Please fill in the name of the event you are preparing this manuscript for.	2021 SPE Annual Caspian Technical Conference
Please fill in your 6-digit SPE manuscript number.	SPE-207004-MS
Please fill in your manuscript title.	The Use of Neural Network Technologies in Prediction the Reservoir Properties of Unconsolidated Reservoir Rocks of Shallow Bitumen Deposits

Please fill in your author name(s) and company affiliation.

Given Name	Middle Name	Surname	Company
Marat	F	Validov	Kazan Federal University
Danis	K	Nurgaliev	Kazan Federal University
Vladislav	A	Sudakov	Kazan Federal University
Timur	A	Murtazin	Kazan Federal University
Kseniya	A	Golod	Kazan Federal University
Albina	R	Galimova	Kazan Federal University
Ruslan	R	Shamsiev	Kazan Federal University
Azat	A	Lutfullin	PJSC «Tatneft»
Marat	I	Amerhanov	PJSC «Tatneft»
Niyaz	A	Aslyamov	PJSC «Tatneft»

Abstract

In the conditions of the dynamically changing conjuncture of the oil and gas market, there is an urgent need to reduce the cost of oil production and increase the efficiency of development, this is especially important for the local ultra-viscous oil. In this regard, it is necessary to optimize costs at all stages, starting from the geological exploration and even at the stage of completion of the development process. For ultra-viscous oil deposits, this is especially relevant at the stage of assessing the resource potential of a separate uplift of any of the fields, when the only reliable way to perform a high-frequency section at shallow depths is to drill appraisal wells with full core sampling. An additional load is exerted by the period between core extraction and obtaining information about the flow properties of each of the samples. By themselves, standard core studies are complicated by the fact that sand rocks of weakly cemented bitumoids can often be destroyed during experiments.

In this regard, the use of new approaches, including digital ones, which allow us to make quick decisions on a part of the geological section in the area of the appraisal well and on the uplift as a whole, are highly in demand.

This article describes the methods that allow the determining of flow properties for uncemented (loose sands) rocks in Permian sediments. More than 25,000 core samples were studied from 805 wells at several fields of the Republic of Tatarstan.

The technology used allows us to calculate a continuous curve of volumetric bitumen saturation in the conditions of complete or partial absence of core at the well.

This paper presents the results of creating an algorithm for automatic prediction of weight bitumen saturation in a sand pack of the Sheshminsky horizon of the Permian system using neural network technologies, as well as using an alternative calculation method.

Introduction

In total, more than 30 ultra-viscous oil deposits have been explored in the Republic of Tatarstan, of which more than 20 deposits are in active development by the use of steam-assisted gravity drainage method. The main object of development in the fields of ultra-viscous oil is the sand pack of the Ufimian tier. As a result of geological and field studies of exploration wells, it was made known that oil/bitumen-saturated sandstones of the Ufimian tier are characterized by a number of geological and geophysical features (Khisamov et al., 2010).

The paleotectonic factor decisively controlled the differentiated nature of bitumen distribution and the scale of bitumen concentration. It also explains the features of their uneven spatial distribution by the presence of paleotraps. Most of the explored bitumen resources are concentrated on the eastern side of the Melekess depression and the western slope of the South Tatar Arch, but their maximum amount falls on specific regions.

The lithofacies factor has a great influence on the distribution of bitumen. The complex situation of sedimentation in the Permian period predetermined significant changes in the volume and composition of bitumen-containing complexes, screening thicknesses and reservoir properties of rocks. Delta formations of the Ufimian tier and shallow-sea (terrigenous-carbonate and carbonate-terrigenous) sediments of the Kazan complex turned out to be favorable for the accumulation of bitumen. Among the rocks of this age, narrower lithofacies are distinguished, where a combination of collectors and seals is noted. The main zones of increased bitumen concentration in the Ufimian and Kazan deposits (the western slope of the South Tatar Arch, the eastern side of the Melekess depression) are associated with them. Lithofacial conditions are less clearly expressed in the west of the region in typical marine carbonate sediments, especially in the zones covered by pre-Tatar and pre-Neogene erosion, due to the" fallout " of a significant part of the reservoir rocks of the Kazan age from the section.

The underground waters of the Permian sediments were the medium for the formation of HC deposits in them, and they have had and continue to have an impact on the degree of their preservation, the physicochemical state of hydrocarbons and the modern hydrodynamic regime of HC deposits.

The increased oil content of the Permian deposits on the territory of Tatarstan is of a secondary nature and is due to the vertical migration of hydrocarbons from the lower and middle carboniferous, since the deposits of the Permian system have not reached the conditions where the generation of hydrocarbons begins.

Ancient uplifts played an important role in the process of directed migration. Their combination with the surrounding paleodepressions provided an intensive inflow and accumulation of hydrocarbons.

The object of the study is the deposits of the Sheshminsky horizon of the Ufimian tier of the Permian system. The Sheshminsky horizon consists of two parts: the lower one is sandy − clayey and the upper one is sandy. The sandstones of the Sheshminsky horizon were mainly distributed on the western slope of the South Tatar Arch.

The porosity of sandstones reaches 30% or more, with varying clay content from 2 to 12%. The mineralization of reservoir waters ranges from 2 to 7 g / l. In addition, due to the high viscosity of naphthides in the formation conditions and a slight depression on the formations created during drilling, a deep zone of drilling fluid penetration is not formed in the intervals of oil-saturated reservoirs, due to which the necessary parameters are determined without significant distortions (Khisamov et al., 2007).

To date, there is no unified method for calculating the flow properties for ultra-viscous oil deposits in Tatarstan, and the presence of loose sands cemented with oil causes problems in quantifying the weight of bitumen saturation (Khisamov, Bazarevskaya et al., 2017).

The main purpose of this work is to develop approaches and mathematical algorithms for automating the processes of predicting weight bitumen saturation and determining the bitumen saturation thickness. To date, most of the methods for determining flow properties are applicable for cemented rocks of reservoirs of shallow depth bitumen deposits, and for uncemented (loose sands) this method is not

applicable for quantitative assessment of parameters and requires core sampling in the interval of interest. The reason for the poor convergence is mainly due to the high content of bound water in loose sands, as well as the substitution of water or bitumen with gas. The presence of a large core sample and a quantitative forecast of parameters using neural network technologies allows us to reliably describe this type of reservoir.

Creating a representative sample of wells

As a result of many years of work on searching for HC oil fields, a unique database of Permian sediment core was accumulated. After a detailed analysis and rejection of a small proportion of substandard studies, a database of 805 wells was constructed according to the general list of deposits, of which, in the end, complete information was provided for 519 that met the sampling criteria and was used in this work.

As the initial data for the development of the technology, the results of the study of core material and GIS at several deposits of ultra-viscous oil of the Republic of Tatarstan were used.

Since the main goal is to obtain a mathematical model for predicting bitumen saturation, in addition to standard core studies, geophysical studies have also been carried out in previously drilled wells, which is represented by electrical, and radioactive methods. In addition, acoustic methods are prescribed in individual wells to clarify the acoustic characteristics of the section. The following geophysical parameters (Table 1) and core data (Table 2) were used to perform this work:

Table 1 – List of geophysical parameters used in the work

№	OH param	Unitless	Description
1	PZ	омм	Potential Resistivity
2	SP	мВ	Spontaneous Potential
3	BK	омм	Lateral Log Resistivity
4	IK	омм	Conductivity
5	GK	u/h	Gamma Ray
6	NGK	u.e.	Neutron gamma
7	RHOB	g/см^3	Bulk Density
8	NNKb	u.e.	Neutron Far
9	NNKm	u.e.	Neutron Near
10	W	%	Neutron Porosity

Table 2 – List of core parameters used in the work

№	Core param	Unitless	Description
1	Kvol.bit	%	The volumetric bitumen saturation coefficient
2	Kweight.bit	%	The weight bitumen saturation coefficient
3	CKH	mD	Permeability
4	CPOR	%	Core porosity
5	CRHOB	g/см^3	Bulk density
6	CSWIRR	%	Core water Irreducible
7	CDENS	gсм^3	Core grain density
8	P_POR	u.e.	Formation Resistivity Factor (F)
9	P_SAT	u.e.	Resistivity Index (RI)

This complex of geophysical parameters is the necessary minimum and characterizes the capacitance, lithological and filtration properties of rocks.

The method of spontaneous polarization (SP) is used to divide the section into clay and non-clay differences, establish the boundaries of the layers and their correlation.

In the database, the parameters of GK and NGK were replaced with double difference parameters of AG, this helped to evaluate clayey and dense components in the section.

To determine the saturation of the collectors, the data listed above were supplemented by electrometry methods: BK, IK, PZ.

The filtration properties of the rock were calculated according to the data of compensatory neutron logging: NNKb and NNKm. The method is based on the joint interpretation of the materials of the CNC, GK and DS, since the neutron porosity is significantly affected by the well diameters and lithology.

To identify gas-containing formations and gas caps in the dome parts of the uplifts, the data of the RHOB were used, which makes it necessary to conduct it at the HO fields. In gas-containing formations, it is necessary to apply a set of methods for calculating porosity, including RHOB and NNKb.

Each geophysical parameter has its own "weight" depending on the problem solved with its use and the features of a particular type of section.

Later, when using equal input parameters, a classification of section types was carried out, which considered not only the location of each individual deposit relative to each other, but also the "weight" of the geophysical parameters.

To establish core-to-core relationships, an exact core binding is not required. It is enough to index the core data by belonging to stratigraphic layers. However, when establishing "core-OH" type links, the core should be linked, since, as a rule, the depths specified during core selection require adjustment and linking to the OH depths.

To link the "core-OH", OH curves were previously interpreted to obtain a continuous porosity curve. At the first stage, the general shift (linear shift) is searched for by linking, then elastic linking occurs, which implies a more change in the distances between the core points for better convergence with OH data, such discrepancies can be explained by different factors ranging from the curvature of wells and ending with the tightening speed when recording the linking GK on the core in the core storage.

Analysis of petrophysical connections for the calculation of flow properties

To identify petrophysical links for calculating flow properties, a correlation analysis was performed to identify the relationship between the data series of the input set. This algorithm is used to estimate the expected dependence of factors.

To do this, an array of core and OH data sets are used by the input in order to identify the correlation relationship for each method separately. The correlation relationship between the OH curves is revealed to determine the relationship between the geophysical parameters, as a result, a matrix is constructed (Table 3). A similar procedure is carried out for core values (Table 4).

Table 3 - Correlation analysis using GIS on one of the deposits

OH	BK	GK	IK	NGK	NNKb	NNKm	PZ	RHOB	PS	W
BK	1.00	-0.40	0.24	0.25	0.29	0.36	0.45	0.24	0.04	-0.21
GK	-0.40	1.00	-0.19	-0.32	-0.40	-0.46	-0.26	-0.30	0.05	0.27
IK	0.24	-0.19	1.00	-0.08	-0.03	0.03	0.61	-0.06	-0.13	0.06
NGK	0.25	-0.32	-0.08	1.00	0.94	0.89	-0.05	0.63	-0.13	-0.91
NNKb	0.29	-0.40	-0.03	0.94	1.00	0.94	0.00	0.71	-0.06	-0.92
NNKm	0.36	-0.46	0.03	0.89	0.94	1.00	0.06	0.70	-0.10	-0.88
PZ	0.45	-0.26	0.61	-0.05	0.00	0.06	1.00	0.05	-0.01	0.07
RHOB	0.24	-0.30	-0.06	0.63	0.71	0.70	0.05	1.00	0.10	-0.67
PS	0.04	0.05	-0.13	-0.13	-0.06	-0.10	-0.01	0.10	1.00	0.07
W	-0.21	0.27	0.06	-0.91	-0.92	-0.88	0.07	-0.67	0.07	1.00

Table 4 - Correlation analysis using core data on one of the deposits

Core	Kvol.bit	Kweight.bit	CKH	CPOR	CRHOB	CSWIRR	CDENS	P_POR	P_SAT
Kvol.bit	1.00	0.86	0.39	0.52	-0.29	-0.70	-0.52	-0.45	0.26
Kweight.bit	0.86	1.00	0.54	0.72	-0.32	-0.71	-0.70	-0.47	0.28
CKH	0.39	0.54	1.00	0.64	-0.26	-0.47	-0.63	-0.38	0.25
CPOR	0.52	0.72	0.64	1.00	-0.35	-0.75	-0.97	-0.63	0.32
CRHOB	-0.29	-0.32	-0.26	-0.35	1.00	0.22	0.46	0.23	0.01
CSWIRR	-0.70	-0.71	-0.47	-0.75	0.22	1.00	0.72	0.64	-0.38
CDENS	-0.52	-0.70	-0.63	-0.97	0.46	0.72	1.00	0.59	-0.34
P_POR	-0.45	-0.47	-0.38	-0.63	0.23	0.64	0.59	1.00	-0.27
P_SAT	0.26	0.28	0.25	0.32	0.01	-0.38	-0.34	-0.27	1.00

The analysis of the information obtained allows us to assess the significance of the difference between certain parameters according to OH and core values. This will make it possible to use closely correlated

methods in the future in tasks for calculating flow properties and isolating lithology. Also, the information received was used to configure features that were missing or were recorded incorrectly.

To calculate continuous porosity and oil saturation curves in the conditions of complete or partial absence of core at the well, it is necessary to determine the correlation between OH curves and core data.

Table 5 below shows an example of correlations between OH curves and the most "weighted" core values separately.

Table 5 - Correlation analysis using core data on one of the deposits

Core/OH	BK	GK	IK	NGK	NNKb	NNKm	PZ	RHOB	PS	W
Kvol.bit	-0.01	0.08	0.29	-0.52	-0.51	-0.48	0.33	-0.35	-0.04	0.57
Kweight.bit	0.00	0.07	0.21	-0.59	-0.60	-0.58	0.22	-0.46	-0.01	0.67
CKH	-0.03	0.09	0.05	-0.42	-0.44	-0.42	0.06	-0.44	-0.05	0.50
CPOR	-0.21	0.23	0.04	-0.63	-0.67	-0.67	0.00	-0.63	0.01	0.70
CRHOB	-0.11	0.09	-0.01	0.08	0.14	0.10	0.07	0.19	0.12	-0.13
CSWIRR	0.08	-0.15	-0.14	0.49	0.51	0.50	-0.15	0.43	0.04	-0.54
CDENS	0.18	-0.21	-0.02	0.62	0.67	0.66	0.04	0.62	-0.03	-0.69
P_POR	0.17	-0.18	0.03	0.43	0.46	0.44	-0.02	0.36	0.03	-0.47
P_SAT	-0.01	0.11	0.10	-0.17	-0.19	-0.19	0.07	-0.18	-0.09	0.22

Data preprocessing

Due to the fact that the OH methods were recorded by different equipment, a preliminary data preprocessing – normalization was applied to bring the measurements to a single scale.

Among the available parameters, those that can have a greater impact on the predicted value are selected. As a numerical characteristic of the probabilistic relationship, the r-Pearson correlation coefficients are used, the values of which vary in the range from -1 to +1.

Table 5 shows that for this field, the curves for OH - IK, BK, PZ, NNKb, NNKm, RHOB and W have a very good correlation with the core.

Since the main desired value was the weight bituminous saturation of the core, the OH methods with an average and high degree of influence are distinguished: W, RHOB, NNKb and NNKm.

Figure 1 shows box-plots that display the spread of Pearson correlation coefficients by OH methods for each field. From this figure, it can be seen that the medians of a number of correlation coefficients and their maximum (or minimum) values are mostly outside the |0.5| level for the already selected methods. This can be traced in 9 out of 11 deposits, which confirms the choice of criteria that are highlighted by the analysis of the average values of correlations.

Such methods as BK and IK have the same behavior on the box-plots. Although the average values of correlations indicate a weak relationship between the values under consideration and the weight of bitumen saturation, it can be noted that in 6 fields out of 11 at certain wells, the maximum value of coefficients of these methods exceeds the level of 0.75, which may indicate a high degree of connection between these criteria and bitumen saturation.

Therefore, in the future, we will use parameters that have shown good convergence with the weight bitumen saturation - W, RHOB, NNKb and BK. As a confirmation of the selected OH methods for each of the criteria, data for all wells were combined into one array, on the basis of which quantile-quantile graphs were constructed (Figure 2).

Some parameters have a strongly asymmetric distribution, close to the log. Since large values, available even in a small amount, strongly affect the results of the neural network, a logarithmic transformation was applied to such parameters. The logarithmic transformation is applied to the data of the BK and NNKb parameters: $\tilde{x} = \ln(x)$, where ln is the natural logarithm (Figure 3).

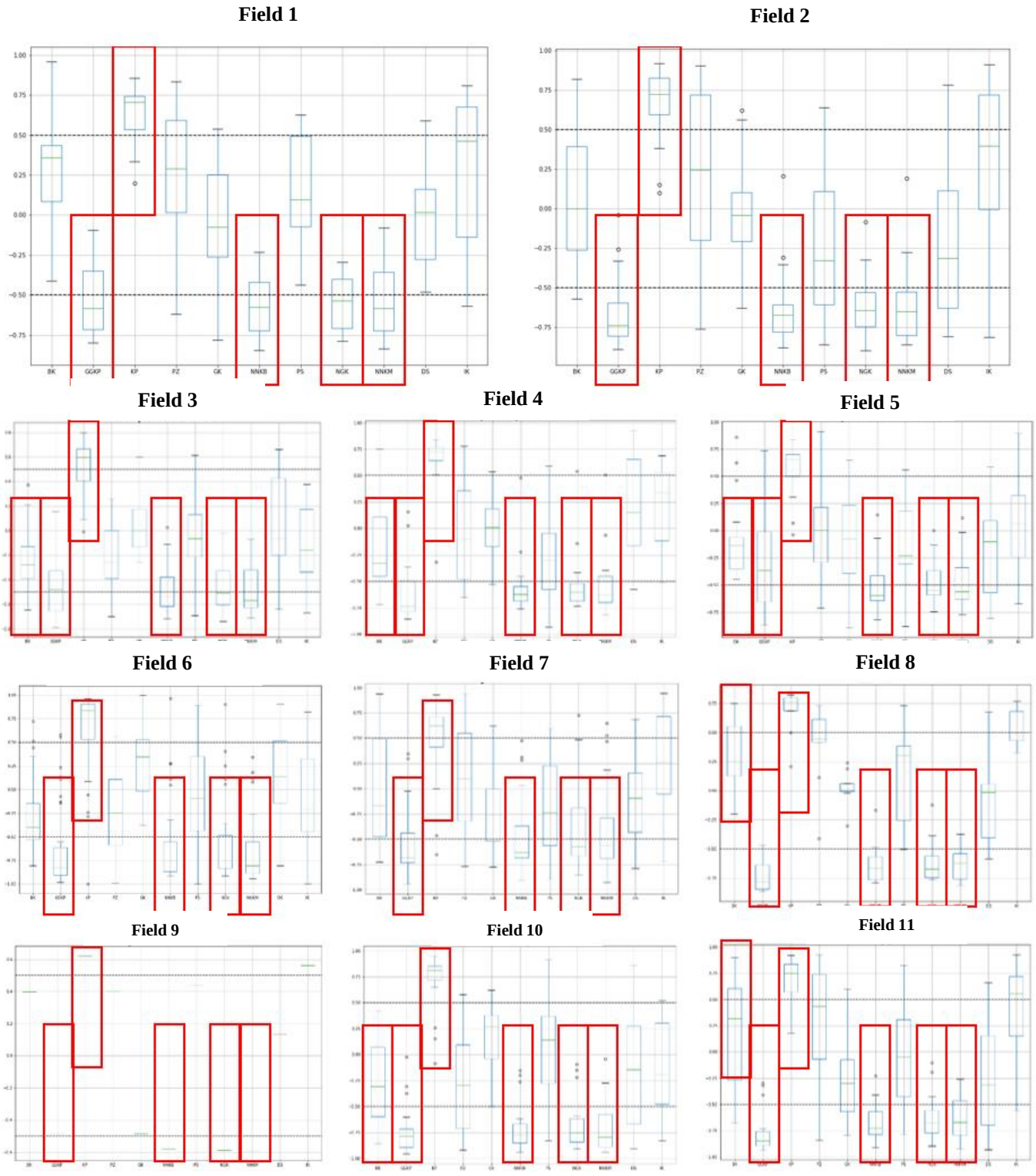


Figure 1-Box-plots of the Pearson correlation coefficients spread by OH methods for each field

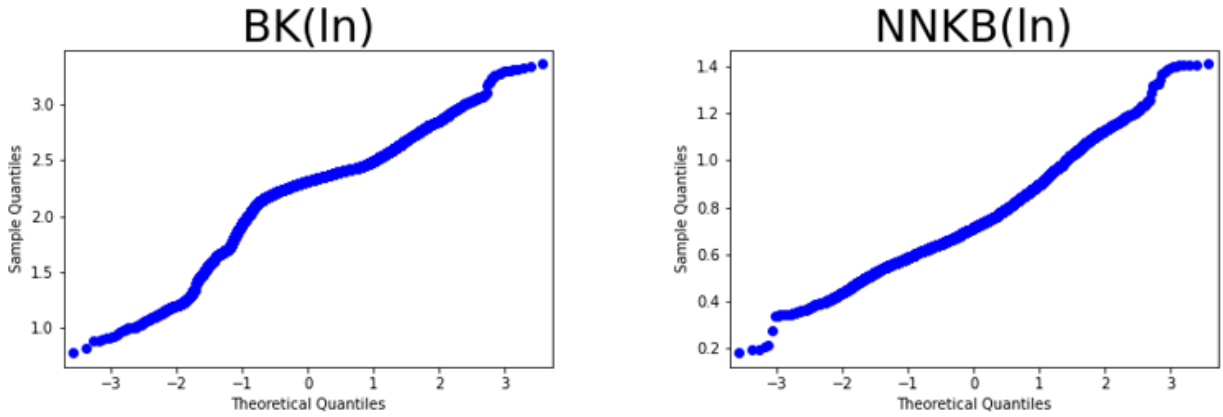


Figure 2 - Quantile-quantile graphs for OH with a high and medium degree of influence

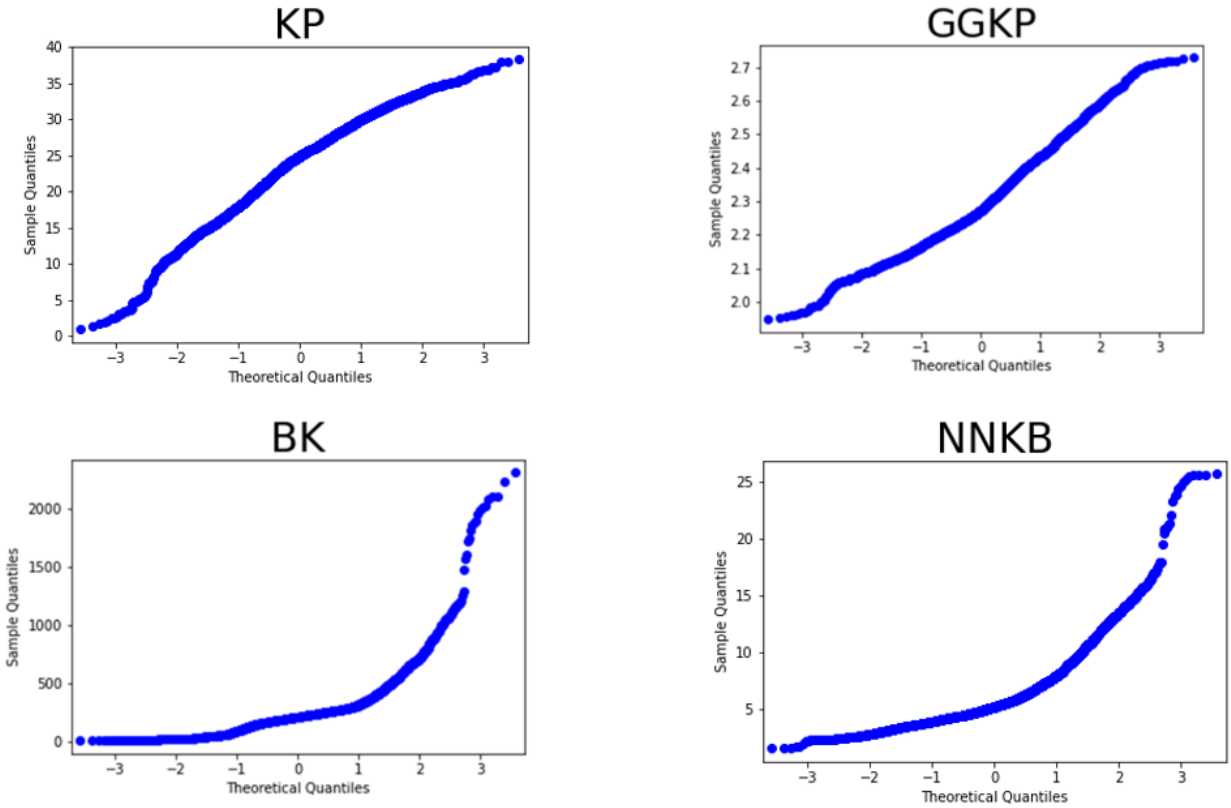


Figure 3-Quantile-quantile graphs for BK and NNKb after logarithmic transformation

Next, for a single parameter, the search for "reference" and "outgoing" wells is performed:

1. For each well, the average value is calculated $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$
2. If x > Q (50%) +0.8*(Q (75%)-Q (50%)) or x < Q (50%)-0.8*(Q (25%)-Q (50%)), then the well with such an average is considered to be out of the general series. Q (50%), Q (75%), Q (25%) - quartiles of a series of averages;
3. 3. The" reference " well is chosen based on the average value of which is located closest to the median of the series of averages (without considering the outlier values) (Figure 4). For the criterion for the reference well, the average and standard deviation are considered:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

4. Then the values of the "departing" wells are corrected according to the formula (Figure 5):

$$\frac{x_{ij}-m_i}{s_i} \cdot s_{\text{ст}} + m_{\text{ст}} \text{ , where}$$

m_i, s_i, mean and standard deviation of the i-th well criterion,

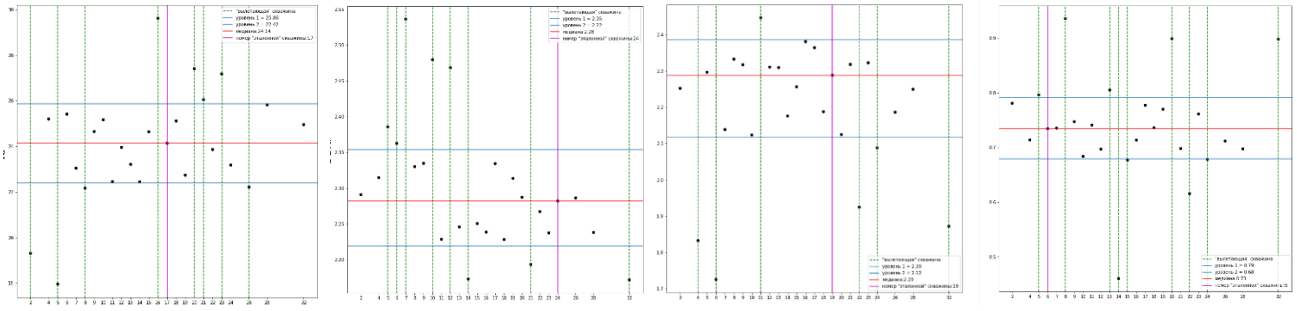
$m_{\text{ст}}, s_{\text{ст}}$, - the mean and standard deviation of the reference well
x_{ij}– the j-th value of the criterion in the i-th well

5. Next, the data is normalized. The average and standard deviation of each parameter is calculated for the entire field at once. Recalculation of the values of the curves in each well according to the formula:

$$\frac{x_{ij} - m_i}{s_i} m_{\text{ст}}, \text{where}$$

m_i, s_i– mean and standard deviation of the lifting criterion,
x_{ij} – the j-th value of the criterion in the i-th well

After normalization, the graphs show (Figure 4) that the data is concentrated relatively close to the average of the "reference well".



W RHOB BK NNKb

Figure 4-Graphs of the average GIS values for wells on the example of one of the fields

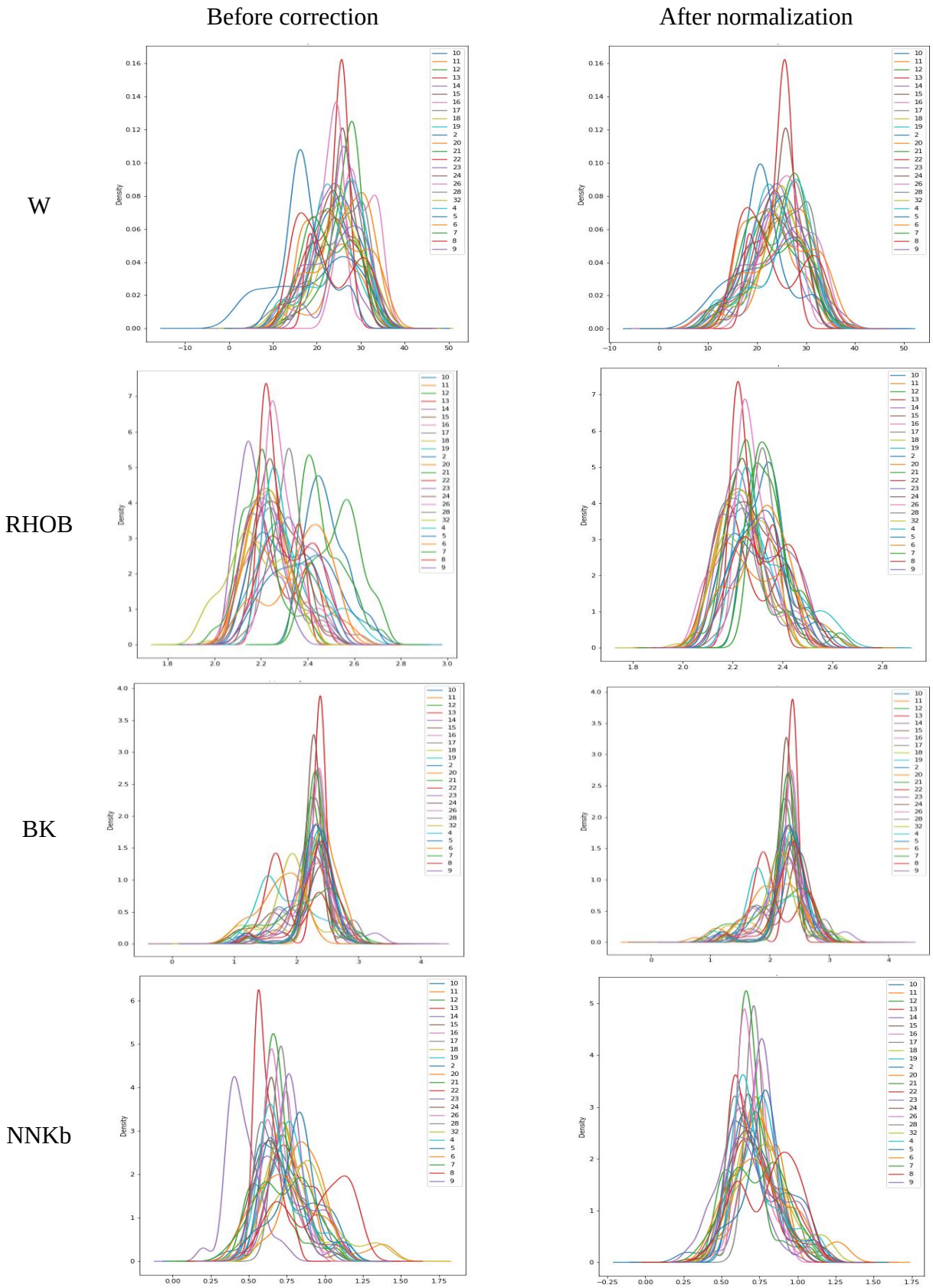


Figure 5 - Graphs of density functions for wells for various GIS methods on the example of one of the fields before and after normalization

Core data rejection

Since a significant amount of core is represented from the interval of interest, the weight bitumen saturation for the core is determined as the most reliable and representative indicator. As a forecast of the weight bituminous saturation according to OH, the K.bit.weight according to the core will be used, therefore, before developing the mathematical apparatus, it is necessary to prepare data on the core – rejection of "outgoing values" according to the K.bit.weight. core, so as not to distort the training sample.

The criteria by which the points on the core were excluded were reduced to the fact that at the initial stage, obvious departures along the W and Kp along the core were visually evaluated. Wells where there were no Kp values for the core were also evaluated, and values were also rejected in these places (Figure 6). For a more reliable assessment, "synthetic" curves of weight bitumen saturation were constructed using multivariate regression for each well and an assessment was made at a qualitative level of how the points behave on the Kweight.bit.core. relative to the synthetic curve. This rejection made it possible to more accurately adjust and calibrate the mathematical model, which ultimately greatly affects the forecast of weight bitumen saturation.

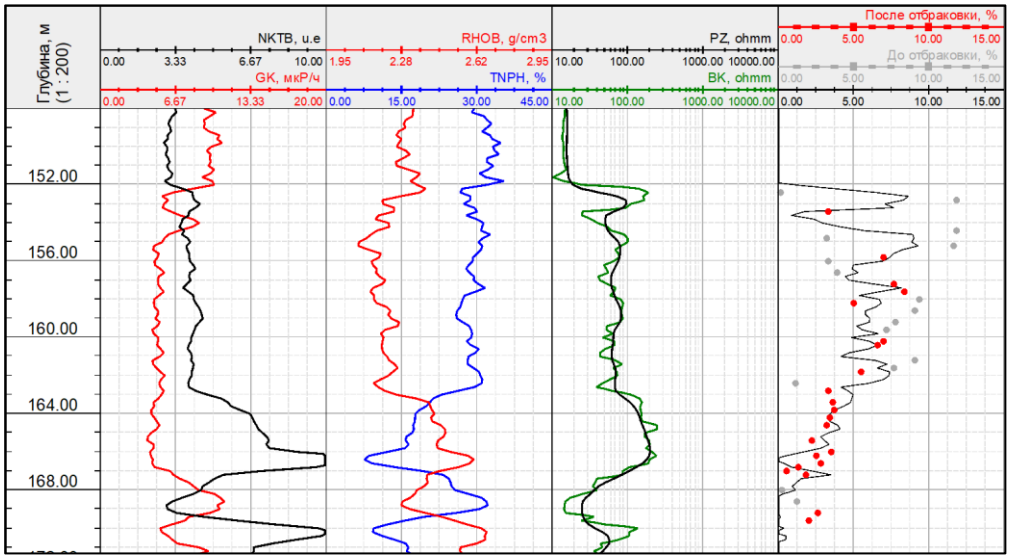


Figure 6 - Example of rejection of weight bitumen saturation by core

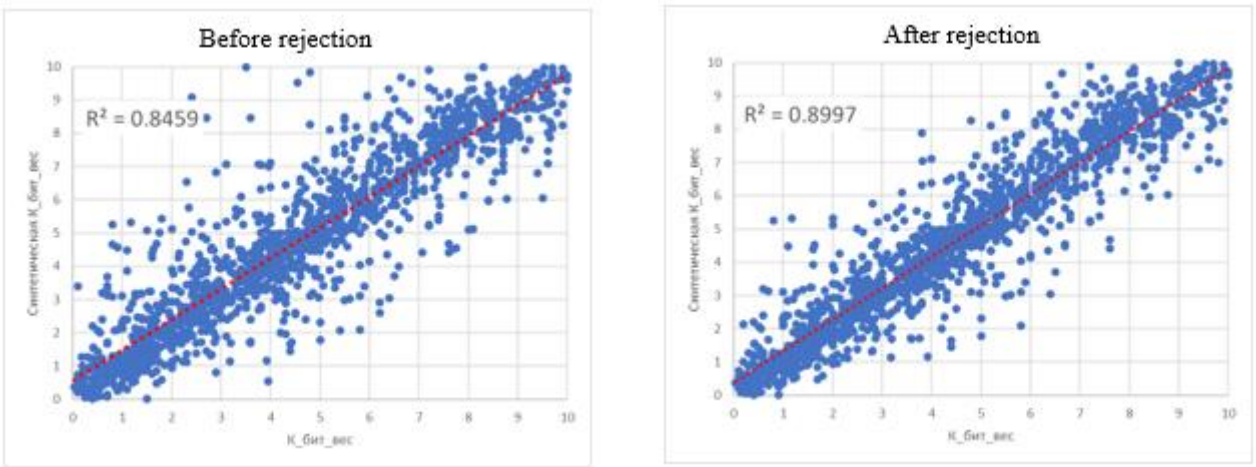


Figure 7 - Cross-plots of comparison before and after rejection of weight bitumen saturation by core

Development of a mathematical algorithm

After preprocessing the data, 4 curves were selected that showed good convergence with the weight bitumen saturation - W, RHOB, NNKb and BK. In the future, work on the forecast of weight bitumen saturation (Kweight.bit.) will be carried out using these OH methods. However, in addition to the K.bit. forecast, there are also tasks for the allocation of stratigraphic chops and for the classification of lithology, where we used the entire set of OH after pre-processing.

Allocation of stratigraphic

The primary task was to automatically determine the zone of interest – stratigraphic chops using a logistic regression algorithm trained on the results of manual correlation of layers. The developed algorithm allows you to automatically allocate stratigraphic chops for the roof and sole of the sand pack of the Sheshminsky horizon.

Stratigraphic chops are distinguished by the curves of the GK, NNKb, BK, for this purpose, training is conducted on a representative set of wells (Validov, 2017).

According to preliminary estimates, the automation of the process of selecting stratigraphic chops is 90%. On a set of 600 wells, the processing time spent by the machine is about 2 minutes. As a rule, about 10% of wells, even after quality control, have some deviations, as a result of which such wells must be processed manually.

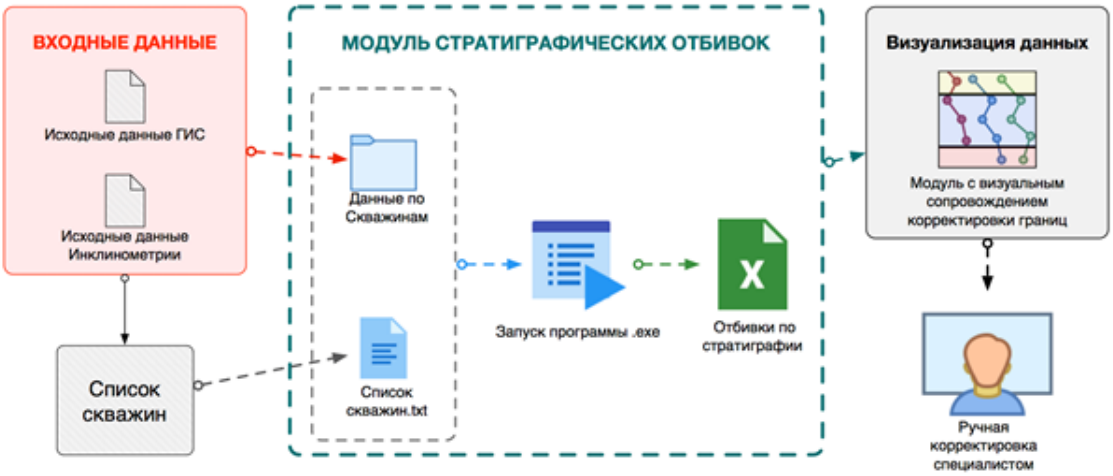


Figure 8 - Diagram of the automatic operation of the module on stratigraphic chops

Selection of lithology

Lithological splitting of the section is one of the conditions for the prediction of weight bitumen saturation. For high-quality automatic selection of rock lithology, it is necessary to prepare a representative set of lithological types in advance and select the optimal neural network algorithm. For this purpose, the lithology was pre-selected manually and the optimal algorithm for selecting the lithology was selected.

To solve the problem of rock classification (lithology selection), the library's machine learning tools were used Keras.py. Keras allows you to build a neural network of direct propagation by adding layers one after another. The neural network architecture is a multilayer perceptron (Figure 9). The first layer accepts input data (Input Layer) in the form of a vector of values of six curves at a specific depth. Next

comes 2 Dense layers (Hidden Layer) with 256 neurons in each, the activation function is ReLU. The Dense layer type means that it is a fully connected layer, where each neuron of the previous layer is connected to each neuron of the next layer. The last layer (Output Layer) is also fully connected with the softmax activation function to display with accuracy that the data belongs to a certain class.

Relu is an activation function that skips values greater than zero, equates negative values to zero. This is done so that the function becomes nonlinear and is able to describe complex patterns. Dropout is turning off part of the neurons in the layer to avoid overtraining, and regularization. Categorical_crossentropy is used as an error function, where categorical means a set of classes in the output, and crossentropy is used to count errors during softmax activation.

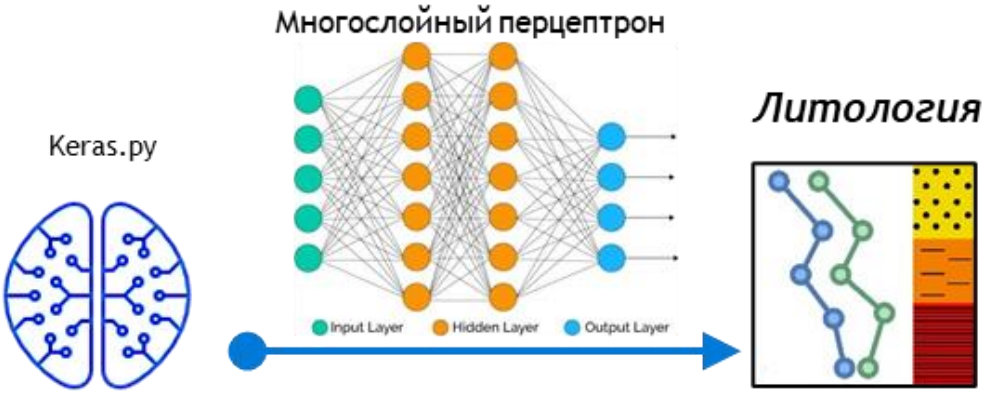


Figure 9 - Basic algorithm for the selection of lithology

For a qualitative assessment of the lithology, the following lithological types of rocks were identified:

1. Clay-code 1;
2. Sandstone-code 2;
3. Carbonated sandstone-code 3;
4. Carbonate rock-code 4;
5. Gas-saturated sandstone-code 5.

The following are used as training data:

1. Normalized data – GK, NNKb, W, RHOB;
2. A discrete lithology curve selected manually.

Figure 10 and Table 6 show the results of the neural network algorithm and a point-by-point comparison of the selected lithotypes when starting at 450 wells. The table shows that the selection of intervals represented by sandstone shows high convergence (the neural network algorithm identifies the lithological type - "sandstone" correctly in 95% of cases).

When identifying lithology interlayers with low reservoir properties, there is an increase in uncertainty (Korolev et al., 2018). This effect is manifested, among other things, when interpreted manually and is associated with the fact that the rocks of the "carbonated sandstone/limestone" series have relatively similar readings on the OH curves of the standard logging complex. Considering the fact that the method of extraction of HC is not carried out from sandstone layers less than 6 meters, this effect is not critical when isolating lithology.

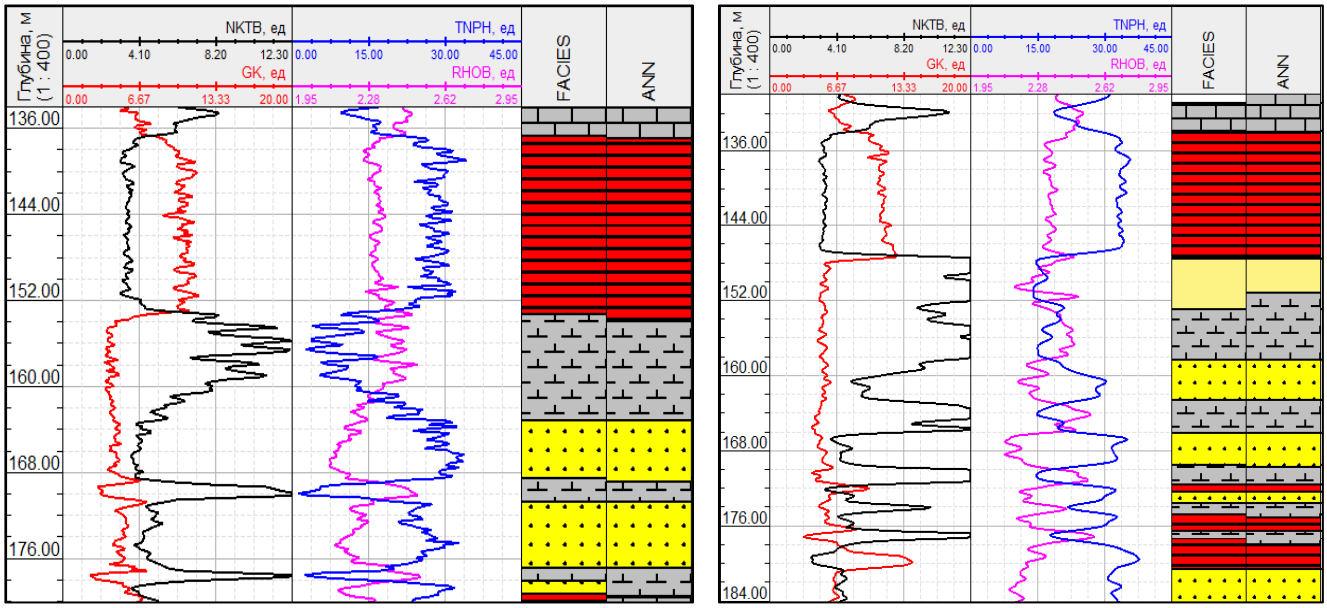


Figure 10-Plates for wells where training and testing of algorithms for the allocation of lithology were carried out (FACIES – lithology was selected manually, ANN – lithology was selected using a neural network algorithm).

Table 6-Classification table for differentiation into 4 classes, R=0.88

			Neuron network algorithm			
Expert interpretation	Lithology		Clay	Sandstone	Carbonated sandstone	Limestone
		Clay	90.93	3.06	0.92	5.1
		Sandstone	2.59	95.03	2.12	0.27
		Carbonated sandstone	1.78	26.41	69.14	2.67
		Limestone	15.65	8.4	0.38	75.57

However, it is quite difficult to describe a large number of wells by using this algorithm alone. This algorithm is sensitive to large shifts in absolute values. To improve the quality of the output data, additional procedures (work around) were used:

1. **Noise processing**

The neural network makes forecasts at each mark (depth), and since the curves sometimes behave erratically, such values appear among the forecasts in the form of" noise "that an expert in this field would immediately exclude (layers less than 0.6 m). If the model predicted the same class for a long time, and then predicted a new class once (twice, three times, but no more than 0.6 m), then it will be written as the previous class.

2. **Claying in the area of obvious values**

It is known that in the zone of "lingul clays "other types of rocks cannot occur except clay. So even if the neural network predicts sandstone or carbonated sandstone, for example, in most cases the value will be overwritten on top of the forecasts for the "clay «class.

3. **Agglomeration of rocks**

This approach is often used where the curves have fairly frequent and large amplitude fluctuations relative to the boundary values of sands and clays.

Calculation of a continuous curve of weight bitumen saturation using neural network technologies

To solve the problem of predicting Kweight.bit., numerous iterations were carried out to select ready-made optimal neural network algorithms. Intermediate results for each of the tested algorithms were not given in this article due to the lack of acceptable results.

At the moment, the most suitable neural network technology is presented, which describes the behavior of Kweight.bit. with a correlation of 0.78 and the smallest error of the COEX.

The regression function is used to predict the weight bitumen saturation, that is, the prediction of a single value (continuous). Using the Keras library, a multi-layer perceiver is built, which receives inputs as a vector of curve values at a specific depth. The input layer is followed by a fully connected layer with 128 neurons, the weights of which are initialized by a normal distribution. The relu activation function, overwrites negative values by zero. We get a nonlinear function. This is followed by a second hidden fully connected layer with 256 neurons, identical to the previous one. The output layer is a single neuron connected to all the neurons of the previous layer. The activation function at this stage is linear, which means that there is no activation function. Since this is a regression task, activation is not needed in the output layer (f (x) = x). As an error function, we use the average absolute error (modulo). The Adam optimizer is used as a replacement for stochastic gradient descent. Adam combines the best properties of the AdaGrad and RMSProp algorithms, which can handle sparse gradients on noise problems. The metric is the average absolute error.

3 such neural networks have been created. Each of them processes separate curves, that is, the original data set was divided according to certain features empirically so that each neural network receives its data with small intersections. During the prediction, each network processes its own data section and makes a forecast, which is summed over all three neural networks and averaged. This approach allows you to identify the most important features in the general pool and give a more accurate forecast.

The following are used as training data:

1. Normalized data – BK, NNKb, W, RHOB;
2. Point-to-point data on Kweight.bit.core

After conducting test calculations of the Kweight.bit. coefficient for a representative sample of wells, the results were evaluated and cross-validated.

Figures 11 and 12 show the results of the Kweight.bit. calculation algorithm of the training and test sample. The cross-plots show that the correlation on the training sample was R = 0.9, on the test sample – R = 0.7, which indicates good convergence and a forecast error of less than 20%.

Training with the neural network

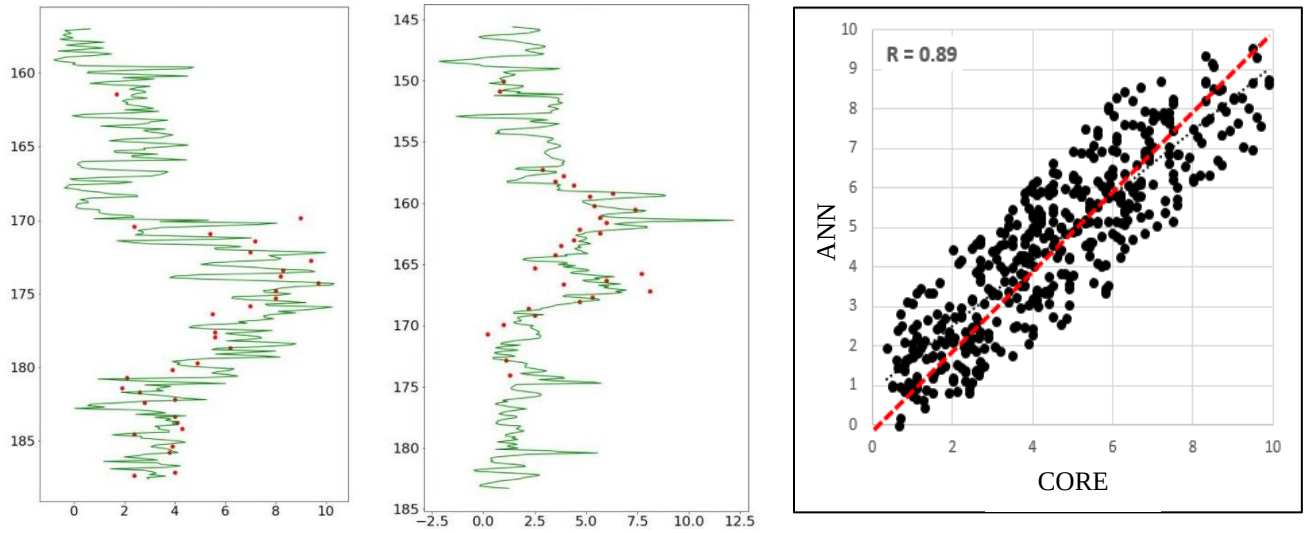


Figure 11-The result of calculating the weight bitumen saturation for the training sample (quantitative assessment)

As a result, the thicknesses of oil-saturated intervals are determined by means of clusters on the section (the boundary is Kweight.bit.> 4.5%). The accuracy of the selection of oil-saturated thicknesses based on the selected method is 89% (Table 7). The working method for calculating flow properties using neural networks gives acceptable results, as work on improvements continues. But we can already say that for the forecast of oil-saturated thicknesses, the neural network gives satisfactory results (Figures 13 and 14).

Forecast using the neural network

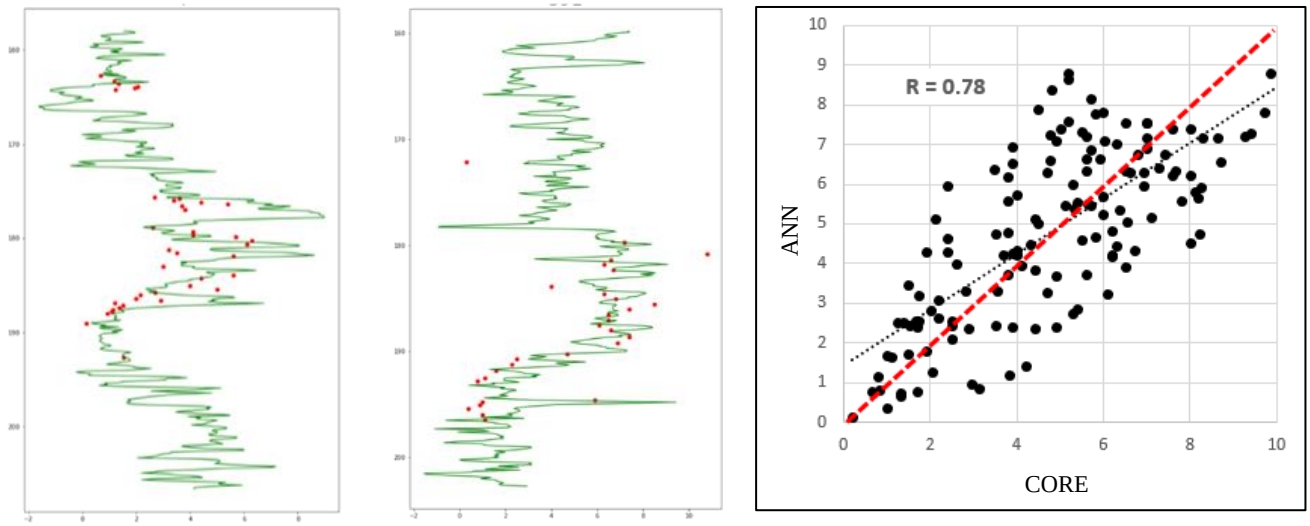


Figure 12 – The result of calculating the weight bitumen saturation for the test sample (quantitative assessment)

Table 7 – Table of correspondence of core thicknesses and the calculated curve using neural network technologies

Frequency matching table (%)			
	Thickness < 4.5 м	Thickness > 4.5 м	Total:
Thickness < 4.5 м	69.85	30.15	**100.00**
Thickness > 4.5 м	11.36	88.64	**100.00**

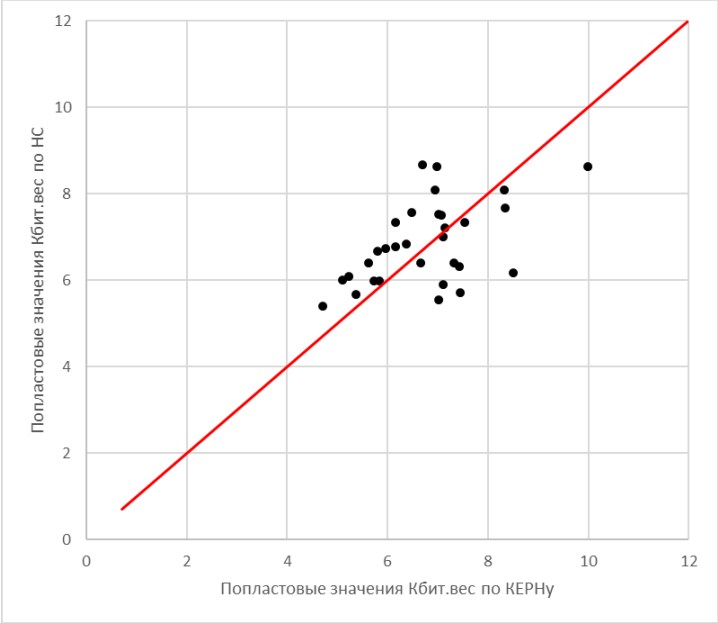
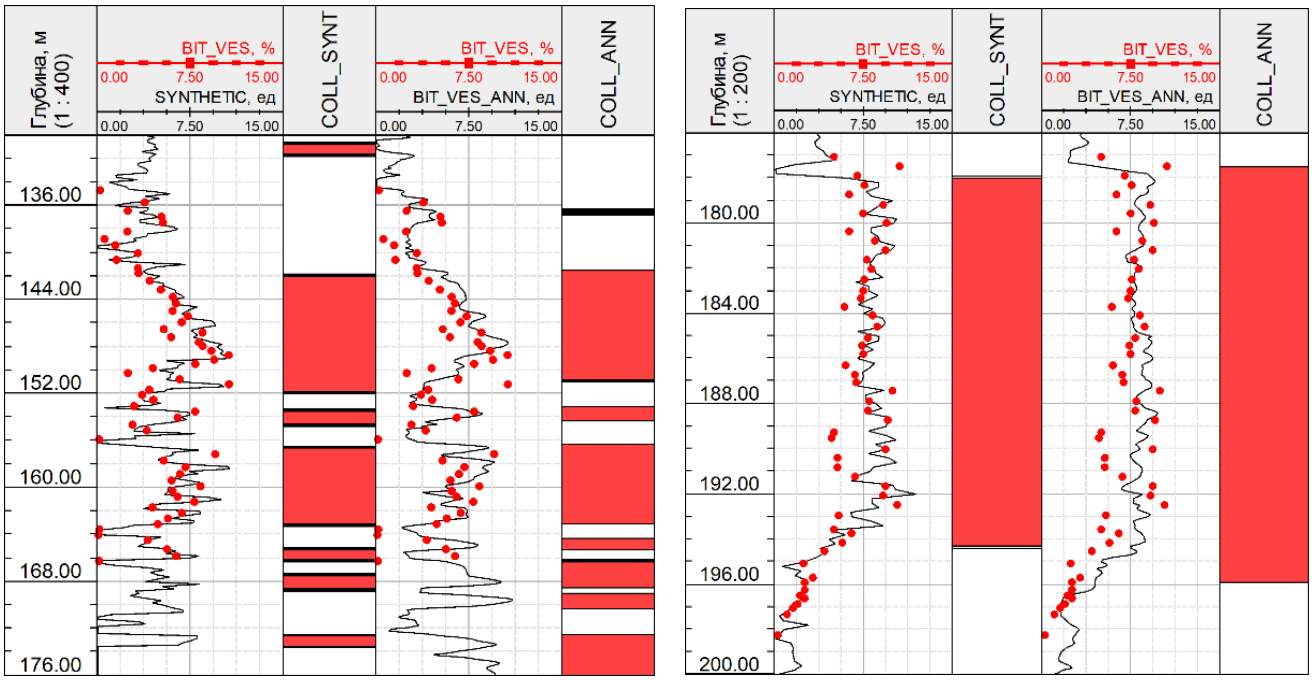


Figure 13-Comparison of the layer-by-layer values of the weight bitumen saturation according to the core and the calculated curve using neural network technologies (quantitative assessment)



Well No. 1 Field 1 Well No. 2 Field 1

Figure 14 – Comparison table of the weight bitumen saturation by core and the calculated curve using neural network technologies (qualitative and quantitative assessment of parameters)

Methodical approach to the calculation of weight bitumen saturation

Let's consider the methodological approach of calculating Kweight.bit. using core data and OH methods. Since the calculation method should be universal and useable, including on wells for which core sampling was not carried out, a common database of OH parameters was created for all wells, and for each uplift separately.

The method of processing and interpreting OH materials is quite time-consuming and includes a number of tasks and stages:

1. Determination of porosity;
2. Determination of volumetric bitumen saturation;
3. Conversion from bulk to weight bitumen saturation;
4. Evaluation of the quality of the calculation of Kweight.bit.

Determination of parameters for porosity calculation

One of the main parameters – the porosity coefficient is calculated from the data of gamma-ray logging (GK) and compensatory neutron gamma-ray logging. The procedure for determining porosity is reduced to the sequential introduction of corrections to the readings of the compensatory neutron gamma-ray logging for the influence of the well diameter and clay content. The first amendment (W') is determined according to the data of well diameter logging, but since this method is not used since 2018, in the future calculations will be made without taking this amendment into account. As a parameter that characterizes the clay content, the double difference parameter AGK is taken:

$$\mathrm{A}_{GK} = \frac{I_{\mathrm{GK}} - I_{\mathrm{GK}}^{\min}}{I_{\mathrm{GK}}^{\max} - I_{\mathrm{GK}}^{\min}}$$

where I_{GK}, $I_{\mathrm{GK}}^{\max}$, $I_{\mathrm{GK}}^{\min}$ are the GK readings, respectively, at the current point, in "lingul clays" and in compacted sandstone with the lowest content of radioactive elements. Judging by the indications of the AGK, the clay content in the sandy productive pack is small, which is confirmed by granulometry data – on average no more than 5 %.

Next, the error on the influence of clay content is determined by the following formula:

$$\Delta W_{\mathrm{sh}} = I_W^{\max} * \mathrm{A}_{\mathrm{GK}}$$

where $I_W^{\max}$ is the hydrogen content of clays that includes the chemically bound water of clay minerals, i.e. the maximum indication of the hydrogen content (95 percentile) in "lingula clays". This indicator is determined automatically for each well individually.

The porosity coefficient is calculated as follows:

$$\mathrm{K}_{\mathrm{p}} = W - \Delta W_{\mathrm{sh}}$$

The sensitivity analysis showed (Figure 15) that the W – hydrogen content has the greatest influence (more than 50%) on the porosity calculation and suggests that for correct calculations of the porosity coefficient, it is necessary to carry out a qualitative interpretation by OH batches. AGK has a slightly smaller influence (about 33%), but since this parameter depends on $I_{\mathrm{GK}}^{\max}$, $I_{\mathrm{GK}}^{\min}$ and is calculated automatically, there is no subjectivity in this parameter.

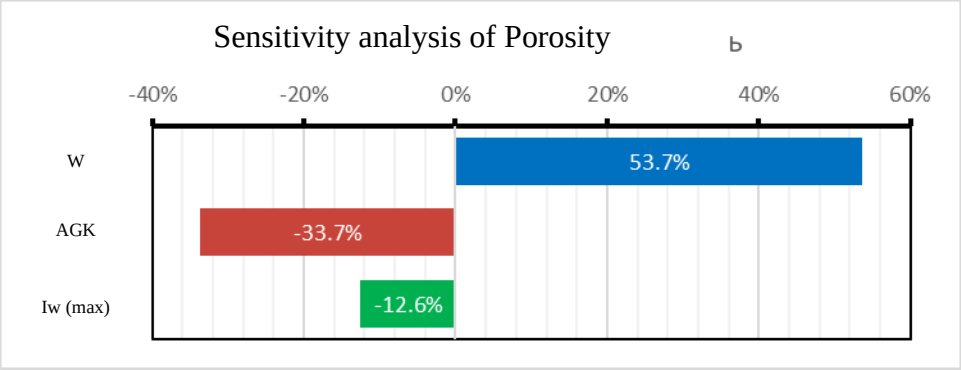


Figure 15 – Sensitivity analysis when calculating the porosity coefficient

During the course of this work, an incorrect recording of the W method and an offset of the curve relative to the depth scale were revealed at a number of wells, which affected the forecast calculations. As a result, an algorithm was written that corrects incorrect W entries before processing and performs linear linking.

Determination of parameters for calculating the volume bitumen saturation

Considering the fact that the readings of neutron logging are greatly overestimated in clays due to the high content of bound water in them and, as a result, the porosity in clay reservoirs is greater compared to core data; apparently, the correction for clay content does not fully compensate for the increase in the readings of the method, therefore, in order to take into account the maximum effect of the presence of clay components, the improved Simandoux equation for terrigenous reservoirs is used to calculate the volume bituminous saturation, which is subsequently converted into weight bituminous saturation. Let's consider the Simandou equation for calculating the volume bitumen saturation at n=2:

$$K_{vol.bit.} = \frac{\left(-\frac{A_{GK}}{I_{IK}^{max}}\right) + \sqrt{\left(\frac{A_{GK}}{I_{IK}^{max}}\right)^2 + 4\frac{K_p{}^m}{a*(1-A_{GK})*R_w*BK}}}{\frac{2K_p{}^m}{a*(1-A_{GK})*R_w}}$$

where I_{IK}^{max} is the maximum value of the resistance of dense rocks (in our case, we use the maximum value of "medium – spirifer limestone"), a is a constant value, which is determined by the cross-plot Rp – Kp, m is the cementation coefficient for this uplift, which is determined by the cross-raft Rp-Kp (takes values from 1.3 to 2.3 depending on the lithological characteristics of the rocks), Rw is the water resistance inside the formation (the average mineralization of reservoir water for the Sheshmin deposits is taken to be 3 g/dm3 is the average for HO deposits, which corresponds to Rw = 2.2 Ohms.m).

Sensitivity analysis shows (Figure 16) that BK has the greatest influence (about 40%) on the calculation of volumetric bitumen saturation, so in the future our algorithm will evaluate the quality of this method in order to obtain correct calculations of Kvol.bit. AGK and the cementation coefficient (m) have a slightly less influence on the Kvol.bit. readings (25% and -18%, respectively), but AGK depends on I_{GK}^{max}, I_{GK}^{min} and is calculated automatically, and the cementation coefficient (m) is calculated from the Rp-Kp core. That is why it was necessary to find the most reliable values of a and m for each field (Table 8) in order to further minimize the calculation error of Kvol.bit.

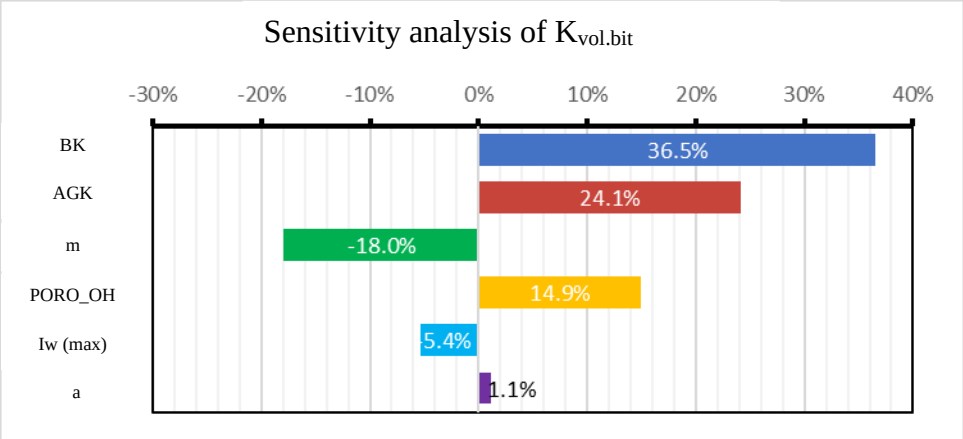


Figure 16 – Sensitivity analysis when calculating the volume bitumen saturation

Table 8 – Table of parameters for calculating Kbit. Ob

Field	a	m	R_w
Field-1	1.7	1.6	2.2
Field-2	2.2	1.4	2.2
Field-3	1.2	1.6	2.2
Field-4	3.8	1.24	2.2
Field-5	1.4	1.56	2.2

Field	a	m	R_w
Field-6	1.7	1.6	2.2
Field-7	2.2	1.4	2.2
Field-8	1.8	1.45	2.2
Field-9	1.37	1.65	2.2
Field-10	3.2	2.2	2.2

Determination of weight bitumen saturation

The coefficient of weight bitumen saturation can be described by the following formula:

$$К_{\text{weight.bit}} = \frac{К_{\text{vol.bit.}} * К_{\text{p}} * \rho_{\text{oil}}}{К_{\text{vol.bit.}} * К_{\text{p}} * \rho_{\text{oil}} + (1 - К_{\text{p}}) * \text{RHOB}}$$

where ρ_{oil} is the density of oil, kg/m3, which we take as 0.956.

Figure 17 shows the sensitivity analysis for the Kweight.bit. forecast. From this figure it can be seen that the most" significant "parameters are the previous calculations and require special attention -W, Agc, m and Rw (Rw studies should be carried out for each field separately, since these parameters have the greatest "weight" in the calculations, but for a qualitative assessment, the influence of other parameters was not presented in the sensitivity analysis).

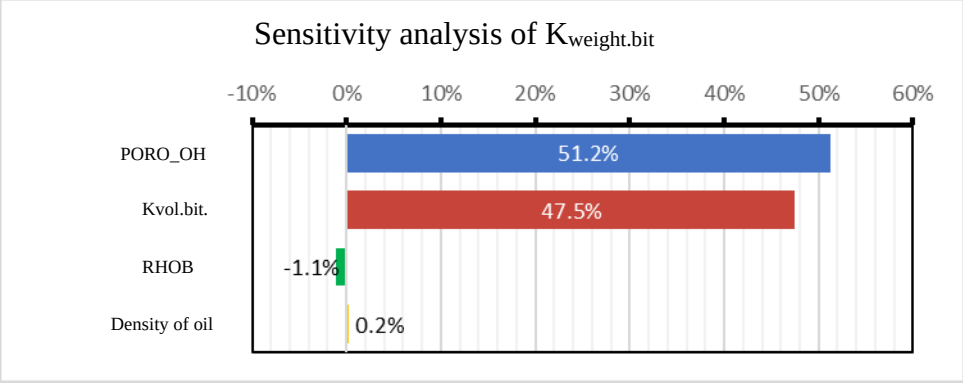


Figure 17 – Sensitivity analysis when calculating the weight bitumen saturation

Assessment of the quality of calculation of weight bitumen saturation

Kweight.bit. data were used to assess the quality of the methodology results calculated using this technique and laboratory studies of the core (Figure 18, 19). The cross-plot shows that the correlation coefficient is ≈ 0.8 on the example of one of the deposits, which is an acceptable result for the Kweight.bit. forecast. The accuracy of the selection of oil-saturated thicknesses based on the selected method is 83% (Table 9).

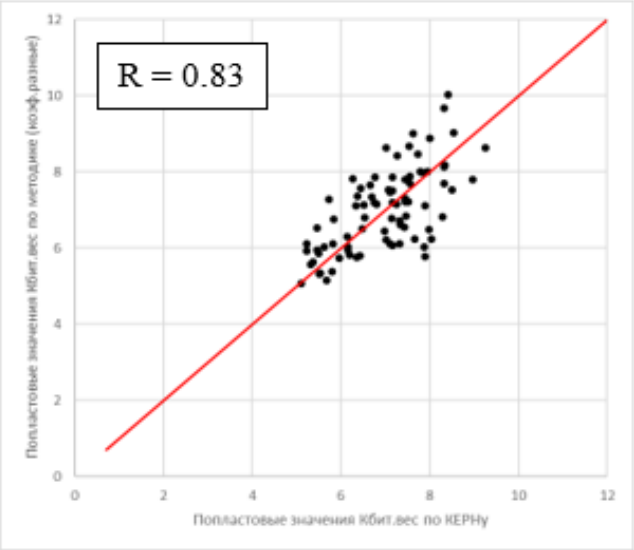


Figure 18 – Comparison of the layer-by-layer values of the weight bitumen saturation by core and GIS (a technique using the eigenvalues a and m for each uplift)

Table 9 – Table of correspondence of core thicknesses and the calculated curve using neural network technologies

Frequency matching table (%)			
	Thickness < 4.5 м	Thickness > 4.5 м	Total:
Thickness < 4.5 м	84.77	15.23	**100.00**
Thickness > 4.5 м	16.97	83.03	**100.00**

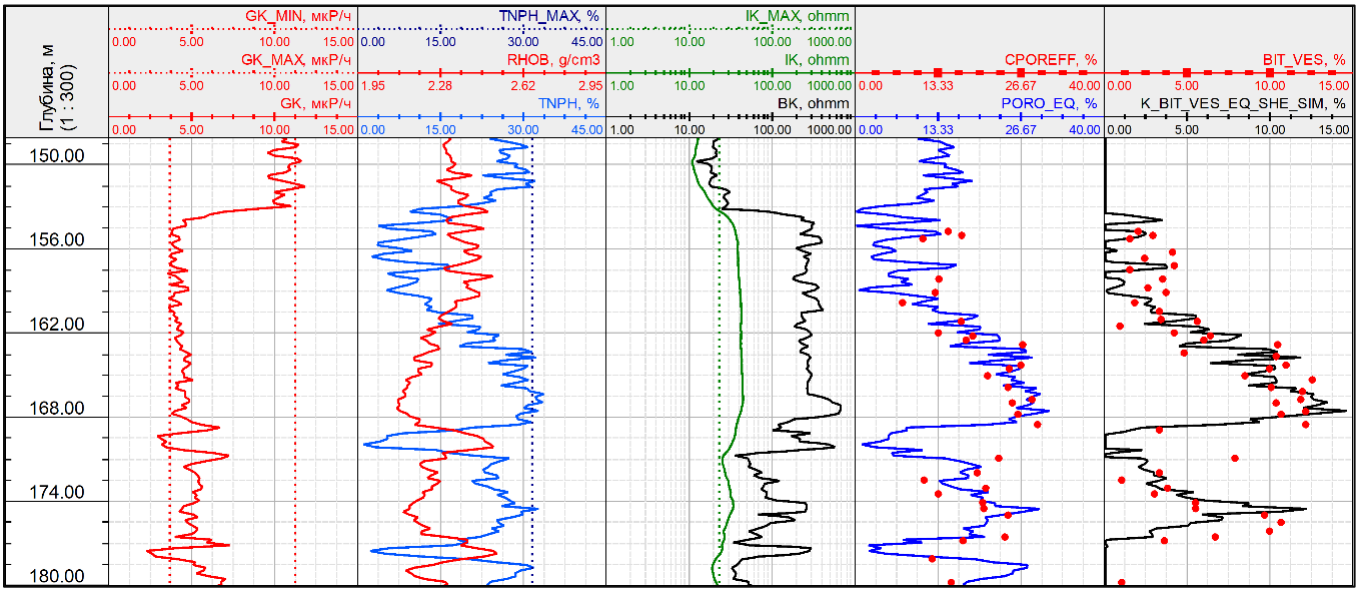


Figure 19 – Results of calculation of the porosity coefficient according to OH, weight bitumen saturation and comparison with core

But since the calculation method should be universal for all uplifts where core sampling is not provided, it will not be possible to find the coefficients a and m, in this case, you can use the average coefficients (Table 10), which were selected using the algorithm.

Table 10 – Table of selected parameters for calculating the volume bitumen saturation

Field	a	m	R_w
Central + Southern group	2.2	1.5	2.2

The accuracy of the selection of oil-saturated thicknesses based on the selected method is 79% (Table 11).

Table 11 – The table of correspondence of core thicknesses and the method using the selected coefficients

Frequency matching table (%)			
	Thickness < 4.5 м	Thickness > 4.5 м	Total:
Thickness < 4.5 м	87.95	12.05	**100.00**
Thickness > 4.5 м	20.52	79.48	**100.00**

The cross-plot (Figure 20) shows that the correlation coefficient = 0.79 on the example of one of the deposits, which is an acceptable result for the Kweight.bit. forecast.

The satisfactory convergence of the obtained results demonstrates the regularity of the selected criteria and the relevance of the developed approach for use in predicting the quantitative values of filtration-capacitance properties without core sampling, and in the presence of a large amount of data, they may have good application prospects and economic significance.

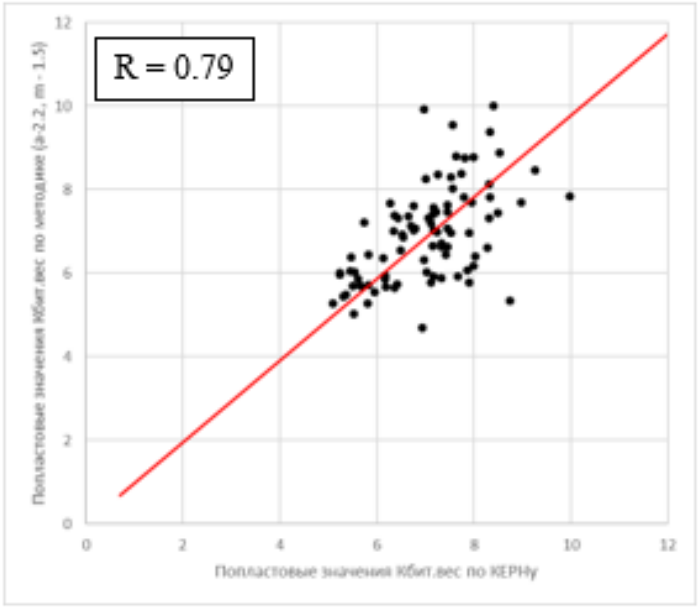


Figure 20 - Comparison of the layer-by-layer values of the weight bitumen saturation by core and GIS (method using the selected coefficients)

Conclusions

To configure the model of the flow properties calculation, information about the correlation relationships between the OH curves and the linked core for each field was used, this allowed more precisely configuring the mathematical model.

For wells that participated in the “training” sample when creating an algorithm using neural network technologies, the porosity and weight of the bitumen saturation in the core do not go beyond the allowed interval (except for intervals where gas-saturated sandstone is present).

In gas-saturated sandstones, a false underestimation of the hydrogen content (W) occurs, which in turn distorts the calculation of the porosity coefficient and the weight bitumen saturation.

The level of correlation with the core of the volume bitumen saturation coefficient calculated using the NS and according to the presented method is 0.8 and 0.83, respectively.

The accuracy of the thickness prediction for bitumen saturation using neural network technologies is 80%, which is a satisfactory result.

In the conditions of “abnormal” behavior of core studies on weight bitumen saturation and to achieve a more accurate forecast of Kweight.bit. on OH is currently not possible due to the lack of a sufficient number of such cases in the training data.

Determination of porosity and bulk bitumen saturation on the core is direct information about the section. Therefore, petrophysical laboratories should strive to minimize errors in determining parameters on the core.

For an accurate Kweight.bit. forecast. according to OH, it is necessary to use calibrated curves, because often the initial data might have incorrect readings of quantitative values.

During the work on the methodological approach to the calculation of weight bitumen saturation, additional procedures were carried out to adjust the OH data, which led to an acceptable result. The accuracy of the forecast of the method used is 83%, which is a satisfactory result.

Acknowledgement

We thank PJSC Tatneft for the data provided and the permission to use it in this publication.

This work was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation in accordance with Agreement No. 075-11-2019-032 dated November 26, 2019 within the framework of the project "Creation of a high-tech hardware and software complex based on neural network algorithms to improve the efficiency of the development of large hydrocarbon deposits at late stage ".

Nomenclature

AGK	= GR index
BK	= Lateral Log Resistivity
OH	= Open hole
GK	= Gamma Ray
RHOB	= Bulk Density
DS	= Caliper
IK	= Conductivity
Kweight.bit	= The weight bitumen saturation coefficient
Kweight.bit.core	= The weight bitumen saturation coefficient by core
Kvol.bit	= The volumetric bitumen saturation coefficient
Kvol.bit.core	= The volumetric bitumen saturation coefficient by core
Kp	= Porosity
CPOR	= Core porosity
NGK	= Neutron gamma
NNKb	= Neutron Far
NNKm	= Neutron Near
ANN	= Artificial Neural Network
PZ	= Potential Resistivity
SP	= Spontaneous Potential
RMS	= Root-mean-square deviation
FES	= Filtration and capacity properties
W	= Neutron Porosity
a	= Tortuosity factor
m	= Cementation exponent
n	= Saturation exponent
Rw	= Formation water resistivity
HC	= Hydrocarbons
HO	= Heavy Oil

References

Deng, L. 2014. A tutorial survey of architectures, algorithms, and applications for deep learning. *APSIPA Transactions on Signal and Information Processing.* Volume 3.

Jones, N. 2014. Computer science: The learning machines. *Nature.* Volume 505. No. 7482: 146-148, 1.

Korolev, E.A., Usmanov, S.A., Nikolaev, D.S., Gabdelvaliyeva, R.R. 2018. Effect of lithological heterogeneity of bitumen sandstones on SAGD reservoir development. *IOP Conf. Ser. Earth Environ. Sci.* Volume 155. https://doi.org/10.1088/1755-1315/155/1/012019

Hagan, Martin T., Demuth, Howard B., Beale, Mark H., De Jesús, Orlando. 2014. *Neural Network Design (2nd Edition*): 800. Martin Hagan.

Vang-Mata, Ruth. 2020. *Multilayer Perceptrons: Theory and Applications*, Computer Science, Technology and Applications Series. Ruth Vang-Mata, editor. Nova Science Publishers, Incorporated: 153.

Validov, M.F., Ismagilov, A.R., Voloskov, D.S., Magdeev, M.S. and Nazarov, A.A. 2017. Development of the Approach for Automatic Well Logging Interpretation for Big Number of Wells with the Use of Machine Learning. *EAGE. Conference Proceedings, Geomodel 2017*. Volume 2017: 1-6. https://doi.org/10.3997/2214-4609.201702257

Volodin, I. N., Simushkin, S. V. 2019. Lectures on probability theory and mathematical statistics. Kazan Federal University. Kazan: 347.

Simushkin, S. V. 2006. *Multivariate statistical analysis. Part 1: Study guide.* Kazan Federal University. Kazan: 98.

Simushkin, S. V. 2009. *Multivariate statistical analysis. Part 2: Study guide.* Kazan Federal University. Kazan: 114.

Khisamov, R.S., Bazarevskaya, V.G., Panina, S.A., Abusalimova, R.R., Abdrashitova, A.F., Grishanina, O.A. 2017. Assessment of reproducibility of estimate parameters based on core studies and well logging data for heavy oil reservoirs. *OIJ*. No. 6: 18-21. https://doi.org/10.24887/0028-2448-2017-6-18-21

Khisamov, R.S., Borovskii, M.Ya., Gatiyatullin, N.S. 2007. *Geophysical methods of prospecting and exploration of shallow bitumen deposits in the Republic of Tatarstan.* «Fan». Kazan: 247.

Khisamov, R.S., Sultanov, A.S., Abdulmazitov, R.G., Zaripov, A.T. 2010. *Geological and technological features of the development of high-viscosity and ultra-viscous oil reservoirs.* «Fan». Kazan: 304.