

Quantitative Evaluation of Accuracy of Digital Microscope System for Automated Petrographic Analysis

Olga E. Baklanova¹, Vyacheslav A. Kornev² and Olga Ya. Shvets²

¹ Department of Mathematical and Computer Modelling, D.Serikbayev East-Kazakhstan State Technical University, 19, Serikbaeva Street, Ust-Kamenogorsk, The Republic Kazakhstan

² Department of Instrument Engineering and Technology Automation, D.Serikbayev East-Kazakhstan State Technical University, 19, Serikbaeva Street, Ust-Kamenogorsk, The Republic Kazakhstan

Keywords: Petrographic Analysis, Digital Microscopy, Image Recognition, Evaluation of Reliability, Machine Vision, a Rock Sample.

Abstract: The purpose of this paper is quantitative estimation of the reliability of digital microscopy systems for automated petrographic analysis. The proposed method consists in measuring the intensity of the three spectral components of the reflected light (red, green, blue) in each pixel of the investigated surface (frame) and using of methods of pattern recognition of images in geological and mineralogical researches of a rock sample. The proposed technology is offered for preparation of a rock sample in system to a quality and quantitative standard of the maintenance minerals with use of technical sight and software for increase of reliability of results. Statement of the problem is the following: it is necessary to develop a model for assessing the reliability of the monitoring process, taking into account the statistical nature of standards for petrographic analysis and using system process control agents. Mathematical models of an assessment and forecasting of reliability measuring system and control of the maintenance useful minerals in studied a rock sample are developed.

1 INTRODUCTION

Petrography is the science that studies the material composition of the rocks. Unlike minerals, rocks are aggregates composed of different minerals (Blatt and Robert, 1995).

Minerals are homogeneous in composition and structure of the rocks and ores. They are natural chemical compounds resulting from various geological processes. Historically minerals initially determined by color and shape. However, the reliability of the color diagnostics minerals traditional visual methods is extremely low and the color differences even can be misleading. In this case, the diagnosis only color minerals do not solve the problem of quantifying the content of useful component in the rock, which is extremely important in practice (Farndon, 2006).

Using math methods and achievements of modern computer technology allow to improve significantly the accuracy of not only the process of monitoring and diagnosing, but also to quantify the mineral content in the sample of rock. But we have

another important task - assessment of reliability process of structural and mineralogical studies (Baklanova and Uzdenbaev, 2013).

Reliability of research depends on several factors:

- Natural variations and mineral color;
- Structures of errors in technology of mineral color parameters measurement;
- Errors in measurement and control standards and control techniques for measuring color diagnostic indicators;
- Uncertainty normative values in quantifying evaluation system for color diagnostic indicators.

Accuracy of control, diagnosis and decision-making system in the petrographic analysis can be enhanced by the presence of a database with information on the location test sample rocks, geological landscapes and associated pore rows. For example, gold often meets with milky quartz. Minerals that occur together are called associated with (Farndon, 2006).

2 MATERIALS AND METHODS

The proposed method consists in measuring the intensity of the reflected spectral components of three colors (red, green, blue) at each pixel of the surface under study (frame) of a rock sample.

Rock sample is subjected to a pretreatment and preparation of the surface to be scanned, for example in the form of ore. There is another way in the literature to expose the sample pulverized and filling in some container (Clarke and Eberhardt, 2002).

Then substance surface in the container milled powder mass is formed by manual compression. Technological prepared surface is scanned by digital optical microscope and conserved in three pixel arrays for each color (Pantelev, Egorova and Klykova, 2005.).

Array elements are equivalents color levels in each pixel. Meets or does not meet the three color pixel for mineral composition will depend on the measured intensities of red, green and blue pixel admissible set for this mineral. The pixels in the frame, in which all three colors after the count rated intensities of color, were in the tolerances for each color, we denote as - to R_i, G_i, B_i .

Lower and upper tolerable color values denoted as - $S_{lr}, S_{hr}, S_{lg}, S_{hg}, S_{lb}, S_{hb}$, where:

- S_{lr}, S_{hr} - the lower and upper values of tolerance red intensity;

- S_{lg}, S_{hg} - the lower and upper values of tolerable intensity of green color;

- S_{lb}, S_{hb} - the lower and upper values of tolerance blue intensity.

Number of pixel in their color tolerable boundaries in j - th frame ($j = 1, L$) is denoted as N_{rj}, N_{gj}, N_{bj} . In this collection necessary to select and register only those pixel, that both were in their color standards and denote their number in the j -th frame as K_j . Then, at a certain scanning area and discontinuity (resolution) becomes aware of the microscope and the total number of pixels in this square - M . This area throughout the study should be constant. Ratio $\frac{K_j}{M} = S_j$ will be intermediate relative diagnostic assessment of mineral content in the j - m frame, the test sample on the basis of the color. Quantity required observations (number of investigated frames) is determined according to known mathematical statistics method depending on the variability of the measured parameter, the resolution of the microscope, and some other factors which can be identified in the study.

Each next observation begins with the mixing of the powder mass analyzed rocks. Upon completion

of the entire cycle of L observations, the results are statistically treated sampling, and calculated the mean value of S_{cp} some numerical estimates of the average mineral content in the sample breed closely related to the actual content of the mineral.

It is necessary to make measuring the content of useful mineral rocks in the volume measurement of mineral known metrological attorneys' means, such as spectral for going from the relative valuation to some absolute. Evaluation of the content of the mineral spectral methods denoted as Y . Based on spectral estimation Y can recalculate color index in conventional bulk or weight.

One of the tasks in the research is validation the values of tolerance, which in the final commercial version will be taken as normative. Mineral ratios for each color will be determined experimentally, using statistical methods.

In this article technique of formation evaluation standards and reliability of the control is considered an example of gold, which is justified for two reasons. The first reason is the reserves of gold and platinum are the most important indicators of the state's economy. The second reason - gold is different low chemical reactivity, on the air and it is not changed, do not act on it, and most acids, so the color of gold indicated a sufficiently high stability and low variability. Gold occurs mainly in the native state in the form of fine grains (Chris, 2002).

Color is lighting from golden yellow to light yellow. Gold grains, meeting in nature, have the form of sheets, glitters, hair like discharges, etc (Shaffer, Herbert and Raymond, 2001).

As etalon of gold color can be used ore with one smooth ground surface or color photo quality. Since gold color varies in a certain range, the experimentally determined ratios will have uncertainty, i.e. standards are random variables. Instrumental measurement means also have the random error, which implies that the results of monitoring the gold content in rock sample will be accompanied by errors that errors in the literature called the first and second kind of risk or the producer and consumer. We have the problem - development formal models estimate the value specified risk in the statistical uncertainty of all agents' process control and decision making. Control process contains the following system agents:

- agent - normative values;
- agent - measured parameters;
- agent - measurement error;
- agent decision-making system.

Agent-based modeling to assess and predict risks at the stage of decision-making in the system diagnostics and quality control process in statistical under uncertainty relies mostly on probability and simulation technique (Borshev, 2008). Agent-based modeling makes it possible to investigate the behavior of the autonomous system objects, agents, and how this behavior determines the behavior of the whole system. In contrast to the system dynamics researcher evaluates the behavior of agents is not at the individual level, and the behavior of the system as a whole - is the result of the activities of all agents, which is called - modeling "bottom up" (Kuleshov and Kornev, 2011).

Control - is a sequence of procedures: measurement, comparison measured values with the regulations and a decision on the principle - "controlled object pass", "fail-controlled object." Final decision "pass-fail" is generally accepted by a man.

Mandatory control procedure is measurement. Measurement regarded as an independent agent has no independent process control characteristics. During the study the multi-agent system, these are subject to change, in order to find the optimal values. This approach also applies to other agents.

Due to the fact that the process is accompanied by measurement errors, we have the control errors. We consider only random errors in this paper.

These events occur in the following cases from the probabilistic point of view (Wentzel, 2002):

- monitored parameter is normal (pass), and the measured value as a result of random error was in the zone - does not pass, it means fail defect;
- monitored parameter is in the "fail", and the measured value as a result of random error was in the area fit that is undetected fail.

Quantitatively, these errors are estimated corresponding probabilities, in this case, P_{fd} - probability of a false defect, and P_{ud} - probability of undetected defect. These probabilities are also worth the risk of the customer's products work and work accordingly.

A lot of works are devoted for problem of risk assessment and decision-making control; however, they assume deterministic normative values. The need to consider regulations as random values indicated in some studies, but mainly in the staging level and qualitative analysis, which deals with general approaches to evaluating these risks (Svinolupov, Kornev and Kuleshov, 2012).

There is often a need for practical purposes with the required reliability quantify and predict the quality of measurement and control procedures in

the application domain, in any combination of the statistical characteristics of these agents in the control system. This may be two problems.

The first problem. Measurement tools are available with known metrological characteristics, and there are regulatory requirements for control precision. It is necessary to evaluate the quality of test and measurement operations.

The second problem. There are regulatory requirements for quality control. It is required to determine the accuracy characteristics of measurement tools to ensure the quality requirements specified control.

It is often used concept to assess the quality of the integrated measurement information - reliability is the main characteristic of the quality control. Reliability of the control - is the degree of confidence that the measured values reflect the true state of the object (Svinolupov, Kornev and Kuleshov, 2012).

Reliability analytically is calculated by the formula:

$$D = 1 - \sum ERR, \quad (1)$$

where ERR - the probability errors.

3 DISCUSSION AND RESULTS

It is often consider on practice the quality control is uniquely determined by the error of measurement and it ignores other system agents' process control. It is assumed that quantifying reliability of inspection results is necessary for the reason that the error control largely determine the technical and economic indicators of the entire quality management system for production processes. In this regard, it was attempting to develop a formal model that takes into account the effect of all the factors, and especially the fact of uncertainty normative values. The model would be having practical importance, not only in this subject area, but also in many other system tasks for quality control processes, diagnosis and decision-making.

Each object of study, economic, social or technical, etc., has its own distinct properties. This fact is confirmed by systematically the most diverse examples in all spheres of human activity. There is a fact of explicitly normative values uncertainty in this subject area. Therefore, the development and evaluating the quality control in statistical uncertainty standards become sufficiently actually.

Hence, rightly hypothesize that normative values-agents system quality management process, random quantities obeying in each case certain statistical distribution laws.

Statement of the problem is the following: it is necessary to develop a model evaluation credibility of the process control based on the statistical nature of standards for different compositions of the laws of distribution of all system agents process control: the controlled parameter, regulations, distributions of errors of measurement means.

Consider as the first case an example of normal distribution of controlled parameter, measurement errors and normative. Assume that the function of grocery distribution normative values of control parameter has the form:

$$\theta_1(S_l) = \frac{1}{\sqrt{2\pi} \cdot \sigma_l} e^{-\frac{(S_l - \bar{S}_l)^2}{2\sigma_l^2}} \quad (2)$$

$$\theta_2(S_h) = \frac{1}{\sqrt{2\pi} \cdot \sigma_h} e^{-\frac{(S_h - \bar{S}_h)^2}{2\sigma_h^2}} \quad (3)$$

where σ_l, σ_h - the standard deviations of the upper and lower standard values;

\bar{S}_l, \bar{S}_h - average values of the lower and upper limits.

We have the following four possible events when measuring means having random errors:

- true value of S_i is within acceptable standards and within normative limits $S_l < S_i < S_h$, and measured value S_i^* is within admissible limits $S_l < S_i^* < S_h$;

- true value of S_i is outside of admissible limits ($S_i < S_l$ or $S_i > S_h$) and measured value S_i^* is outside of admissible limits ($S_i^* < S_l$ or $S_i^* > S_h$);

- true value S_i is within admissible limits ($S_l < S_i < S_h$) and the measured value S_i^* is exceeds the upper limit or outcome beyond the lower limit ($S_i^* < S_l$ or $S_i^* > S_h$). In this course there is a case when the true value of the monitored parameter is in the zone of acceptance - "pass" and "device" erroneously fixes it beyond the norm - "fail." This case was called "false defect", and the probability of its occurrence - the probability of a false defect P_{fd} ;

- true value S_i is outside of ($S_i < S_l$ or $S_i > S_h$) and the measured value S_i^* within the tolerance ($S_l < S_i^* < S_h$). This event is called the undetected defect and the probability of its occurrence - probability of undetected defect P_{ud} .

The above group of events in probability theory called complete group of mutually exclusive events.

The first two cases of this group are unmistakable outcomes and they do not represent some interest.

The last two cases are control errors are possible due to the fact that each measurement to a greater or lesser extent, but always accompanied by the random error of measure.

And this is the question, how to estimate the probabilities of these errors, in dimensional case P_{fd} and P_{ud} , and how to manage their level.

It was developed probabilistic model to assess the quality of decisions and determine the probable types of defect. Figure 1 illustrates the analysed case.

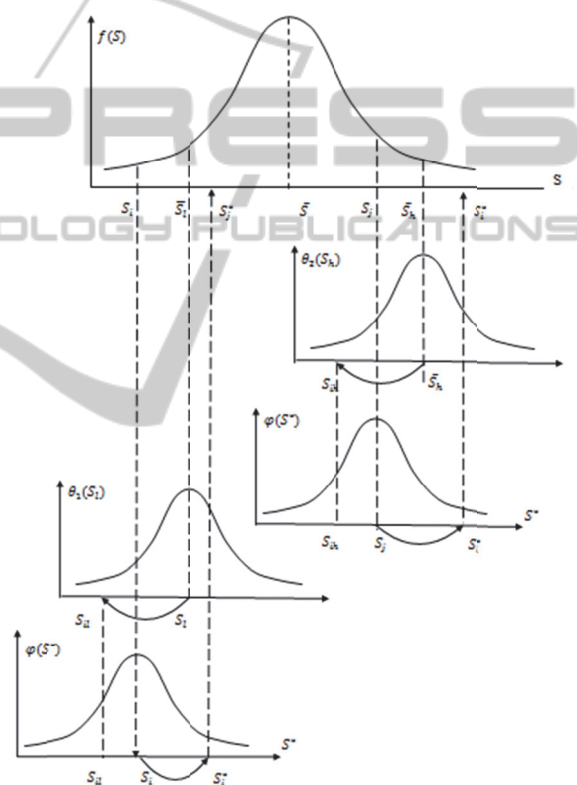


Figure 1: Graphical model of control for random errors in admissible limits.

Consider the derivation of the mathematical expression of risk assessment P_{fd} originally an example of the standard for the lower admissible limit parameter S .

Diagram $\theta_1(S_l)$ is the density distribution of the lower limit. The average value of \bar{S}_l is a center of the region of uncertainty (scattering) of the lower limit.

Divide the interval of variation of the norm of $\bar{S}_l - 3\sigma_l$ to $\bar{S}_l + 3\sigma_l$ by m plots. So from the N -total

number of monitored objects, N_i - probable number of objects have a standard value S_{ni} will be

$$N_i = P_i \cdot N = N \int_A^B \theta(S) dS, \quad (4)$$

where:

$$\begin{aligned} P_i &- \text{interval } i\text{-th probability;} \\ A &= \bar{S}_l - 3\sigma_l + (i + 1)\Delta S_l; \\ B &= \bar{S}_l - 3\sigma_l + i\Delta S_l; \\ \Delta S_l &= \frac{3\sigma_l}{m}. \end{aligned}$$

P_{jfd} probability expressions for the i -th interval value standard S_l and j -th value of the parameter S has the following form:

$$P_{ijfd} = \frac{1}{\sqrt{2\pi}} \int_{M_i}^{D_i} e^{-\frac{t^2}{2}} dt \cdot \frac{1}{\sqrt{2\pi}} \int_C^3 e^{-\frac{z^2}{2}} dz, \quad (5)$$

where:

$$\begin{aligned} M_i &= \frac{S_{li} - \bar{S}}{\sigma_s} + \frac{3(j-k)}{k}, \\ D_i &= \frac{S_{li} - \bar{S}}{\sigma_s} + \frac{3(j+1-k)}{k}, \\ \Delta S &= \frac{3\sigma_s}{k}, \\ C &= \frac{3k-j}{k}. \end{aligned}$$

The total probability for P_{fd} limit S_{li} is:

$$P_{ifd} = \sum_{j=0}^k P_{jfd} = \sum_{j=0}^k \frac{1}{\sqrt{2\pi}} \int_{M_i}^{D_i} e^{-\frac{t^2}{2}} dt \cdot \frac{1}{\sqrt{2\pi}} \int_{\frac{3k-j}{k}}^3 e^{-\frac{z^2}{2}} dz, \quad (6)$$

The probable number of wrong decisions of a set N_i , for the case of a false defect will be equal

$$N_{ifd} = N_i(P_{fd}), \quad (7)$$

The probable number of wrong decisions of the set N , expressed by the formula

$$N_{fd} = \sum_{i=0}^m N \int_{L_i}^{H_i} \theta_1(z) dz \left[\sum_{j=0}^k \frac{1}{2\pi} \int_{\theta_i}^{\lambda_i} e^{-\frac{t^2}{2}} dt \cdot \frac{1}{\sqrt{2\pi}} \int_{\frac{3k-j}{k}}^3 e^{-\frac{z^2}{2}} dz \right], \quad (8)$$

where:

$$\begin{aligned} L_i &= (\bar{S}_l - 3\sigma_l + j\Delta S_l)_{\sigma_l}; \\ H_i &= (\bar{S}_l - 3\sigma_l + (j + 1)\Delta S_l)_{\sigma_l}. \end{aligned}$$

New variables z and t are centered and normalized integration variables.

A similar algorithm are an expression to calculate the probable number of objects in the case of undetected marriage, as well as a whole to calculate the values of these probabilities on the upper normative value in S_h .

Let us consider that p et case of the distribution of normative values of S_l by the normal law and monitored parameter S by Weibull law, which, as studies have shown, is not only one of the most common of the known laws, but also the most appropriate distribution law for modeling purposes. Weibull law is three-parameter law. Common laws, as the normal law can be considered in a certain approximation, as a special case of this law (Law and Kelton, 2000).

Density function of the Weibull law is as follows:

$$f(S, \alpha, \beta, \gamma) = \frac{\beta}{\alpha} (S - \gamma)^{\beta-1} e^{-\frac{(S-\gamma)^\beta}{\alpha}}, S \geq \gamma \quad (9)$$

where:

α - the scale parameter;
 β - the shape parameter;
 γ - location parameter.

At distribution shape parameter $\beta=1$, it simulates an exponential law, with $\beta=2.5$ - approximates the Rayleigh law, while $\beta=3.25$, the shape of the Weibull distribution is close to the normal distribution, which is quite often used in research practice. However, the dignity of the Weibull law consists in the fact that he, unlike the normal law has an analytical form of the integral of the distribution law, which has the form

$$F(S) = 1 - e^{-\frac{(S-\gamma)^\beta}{\alpha}}, \quad (10)$$

So the probability P_{ifd} takes the following form:

$$P_{ifd} = \sum_{i=0}^k (P_{jfd})_i = \sum_{i=0}^k \left[e^{-\frac{S_1^\beta}{\alpha}} - e^{-\frac{S_1^{\beta+1}}{\alpha}} \right] \frac{1}{2\pi} \int_{\frac{S_{li}-S_j}{\sigma_\varphi}}^3 e^{-\frac{z^2}{2}} dz, \quad (11)$$

Then, the total number of false defect of N samples will be

$$N_{fd} = \sum_{i=0}^k (\Delta N_i) = \sum_{i=0}^m N \int_{L_i}^{H_i} \theta(S) ds \cdot$$

$$\sum_{i=0}^k \left[e^{-\frac{\left(\frac{1}{\alpha\beta}\Gamma(1+\frac{1}{\beta})+n\sqrt{\frac{1}{\alpha\beta}\left[\Gamma(1+\frac{2}{\beta})-\left(\Gamma(1+\frac{1}{\beta})\right)^2\right]-3\sigma_\varphi+i\Delta S}\right)^\beta}{\alpha}} - e^{-\frac{\left(\frac{1}{\alpha\beta}\Gamma(1+\frac{1}{\beta})+n\sqrt{\frac{1}{\alpha\beta}\left[\Gamma(1+\frac{2}{\beta})-\left(\Gamma(1+\frac{1}{\beta})\right)^2\right]-3\sigma_\varphi+(i+1)\Delta S}\right)^\beta}{\alpha}} \right] \cdot \frac{1}{2\pi} \int_{3-\Delta S}^3 e^{-\frac{z^2}{2}} dz, \quad (12)$$

According in formula (12), the average and standard deviation of the Weibull law expressed in (Law and Kelton, 2000):

$$\bar{S} = \frac{1}{\alpha\beta} \Gamma\left(1 + \frac{1}{\beta}\right), \quad (13)$$

$$\sigma_s^2 = \frac{1}{(\alpha\beta)^\beta} \left[\Gamma\left(1 + \frac{1}{\beta}\right) - \left(\Gamma\left(1 + \frac{1}{\beta}\right)\right)^2 \right], \quad (14)$$

For the case of defect, the number of undetected decisions in the sample N will be

$$N_{ud} = \sum_{i=0}^m N \int_{L_i}^{H_i} \theta(S) ds \cdot \sum_{i=0}^k \left[e^{-\frac{\left(\frac{1}{\alpha\beta}\Gamma(1+\frac{1}{\beta})+n\sqrt{\frac{1}{\alpha\beta}\left[\Gamma(1+\frac{2}{\beta})-\left(\Gamma(1+\frac{1}{\beta})\right)^2\right]-3\sigma_\varphi-(i+1)\Delta S}\right)^\beta}{\alpha}} - e^{-\frac{\left(\frac{1}{\alpha\beta}\Gamma(1+\frac{1}{\beta})+n\sqrt{\frac{1}{\alpha\beta}\left[\Gamma(1+\frac{2}{\beta})-\left(\Gamma(1+\frac{1}{\beta})\right)^2\right]-3\sigma_\varphi-i\Delta S}\right)^\beta}{\alpha}} \right] \cdot \frac{1}{2\pi} \int_{3-\Delta S}^3 e^{-\frac{z^2}{2}} dz, \quad (15)$$

Use simulation algorithm is shown on Figure 2 to calculate the values of the probable risks expressions

(12) and (15) rationally. Algorithm works as follows:

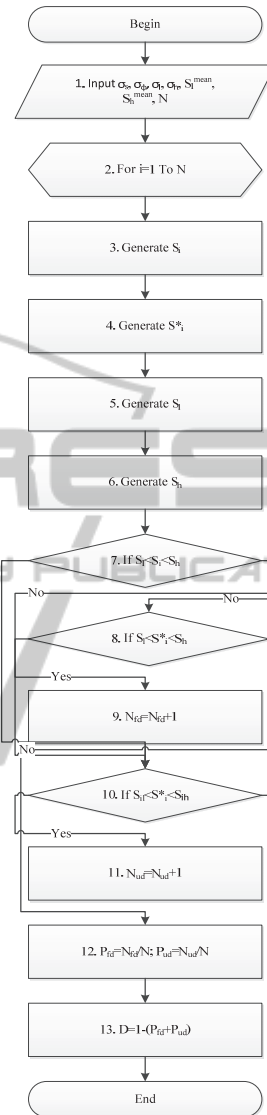


Figure 2: The simulation algorithm for estimating the quality control in the conditions of uncertainty normative values.

In the unit 2 opens the cycle 1 to N. In blocks 3, 4, 5, 6 are generated ("played out"), the random values of the lower S_{il} and upper S_{ih} ratios in the controlled parameter.

Section 7 contains a logical condition IF (branching) $S_{il} < S_i < S_{ih}$.

If provided, $S_{il} < S_i < S_{ih}$. value S_i will be within a limits (the condition is true - YES), the condition is now following analysis of the

measurement result $S_{il} < S_i^* < S_{ih}$ in (block 8) and in the case of YES - a true outcome, control is passed to a new cycle unit 2. If the condition is false block 8 - NO, there was an error - a false defect, in block 9 counter is triggered these cases and will return to the beginning of the next cycle of unit 2.

If block 7 the condition is false - it is not, at block 10 S_i analysis of the condition $S_{il} < S_i^* < S_{ih}$ and at the right end of the control is passed to the beginning of a new cycle (block 2), otherwise (NO) an error undetected defect and in block 11 N_{ud} counter is triggered and a new cycle begins in block 2.

Upon completion of the specified number of simulations equal to N , in block 12 is evaluated and the probability of undetected false defect by (probable error control)

$$P_{fd} = \frac{N_{fd}}{N}, \quad (16)$$

$$P_{ud} = \frac{N_{ud}}{N}, \quad (17)$$

where N_{fd} is the number of events false defect;

N_{ud} - number of events undetected defect;

N - total number of simulation events.

The integral indicator of quality control - the accuracy of D in block 13 calculates by the formula

$$D = 1 - (P_{fd} + P_{ud}). \quad (16)$$

Similar calculations can be made for different combinations of distribution laws, and compare the results to determine the degree impact on quality control and decision-making process of distribution laws.

4 CONCLUSIONS

We can make the following conclusions on simulation results: the formation of the total error control should be considered as a system of interconnected agents, expression quantitative assessment of the likely risks of undetected and false defect laws are functions of distribution, the statistical characteristics of agents and average standard deviations of the statistical characteristics of standards.

REFERENCES

- Blatt, H., Robert J. T., 1995. *Petrology: Igneous, Sedimentary, and Metamorphic*, 2nd ed. New York: W.H. Freeman.
- Farndon, J., 2006. *The practical encyclopedia of rocks and minerals. How to Find, Identify, Collect and Maintain the World's best Specimens, with over 1000 Photographs and Artworks*. London: Lorenz Books.
- Baklanova, O. E., Uzdenbaev, Z.S., 2013. Development of methodology for analysis of mineral rocks in the mining industry. *Joint issue of the Bulletin of the East Kazakhstan state technical University and Computer technology of Institute of computational technologies, Siberian branch of the Russian Academy of Sciences, Part 1, September, 2013. - P.60-66.*
- Clarke, A. R., Eberhardt, C. N., 2002. *Microscopy Techniques for Materials*. Science Woodhead Publishing, CRC Press, 459 p.
- Pantelev, C., Egorova, O., Klykova, E., 2005. *Computer microscopy*. Technosphere, 304 p.
- Chris, P., 2002. *Rocks and Minerals. Smithsonian Handbooks*. New York: Dorling Kindersley.
- Shaffer, P. R., Herbert S. Z., Raymond P., 2001. *Rocks, Gems and Minerals*. Rev. ed. New York: St. Martin's Press.
- Borschev, A. V., 2008. How agent-based modeling and its place in the arsenal of the analyst. *Exponenta Pro, 2008. № 3-4 [Electronic resource] //URL: <http://www.xitek.ru/anylogic/articles/>*
- Kuleshov, V.K., Kornev, V. A., 2011. *Modeling process control and decision-making*. Tomsk Polytechnic University, 295 p.
- Wentzel, E. S., 2002. *Probability theory*. Moscow: Higher School, 575p.
- Svinolupov, U., Kornev, V.A., Kuleshov, V. A., 2012. Methodology for assessing the quality of measuring instruments (Case study: Pressure sensors) // *Russian Journal of Nondestructive Testing* 48 (2) 2012, P. 129-136.
- Law, A. M., Kelton, W. D., 2000. *Simulation Modeling and Analysis, Third Edition*, McGraw-Hill, New York.