

Procedure of Analysis and Digital Processing of Large Data Array in Real Time upon Intellectualization of Roller Drilling

A.A. Shigina¹, S.M. Antipina¹, A.A. Stupina^{1,2}, V.S. Tynchenko², N.N. Dzhioeva¹

¹Institute of Business Process Management and Economics, Siberian Federal University, Krasnoyarsk, Russia

²Institute of Informatics and Telecommunications, Siberian State University of Science and Technology named after academician M.F. Reshetnev, Krasnoyarsk, Russia

ABSTRACT

Principles and tools of efficient acquisition, analysis, and processing of large data array have been determined using modern information technologies and confidence estimation of data from sensors in systems of automatic and automated control of drilling. Procedure of analysis and digital processing of large data array in real time has been developed on the basis of intelligent control of roller drilling operating under conditions of information uncertainty and related with unpredictable variations of rock properties during drilling. Using scatter plot, the structure of rock bedding in 3D space is presented. Classification rule has been developed on the basis of acquired information to solve the problem of automatic identification of rocks. Procedure of total tree walking has been used aiming at minimization of errors upon classification of samples. In order to estimate the classification strength of the developed tree, classification matrix has been proposed containing data of correctly and incorrectly classified samples. Application of intelligent data analysis (Data mining) has made it possible to detect usable data about properties of drilled rock and process variables, digital data processing, recognition of regularities and trends existing in these data. The problem of recognition of usable data in large data arrays and estimation of interrelation between various factors has been solved, regularities and trends existing in these data have been recognized with regard to roller drilling during intellectualization of roller drilling.

Key words: digital processing, data mining, data array, intelligent control system, roller drilling.

1. INTRODUCTION

1.1 Formulation of the Problem

Taking into account the increasing demand for automation of acquisition, processing, and preparation of large data arrays, real-time recognition and forecasting of difficulties during drilling, improvement of control efficiency of roller drilling, it is required to apply more advanced information technologies of processing and simulation for large data arrays with complex heterogeneous or uncertain structure: Big Data technologies, capable to optimize roller drilling and

to improve its efficiency in mining, oil and gas, and construction industries [1–3]. Improvement of the combination of heterogeneous indicators of control efficiency of complex systems by methods of system analysis becomes difficult due to necessity of their comprehensive estimation [4].

Roller drilling control is a complicated nonstationary process with random alternation of disturbances. The existing control systems are based on various criteria and various information technologies. However, application of such control systems becomes more difficult due to certain external and internal factors, negatively affecting drilling:

- information uncertainty related with unforecastable variation of rock properties during drilling;
- difficult and sometimes impossible acquisition and processing of large drilling data arrays;
- mismatch between current predicted drilling variables and actual process due to duration of fuzzy model applied in control system based on statistic data processing;
- hardware and control system algorithm do not consider for quick and short shock impacts upon variation of rock properties during drilling;
- complicated optimization of roller drilling in real time;
- risk of complications and their difficult forecasting during drilling in the case of invalid solution and selection of control actions in real time;
- significant economic loss due to improper analysis of large data arrays and acquisition of wrong information about rock properties and process variables using instrumentation of control system, as well as due to impossibility to forecast complications during drilling under the conditions of information uncertainty leading to faults of hard- or software with subsequent process shutdown;
- human factor.

The mentioned problems occurring upon roller drilling control could be solved by implementation of modern procedure, being the basis of information technologies of simulation, analysis and digital processing of large data arrays with complex heterogeneous or uncertain structure, into intelligent process control system [5]. This procedure would allow to adequately estimate reliability of acquired

information during drilling, to automate interpretation of results of digital processing of large data arrays, to optimize roller drilling under the conditions of uncertainty in initial data and to forecast process efficiency in real time.

Importance of the formulated problem is determined by necessity to obtain accurate results of analysis and digital processing of large arrays of heterogeneous and rapid information about rock properties, varying drilling variables transferred from sensors of intelligent system of roller drilling in real time, which cannot be performed using conventional instruments. Data for computations from sensors and instrumentation are always characterized by significant error [6]. Faults of control system of this process can be caused by minor oscillation of the variables. In order to solve this problem, it is required to apply Data mining technologies aiming at detection of usable data in large data arrays, detection of regularities and trends in the existing data applicable to roller drilling [7–9], as well as the technologies of predictive analytics aimed at achievement of reliable forecasts on the basis of processed and structured digital information about rock properties using intelligent system of roller drilling control and estimation of interrelation between various factors, their interpretation and estimation of risks for prevention of faults in operation of control system.

Using Data mining technologies for digital processing of information about rock properties by control system of roller drilling would allow to reveal regularities and trends in data, which usually cannot be detected using conventional analysis of data due to complex bonds or excessive data amount. The revealed bonds between various factors would allow to make proposals for development of regression, prognostic and/or neural models of roller drilling. Using these technologies, the intelligent system of roller drilling control can recognize and estimate situation in order to forecast possible pattern of production process and to adopt optimum solution concerning selection of control actions and drilling mode with consideration for revealed risk factors aiming at process optimization and efficient drilling. In the case of faults of drilling equipment, decrease in accuracy and rate of drilling, it is possible to apply such data analysis as predictive analytics used for reduction of cost and frequency of equipment overhauls and, therefore, decrease in drilling costs.

Therefore, this work is aimed at improvement of efficiency of roller drilling due to development of procedure of analysis and digital processing of large structured and nonstructured arrays of data about rock properties and roller drilling variables, which can be applied to Big Data technologies on the basis of intelligent control system upon drilling of holes of various purposes (oil, gas, and blast holes in quarries) under the conditions of information uncertainty.

1.2 Review of Existing Procedures

At present stage of development of drilling technology, it becomes necessary to apply reserves stipulated in

mathematical basis and software of drilling variables. Many oil and gas producing companies started development of integrated control centers or centers of hi-tech, communications, visualization, and control rooms, which provide more efficient data acquisition, monitoring, communication and data exchange. These integrated control centers are aimed at adoption of more substantiated solutions and execution of respective enterprise-wide actions. In this regard, development and commercial implementation of optimization procedures, as well as continuous improvement of existing software, tools and methods of large data array processing aimed at searching for optimum drilling variables and improvement of its efficiency are an urgent R&D problem.

Databank should be developed for estimation of overall data array [10–11]. Recently the mentioned problem has been solved by development of storage in relational database management system (RDBMS) for digital structured data, nonstructured data were stored in file directories. RDBMS is characterized by one major disadvantage: rigid link between the storage structure and the structure of initial data. Due to high uncertainty of data search, the use of SQL queries to rigidly structured database is inefficient. This disadvantage is eliminated in the case of storage based on NoSQL in combination with searching system. Application of "key-value" databases provides unified indexation of all data arrays irrespective of their structure. Using Apache Lucene search engine as linking element between the object of interest and data array provides extended opportunities to find the required data under the conditions of high uncertainty.

1. "Drilling Optimizer" developed by Geonavigation Technologies company involves large data arrays in geoengineering: the technology improves drilling efficiency, for instance, correctly estimating flow rate and weight of boring fluid. Gazprom Neft company considers the project as promising for cooperation. In 2010, at initial stage of the project, the main activity was aimed at physical simulation of drilling. The main target was to determine the position of horizontal well shaft at any specific time inside the target formation. Such technology and visualization of overall process were required. The concept of "Drilling Optimizer" is to add statistic information to physical simulation, which determines position of well shaft in bed, so that to supplement the information about drilling direction, rate, and weight of boring fluid with the data of difficulties occurring in all wells at given depth in any specific bed. All this information should be available to an operator with a single click. Another problem in this project is development of predictive models.

2. MIPT Engineering Center for hard-to-recover reserves in cooperation with experts of R&D center of Gazprom Neft initiated project of technology based on machine learning.

During development of deposits, the solutions on application of various methods to increase extraction rate are based on operational data from bore wells. Measured operation data

(discharge of fluid, oil, water content in product, bottom hole pressure) are transferred from all bore wells of the company, in addition, monthly data on production are arranged, as well as the information is available of performed studied and physical properties of bed and mined fluid and gas.

At the same time, the quality of these data sometimes does not allow to perform comprehensive analysis: information for certain time ranges can be unavailable, some measurements sometimes mismatch physical model or do not agree with each other. Invalid data in reports can be attributed both to faults of measuring devices and to human factor.

In the scope of this project, the searching algorithms of invalid data and recovery of omitted data were developed, as well as detection of mutual interaction of bore wells, classification of bore wells in terms of deviation of current output from achievable for bore wells in similar geological environment. The algorithms of machine learning are developed by Gazprom Neft in the frames of the engineering strategy: Electronic asset development (ERA). This strategy of development of IT projects by Gazprom Neft in the field of exploration and production includes all activities: exploration, geology, boring, production, oil field development.

The developed tools based on machine learning would improve the rate of processing and analysis of large data arrays from oil fields. In addition, the applied methods of machine learning would allow to integrate heterogeneous data, to analyze in details all available information, which would lead to new conclusions capable to reach a new level of data quality. Implementation of the developed methods would allow to halve the time of online analysis of operation data, to consider rapidly for new regularities upon further development of oil fields, to select optimum methods aiming at increase in oil output with simultaneous reduction of expenses.

3. IBM Research proposes an innovative engineering solution for oil and gas industry: "Hybrid Drilling Advisor" (Moscow). The IBM drilling advisor is a combination of physical and statistical models for early detection of problems on the basis of cognitive technologies and machine learning upon selection of modern optimum solutions during drilling. This approach allows to forecast risks of drilling using bore well physics and advanced analytical methods of data processing; to inform experts about problems during drilling with sufficient time reserve to avoid these problems; to provide optimum solutions to a problem on the basis of previous experience and continuous improvement of search using cognitive technologies and methods of machine learning. The solution is integrated with existing systems of acquisition, processing, and visualization of data used during drilling.

The IBM "complication" forecast indicates at high risk of complications caused, for instance, by stepwise pressure increase upon constant pump stroke. In order to determine optimum solution to this problem, the Watson advisor processes thousands of reports and records of knowledge base;

determines similar risks; analyzes results of adopted solutions; proposes variants of solutions; recommends solutions based on results of analysis; requests for response from engineers concerning the given variants aiming at quality improvement of analysis and recommendations.

4. International Science and Technology Centre in cooperation with MPTI and Ecole des Mines de Paris developed CAD system for blast and drilling activities (CAD BDA). The system is based on advanced microprocessor technologies and software. CAD BVA is a complex combining technical means of data acquisition during drilling and excavation with software: automated workstation of blast and drilling engineer. The developed complex is a practical implementation of blast and drilling operations with continuous adjustment of strength properties of developed deposit by measurement of energy consumption by well drilling.

CAD BDA Blast Maker is the most complete solution to optimum designing of blast and drilling operations at mining sites without modification of the applied blast and drilling technology. Full-scale implementation of the system would allow not only to decrease costs of blast and drilling operations but to improve their quality as well as control efficiency of equipment and processes at mining sites. Completeness of information about previously designed and executed blasts, about drilling variables and modes at all processed bore wells provides actual opportunity for adaptive control of blast and drilling. This reduces costs by 10-15% of overall amount required for blast and drilling operations, which equals to one third of overall expenses of a mining company.

During system operation, data from sensors of drill rigs are continuously recorded and processed. The data acquisition system of drilling rigs is based on the determination of strength properties of rocks by measurement of energy consumption for drilling. A useful energy indicator is specific energy of drilling. On the basis of appropriate processing of the acquired data, it is possible to predict distribution of energy consumption of drilling, hence, the strength properties of rocks and cracks along total depth of bore well; in addition, after accumulation of sufficient statistic information, the rock type can be identified. Moreover, the measured drilling variables reflect not only the properties of rocks but also the state of drilling equipment. The acquired data on all processed bore wells are transferred to system database, then they are used by digital model for adjustment of geological structure of designed block.

It should be mentioned that the applied operation algorithm and the intelligent control system include inputs formed under the conditions of continuous monitoring of required variables [12]. In this regard, the proposed intelligent system is comprised of specialized software and hardware used for adjustment of variables on the basis of certain values of external factors [5]. Implementation of intelligent automated

control systems of data acquisition and processing provides detection of actual regularities with consideration for informative heterogeneity, development of solution under uncertainty with reproduction of nonformalized managerial expert practice. Database regarding operation of such system is obligatory supplemented and saved in automatic mode, training of highly efficient managerial practices is carried out.

Modern and optimum well-proven intelligent systems are characterized by advanced communicative capabilities, they can solve complicated and poorly formalized problems, they can be self-trained and have high level of adaptiveness [13–14]. The core of such adaptive information system is presented by actively developed model in problem areas. Therefore, the procedure of designing and adaptation of all automated control systems should be reduced to development of model of the most problematic areas with obligatory timely adjustment. At present, there is no distinct and unified classification of such intelligent systems, because their common definition is unavailable. Herewith, high efficiency of modern automated intelligent control systems is determined by stringent requirements to such properties as data exchange and quality of data transfer [15–16].

2. METHODS

Procedure of estimation and processing of large data arrays upon intellectualization of roller drilling assumes solution of problems aimed at development of control function of dynamic model by application of modern adaptive electromagnetic rotary–feeding mechanism (ARFM) [12]. Such mechanism operates as a sensor with actuator performing timely response to any sudden short disturbances in the form of drilling shock loads. In that respect, control loss can be observed during drilling as well as inefficient energy consumption for control, thus, it is important to pay attention to correction and smoothing of shock loads with obtaining of maximum rapid back response. This response should describe such variables as time and shock value; hence, this element is obligatory included into control hardware in the form of sensor.

Therefore, such modern drilling rig with intelligent automated control significantly improves process efficiency [17]. Adjustment of modes in the executed processes involves numerous input and output variables, therefore, decrease in important indicators presented by the value, duration, and frequency of transition processes under the conditions of conversion of automated control variable. In such situation it would be reasonable to apply corrections under the conditions of long conversions of uncontrolled sudden perturbations, which are the properties of drilled rock. Short sudden (unpredictable) perturbations require for application of nondelay systems. With the aim of highly efficient and timely

response to unpredicted shock impacts, it is required to apply adaptive rotary–feeding mechanisms, which are capable to smooth all such disturbances and be embodied in the form of internal system sensor to control roller drilling.

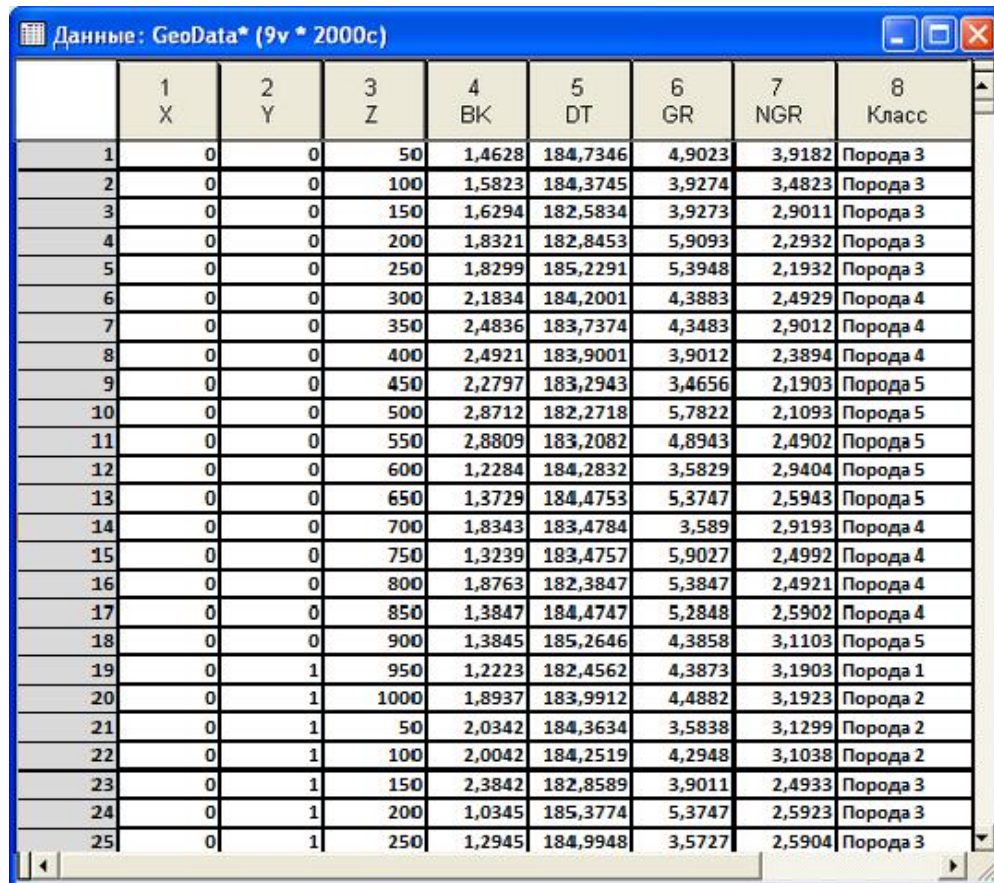
Generation of control action on controlled objects means selection of appropriate algorithm aiming at obtaining of the most optimum output variables. As shown, a valid solution to the problem is to add special regulator into such intelligent system in the form of controller operating without direct participation of an operator. Regulator controls operation of such variables as drilling rate and rotation frequency with generation of data signals qualifying variations of rock properties [12]. Adjusting values of unpredicted perturbations in this case are determined by computer and controller, and preset variables for adaptive constituents and sensors depend on data relating to variations of physical and mechanical determinants of drilling rock. It should be mentioned that a set of random uncontrolled variables is presented by a set of effects on drilling in the form of random noises in controlling channels not subjected to highly accurate measurements [18–20].

Adherence to the procedure of estimation and processing of large data arrays upon intellectualization of roller drilling is based on close interrelation of drilling model variables in combination with addition of regulator in the form of intelligent system (adaptive constituent) [12]. Such element provides all input control actions in the form of adjusted variables presented by axial force and rotation frequency, which modifies output variables, presents analysis of drilling efficiency and equipment resources with consideration for all key factors.

Conventional statistics operate with averaged properties of selection, which are often dummy variables. Therefore, the methods of mathematical statistics are useful mainly for verification of previously formulated hypotheses, whereas determination of hypothesis is sometimes sufficiently complicated and labor consuming problem. Modern Data Mining technologies process data aiming at automatic searching for patterns peculiar for any fragments of heterogeneous multidimensional data.

Selection of Data Mining method often depends on the type of available data and on information which should be obtained. All this makes it possible to assume efficient application of such Data Mining methods for solution of geophysical problems as neural networks, solution tree, and others, since these methods are characterized by adaptiveness, generalization, knowledge acquisition, and simulation of complex nonlinear dependences in data arrays.

This work is aimed at description and analysis of data structure of roller drilling in real time, at development of automatic model of rock classification. The data are analyzed obtained upon drilling of bore wells to 1,000 m (Fig. 1).



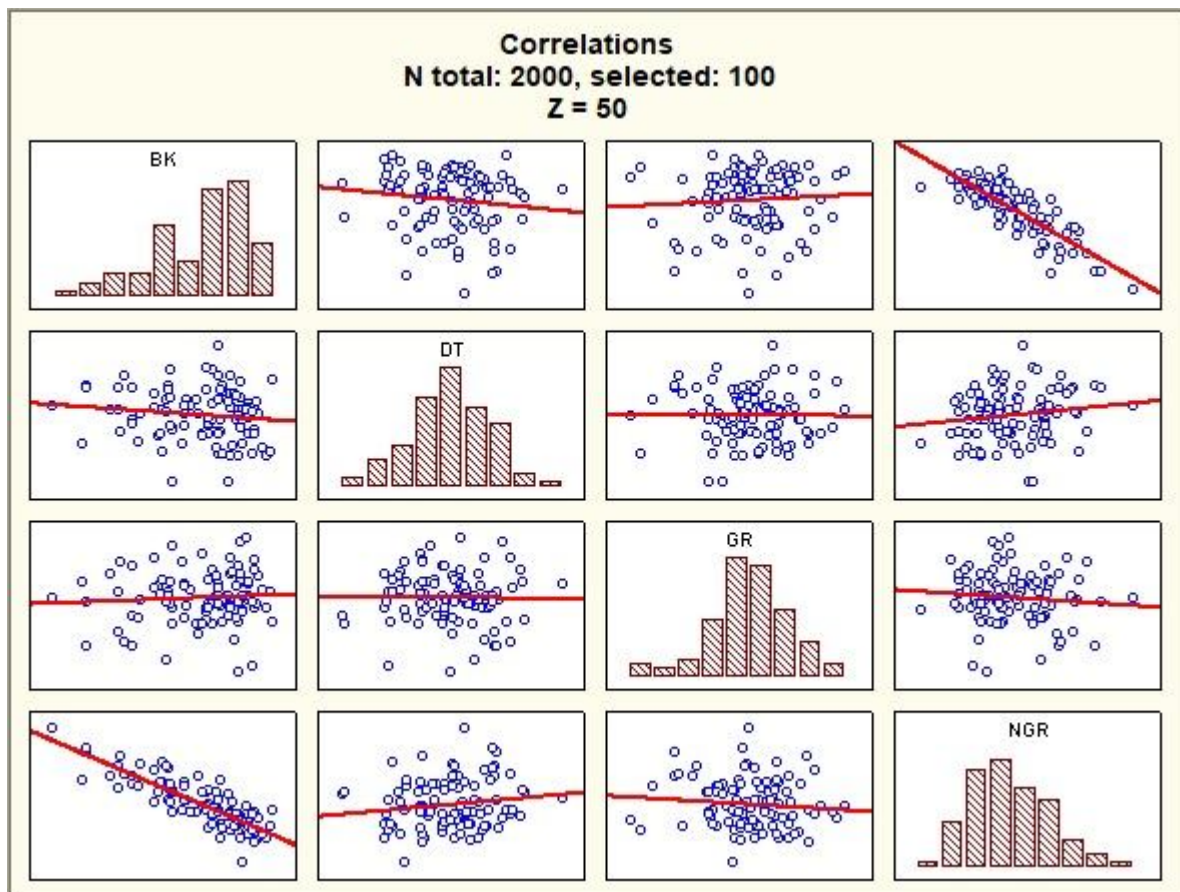
	1 X	2 Y	3 Z	4 BK	5 DT	6 GR	7 NGR	8 Класс
1	0	0	50	1,4628	184,7346	4,9023	3,9182	Порода 3
2	0	0	100	1,5823	184,3745	3,9274	3,4823	Порода 3
3	0	0	150	1,6294	182,5834	3,9273	2,9011	Порода 3
4	0	0	200	1,8321	182,8453	5,9093	2,2932	Порода 3
5	0	0	250	1,8299	185,2291	5,3948	2,1932	Порода 3
6	0	0	300	2,1834	184,2001	4,3883	2,4929	Порода 4
7	0	0	350	2,4836	183,7374	4,3483	2,9012	Порода 4
8	0	0	400	2,4921	183,9001	3,9012	2,3894	Порода 4
9	0	0	450	2,2797	183,2943	3,4656	2,1903	Порода 5
10	0	0	500	2,8712	182,2718	5,7822	2,1093	Порода 5
11	0	0	550	2,8809	183,2082	4,8943	2,4902	Порода 5
12	0	0	600	1,2284	184,2832	3,5829	2,9404	Порода 5
13	0	0	650	1,3729	184,4753	5,3747	2,5943	Порода 5
14	0	0	700	1,8343	183,4784	3,589	2,9193	Порода 4
15	0	0	750	1,3239	183,4757	5,9027	2,4992	Порода 4
16	0	0	800	1,8763	182,3847	5,3847	2,4921	Порода 4
17	0	0	850	1,3847	184,4747	5,2848	2,5902	Порода 4
18	0	0	900	1,3845	185,2646	4,3858	3,1103	Порода 5
19	0	1	950	1,2223	182,4562	4,3873	3,1903	Порода 1
20	0	1	1000	1,8937	183,9912	4,4882	3,1923	Порода 2
21	0	1	50	2,0342	184,3634	3,5838	3,1299	Порода 2
22	0	1	100	2,0042	184,2519	4,2948	3,1038	Порода 2
23	0	1	150	2,3842	182,8589	3,9011	2,4933	Порода 3
24	0	1	200	1,0345	185,3774	5,3747	2,5923	Порода 3
25	0	1	250	1,2945	184,9948	3,5727	2,5904	Порода 3

Figure 1: Table with initial data. Column 8: "Class", identifying class of rocks

The variables X and Y provide conventional coding of bore well coordinate on land, the variable Z determines bore well depth. Soil sample was taken each 50 m and the following variables were analyzed:

- BK – lateral log. Measurement of specific electric resistance of rocks.
- DT – acoustic log. Measurement of travel time interval of longitudinal acoustic wave.
- GR – gamma method. Intensity of natural gamma radiation of rocks.
- NGR – neutron gamma method. Measurement of neutron absorption and scatter by rocks.

Then, on the basis of combination of these indices, the conclusion is obtained ("Class" variable) about structural composition of rocks. At the first stage, Data Mining performs interactive data studies while determining initially the types of applied variables and possible interactions between them using STATISTICA software and STATISTICA Data Mining module. During the studies specific observations were selected from large data array on the basis of highlighted subgroups characterized by certain values or ranges of variables. Data are analyzed for samples from the depth of 50 m (Figs. 2, 3).

**Figure 2:** Correlations

Descriptive Statistics (GeoData2)

N total: 2000, selected: 100

Z = 50

	Average	Standard deviation	Minimum	Maximum	Observations	Missed
BK	1.3086	0.550074	0.1065	1.8465	100	0
DT	184.5221	0.894564	193.0654	195.0843	100	0
GR	4.4956	0.709845	1.8562	6.3561	100	0
NGR	2.3045	0.276484	1.7651	3.2024	100	0

Figure 3: Descriptive statistics

While analyzing average values, scatters and dependence between the indices, it is possible to estimate the chance of detection of oil collector at a given depth. Using maximum and minimum, the number of observations, it is possible to reveal whether the instrument readings contain faults or omissions.

The indicators are analyzed not only along the bedding depth, the results can be analyzed for each bore well. It is sufficient to select conventional coordinates (variables X and Y) as drilling variables. After boring, the plot indicator as a function of depth (log chart) was obtained for this bore well (Fig. 4).

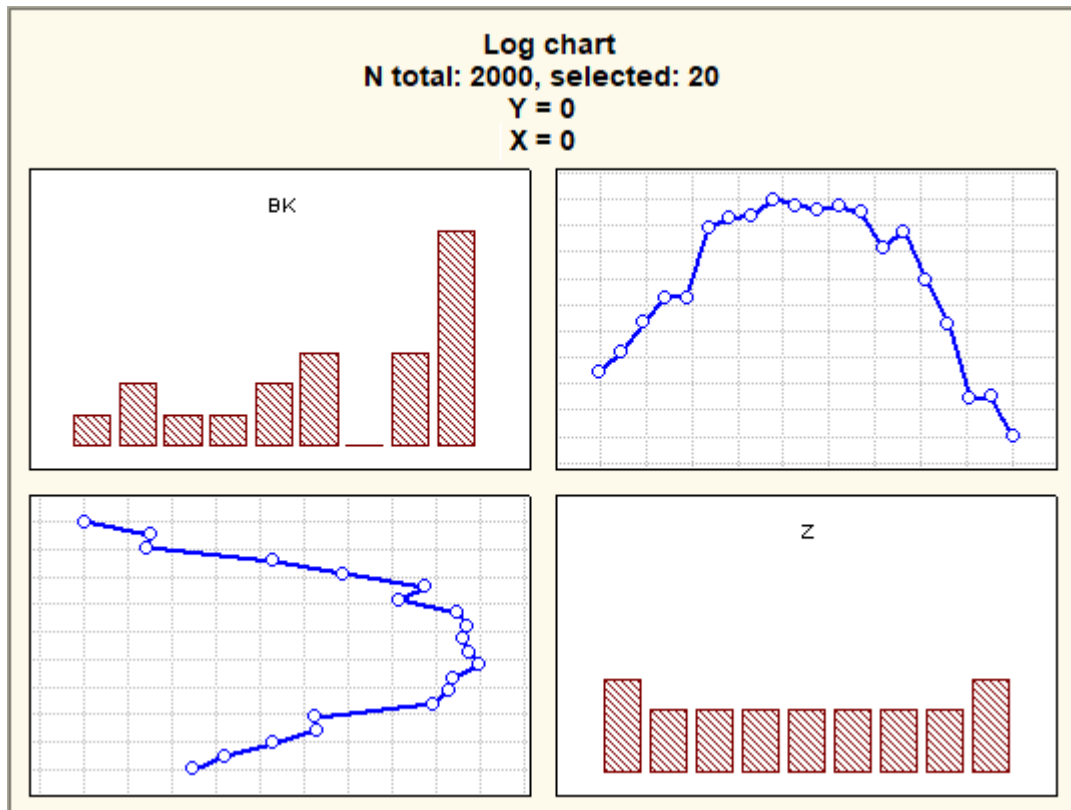


Figure 4: Log chart

Pie chart is obtained which reflects structural composition of the obtained samples from bore well (Fig. 5).

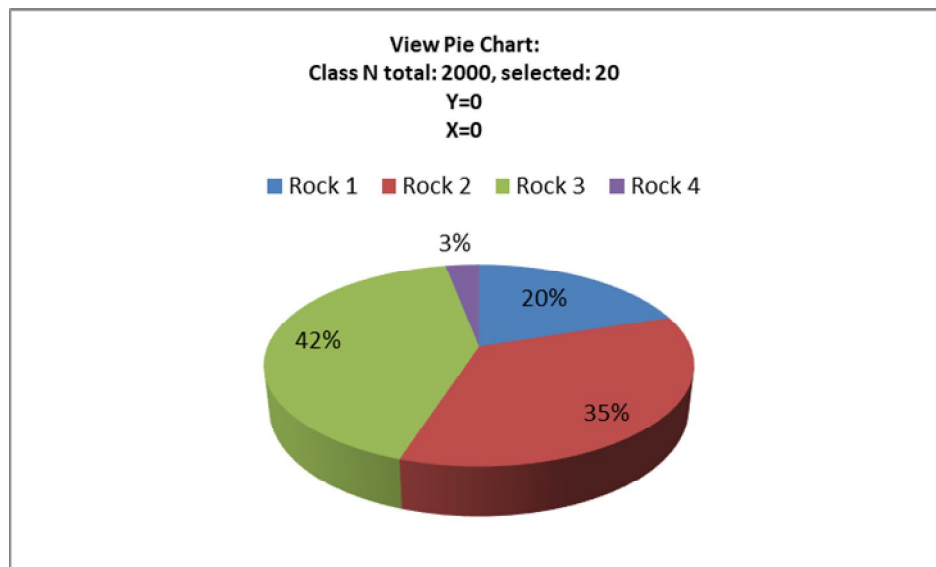


Figure 5: Pie chart

STATISTICA software provides wide range of graphical methods for visual presentation of experimental results. Using the graphical means in STATISTICA, it is possible to select analytical online method of analysis. They contain various presets allowing interactive on-screen data presentation. Flexible controls of several plots make it possible to combine various images and to carry out dynamic

exchange between applications (using OLE).

Using STATISTICA Data Miner, the research problem is solved step-by-step:

- 1) Data import
- 2) Data conversion
- 3) Selection of visualization method
- 4) Analysis of results.

1) Data import

The table with GeoData.sta initial data (Figure6) was selected As Data source.

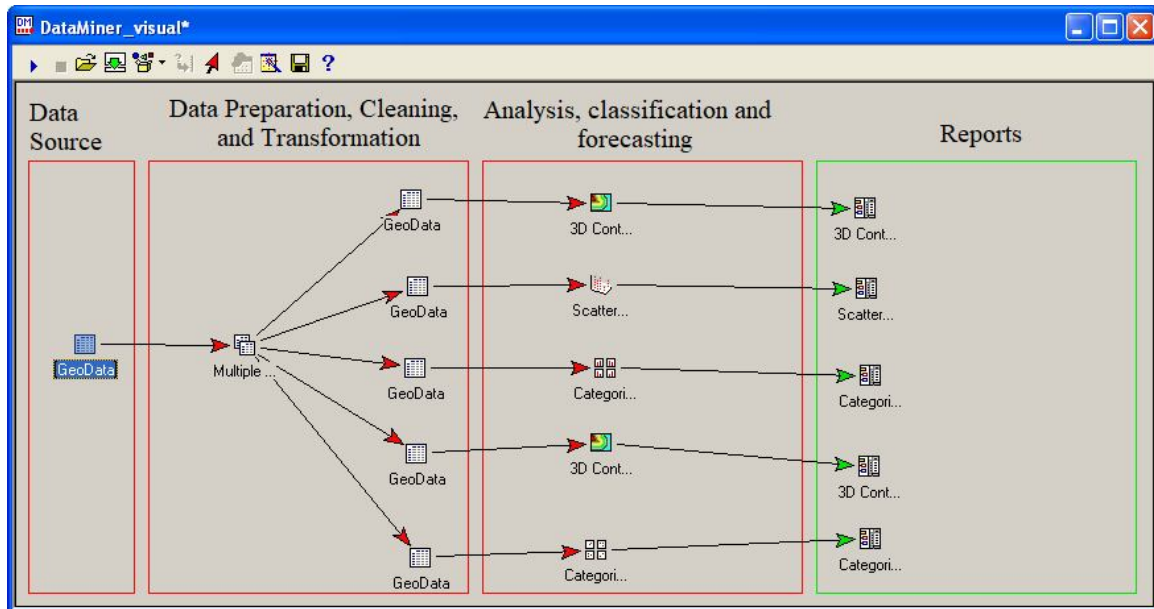


Figure 6: Data Miner workspace

2) Data conversion

At the conversion stage, the data required for analysis are prepared and selected.

b) 3D and 2D scatter plots;

c) Linear plots.

d) Analysis of results.

3) Selection of visualization method

Various visualization methods are available in "Analysis, classification and forecasting" window in addition to data analysis methods. The following plots were used in the studies:

b) Scatter plots

The available data can be presented not only in numerous cross sections but also be visualized in 3D space using 3D scatter plots (Fig. 7).

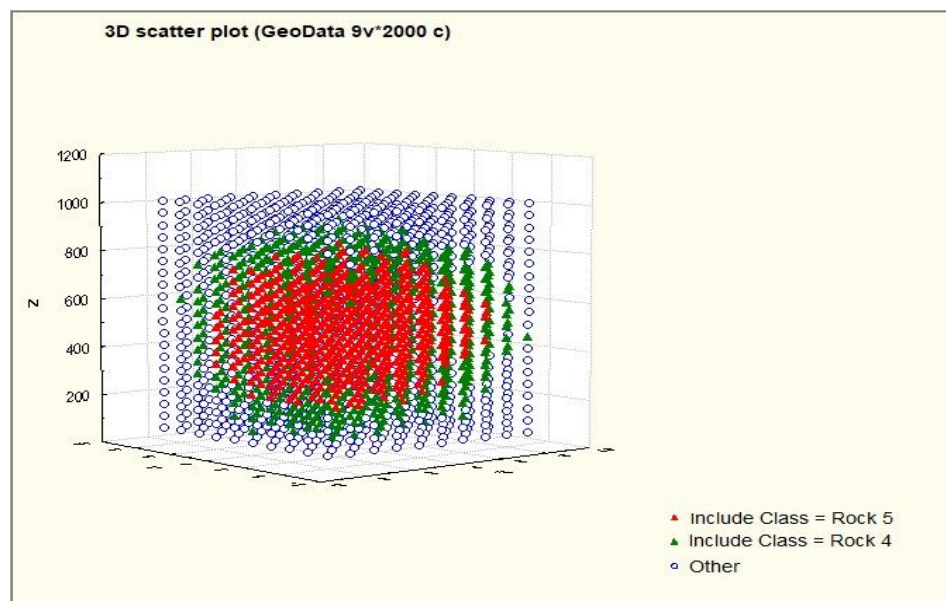


Figure 7: 3D scatter plot

Using this plot, the structure of rock bedding can be presented already in 3D space.

With such amount of data, it is proposed to apply Classification trees developed in STATISTICA Data Miner. Then the classification rule was on the basis of the acquired data, so that rock could be classified automatically. Using the same menu in STATISTICA Data Miner, the module of

classification tree plotting was added (Fig. 8). Complete enumeration of trees is used aiming at minimization of errors during classification of samples.

Classification trees are efficient when the predictors are different and present significantly different information for classification of objects.

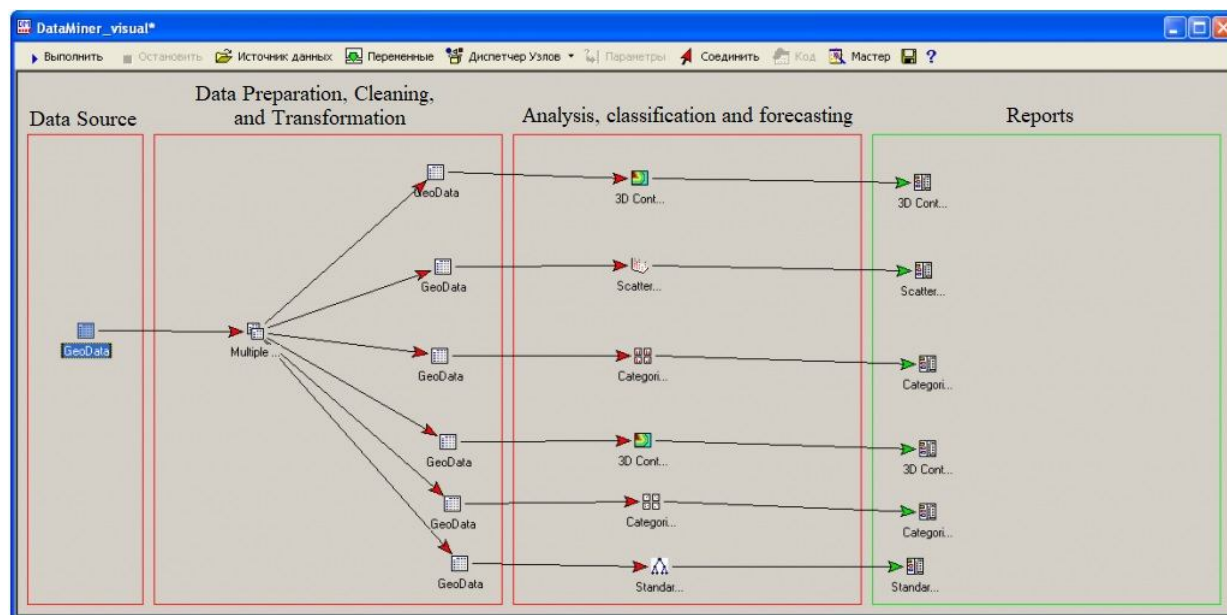


Figure 8: Data Miner workspace

After execution of the module, a model for classification is obtained. The classification tree shown below makes it possible to analyze online newly obtained samples by comparison of the indices (Fig. 9). All terminal tops (highlighted by red frame) are marked with one of five rocks contained in 'Class' variable. Finally, each new sample is referred to one rock type.

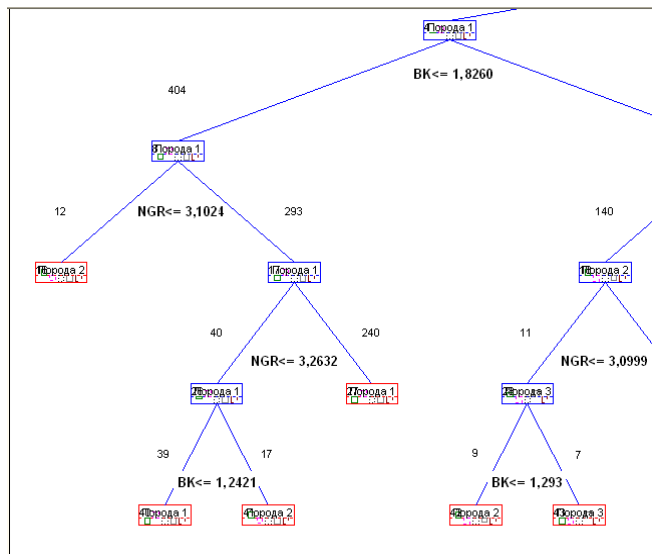


Figure 9: Classification tree with scrolling

The strength of classification tree can be estimated using the classification matrix which contains information of correctly and incorrectly classified samples (Figs. 10, 11). Large columns in the diagonals evidence sufficiently accurate classification of the model.

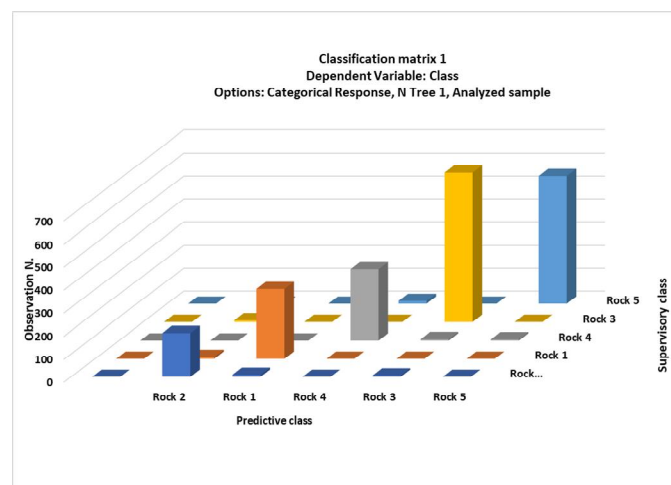
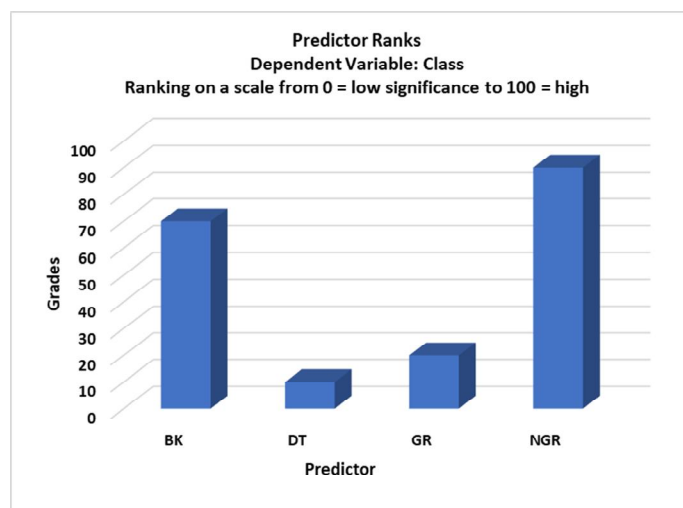


Figure 10: Classification matrix

Predicted and observed objects (GeoData)					
Predicted (rows) and observables (columns)					
Learning sample size N=2000					
Class	Class				
	Rock 2	Rock 1	Rock 4	Rock 3	Rock 5
Rock 2	190	9	0	6	0
Rock 1	8	302	0	0	0
Rock 4	0	0	309	6	5
Rock 3	7	0	0	650	0
Rock 5	0	0	12	0	556

Figure 11: Predicted and observed objects.

In order to determine the criteria, in terms of which the rocks in the selection are different, it is sufficient to analyze the predictor ranks in the report of STATISTICA Data Miner (Fig. 12).

**Figure 12:** Predictor ranks

3. RESULTS AND DISCUSSION

Visual analysis performed in STATISTICA Data Miner has demonstrated that some indicators strongly correlate between them. Moreover, the regions of homogeneity and structure of variation of these indicators were recognized. To some extent, it is possible to reach the deepest layers of data, while analyzing narrower subsets of observations on the basis of more and more complicated conditions of selection.

In the course of the studies, the procedure of real time analysis and digital processing of large data array was developed on the basis of intelligent control system of roller drilling operating under the conditions of information uncertainty related with unpredictable variations of rock properties during drilling. Development of this procedure applicable to Big Data technologies during intellectualization of roller drilling, estimation of interrelation between various factors was based on the methods of probability theory and mathematical statistics, mathematical simulation, data mining. The Data mining technology allowed to solve the problem of recognition of usable data about properties of

drilled rocks and process variables, digital data processing, detection of regularities and trends in these data. The problem of recognition of usable data in large arrays was solved including estimation of interrelation between various factors, regularities and trends in the data were revealed with regard to roller drilling upon its intellectualization.

4. CONCLUSION

Estimation and processing procedure is quite simple and efficient mechanism intended for predicted resources. Its application proved to be successful during drilling characterized not only by different stratification but also by fracturing indices. From this point of view, multiparametric predictions are reasonable in the case of specialized software reflecting residual forecasts of drilling resources. Operators in this case can follow key values in the form of ratios of such variables as axial force, rpm, drilling rate, as well as residual predicted resource. Currently developed mathematical models of roller drilling are characterized by sufficient accuracy, which makes it possible to estimate their adequacy to actually applied processes, as well as possibilities of practical application. Drilling processes should be reasonably controlled by specific constituent presented by special adaptive electromagnetic rotary–feeding mechanism [1]. It should be mentioned that roller drilling is a complex multiparametric process directly depending on several variables, including specifications of drilling bit and tools, physical and mechanical properties of rocks, as well as all process variables. Process control becomes more difficult due to impossibility to monitor variations of rock properties, which is stipulated by unavailability of facilities providing accurate preliminary information about strength and structure [21].

Therefore, it can be concluded that at present, highly efficient operation of the automatic systems applied for drilling requires for only rational solutions aimed at optimum values of controlled variables. The most important feature of such projects is the existence of intelligent control system of roller drilling on the basis of large arrays. From this point of view, only correctly selected intelligent system in steady mode can apply administrative action varying key process variables. Herewith, the current indicators are the closest to optimum variables. Thus, any simulation of dynamic drilling processes is carried out aiming at its optimization, and integrated approach to solution of problems aimed at development of mathematical models, system of acquisition and subsequent processing of primary data, makes it possible to obtain the required result which is characterized by high practical importance.

Some tools of Data Miner (for instance, interactive drilling), regularly required at initial stages of investigations, allow to obtain quickly primary concept of data dependences. The obtained bonds between various factors make it possible to make proposals for development of regression, prognostic

and/or neural models. STATISTICA Data Miner makes it possible to analyze data of geoinformation systems of complex multidimensional structure, to substantiate expert solutions, to develop complex prognostic models.

The new procedure differs from the known procedures in that it is based on intelligent control system of roller drilling, operating under the conditions of information uncertainty; it allows to obtain in real time accurate results of analysis and to process large amount of heterogeneous digital data rapidly received from sensors of intelligent control system regarding the properties of drilled rocks, varying drilling parameters, to structure this information and to reveal regularities and trends existing in the data.

This procedure applied in intelligent control system made it possible to reveal regularities and trends existing in the data regarding properties of drilled rocks and roller drilling variables. In the future the proposed procedure would allow to obtain reliable predictions of variations of these data and process complications aiming at further optimization of roller drilling and its improvement in mining, oil and gas, and construction industries.

ACKNOWLEDGEMENT

This article was prepared as part of a research carried out with the financial support of the Russian science Foundation according to the research project No. 19-71-00028 within the framework of the Competition of 2019 "Conducting initiative research by young scientists".

REFERENCES

1. F.J. Mohammed. Using big data technologies in petroleum industry. AIP Conf. Proceed. 2207(1), 2020.
2. J. Johnston, A. Guichard. New Findings in Drilling and Wells using Big Data Analytics. Offshore Technology Conf., 2015.
<https://doi.org/10.4043/26021-MS>
3. P. Dittrick. Data analytics optimize drilling technology. Oil&Gas Journal vol. 114 no. 11, 2016.
4. [4] A.A. Shigina, A.O. Shigin, A.A. Stupina. System of indicators for estimating the efficiency of boring rigs. Global Science and Innovation Materials of the I International Conf., pp 471-480, 2013.
5. A.A. Stupina, A.A. Shigina, M.V. Karaseva, L.N. Korpacheva, A.O. Shigin. Control by technological mode parameters with an intellectual automated system. IOP Conf. Series: Materials Science and Engineering, 2016.
6. A.A. Stupina, A.A. Shigina, A.O. Shigin. Mathematical formulation of technological processes optimization problem. IOP Conf. Series: Materials Science and Engineering, "International Scientific and Research Conf. on Topical Issues in Aeronautics and Astronautics", 2015.
<https://doi.org/10.1088/1757-899X/94/1/012001>
7. S. Ferreiro, R. Arana, G. Aizpurua, G. Aramendi, A. Arnaiz, B. Sierra. Data mining for burr detection (in the drilling process). International Work-Conf. on Artificial Neural Networks IWANN, pp. 1264-1273, 2009.
8. Q. Yin, J. Yang, B. Zhou, M. Jiang, X. Chen, C. Fu, L. Yan, L. Li, Y. Li, Z. Liu. Improve the Drilling Operations Efficiency by the Big Data Mining of Real-Time Logging. SPE/IADC Middle East Drilling Technology Conf. and Exhibition, 2018.
9. G.A. Saptawati, G.N. Nata. Knowledge discovery on drilling data to predict potential gold deposit. International Conf. on Data and Software Engineering (ICoDSE), 2015.
10. L. Bellatreche, S. Chakravarthy. A special issue in extending data warehouses to big data analytics. Distributed and Parallel Databases vol. 37, pp. 323-327, 2019.
<https://doi.org/10.1007/s10619-019-07262-1>
11. M. Isqour, K. Matouk, M.L. Owoc. A survey of data warehouse architectures – Preliminary results. Computer Science and Information Systems (FedCSIS), 2012.
12. A.A. Shigina, A.A. Stupina, N.N. Dzhiyeva, V.S. Tynchenko, O.A. Antamoshkin, V.V. Kukartsev. Adaptive model of the control of roller bit drilling process. Journal of Physics: Conf. Series, 2019.
13. R. Marinov. New Intelligent Technologies – Interactivity and Information Issues. Advances in Journalism and Communication vol. 7, pp. 94-108, 2019.
14. W. Fritz, R.G. Martínez, J.R. Blanqué, R. Adobbati, A. Rama, M. Sarno. The autonomous intelligent system (Robotics and Autonomous Systems) vol. 5 no. 2, pp. 109-25, 1989.
[https://doi.org/10.1016/0921-8890\(89\)90004-3](https://doi.org/10.1016/0921-8890(89)90004-3)
15. Y.B. Zhao, Z. Man, J. Kim, C.Q. Ma. Using Communication Networks in Control Systems: The Theoretical and Practical Challenges. Journal of Control Science and Engineering, 2018.
16. D. Fisch, M. Jänicke, E. Kalkowski, B. Sick. Learning from others: Exchange of classification rules in intelligent distributed systems. Artificial Intelligence vol. 187-188, pp 90-114, 2012.
<https://doi.org/10.1016/j.artint.2012.04.002>
17. A.O. Shigin, A.V. Gilyov, A.A. Shigina. Automation of rotary blasthole drilling in open pit mines. Gornyi Zhurnal vol. 2, pp. 79-82, 2017.
18. M. Sorg, J. Osmer, A. Fischer. Methodical Approach for Determining the Length of Drill Channels in Osteosynthesis. Sensors (Basel) vol. 19 no. 16, pp. 3532-3541, 2019.
19. E.A. Jantunen. Summary of methods applied to tool condition monitoring in drilling. International Journal of Machine Tools and Manufacture vol. 42 no. 9, pp. 997-1010, 2002.
20. J. He, M. Luo. Drilling Rig Hoisting Platform Security Monitoring System Design and Application. Machines vol. 5 no. 19, 2017.

21. J.Choia, J. Demmel, I. Dhillon, J. Dongarra, S. Ostrouchov, A. Petitet, K. Stanley, D. Walker, R. C. Whaley. 1996 ScaLAPACK: a portable linear algebra library for distributed memory computers – design issues and performance. Computer Physics Communications, High-Performance Computing in Science vol 97 no 1-2, pp. 1-15, 1996.
[https://doi.org/10.1016/0010-4655\(96\)00017-3](https://doi.org/10.1016/0010-4655(96)00017-3)