sciendo

RIGA TECHNICAL
UNIVERSITY

Suitability Determination of Machine Learning Techniques for the Operational Quality Assessment of Geophysical Survey Results

Kirill Abramov[1][*], Janis Grundspenkis[2]
[1]*Branch Office "Geotehnocenter" of JSC Volcovgeology, Almaty, Kazakhstan*
[2]*Riga Technical University, Riga, Latvia*

Abstract – **Well logging, also known as a geophysical survey, is one of the main components of a nuclear fuel cycle. This survey follows directly after the drilling process, and the operational quality assessment of its results is a very serious problem. Any mistake in this survey can lead to the culling of the whole well. This paper examines the feasibility of applying machine learning techniques to quickly assess the well logging quality results. The studies were carried out by a reference well modelling for the selected uranium deposit of the Republic of Kazakhstan and further comparing it with the results of geophysical surveys recorded earlier. The parameters of the geophysical methods and the comparison rules for them were formulated after the reference well modelling process. The classification trees and the artificial neural networks were used during the research process and the results obtained for both methods were compared with each other. The results of this paper may be useful to the enterprises engaged in the geophysical well surveys and data processing obtained during the logging process.**

Keywords – **Classification trees, machine learning, neural networks, well logging.**

I. INTRODUCTION

Uranium energy is one of the most advanced, efficient and environmentally-friendly types of energy today. There is no need to burn tons of coal or million cubic meters of gas for its operation. The enriched uranium ore is fuel for uranium energy. This ore is extracted with the help of in situ leaching methods. The uranium ore is extracted by this method from the ore deposits of the Republic of Kazakhstan by the National Atomic Company Kazatomprom (NAC Kazatomprom). The NAC Kazatomprom is the world's largest uranium company, the national operator of the Republic of Kazakhstan for the exploration, production, processing and export of uranium and its compounds, rare metals, nuclear fuel for nuclear power plants, special equipment, technologies and materials of dual-purpose [1]. The NAC Kazatomprom has a full nuclear fuel cycle – the entire sequence of recurring production processes, from fuel production (including electricity generation) to radioactive waste disposal [2].

The geophysical service, for which one of the authors of this paper is working, has one of the key places in the nuclear fuel cycle. The geophysical service follows directly after the drilling one, which is engaged in the construction of the wells. The geophysical surveys allow for determination of the location of ore bodies and the lithological composition of rocks in which these ore bodies lie. However, the operational assessment of the geophysical study results immediately after logging is carried out only with the help of tacit knowledge – the experience of the logging station operator. In other words, the operational assessment of the quality of the geophysical surveys depends on the so-called "human factor" and all the power of modern methods of data processing is not used.

That is why the main aim of the present research is to study the feasibility of applying machine learning techniques to model a reference well for the ore deposits and to perform operational quality assessment of well logging results.

There are many methods of machine learning [3]. These methods have been successfully applied to solve problems of interpreting geophysical data [4]–[7]. However, to assess the quality of these data themselves, machine learning methods have not been applied yet.

Since this article shows only the first attempt to study the possibility of using machine learning methods for operational quality assessment of the geophysical surveys, it describes the application of the most common and simple techniques – classification trees and neural networks.

As the study of the machine learning methods applicability presupposes the existence of a reference well, the initial task that needs to be accomplished to achieve the aim of the research is the construction of the reference well. The term "reference well" that is used here means a well according to which further comparison of the logging results would be made.

Next tasks that need to be accomplished in the framework of the research are the formation of data sets for usage in the machine learning process and the formulation of rules for the operational quality assessment of the results of the geophysical surveys.

---

* Corresponding author's email: abramc@mail.ru

The final task that must be accomplished is the construction of the models, which are planned to be used in the study process.

The comparison of the results gained with the help of these models will lead to a conclusion about the possibility of applying the machine learning methods for the operational quality assessment of the geophysical study results.

The paper is organised as follows. The introduction describes the current state of the problem of operational quality assessment of the geophysical surveys and proposes examining the possibility of using machine learning methods to solve it. The next section introduces the methods of geophysical measurements, the results of which are planned to be applied to solve the indicated problem. Then, the authors formulate the problem of operational quality assessment of the geophysical surveys, which is planned to be solved with the help of the applied machine learning methods. The subsequent section describes the results and their comparison. The conclusion presents the obtained results of the study.

II. Geophysical Methods of the Earth's Crust Study

Geophysical methods of the Earth's crust study, which have different names such as applied or exploratory well logging or just logging, are the area of applied geophysics that studies the Earth's crust ranging from 35 km to 70 km on land and 5 km up to 10 km at the seabed [8].

The main aim of geophysics is information restoration about the composition, structure, and history of the objects of the crust. This information can be obtained using indirect information about the various physical fields in the Earth`s crust [8].

The geophysical studies of the crust are produced by indirect studying of the Earth's physical fields. Formally, such studies are reduced to the detection and evaluation of geological objects for geometry and physical properties to determine their nature [9].

The geophysical studies of the crust divide into some technological methods according to the method of work. There are the following methods of geophysical studies: remote studies, land or sea studies, and geophysical well logging surveys (GWLs) [8].

GWLs occupy a special place in geophysics. They are characterised by specialised equipment and technology. In addition, GWLs are of great practical importance in the documentation sections of wells in the Earth's crust. These methods are referred to as operational well logging [8].

The possibility of geophysical studies is based on various physical fields of the Earth. This possibility is determined by the fact that the distribution of different parameters in the Earth's crust depends not only on the natural origin but also on the lithologic and geometric inhomogeneities of the Earth's crust, which create these anomalous fields. The anomaly in geophysics is the deviation of the measured field parameter from the normal one. The normal parameter is a parameter for which the field over a homogeneous half-space most often takes place. Anomalies appear because the search object, called the source of anomalies or disturbances, either creates the field itself for natural reasons, for example, a natural constant electric field or distorts the field due to the differences in physical properties, for example, a reflection of seismic or electromagnetic waves from contacts of different rocks. The contrast of physical properties, the relative depth of the object, and the level of interference determine the intensity of anomalies [9].

With the help of the measurement of the physical parameter inside the systems of usually parallel profiles and revealing anomalies, one can judge about the properties of the rocks and obtain information about the geological structure of the investigated rock arrays [8].

A. The GWL Methods Used in the Uranium Deposits of Kazakhstan

Uranium is one of the most important energy carriers now. The uranium deposits are the main source of uranium ore to meet the growing needs of nuclear energy. These deposits are the natural accumulation of uranium ores in such quantities and mineral forms that make their commercial mining economically feasible [10].

There are two main groups of deposits, considering the structural environments of uranium mineralization, geochemical features of migration and uranium content [11]:

- the endogenous group (hydrothermal, hydrothermal-metasomatic) of deposits with the zones of age-related tectonomagmatic activation associated with hydrothermal processes (alkaline or acid metasomatism) [11];
- the exogenous group, including infiltration (sandstone – according to the IAEA classification) of deposits that resulted from the activity of underground (reservoir and groundwater) waters in the sedimentary sequences of activated depressive structures of the Earth's crust (platform covers, paleolines, orogenic depressions) [11].

Specific features of the infiltration deposit formation predetermine their radiological situation and elemental composition. In this case, a lack of radium is observed in the ore deposit in comparison with the equilibrium state, and in its frame, the radioactive equilibrium is violated toward the excess of radium, which is the result of the formation of the so-called "residual" and "diffusion" areas of radium [12].

The GWL on uranium deposits, especially in conditions where it is difficult to perform the core output when drilling wells, is practically the main supplier of measurement information about the composition and properties of the productive rock horizons and the parameters of uranium-ore intervals at all stages of work [13].

These studies are used to solve technologically (determining the parameters of ore intervals and assessing the properties of the productive rock horizons), technical (evaluation of the state of the well) and ecological problems (control over the state of the environment and study of the dynamics of recovery of the environment after the completion of the process). The complex of GWLs, their types, and frequency depend on the tasks that need to be addressed at each stage of the work, as well as the conditions for conducting them [11].

B. The Gamma Logging Method

The mandatory logging methods at all stages of work in uranium fields are gamma logging (GK), caliper logging (KM), and standard electrical logging.

The GK method is the main method of the uranium ore geophysical study. The thickness of the ore interval and the mass fraction of uranium content in it are the main parameters of uranium ore intervals, determined by it. These parameters are used for the calculation and accounting of uranium reserves in the subsoil at the stages of prospecting, exploration, and exploitation of uranium deposits [10].

The GK is based on the fact that different rocks have different degrees of natural radioactivity [14].

Natural radioactivity of the rocks depends mainly on the content of radioactive elements of the uranium-thorium series and radioactive potassium in the rocks. The content of other radioactive elements in rocks is negligible and amounts to less than 1 % of the total natural radioactivity. That is why the radioactivity of these elements can be neglected when interpreting the results of gamma logging for the considered purposes. The natural radioactive elements emit alpha, beta, and gamma radiation during decay. However, the gamma radiation of these elements is interesting for engineering geology and hydrogeology [15].

The counting rate of pulses, expressed in terms of the exposure dose rate of gamma radiation from a point source of 226 Ra, is the measurement information for the GK. The separation of ore intervals and the determination of uranium contents in them are carried out as a result of processing (interpretation) of log data [11].

Since the main sources of gamma radiation in the uranium series are the short-lived products of the radon decay, the measurement result for the GK will depend not only on the uranium content in the ore interval but also on the degree of preservation of the radioactive equilibrium state between the maternal and subsidiary radionuclides that form this system. Therefore, the most correct results of the GK are provided in the study of uranium ores with undisturbed radioactive equilibrium, to which mainly endogenous (hydrothermal) type deposits belong [11].

The exogenous deposits (the infiltration deposits or the uranium deposits of sandstone type according to international classification) are characterised by a violation in the radioactive equilibrium between radium and uranium and sometimes between radon and radium, which significantly limits the possibilities of the GK method. For the uranium deposits of the infiltration type, the mining of which is carried out mainly by the in situ leaching method (ILM), it is necessary to have a complete understanding of the lithologic composition of the productive horizon rocks for the optimal organisation of the in situ leaching process. The information about the physical properties of the ore intervals is currently obtained mainly from the results of standard electric logging using apparent resistance and/or spontaneous polarization methods. In this case, the subject of detailed studies of these deposits is not only ore but the entire productive horizon, as well as water imperatives –

impermeable rocks that overlap and underlie the productive horizon [16].

The interpretation of the GK results is to determine the parameters of the uranium-ore intervals (power and mass fraction). It is based on the connection between the shape and intensity of the gamma anomaly over the ore interval with the mass fraction [10].

The digital well log data file used for interpretation, which is expressed by its values with the quantization step, should be expressed in units of µR/h of the point source radium and corrected for depth, taking into account the points recorded during the logging process [11].

To recalculate the values measured in the GK to the mass fraction values, it is necessary to know the conversion factor and the corrections for the absorption of gamma radiation by washing liquid and casing, the violation of radioactive equilibrium between radium and uranium, and between radon and radium, and the presence of thorium and potassium in ores [12].

C. Electrical Logging Methods

Electrical logging is the most common method for the study of wells. In this method, a well device is lowered into the well, which allows determining the electrical properties of the rocks traversed by the well. The results of the measurements are represented in the form of electric logging diagrams. Deciphering them, it is possible to determine the depth of occurrence of permeable layers with high electrical resistance [17].

The electrical logging methods serve as a basis for the construction of the geoelectric sections. These methods are based on observing and studying the features of the distribution and characteristics of electromagnetic fields of natural and generated origin. These features are mainly due to the separation of rock from electromagnetic properties (electrical conductivity, electrical resistance, and inductance) by a variety of sources used in logging and recorded characteristics [12].

Resistivity logging (KS) is the most used method of electrical well logging. This logging method is based on the differences in the resistivity of different types of sedimentary rocks. Specific electrical resistance is the resistance of a homogeneous medium with a length of 1 meter and a cross-sectional area of 1 m^2. In the most general case, it is determined by the presence of natural water in the pore space [12].

The measurement of the parameters which characterise the lithological and filtration properties of productive rock horizons in infiltration-type deposits is based on these parameter relationships to the parameters that characterise their electrical properties, namely the magnitude of the potential of the induced polarization and the electrical resistance of rocks [12].

The processing and interpretation of electric logging results are carried out within the productive horizon, which is extracted from the reference horizons separated by correlation sections. The processing software must be set up in such a way as to distinguish intervals with various electrical properties against the background of interference due to hardware fluctuation and the inhomogeneity of the measured electric field.

The quality of the primary materials is assessed by comparing the main and control logs. The volume and requirements for the performance of the control electrical logging are determined by the volume and requirements when performing the GK. The discrepancy between the results of the main and control measurements of the signals of the KS calculated at intervals of not less than 1 m should not exceed 10 % [10].

The main part of geophysical research is carried out immediately after the new well construction. Based on the results of these studies, the geological service should determine where the productive horizons are located, what their filtration properties are, and how it is necessary to cope with the constructed well for its further use. Any inaccuracies in the results of these studies usually lead to the fact that the constructed well will be defective and it will be necessary to build a new one. That is why the operational quality assessment of the logging result is one of the most pressing problems in surveys of geophysical wells.
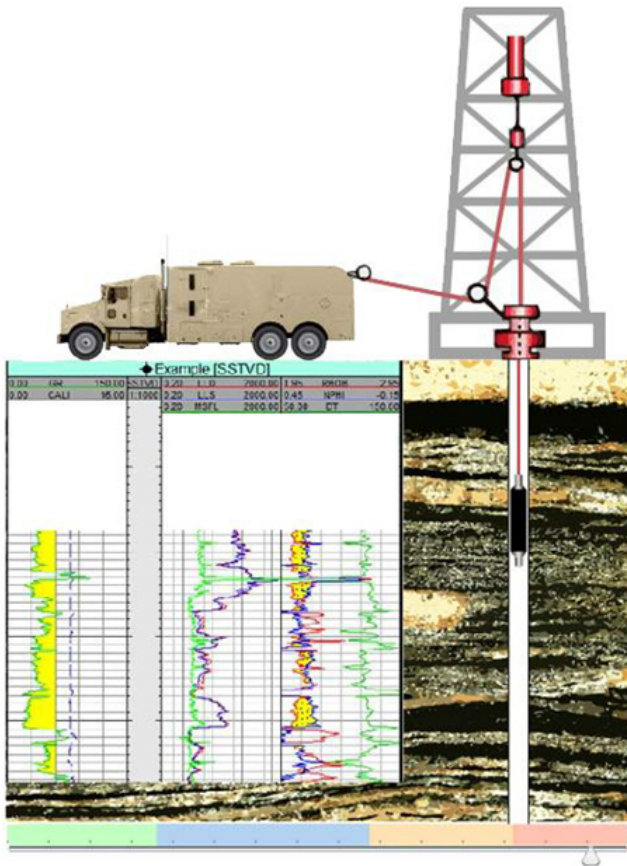

Fig. 1. Geophysical surveys.

The GK and the KS were used within the framework of the research. These logging methods are carried out by one well device made for each constructed well, and the main ones in the development of the infiltration type uranium deposits.

There is no way to simulate the geological section for each uranium deposits of the Republic of Kazakhstan. This is not possible because there are no modelled reference wells for each logging type that could be used for this purpose. Moreover, the

concept of a "reference well" can be very controversial. There is no sufficiently clear understanding of what a reference well is and how to obtain it. There are only logging data from past years that are stored in the database.

There is currently no way to automatically evaluate the quality of logging data. The main criterion for assessing the logging quality is still the experience of the operator who conducts these loggings. Thus, it can be argued that the rapid quality assessment of the logging result is highly dependent on the human factor.

The present study examines the possibility of using machine learning methods for the reference well modelling and for the rapid quality assessment of the logging result.

There are only typical sections for the values of the KS for many uranium deposits in Kazakhstan, but they cannot be used as reference wells. These typical sections are presented only in graphical form and their digitisation, on the one hand, takes a very long time, and on the other hand, a very small number of points are obtained as a result of such digitisation. Such incisions are also not available for the GK, since such studies have not been conducted for this logging method. Thus, it can be argued that there is no way to simulate the entire geologic section for each of the uranium deposits.

III. PROBLEM STATEMENT AND DESCRIPTION OF THE METHODOLOGIES TO SOLVE IT

The results of control measurements are the main ways for the GWL quality assessment. However, such measurements are not carried out for each well and constitute no more than 10 % of the total number of investigated wells [11].

The engineers with special software process the control measurements and the processed data are not available online.

Therefore, it turns out that such an important parameter as the operational quality assessment of logging results is completely dependent on implicit knowledge, an estimate that the logging station operator can provide in situ.

Within the framework of the research, the operational quality assessment is carried out by comparing the logging results with the reference well. Therefore, the first question that needs to be clarified is how this reference well should be formed.

No attempts have been taken earlier to simulate the reference well. The term "reference well" means here not only a longitudinal section, wherein the geological structure horizons are reflected (this section structure is obtained by the electric logging data) but also the location and power of the ore interval (this information is obtained according to the GK). The term "reference well" can be significantly expanded during further studies to include the results of other geophysical research methods.

The attempts to simulate a geological section based on the results of the electric logging data have been made earlier. The results of this simulation can be seen, for example, in the appendices [12]. However, these attempts have been reduced to finding the standard average value for several wells in each uranium deposit of South Kazakhstan.

The attempts to simulate the location and thickness of ore intervals over the uranium deposits have also been undertaken

earlier. The results can be seen in [18]. However, such ore interval simulation has been carried out for geological purposes and these models cannot be used for comparison with new geophysical studies.

The idea that underlies how it is necessary to conduct the rapid quality assessment of logging results is shown in the figure below.
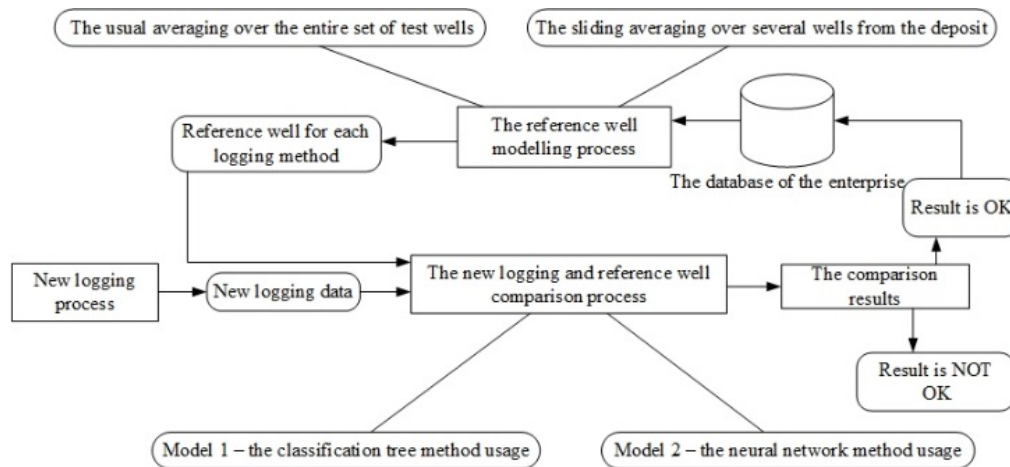

Fig. 2. The scheme of the rapid quality assessment of logging results.

This idea has been proposed by the geophysical service and realised by one of the authors of this paper.

According to this idea, each new logging done by the logging station operator is compared with the reference well for this type of geophysical study, and a conclusion is made about how likely the completed logging is a mistake.

Two basic questions must be solved for the operational quality assessment of logging results (see Fig. 2). The first question – How the reference well for the deposit is formed? The second question – What rules and parameters should be used for the comparison of geophysical studies?

Reference well forming is one of the key issues addressed in the present study. There are the following models proposed to organise the construction of a reference well.

- Usual averaging over the entire set of test wells is proposed to be used as the basic model for each type of geophysical studies. The idea of the geological section simulation for each deposit is used here. The only difference is to make such averaging not only for the KS but also for the GK method.
- Sliding averaging over several wells from the deposit is proposed in addition to the basic model. This will allow operating with several average curves for the deposit and will lead to an increased inaccuracy.

The curves of a reference well for each logging type obtained at this stage of the research can be used in further studies. These studies involve the training sample formation and their usage with machine learning algorithms. The correlation coefficients with the reference wells have been calculated for each test curve. These correlation coefficients are one of the key parameters for which the new curve will be assessed as a defect one.

The next question to be clarified in this paper is the definition of the necessary parameters that can be used for further research.

The dependencies of the rock filtration coefficients on the magnitude of the apparent electrical resistance of the formation are also determined. It should also be mentioned that all values of such dependencies for the KS are reduced to a single standard recording level for each of the deposits. These data are calculated, approved, and mandatory for use [18].

This information allows formulating the parameters that must be taken into account for further research. The parameters for each type of geophysical studies used in this paper are given below:

- the parameters used for electric logging are the depth, the resistivity (pk) value, the resistivity value averaged over the horizon, the correlation coefficient with the standard well for the electrical logging method, the defect classifier;
- the parameters used for the gamma logging are the depth, the intensity, the location of the gamma anomaly in the general ore interval, correlation coefficient with the standard well for the gamma logging method, the defect classifier.

The finding of each reference well and the identification of research parameters with which it is necessary to generate samples of geophysical survey data identification are preparatory steps for the present research.

The main part of the research is the application of the machine learning methods to the operational quality assessment of geophysical study results.

According to the definition of big data [19], [20], it can be argued that geophysical studies, which consider dozens of different parameters accumulated over several decades, are a classic example of such "big data", for the processing which the special machine learning and data mining methods have been developed [21]–[23]. Therefore, the machine learning methods have been chosen as a tool for achieving the aim of the present research.

The following models are proposed to study the possibility of the operational quality assessment of geophysical study results:

- the first model involves the usage of classification tree algorithms for the defect prediction;
- the second model involves the usage of neural network algorithms for the defect prediction.

The unified sample set of 220 wells from one of the uranium deposits in South Kazakhstan has been used for the classification trees and the artificial neural networks within the framework of the research.

This sample set for the classification trees has been divided into the training sample set consisting of 180 wells; the test sample set consisting of 20 wells, and the remaining 20 wells have been used as data for verification.

For the artificial neural networks, the sample set has been divided as follows: 70 % of total data represent the training set (154 wells), 15 % – test set (33 wells), and 15 % (33 wells) – verification data.

The method of classification trees allows predicting whether the observations or objects belong to a class of the categorical dependent variable. This classification is made in relation to the corresponding values of one or more predictor variables.

The method of classification tree has the following structure in this paper. Each logging curve for resistivity (KS) and gamma (GK) logging has been considered separately. The studies have been carried out using different split selection methods, prior probabilities, and goodness of fit, for each examined curve. In addition, a different number of counted parameters have been considered for each logging curve. A FACT-style direct stopping has been used as a stopping option for the classification tree creation. The parameters of the classification tree method used in this paper for both resistivity (KS) and gamma (GK) logging are presented in Table I.

TABLE I

THE PARAMETERS OF THE CLASSIFICATION TREE ALGORITHM

A	Split selection method	Discriminant based univariant split for categorical and ordered predictors
	Prior probabilities	Estimated
	Stopping options	FACT-style direct stopping
B	Split selection method	Discriminant based univariant split for categorical and ordered predictors
	Prior probabilities	Equal
	Stopping options	FACT-style direct stopping
C	Split selection method	C&RT-style exhaustive search for univariant splits
	Goodness of fit	Gini measure
	Prior probabilities	Estimated
	Stopping options	FACT-style direct stopping
D	Split selection method	C&RT-style exhaustive search for univariant splits
	Goodness of fit	G-square
	Prior probabilities	Estimated
	Stopping options	FACT-style direct stopping
E	Split selection method	C&RT-style exhaustive search for univariant splits
	Goodness of fit	Gini measure
	Prior probabilities	Equal
	Stopping options	FACT-style direct stopping
F	Split selection method	C&RT-style exhaustive search for univariant splits
	Goodness of fit	Chi-square
	Prior probabilities	Equal
	Stopping options	FACT-style direct stopping
G	Split selection method	C&RT-style exhaustive search for univariant splits
	Goodness of fit	G-square
	Prior probabilities	Equal
	Stopping options	FACT-style direct stopping

The parameter sets listed in Table I are indicated to those depicted in Fig. 4 and 6 in the form of letter designations (A, ..., G).

A multilayer perceptron (MLP), used in this paper, is a feedforward neural network architecture with unidirectional full connections between successive layers. However, this does not uniquely determine the property of a network. In addition to network architectures, the neurons of a network have activation functions that transform the incoming signals from the neurons of the previous layer using a mathematical function. The type of this function is the activation function itself and can profoundly influence the performance of the network. Thus, it is important to choose a type of activation function for the neurons of a neural network. The softmax activation function mainly applied to classification tasks has been used in this paper. This function is useful for constructing neural networks with normalized multiple outputs, which makes it particularly suitable for creating neural network classifiers with probabilistic outputs. The general view of the softmax function is the following:

$$\frac{\exp(a_i)}{\sum \exp(a_i)},$$

where a is the weighted sum of neuron inputs.

The number of neurons in the hidden layer measures the complexity of a neural network. The more neurons in a neural network, the larger the flexibility and complexity of the system. The MLP with the number of hidden layers 15 and the number of neurons in it from 5 to 14 has been used for the GK curve in this paper. For the KS curve – the MLP with the number of hidden layers 18 and the number of neurons in it from 5 to 16.

IV. THE STUDY OF THE SUITABILITY OF MACHINE LEARNING TECHNIQUES TO ASSESS THE QUALITY OF WELL LOGGING RESULTS

A. Reference Well Modelling and Determination of the Significant Parameter

The results of the first model study are the standard wells curves for the methods of the KS and the GK logging obtained by means of the standard average method.

The results of the first model studies (using routine averaging over the entire set of selected test wells) can be seen in Fig. 3.
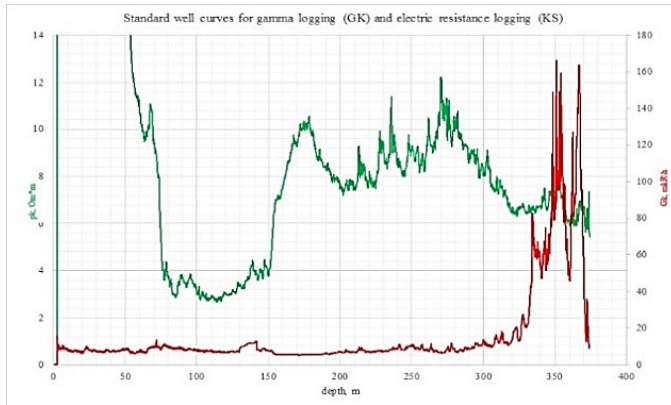
Fig. 3. Standard well curve for the KS and the GK loggings.

The reference well for the KS, obtained with the help of routine averaging over the entire set of test wells, corresponds to an approximate geological section for the selected deposit and can be used for further research according to the geophysical service experts.

The location of the ore body in the standard well for the GK, obtained with the help of routine averaging over the entire set of test wells, corresponds to the standard for the selected deposit. However, the so-called "power" of the ore body, i.e., its thickness is overstated. Therefore, the location of the anomaly in the common ore interval has been chosen as one of the significant parameters for comparing the GK data with a reference well.

The studies of the second model (using sliding averaging over the entire set of selected test wells) have been carried out as the next step. The average curves for a different number of wells have been calculated here. Each average curve for the KS, obtained at this step, corresponds to an approximate geological section for the selected deposit and can be used for further research according to the geophysical service experts.

After consultations with the geophysical service experts and considering the results of the longstanding research carried out to determine the dependence of the rock filtration coefficients on the apparent electrical resistivity of the formation [15], the following parameters for the above-mentioned geophysical study methods have been used in this paper:

- there are the following parameters for the KS: well number, correlation coefficients with the reference well for this method, the value of the resistivity averaged over the horizon, the level of values for the reference clay horizon in the deposit, the fault classifier;
- there are the following parameters for the GK: well number, correlation coefficients with the reference well for this method, the anomaly location in the general ore interval, the nominal background values, the fault classifier.

B. Summary of the Methods of Classification Trees and Neuron Networks

The accuracy for the GK curve using the method of classification trees is provided in Table II.

TABLE II
ACCURACY FOR THE GK CURVE USING THE METHOD OF CLASSIFICATION TREES

The variable sets	Accuracy
The 1st variable set (correlation coefficients for standard averaging)	0.8
The 2nd variable set (the 1st variable set + background variable)	0.99
The 3rd variable set (correlation coefficients for sliding average)	0.6–0.8
The 4th variable set (the 3rd variable set + background variable)	0.99

The main contribution to the constructed tree has the parameter "in range" that corresponds to the presence of a gamma anomaly in the general interval obtained by averaging.
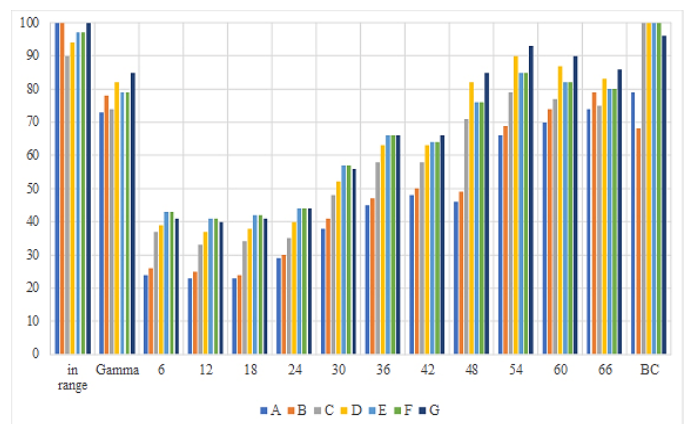

Fig. 4. Predictor variable importance rankings for the GK curve.

The resulting classification tree for the GK curve, including the background variable, is presented in Fig. 5.
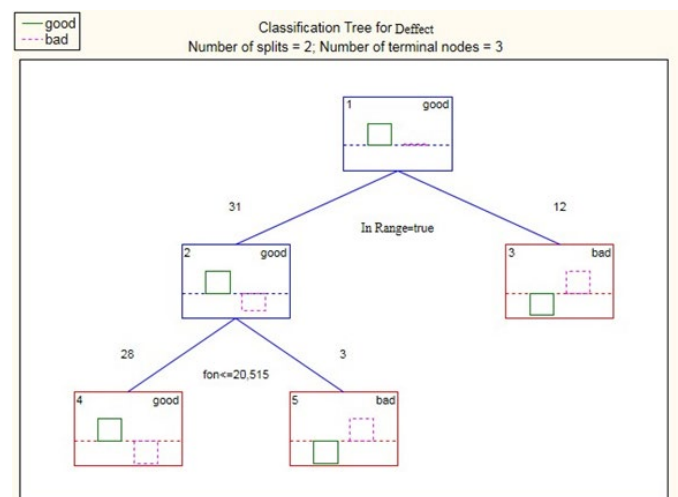

Fig. 5. Resulting classification tree for the GK curve.

The accuracy for the KS curve using the method of classification trees is provided in Table III.

The variable sets	Accuracy
The 1st variable set (correlation coefficients for standard average)	0.5–0.66
The 2nd variable set (the 1st variable set + the clay line (minimal value))	0.9
The 3rd variable set (the 1st variable set + the real average for the horizon)	0.8
The 4th variable set (correlation coefficients for sliding average)	0.6
The 5th variable set (the 4th variable set + the clay line (minimal value))	0.99
The 6th variable set (the 4th variable set + the real average for the horizon)	0.8-0.9
The 7th variable set (the 4th variable set + the clay line (minimal value) and the real average for the horizon)	1.0

The most significant contribution to the classification tree is made by the "clay line" (minimal value) parameter and the real average for the reference horizon parameter. The absence of these parameters can decrease the accuracy of classification trees by 0.5.
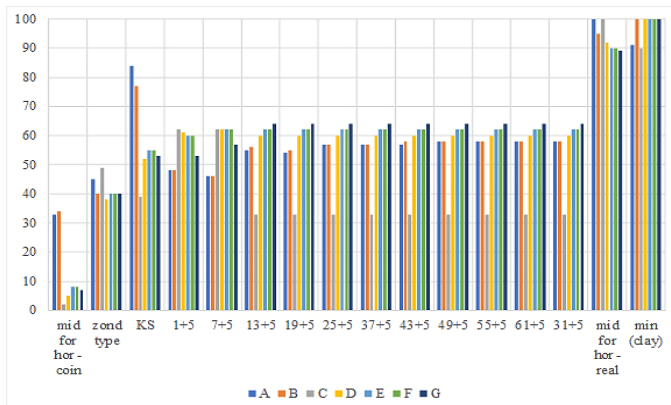

Fig. 6. Predictor variable importance rankings for the KS curve.

Figure 7 presents the resulting classification tree for the KS curve, including the clay line (minimal value) and the real average for the reference horizon variables.
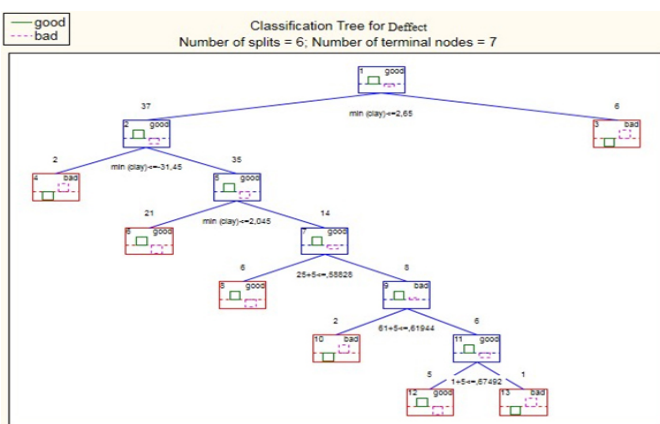

Fig. 7. Resulting classification tree for the KS curve.

The total accuracy of real and predicted results for the GK curve in the case of using the method of artificial neural networks is summarised in Table IV.

The variable sets	Accuracy
The 1st variable set (correlation coefficients for standard averaging)	0.88–0.92
The 2nd variable set (the 1st variable set + background variable)	0.9–1.0
The 3rd variable set (correlation coefficients for sliding average)	0.92
The 4th variable set (the 2nd variable set + background variable)	0.86–0.92

It is possible to organise the verification of solutions for the GK curve using the MLP algorithm for the 2nd set of parameters.

The accuracy of real and predicted results for the KS curve in the case of using the method of artificial neural networks is shown in Table V.

The variable sets	Accuracy
The 1st variable set (correlation coefficients for standard average)	0.68–0.81
The 2nd variable set (the 1st variable set + the clay line (minimal value))	0.84
The 3rd variable set (the 1st variable set + the real average for the horizon)	0.84
The 4th variable set (correlation coefficients for sliding average)	0.84
The 5th variable set (the 2nd variable set + the clay line (minimal value))	0.78–0.81
The 6th variable set (the 2nd variable set + the real average for the horizon)	0.81
The 7th variable set (the 2nd variable set + the clay line and the real average for the horizon)	0.85

The usage of the neural network method to predict the values of the KS curve involves certain difficulties. It is necessary to organise a rather large sample dataset for determination of more accurate results. The best thing to achieve is accuracy at level 0.85, which is not bad, but significantly lower than the accuracy in the case of classification trees (0.99).

C. The Comparison of the Results of Classification Trees and Neural Networks

The results of the methods of classification trees and neural networks are demonstrated in the Figs. 8 and 9.
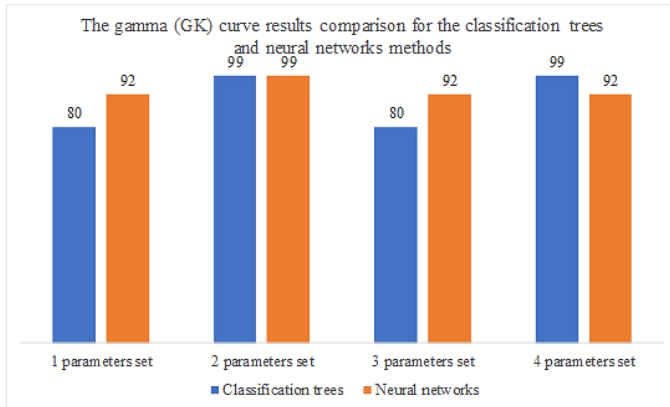
Fig. 8. The comparison of the GK curve results for the classification trees and neural networks.

The results for the GK curve predicted by the method of neural networks are better for the 1st and 3rd sets of parameters, as it can be seen from the figure above. For the 4th set of parameters, the best accuracy has been shown by the method of decision trees. There is equal accuracy provided by classification trees and neural networks for the 2nd set of parameters for the GK curve.

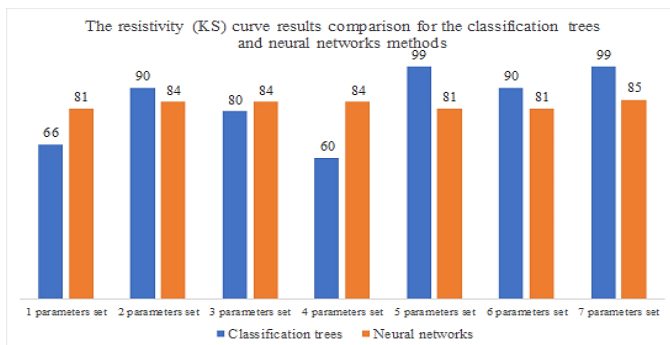The comparison results for the KS curve are provided in Fig. 9.

Fig. 9. The comparison of KS curve results provided by the classification trees and neural networks.

The results for the KS curve predicted by the method of neural networks are better for the 1st, 3rd, and 4th sets of parameters. The results for the 2nd, 5th, 6th, and 7th sets of parameters are better using the method of decision trees, as it can be seen from the figure above.

It is worth noting that the usage of the classification trees is preferable for those cases where the so-called "limiting parameters", i.e., either the "background" value for the GK curve or the "clay line" and the "real average for the reference horizon" for the KS curve, are used.

As can be seen from the figures above, for the GK curve, if the "background" variable is used, the accuracy of the results by the method of classification trees is better than that by the neural networks (0.99 vs. 0.92). In the case that the "background" variable in the data set is not used, then the accuracy is better for neural networks (0.92 vs. 0.80).

The same is observed for the KS curve. When using the "clay line" and the "real average for the reference horizon" variables, the accuracy by the method of classification trees is better than that by the neural networks (0.90–0.99 vs. 0.81–0.85). If none of these parameters is used, then accuracy by the method of neural networks is much worse (0.8 vs. 0.6).

V. Conclusion

The main aim of this paper has been to study the possibility of applying machine learning techniques for the operational assessment of the geophysical research quality.

The idea of how to carry such an assessment has been proposed by the geophysical service of NAC Kazatomprom and realised by one of the authors of this study. This idea is that for such kind of assessment it is necessary to compare the new logs with a certain well that will be a reference for the deposit and give an opinion on the probability of the fact that the new log is incorrect.

There are some questions that need to be addressed to clarify this idea. The first question is "How the reference well for the deposit is formed?" The second one – "What rules and parameters should be used for the comparison of geophysical studies?"

The present study has been performed in three stages. The first stage has been devoted to the reference well modelling for the KS and the GK curves. The second stage has allowed for the determination of the significant parameters by which the results of geophysical studies should be rejected. In addition, the third stage has been concerned with the possibility of applying machine learning techniques for the operational quality assessment of the geophysical research results.

The main conclusion that can be drawn is that machine learning techniques can be used to a full extent for operational quality assessment of well logging results.

The study of all models has shown that it is necessary to more carefully formulate the comparison rules and criteria for defecting the wells. These issues need to be addressed in very close interaction with the geophysical service because the employees of this service are the main stakeholders in these studies. In addition, the study has demonstrated that it is necessary to involve more complex algorithms for in-depth investigation of the issue.

The following areas can be recommended for further research. To improve the reference curve quality, especially for the GK logging, it may be proposed in future research to increase the number of the test data sets for which averaging should be performed. That is the first area.

Secondly, instead of directly increasing the number of wells for samples, it is possible to suggest including other logging methods – caliper, induction, and neutron logs. These logs could be examined more comprehensively both in the interaction and in impact on each other. It will also be necessary to determine the significant parameters for which to study these logging methods.

Finally, one can suggest using other machine learning techniques to consider the test models.

R EFERENCES

[1] Kazatomprom, Website of "Kazatomprom" National atomic company. Available at: http://www.kazatomprom.kz/en

[2] M. Dzhakishev, Chto takoe yaderno-toplivnyj cikl", Aug. 04, 2009. Available at: http://www.liveinternet.ru/users/3362230/post107753670/

[3] R. I. Muhamedyev, "Machine learning methods: An overview," *Computer modelling & new technologies,* vol. 19, no. 6, pp. 14–29, 2015.

[4] A. F. Kobussen, P. D. Agnew, and G. Broadbent, "Application of Machine Learning Techniques to Exploration: An Example Using Self-Organising Maps for Garnet Data," *11th International Kimberlite Conference*, Extended Abstract No. 11IKC-4917, 2017.

[5] A. Varley, A. Tyler, L. Smith, and P. Dale, "Development of a neural network approach to characterise ^{226}Ra contamination at legacy sites using gamma-ray spectra taken from boreholes," *Journal of Environmental Radioactivity*, vol. 140, pp. 130–140, 2015.
https://doi.org/10.1016/j.jenvrad.2014.11.011

[6] K. Heibig and S. Treitel (Eds). *Handbook of geophysical exploration. Seismic exploration.* Elsevier B.V., 2010.

[7] Y. Kuchin and J. Grundspeņķis, "Machine Learning Methods for Identifying Composition of Uranium Deposits in Kazakhstan," *Applied Computer Systems*, December 2017, vol. 22, no. 1, pp. 21–27.
https://doi.org/10.1515/acss-2017-0014

[8] V. K. Hmelevskoj, "Geofizicheskie metody issledovaniya zemnoj kory" Chast 1. Mezhdunarodnyj universitet prirody, obshhestva i cheloveka, Dubna, 1997

[9] M. V. Kalinnikova, B. A. Golovin, and K. B. Golovin, "Uchebnoe posobie po geofizicheskim issledovaniyam skvazhin", Saratov, 2005.

[10] ZAO NAC "Kazatomprom", TOO "IVT", "Metodicheskie rekomendacii po kompleksu geofizicheskih metodov issledovaniya skvazhin pri podzemnom vyshhelachivanii urana", Almaty, 2003.

[11] ZAO NAC "Kazatomprom", TOO "IVT", "Instrukciya po gamma karotazhu pri podgotovke k ekspluatacii i ekspluatacii plastovo infiltracionnyh mestorozhdenij urana", Almaty, 2009.

[12] TOO "Gornorudnaya kompaniya", TOO "Geotehnoservis", "Tehnicheskaya instrukciya po provedeniyu issledovanij v skvazhinah na plastovo infiltracionnyh mestorozhdeniyah urana", Almaty, 2010.

[13] MPR RF, "Metodicheskie rekomendacii po geofizicheskomu oprobovaniyu pri podschete zapasov mestorozhdenij metallov i nerudnogo syrya". Oct. 13, 2007, № 37.

[14] V. M. Muravev, "Spravochnik mastera po dobyche nefti", Moskva: Nedra, 1975.

[15] V. I. Ferronskij, "Radioizotopnye metody issledovaniya v inzhenernoj geologii i gidrogeologii", 1972.

[16] N. I. Buyalov, "Poiski i razvedka neftyanyh i gazovyh mestorozhdenij", Moskva: Gostoptehizdat, 1960.

[17] A. A. Korshak, "Osnovy neftegazovogo dela", Ufa: DizajnPoligrafServis, 2005.

[18] AO NAC Kazatomprom, "Popravochnye koefficienty primenyaemye pri avtomatizirovannoj interpretacii gamma i elektrokarotazhej na mestorozhdeniyah AO NAC Kazatomprom", AO NAC Kazatomprom, Astana, 2014.

[19] Promdevelop Editors Team, "Big Data - chto takoe sistemy bolshih dannyh? Razvitie tehnologij Big Data", Oct. 5, 2017. Available at: https://promdevelop.ru/big-data/

[20] I. Chubukova, "Data mining" Binom. Laboratoriya znanij, 2008.

[21] A. M. Turing, "Computing machinery and intelligence," *Mind,* vol. LIX, no. 236, pp. 433–460, October 1950.
https://doi.org/10.1093/mind/LIX.236.433

[22] V. Dyuk, and A. Samojlenko, "Data mining: uchebnyj kurs", Spb: Piter, 2001.

[23] Z. Markov and D. T. Larose. *Data-mining the Web: uncovering patterns in Web content, structure, and usage*. John Wiley & Sons Inc., 2006.
https://doi.org/10.1002/0470108096

Kirill Abramov received the Master degree in Computer Systems from Riga Technical University in 2018. Since 2019, he has been a Doctoral student at Riga Technical University. Since 2005, he has been working at the Branch Office "Geotechnocenter" of JSC Volkovgeology, Kazakhstan. He is the Leader of the Software Development Group for registration of geophysical data. Research interests include data mining, machine learning, artificial intelligence and well logging surveys.
E-mail: abramc@mail.ru
ORCID iD: https://orcid.org/0000-0002-0646-3714

Janis Grundspenkis received the qualification of Electrical Engineer in Automation and Telemechanics in 1965 and the degree of Candidate of Technical Sciences in 1972 (recognised in 1992 as the degree of Doctor of Engineering Sciences) from Riga Polytechnical Institute, Latvia. In 1993, he received the degree of Habilitated Doctor of Engineering Sciences from Riga Technical University, Latvia. Since 1994, he has been a Professor of Systems Theory at Riga Technical University. His research interests include agent based and multiagent intelligent systems, knowledge acquisition and representation, causal domain models for complex cyber-physical systems, and structural modelling. He is a full member of the Latvian Academy of Sciences, senior member of IEEE, and member of ACM.
E-mail: Janis.Grundspenkis@rtu.lv
ORCID ID: https://orcid.org/0000-0003-2526-4662