

$$S_{i1}^2 = \frac{S_{i1}^2 + \dots + S_{il}^2}{l-1} \quad (1)$$

$$S_{ij}^2 = \left(\xi_{ij} - \frac{1}{l} \sum_{k=1}^l \xi_{ik} \right)^2 \quad (2)$$

Which can be transformed as follows:

$$C = \frac{\max_{i=1...l} S_i^2}{S_1^2 + \dots + S_l^2} = \max_{i=1...l} C_i \quad (3)$$

$$C_i = \frac{S_i^2}{S_1^2 + \dots + S_l^2} \quad (4)$$

Under the assumption of the normality of the distribution of the numerator in C_i distributed on χ^2 with a degree of freedom $np-l$, and the denominator represents the sum l sizes with such distribution and therefore, it is distributed on χ^2 with a degree of freedom $l(np-l)$. From here

follows, that statistics $F_i = lC_i = \frac{S_i^2}{\frac{np-1}{S_1^2 + \dots + S_l^2}}$ has an f-Fisher distribution with degrees of

freedom $np-l$ and $l(np-l)$. Statistics Kohren takes, thus, the form $C = \frac{F_M}{l}$, where $F_M = \max_{i=1...l} F_i$.

Similarly, the Grubbs criterion for normalized student relations is constructed. These relations form a variational series ξ_1, \dots, ξ_p , the decisive statistics in which are the minimum and maximum Student ratios: $G_{\min} = \xi_1$, $G_{\max} = \xi_p$. In the case of normality of the sample student relations are distributed by Student, and serial statistics G_{\min} and G_{\max} have distributions of a kind $T_{n-1}^p(-G_{\min})$ and $T_{n-1}^p(-G_{\max})$, where T_{n-1} Student distribution with $n-1$ a degree of freedom. The decision on absence of reproducibility is accepted, if the condition

$T_{n-1}^p(-G_{\min}) > P$ or $T_{n-1}^p(-G_{\max}) > P$, where P – the set confidential probability is satisfied. Otherwise it is considered, that measurements are reproduced.

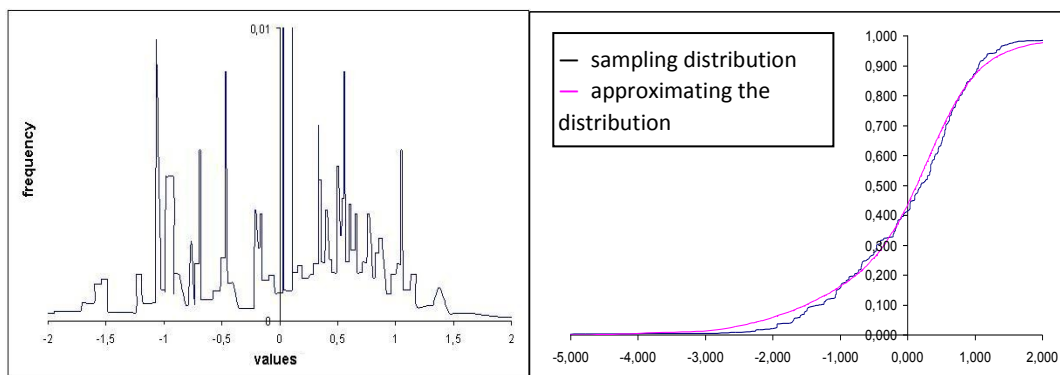
Is equivalent to the Grubbs criterion is the criterion Hawkins, crucial statistics which H is associated with statisticians G_{\max} , G_{\min} ratio:

$$H = \max(|G_{\max}|, |G_{\min}|). \quad (5)$$

In view of that H is an ordinal statistic, its distribution under the assumption of normality of the General set of measurements has the form $(2T_{n-1}(H) - 1)^p$. The relevant decision rule is as follows: the decision on the absence of reproducibility of measurements, if $(2T_{n-1}(H) - 1)^p > P$, and about their reproducibility otherwise.

The important component of interlaboratory measurements of characteristics of mineral oil is identification of distribution of their error defining a technique of the further processing of measurements. The article [4] presents an analysis of the distribution of these errors, including measurements of density and viscosity is lead at temperatures 20 °C and 100 °C, and also temperatures distillate and flashes. The analysis of errors of measurements of these characteristics, lead by a method χ^2 , has shown, that the assumption used usually about normality this distribution is represented superficial. In these conditions there is a necessity to lead identification of the specified law of distribution and on this basis to develop algorithm of rational processing of measurements.

Below some methods of construction of approximating distributions and their parametrization are considered. A starting point of construction of approximating distribution of an error of measurement of density, construction of the frequency histogram of the specified error displayed on Fig. 2.



Frequency histogram of density error

Fig. 2. Approximation of distribution of an error of measurement of density

It obviously specifies on bimodal character of distribution of an error of corresponding measurements. For approximation of this distribution it is offered to use distribution of a kind:

$$f(x) = \frac{\exp\left(-\frac{(x - m_1)^2}{2\sigma_1^2}\right)}{2\sigma_1\sqrt{2\pi}} + \frac{\exp\left(-\frac{(x - m_2)^2}{2\sigma_2^2}\right)}{2\sigma_2\sqrt{2\pi}}. \quad (6)$$

Corresponding integrated distribution is represented thus in the form of:

$$F = \frac{1}{2}\Phi\left(\frac{x - m_1}{\sigma_1}\right) + \frac{1}{2}\Phi\left(\frac{x - m_2}{\sigma_2}\right). \quad (7)$$

From expressions 6 and 7 it is visible, that approximating distribution includes following parameters: m_1, m_2 – population means of components of distribution; σ_1, σ_2 – root-mean-square errors of a deviation components of distribution.

For an estimation of these parameters it is offered to use a method of the moments which allows to receive for them system of the algebraic equations:

$$\left\{ \begin{array}{l} \frac{m_1 + m_2}{2} = m \\ \sigma_1^2 = \sigma^2 - \frac{m_2 - m_1}{2} - \frac{A}{3 \left(\frac{m_2 - m_1}{2} \right)} \\ 3 \left(\frac{m_2 - m_1}{2} \right) \frac{\sigma_2^2 - \sigma_1^2}{2} = A \\ 3 \frac{\sigma_1^4 - \sigma_2^4}{2} + \left(\frac{m_2 - m_1}{2} \right)^4 + 6 \left(\frac{m_2 - m_1}{2} \right)^2 \frac{\sigma_1^2 - \sigma_2^2}{2} = E \end{array} \right. \quad (8)$$

Here A and E – selective asymmetry and an excess for sample of errors of measurements. The decision of the constructed system of the equations was carried out by a method of search on parameter m_1 in a range $(0, m)$. Value m_2 was for data m_1 from the first equation of system (8), σ_1^2 – from the second on set m_1 and calculate dm_2 , and σ_2^2 – from the fourth equation which at set m_1 and calculated m_2, σ_1^2 represents a quadratic rather σ_2^2 . During a variation the parameter m_1 stole up so what to provide the minimal deviations of the calculated parameters from selective. Result of the calculation lead thus were following values of parameters: $m_1 = -0.353$; $m_2 = 0.314$; $\sigma_1 = 1.384$; $\sigma_2 = 0.519$.

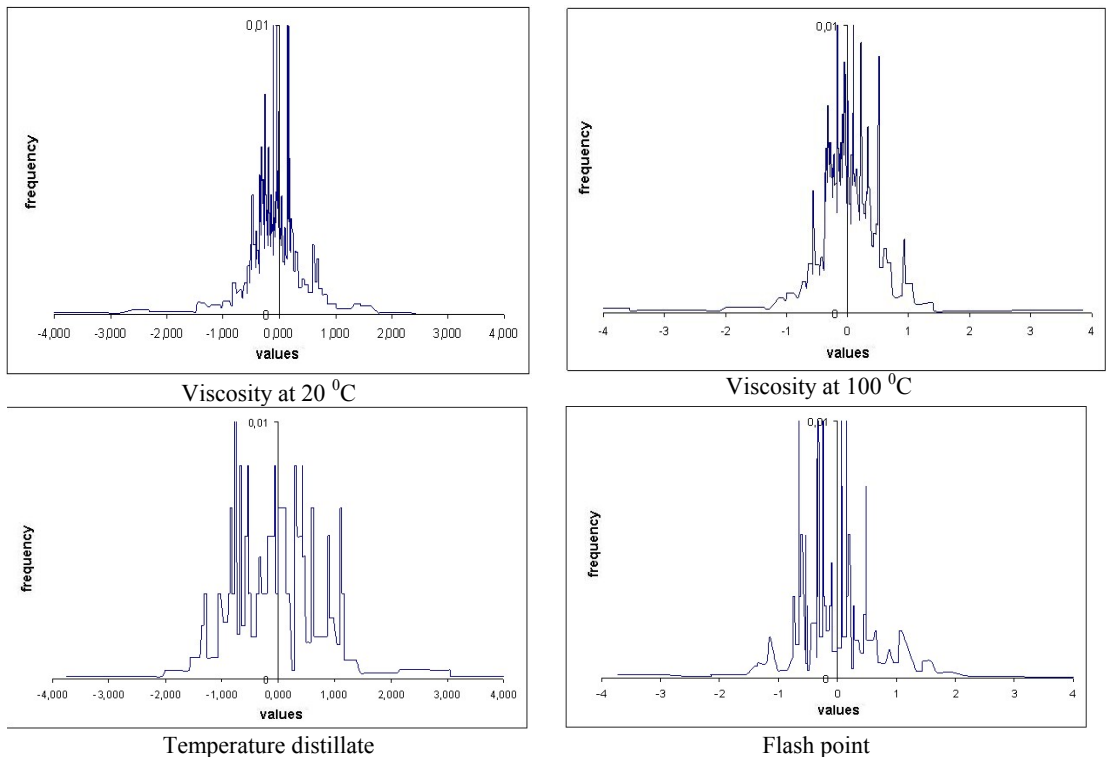


Fig. 3. Frequency histograms of an error of measurements

Substituting these values in expressions (6) and (7), we receive approximating distributions of an error of measurements of density. The result of comparison of integrated approximating distribution with selective is resulted on Fig. 2 and specifies qualitatively satisfactory approximation of distribution of the specified error.

Frequency histograms for the standardized error of measurements of other characteristics of mineral oil – viscosity at temperatures 20 °C and 100 °C, and also temperatures distillate and flashes – are resulted on Fig. 3 and specify unimodal character of distributions of an error of their measurements. As possible approximations of these distributions in work [4] were considered alongside with normal Pearson and Weibull approximation, and also lognormal. Parametrization of these distributions also was carried out by a method of the moments. The lead comparative analysis of the constructed approximations of unimodal distributions of an error shows, that the best on accuracy is approximation by the Weibull. Comparison of the received distributions of an error with bimodal approximation for an error of measurement of density, and Weibull approximation for an error of measurement of other characteristics specifies sufficiency of used approximations.

The modified algorithm of the control of reproducibility of interlaboratory measurements is based on alignment of the last by transition to Student fractions that allows to carry out the control of reproducibility over measurements of changing sizes. However in case of normal sample of measurements distribution of Student fractions appears standard and cannot be used for the control of reproducibility. Therefore Student's relation is expedient for defining in the form of $\xi = \frac{x - m}{s(m)}$, where $s(m)$ – approximation of dependence standard deviation from a population mean. Thus transformation of measurements in Student's relation provides precomputation mathematical expectations and standard deviation, corresponding different levels. They are used to construct the function $s(m)$ of linear regression which together with settlement population means is used for construction of fractions of Student's relation.

The further statistical conclusions base on the analysis of constructed Student fractions which are analyzed by Grubb's criterion. In case of non-reproducibility measurements by Grubb's criterion, a series corresponding extreme statistics, is rejected. The defective series leaves from sample of measurements, and solving statistics recalculate for the remained levels in the assumption, that approximation $s(m)$ remains constant. It can be interpreted as replacement of the defective series such which does not change approximation, i.e. we neglect influence of an abnormal series on last. If and in the truncated sample it appears possible to allocate an abnormal series, the assumption of weak influence of the rejected series on approximation is considered insolvent and it is necessary to carry out recalculation of approximation without taking into account the defective series.

On the samples of Student's relation generated thus dispersions, which then are analyzed by Cochran's criterion. It is serial and has distribution $F^l(F_m, np - 1, l(np - 1)) = F^l(lC, np - 1, l(np - 1))$, where F – Fisher's distribution. The condition of heterogeneity in these conditions looks like $F^l(lC, np - 1, l(np - 1)) > P$, where P – confidential probability of the decision about non-reproducibility. At default of this condition the decision on uniformity of sample, and, hence, makes of reproducibility. Thus, the modified algorithm of the control of reproducibility is applied in conditions of possible change of measured sizes. In this sense it differs universality from a usual technique of the classification focused on an invariance of measured values. Use of alignment of measurements by transition to Student fractions allows to apply this technique without any restrictions.

According to [5] criteria model allows to carry out forecasting by results of the qualifying tests spent in set conditions. There is, however, a question on a choice of such conditions which would provide the guaranteed forecast – for example, its top border. The problem of a preliminary choice of extreme conditions of qualifying tests for which decision it is offered to use a method of factorial planning of experiment from here follows. The statistical processing of these

measurements displayed on Fig. 4, includes: an estimation of reproducibility of multiple measurements by the analysis of their disorder; construction linear-regression models and an estimation of the importance of the entered factors; the analysis of adequacy of model and an estimation of completeness of system of the entered factors.

For an estimation of reproducibility of measurements Cochran's criterion is used. It provides calculation of dispersions of multiple measurements, formation on them Cochran's statistics and calculation of probability of its realization. This probability is compared to confidential probability non-reproducibility measurements. If the calculated probability appears below confidential, the hypothesis about reproducibility of measurements is accepted. Their dispersions thus are averaged, forming a dispersion of reproducibility. Otherwise a series of measurements is considered abnormal and should be repeated.

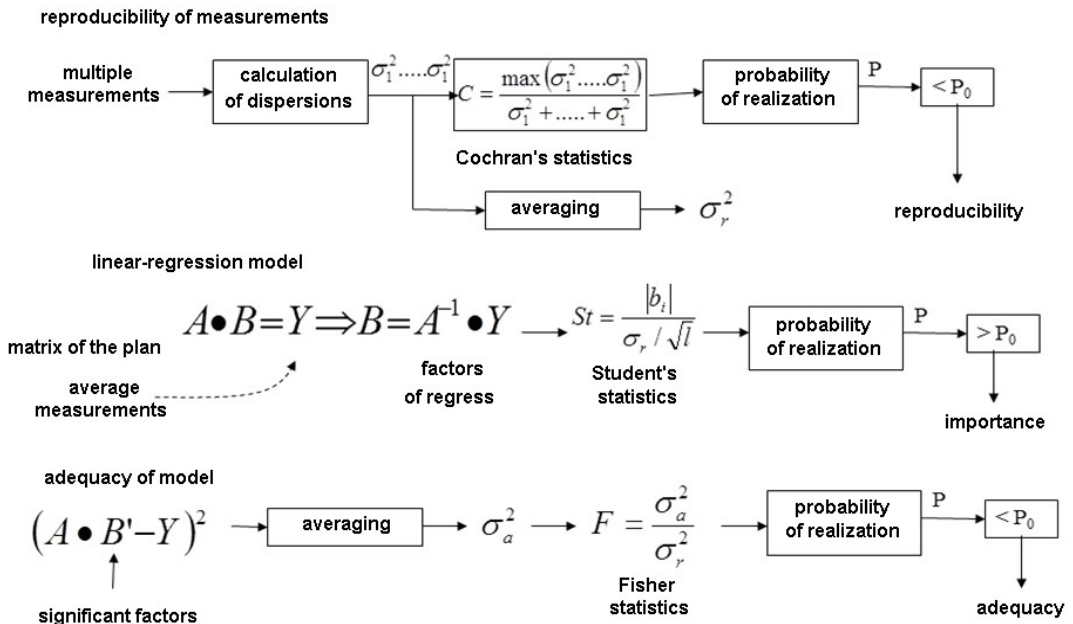


Fig. 4. Statistical processing of measurements

Calculation of factors of regress is carried out from following reasons. The matrix of the plan for not fictitious factors is nondegenerate matrix of system of the linear equations which right part is the vector of the average measurements. The vector of the factors interpreted as loading of factors, represents the decision of the specified system.

For an estimation of the importance of factors are used Student relations, the decision on which is accepted by Student criterion. For this purpose, student statistics are calculated for factors of regress and probability of their realization. Significant factors are considered, probabilities of realization of statistics Стьюдента for which exceed confidential. Other factors of regress are considered insignificant and in model should be nulled.

The control of adequacy linearly-regression models (completeness of system of the entered factors) is carried out by Fisher's criterion with use of the relation of a dispersion of adequacy to a dispersion of reproducibility. For this purpose the average square of a deviation of the average measurements from regression models, in which significant factors of regress. This size is interpreted as a dispersion of adequacy. After it and a dispersion of reproducibility Fisher's statistics is under construction and the probability of its realization. The decision on adequacy is

accepted, if the specified probability below confidential. Otherwise the model is considered inadequate, that, as a rule, means loss of essential factors at its construction.

Two-level factorial planning can be used for consecutive optimization of parameters of researched object or qualifying experiment. As local optimum thus are reached in tops multidimensional simplices, corresponding lines of the two-level plan, search of a global extremum is offered to be carried out a method of successive approximation. Consecutive optimization provides use of points of the reached local extremum as the centers of new areas in parametrical space.

When using a linear regression models search of an extremum represents a problem of linear programming for the closed area of factorial values in the form of a multivariate simplex. The local extremum of linear regress is reached thus in one of tops of area and, hence, is equal to the extreme average measurement corresponding one lines of the plan of experiment. The simplex-planning provides inclusion in factorial area also the simplex symmetric considered.

It corresponds to the plan of the experiment with the inverted matrix of factors, and the common area of search of an extremum according to Fig. 5 is formed initial, symmetric and additional simplices, adjoining to the beginning of coordinates.

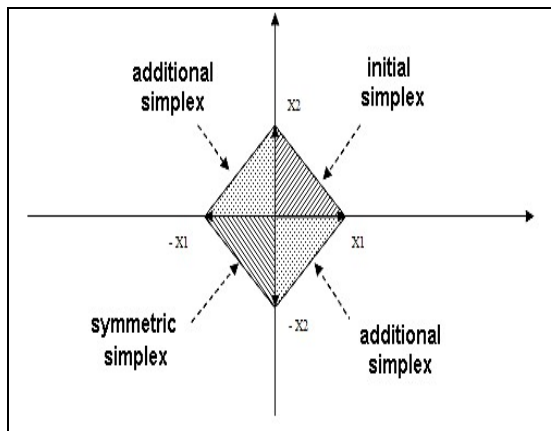


Fig. 5. Expansion of factorial area

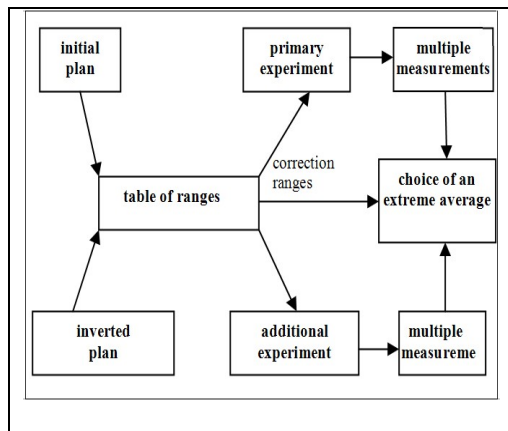


Fig. 6. Statistical processing of measurements

At the expanded planning displayed on Fig. 6, for formation of tables of parameters the initial and inverted plans are used. According to these tables qualifying tests during which multiple measurements are carried out are spent. Processing of these measurements is carried out in two stages. During primary processing search of the extreme average measurement which parameters are compared with the new center of ranges is carried out. To this center there corresponds the corrected range of parameters used for realization of following approach of procedure of search. If the extremum reached in following approach is more extreme" previous, procedure proceeds. Otherwise it stops also to a global extremum value previous local is compared. Parameterization of the criterion mode by results of qualifying tests in extreme conditions allows to lead forecasting of frictional deterioration in these conditions and to generate on this basis the guaranteed norms of deterioration.

Interesting example of application of statistical methods in problems virtualization is factorial planning in the analysis of criteria of similarity [6.] Criterial models are natural base for virtualization qualifying tests as the last are based on physical model of researched object or process in the form of laboratory installation [7]. The experiments spent on it, allow to carry out

parameterization of the criterion model that then to use it for recalculation of laboratory results in operational characteristics, and also for return recalculation of operational conditions in parameters of qualifying tests.

The important problem at a substantiation of invariancy of used criteria of similarity is their representation in the form of functions from standard numbers of similarity as their invariancy provides invariancy of the constructed criterion. Often this function is multiplicative-power, i.e. it is represented in the form of:

$$K = K_0 N_1^{\beta_1} \dots N_l^{\beta_l}, \quad (9)$$

where N_1, \dots, N_l – used numbers of similarity, and β_1, \dots, β_l – their degrees in structure of criterion.

The logarithm of criterion (3) transforms the multiplicative-power function into a linear combination $\ln K_0 + \beta_1 \ln N_1 + \dots + \beta_l \ln N_l = \ln K$. Logarithms of numbers of similarity are interpreted as natural parameters, standard image converted in two-level factors, and the linear combination constructed above for natural parameters with a priori uncertain factors is transformed to linear regress for these factors. Its factors are a subject to definition according to the results of calibration experiments. For the decision of this problem two-level factorial planning is traditionally used. The factors defined on the basis of natural parameters, are not fictitious, but can be added by fictitious factors as part of the linear regression model.

The matrix of two-level planning is a matrix of factors in linear system of the equations for factors of regress, including fictitious. The right part of this system are «measurements» – values of the logarithm of criterion the similarity corresponding borders of natural parameters in structure of a linear combination. The analysis of the importance of counted factors of regress and adequacy of the linear regression model as a whole are carried out by standard statistical methods.

As an example of the offered approach the analysis of criterion of frictional deterioration in a greasing liquid is considered. The given criterion looks like [8]:

$$Vw = \frac{pv\Delta V}{\mu Nt} \quad (10)$$

where p – density of a wearing out material, v – speed of sliding, ΔV – size of volumetric deterioration, μ – factor of friction, N – normal loading on a rubbing surface, t – duration of deterioration. Physical sense of this criterion – the ratio of the wear material pulse to the friction force pulse.

Let's present the given criterion in the form of linear function of numbers of similarity:

$$\ln(Vw) = \beta_0 + \beta_1(\ln St) + \beta_2(\ln S_0) + \beta_3(\ln Re) + \beta_4(\ln De) \quad (11)$$

where – factor of proportionality; $St = \frac{vt}{D}$ Strouhal number (homochronity); D – maximum diameter of the wear band of the ball; $S_0 = \frac{\eta v}{\mu \frac{N}{D^2} \Delta L}$ – number of Sommerfeld; η – dynamic viscosity of the liquid; ΔL – depth of wear of the ball; $Re = \frac{p_f v \Delta L}{\eta}$ – Reynolds number, ρ_f – liquid density; $De = Re \sqrt{\frac{d}{D}}$ – Dean's number; d – minimum diameter of the wear band of the ball.

For recalculation of natural parameters in corresponding factors we use ratio:

$$\begin{aligned}
x_1 &= 2 \frac{\ln St - (\ln St)_{\min}}{(\ln St)_{\max} - (\ln St)_{\min}} - 1 \\
x_2 &= 2 \frac{\ln So - (\ln So)_{\min}}{(\ln So)_{\max} - (\ln So)_{\min}} - 1 \\
x_3 &= 2 \frac{\ln Re - (\ln Re)_{\min}}{(\ln Re)_{\max} - (\ln Re)_{\min}} - 1 \\
x_4 &= 2 \frac{\ln De - (\ln De)_{\min}}{(\ln De)_{\max} - (\ln De)_{\min}} - 1
\end{aligned} \tag{12}$$

The borders of the natural parameters calculated by results of mentioned above experiments, are resulted in [6]. Thus to boundary values of natural parameters there correspond values of factors -1 and 1. Return recalculation of factors in natural parameters is carried out according to ratio:

$$\begin{aligned}
\ln St &= (\ln St)_{\min} + ((\ln St)_{\max} - (\ln St)_{\min}) \frac{x_1 + 1}{2} \\
\ln So &= (\ln So)_{\min} + ((\ln So)_{\max} - (\ln So)_{\min}) \frac{x_2 + 1}{2} \\
\ln Re &= (\ln Re)_{\min} + ((\ln Re)_{\max} - (\ln Re)_{\min}) \frac{x_3 + 1}{2} \\
\ln De &= (\ln De)_{\min} + ((\ln De)_{\max} - (\ln De)_{\min}) \frac{x_4 + 1}{2}
\end{aligned}$$

Substituting them in (12), we receive for regress of criterion concerning not fictitious factors $\ln(Vw) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4$, where:

$$\begin{aligned}
b_0 &= \beta_0 + \frac{(\ln St)_{\max} + (\ln St)_{\min}}{2} + \frac{(\ln So)_{\max} + (\ln So)_{\min}}{2} + \frac{(\ln Re)_{\max} + (\ln Re)_{\min}}{2} \\
&\quad + \frac{(\ln De)_{\max} + (\ln De)_{\min}}{2} \\
b_1 &= \beta_1 \frac{(\ln St)_{\max} - (\ln St)_{\min}}{2} \\
b_2 &= \beta_2 \frac{(\ln So)_{\max} - (\ln So)_{\min}}{2} \\
b_3 &= \beta_3 \frac{(\ln Re)_{\max} - (\ln Re)_{\min}}{2} \\
b_4 &= \beta_4 \frac{(\ln De)_{\max} - (\ln De)_{\min}}{2}
\end{aligned} \tag{13}$$

Construction of the regular plan for the analysis of influence of the entered factors is carried out on the basis of a fractional retort of full factorial experiment. Thus the number m factors in structure of this experiment should satisfy the condition $2^{m-1} < l + 1 \leq N = 2^m$, where l – the number of not fictitious factors certain above, and N – the common number of factors equal nearest dominating to the degree of the two. Considering this circumstance, it is necessary to enter into a ratio for regress additional fictitious factors x_5, x_6, x_7 . As a result it gets a kind:

$$\ln(Vw) = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 + b_5 x_5 + b_6 x_6 + b_7 x_7 \tag{14}$$

The matrix of planning thus corresponds to a fractional retort 2^{7-4} full factorial plans which is resulted on Fig. 7. Characteristic property of this matrix is centeredness its columns, except for the first. From here follows, that the free member b_0 in expression for regress is equal to an average of the right parts of the equations of linear system. Proceeding from it, we shall accept it in the further equal to result of averaging $\ln(Vw)$ on all lead experiments. The truncated matrix of the plan concerning factors of regress for the remained factors is allocated on Fig. 7.

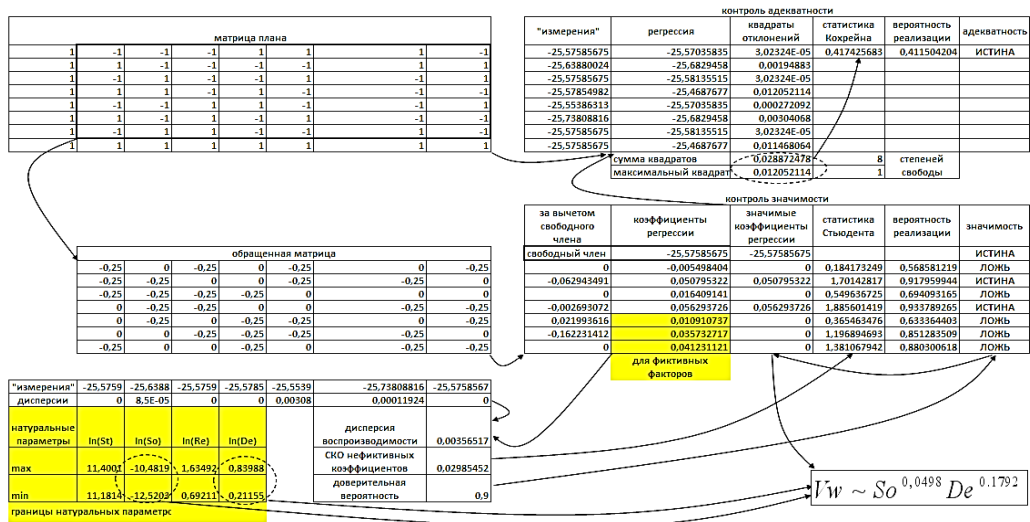


Fig. 7. Statistical analysis of criterion of similarity

Thus, for calculation of factors of regress b_1, \dots, b_7 the truncated matrix of the plan is used, and the right part of linear system represents average "measurements" minus a free member. The vector of factors of regress represents the decision of this linear system calculated by multiplication of the turned matrix of the plan to the vector of the right parts certain above. Thus it is necessary to consider, that factors x_5, x_6, x_7 are fictitious and the sum of squares of factors corresponding them is summarized with the average dispersion calculated on repeating «measurements», giving a dispersion of reproducibility σ^2 . Carrying it to quantity of not fictitious factors $l = 4$, we receive a dispersion of an estimation of corresponding factors of regress and its root-mean-square deviation $s_b = \frac{\sigma}{\sqrt{l}}$.

For an estimation of the importance of factors of regress at not fictitious factors $b_1...b_4$ are used the statistics $\frac{|b_1|}{s_b} \dots \frac{|b_4|}{s_b}$, distributed over the Student with the number of degrees of freedom l .

The probabilities of implementing these statistics are $T_l\left(\frac{|b_1|}{s_b}\right) \dots T_l\left(\frac{|b_4|}{s_b}\right)$, where T_{il} is the Student distribution with l degrees of freedom. The decisive rule for confirming the significance of the i -th coefficient is as follows $T_l\left(\frac{|b_i|}{s_b}\right) > P = 0,9$, where P is the confidence probability of making a decision on significance. Factors of regress for which these conditions are not carried out, considered insignificant and are equated to zero. The results of calculations displayed on Fig. 4, show, that significant, i.e. nonzero, factors b_2 and b_4 , corresponding to the similarity numbers of Sommerfeld and Dean. Other factors are insignificant and consequently dependence on the Strouhal and Reynolds criteria is ignored.

For the control of adequacy of the received regress the mismatch between average «measurements» and values of regress in each line of the plan. For calculation of these values the line of the plan is multiplied by a vector of significant factors of the regress, displayed on Fig. 7.

The decision on the adequacy of the linear-regression model is taken by the Cochran criterion. The decisive statistic is Cochran C – the ratio of the maximum square of deviation to the sum of the squares of deviations. The probability of realization of this statistics makes $F^N(C, l, N)$, where F – Fisher's distribution. The decisive rule in this case is as follows: a decision is made on

the adequacy if the probability of realization is lower than the confidence value 0.9, and on the lack of adequacy otherwise. Results of calculation of statistics C and probabilities of its realization also are resulted on Fig. 7. As the received probability below confidential value, is accepted a hypothesis about adequacy constructed linear-regression model.

For formation of a final kind of criterion of similarity it is necessary to transform preliminary significant factors b_2 and b_4 to factors β_2 and β_4 , entering in (9). It is carried out by the reference of corresponding expressions in (13):

$$\begin{aligned}
 & V_W \sim S_0^{\beta_2} D e^{\beta_4} \\
 \beta_2 &= \frac{2b_2}{(\ln S_0)_{max} - (\ln S_0)_{min}} = 0.0498 \\
 \beta_4 &= \frac{2b_4}{(\ln De)_{max} - (\ln De)_{min}} = 0.1792
 \end{aligned}
 \tag{15}$$

The lead consideration shows, that factorial planning can be used for a substantiation of invariancy of a wide class of entered criteria of similarity, and also for the proof of adequacy of their representation depending on standard numbers of similarity. Advantage of the offered approach is that it can be used practically without restrictions of a generality and is based on well developed methods of the factorial and statistical analysis.

The usual technique estimation precisions corresponds to the standard approach to processing complex experiment. It differs possible distinction of the sizes measured in consecutive series. The basic in these conditions is the research problem of influence of considered factors on the measured size. Research of reproducibility is auxiliary in these conditions and is intended for branch of influence of an error of measurements from influence of considered factors. Other situation arises at the analysis precisions interlaboratory researches. Research of influence of factors thus becomes secondary, and the main thing becomes research of interlaboratory reproducibility [9]. A problem of interlaboratory comparative tests is the analysis of competence of laboratories, and also them rejection on the basis of incompetence. Below the statistical analysis algorithm of competence of the laboratories participating in specified tests is developed. The algorithm is based on preliminary standardization of spent measurements by their transformation to corresponding Student. Conditions of incompetence of laboratory are formulated in case of is abnormal the big error of measurements and in case of the several moderate errors.

In a basis of the given analysis lays the approach providing comparative measurements of polytypic standard samples of substances and materials, and also their rational processing with use of methods of the dispersive analysis. The purpose of the last is decomposition of a total dispersion on two sizes: a dispersion caused by technics of experiment, and a dispersion caused by action of the researched factor. In case of multifactorial experiment the dispersions corresponding various factors and their interactions are defined, and also the statistical importance of these sizes in view of an error of reproducibility is estimated.

Competence of laboratory is defined by reproducibility of results of comparative tests which is provided in turn with statistical uniformity of sample of the lead measurements. Traditional methods of the analysis of this uniformity are the comparative analysis of serial dispersions of researched sample or serial Student relations.

Quite often, the following parametric criteria are used to solve this problem: Cochran, based on the comparison of the maximum of serial variances with their sum, and Grubbs (or Hawkins), based on the analysis of anomalies of extreme values in serial samples of student relations for measurements.

Preliminary stage of processing is standardization of made measurements. For the given standard sample it is carried out by transition from measurements of the characteristic of this sample to Student fractions, constructed according to these measurements. The last are formed according to ratio:

$$\begin{aligned}\xi_i &= \frac{x_i - m}{s} \\ m &= \frac{\sum_{k=1}^L x_k}{L} \\ s &= \sqrt{\frac{\sum_{k=1}^L (x_k - m)^2}{L - 1}}\end{aligned}\tag{16}$$

where x_i – interlaboratory measurements of the standard sample; m , s – sizes of an average and standard deviation measurements; L – quantity of laboratories-participants.

In these conditions for the decision of a problem of processing of the standardized measurements it is natural to use criterion Hawkins with selective averages and standard deviation, counted on serial samples. Crucial statistics in this case are $Y = \max_{i=1\dots n} \xi_i$ and $y = \min_{i=1\dots n} \xi_i$, and conditions of anomaly of extreme measurements in a series become:

$$\begin{aligned}(2T_{L-1}(H) - 1)^L &> P \\ H &= \max(|Y|, |y|)\end{aligned}\tag{17}$$

This condition can be interpreted as the first condition of incompetence: the laboratory in which corresponding measurement is lead, is considered incompetent.

One more condition of incompetence is connected with calculation of unsatisfactory measurements from all lead in given laboratories. Satisfactory the measurements close to an average m , and unsatisfactory – strongly deviating from it are considered. As these measurements concern to various samples, they also are preliminary standardized by transformation to Student relations. At transition to them satisfactory the values of fractions located in a vicinity θ , and unsatisfactory-laying outside of this vicinity are considered.

We shall be set by probability p insufficiency of measurement. Considering, that Student fractions have T – distribution c $L-1$ a degree of freedom, we find for threshold H of decision-making on the specified insufficiency:

$$2T_{L-1}(H) - 1 = 1 - p\tag{18}$$

Satisfactory it is considered value Student fraction smaller H , and unsatisfactory – greater H . Distribution of probabilities of quantity of unsatisfactory measurements n from N lead in given laboratories is binomial with parameters p and N . Being set by confidential probability P_{II} rejectionson number of unsatisfactory measurements, we receive solving rule for specified rejections:

$$\sum_{k=0}^n C_N^k p^k (1 - p)^{N-k} > P_{II}\tag{19}$$

The condition (19) is interpreted as the second condition of incompetence. The laboratory in which from N the lead measurements n are unsatisfactory, is considered incompetent at performance of this condition and competent otherwise.

Let's consider sense of the deduced conditions of incompetence. For this purpose we shall compare conditions (17) and (18) in the assumption, that $P_I = 1-p$. Assuming also, that in (17) equality takes place, we find from this comparison, that $\xi_i > H$. Thus, the first condition of incompetence is carried out in case of an abnormal error even one of the lead measurements. The second condition means, that the laboratory is considered incompetent in case of the several moderate errors in the lead measurements.

factor at t^3 , i.e. a set of plans of the kind $x_1x_2x_3 + x_1x_2x_4 + x_1x_3x_4 + x_2x_3x_4 + x_1x_2x_5 + x_1x_3x_5 + x_1x_4x_5 + x_2x_3x_5$ corresponding replicas 2^{5-3} . For each of the plans chosen by such principle a series of three times repeating experiments is spent. Factors of regress for each of these plans unequivocally according to (20), where y_1, \dots, y_N are interpreted as a median of repeating measurements. In turn, on the found factors loadings for the factors covered by a given series. As each factor is covered $P-1$ by a series of experiments, for it the same number of relative loadings L_{i1}, \dots, L_{iP-1} , counted in view of change of a range of measurements from a series to a series turns out. The maximal loadings from received $L_{imax} = \max(L_{i1}, \dots, L_{iP})$, $i = 1 \dots k$, are considered as characteristics of the importance of factors... k on which ranging is carried out.

Relative maximal loadings $L_{1max}, \dots, L_{kmax}$ can be used also for rational elimination of insignificant factors. We shall assume without restriction of a generality, that number $L_{1max}, \dots, L_{kmax}$ represents result of ranging of loadings on decrease, and we form an equivalent number $\frac{L_{1max}}{L} \dots \frac{L_{1max}}{L}$, where $L = L_{1max}$. We shall expose a threshold on the importance, for example,

$h=0,05$, and all the factors, satisfying to a condition $\frac{L_{1max}}{L} < h$, we shall consider insignificant. This intuitive method of elimination is more simple in comparison with traditional statistical elimination. Its basic advantage, alongside with simplicity of realization, use is at the analysis only not fictitious factors. The basic lack, alongside with intuitivism, absence of a regular technique of a choice of a threshold of the importance which should be chosen intuitively is.

The lead consideration shows, that the statistical component of virtual toolkit covers the broad audience of problems of processing of results of qualification. Procedures used at it are led to an analytical kind and provide sufficient severity of the analysis. At the same time, they are based on standard approaches and are rather simple in circulation.

The Important component virtualization toolkit is the statistical component for the decision of following problems: identifications of statistical models by results of qualifying tests, parametrization criteria models for carrying out of forecasting within the limits of virtual qualification, synthesis regression models, the statistical analysis of their adequacy and consecutive optimization during factorial planning. Thus linearity regression models allows to carry out the control of the importance of factors and completeness of their system over use of traditional statistical methods, and also optimization of parameters by means of methods of linear programming.

Identification of distributions of measurements of various parameters of mineral oil specifies their abnormality. As alternative approximation bimodal the multipleparameter distribution representing the sum differing normal was offered to use distributions of density. Unimodular distributions of other characteristics are approximated are approximated by the Weibull. Parametrization of the constructed distributions was carried out by a method of the moments that provides sufficiency of used approximations.

For an estimation of reproducibility of results of qualifying experiments have been constructed analytical distribution of Cochran's statistics on the basis of F -Fisher's distribution and Grubbs statistic on the basis of T -Student's distribution. It is shown, that preliminary normalization allows to avoid the unreasonable conclusion about heterogeneities of the studied sample. That the probability of an error of the first sort decreases at decision-making on reproducibility of results of qualifying experiment.

Parametrization criteria model by results of qualifying tests is offered for spending models by results of forecasting effect in extreme conditions that allows to generate the guaranteed qualifying specifications. The preliminary choice of extreme test specifications at qualifying normalization is offered for carrying out a method consecutive approximations with use of factorial planning.

The approach for a substantiation of invariancy of a wide class of criteria of similarity is offered. The Approach is based on interpretation of the logarithm of criterion as linear form from the logarithms of numbers of similarity treated as independent factors. The estimation of factors of the specified form – loadings of factors – is carried out with use of the factorial analysis and includes also the proof of adequacy of the specified representation. Advantage of the offered approach is the opportunity of its application without restrictions of a generality.

The intuitive method of ranging of factors in the linear-regression model with use of the sated plans and the subsequent elimination of the least loaded factors is developed. Its basic advantage, alongside with simplicity of realization, use is at the analysis only not fictitious factors, and the basic lack – absence of a regular technique of an establishment of thresholds of the importance which should be chosen intuitively.

РЕФЕРАТ

*Олександр Моїсєєв, Микола Грішин, Олександр Орешенков
ФАН «25 ДержНДІ хімматології», ikmvk@yandex.ru, 25gosniihim@mil.ru*

СТАТИСТИЧНІ МЕТОДИ В ЗАДАЧАХ ВІРТУАЛІЗАЦІЇ

Показана важливість віртуалізаційного інструментарію вирішення завдань ідентифікації статистичних моделей за результатами кваліфікаційних випробувань.

В ході ідентифікації встановлена ненормальність розподілів похибки вимірювань характеристик нафтопродуктів. Для побудови алгоритмів обробки вимірювань синтезовано ряд альтернативних апроксимацій. Для оцінки відтворюваності результатів віртуальної кваліфікації побудовані розподіли статистики Кохрейна і статистики Граббса. Параметризацію критеріальної моделі запропоновано проводити за результатами прогнозування ефекту в екстремальних умовах. Вибір умов запропоновано здійснювати методом послідовних наближень в рамках факторного аналізу.

Ключові слова: інструментарій, ідентифікація, параметризація, критеріальна модель, прогнозування, регресійна модель, факторне планування, лінійне програмування, бімодальний розподіл, апроксимація, статистика Кохрейна, статистика Граббса, кваліфікаційні нормативи, критерій подібності, число подібності, фіктивний фактор, значущість, ранжування.

РЕФЕРАТ

*Александр Моисеев, Николай Гришин, Александр Орешенков
«НИИ химмотологии», ikmvk@yandex.ru, 25gosniihim@mil.ru*

СТАТИСТИЧЕСКИЕ МЕТОДЫ В ЗАДАЧАХ ВИРТУАЛИЗАЦИИ

Показана важность виртуализационного инструментария решения задач идентификации статистических моделей по результатам квалификационных испытаний.

В ходе идентификации установлена ненормальность распределений погрешности измерений характеристик нефтепродуктов. Для построения алгоритмов обработки измерений синтезирован ряд альтернативных аппроксимаций. Для оценки воспроизводимости результатов виртуальной квалификации построены распределения статистики Кохрейна и статистики Граббса. Параметризацию критерияльной модели предложено проводить по результатам прогнозирования эффекта в экстремальных условиях. Выбор условий предложено осуществлять методом последовательных приближений в рамках факторного анализа.

Ключевые слова: инструментарий, идентификация, параметризация, критерияльная модель, прогнозирование, регрессионная модель, факторное планирование, линейное программирование, бимодальное распределение, аппроксимация, статистика Кохрейна, статистика Граббса, квалификационные нормативы, критерий подобия, число подобия, фиктивный фактор, значимость, ранжирование.

ABSTRACT

*Alexandr Moiseev, Nikolay Grishin, Alexandr Oreshencov
The Scientific-Research Institute of Chemmotology, ikmvk@yandex.ru, 25gosniihim@mil.ru*

STATISTICAL METHODS IN VIRTUALIZATION

The importance of virtualization tools for solving the problems of identification of statistical models based by the results of qualification tests is shown.

During identification abnormality of distributions of measurements error of mineral oil characteristics was established. For construction of algorithms of processing of measurements a number of alternative approximations is synthesized. To assess the reproducibility of the virtual qualification results, the distribution of Cochran statistics and Grubbs statistics is constructed. Parameterization of the criterion model is proposed based on the results of predicting the effect in extreme conditions. The choice of conditions is proposed to be carried out by the method of successive approximations in the framework of factor analysis.

Key words: toolkit, identification, parametrization, criteria model, forecasting, regression model, factorial planning, linear programming, bimodal distribution, approximation, Cochran's statistics, Grubbs statistics, qualifying specifications, criterion of similarity, number of similarity, fictitious factor, importance, ranging.

REFERENCES

1. *Moiseev A.* Virtualization's tool and its application / A. Moiseev // Industrial Automatic Control Systems and Controllers.– 2016. – № 11. – P. 16.
2. *Grishin N.N.* Encyclopedia chemmotology / N.N. Grishin, V.V.Sereda. – M.: Publishing house "Pero", 2016. – P. 960.
3. *Moiseev A.* Modified criteria of statistical uniformity / A.Moiseev // Industrial Automatic Control Systems and Controllers. – 2011. – № 15. – P. 20.
4. *Moiseev A.* Error distributions approximations for measurements of mineral oils characteristics / A.Moiseev // Industrial Automatic Control Systems and Controllers. – 2015. – № 9. – P. 30.
5. *Moiseev A.* Criteria modeling at qualification normative forming / A.Moiseev // Proceedings of XI international scientific and technical conference «Tribology – to machine-building». – M., 2016. – P. 161.
6. *Moiseev A.* Factor's planning at similarity criteria analysis / A.Moiseev // Vestnik NIYU "MIFI". – 2016. – Vol. 5. – P. 303.
7. *Grishin N.N.* Condition of theoretical base chemmotology. Monography is prepared on the basis of materials, given by authors participants VI International scientific and technical conference. Under the common edition of professor Sergey Bojchenko / Grishin N.N., Oreshencov A.V. – Kiev: "Publishing house "Center of the educational literature", 2016. – P. 217.
8. *Moiseev A.* Criteria model of qualifying tests for deterioration / A.Moiseev // Engineering physics. – 2015. – № 12. – P. 30.
9. *Moiseev A.* Statistical analysis of competence at inter-laboratory tests / A. Moiseev // Modelling, optimization and information technologies [Electronic reference]. – 2017. – No 1 (16). – Access mode: <http://moit.vivt.ru/>.
10. *Moiseev A.* Intuitive method of factors ranking on significance / A. Moiseev // Modelling, optimization and information technologies [Electronic reference]. – 2017. – No 1 (16). – Access mode: <http://moit.vivt.ru/>.