Bazhenov Fm Classification Based on Wireline Logs

D A Simonov[1], V Baranov[2] and N Bukhanov[3]

[1, 2, 3] Tomsk Polytechnic University, Tomsk 634034, Russia

E-mail: [2]BaranovVE@hw.tpu.ru

Abstract. This paper considers the main aspects of Bazhenov Formation interpretation and application of machine learning algorithms for the Kolpashev type section of the Bazhenov Formation, application of automatic classification algorithms that would change the scale of research from small to large. Machine learning algorithms help interpret the Bazhenov Formation in a reference well and in other wells. During this study, unsupervised and supervised machine learning algorithms were applied to interpret lithology and reservoir properties. This greatly simplifies the routine problem of manual interpretation and has an economic effect on the cost of laboratory analysis.

1. Introduction

Today one of the critical tasks in oil and gas geology is to study unconventional reservoirs being a source rock for tight reserves. The Bazhenov Formation is exactly a reservoir of that kind. The Bazhenov Formation is the analog of the Bakken field that is successfully developed now [1]. The commercial development of the Bazhenov Formation was proved in the 1960s, but its comprehensive commercial development has started only recently. The Bazhenov Formation is predominantly commercially developed in the central part of Western Siberia. This is a group of fields located in the territory of the Salym megabank. But even this field that is comprehensively studied with drilling has areas where well productivity is very variable. There can be a ten-time difference in oil rates where highly productive wells are located near non-productive ones. This can be explained by the highly heterogeneous structure of the formation and the absence of a unified theory of formation research. Detailed geophysical data, modern core sampling and laboratory analysis allow obtaining a lot of data about the formation. However, very often core data on many wells are absent due to their high cost. Therefore it is difficult to interpret the section of the Bazhenov Formation, especially if it refers to productive intervals consisting of carbonate and siliceous rocks. Formation interpretation is a routine requiring a technique that would help interpret lithology and petrophysical data in wells that are not characterized with core. This task is a machine learning procedure that is gaining momentum now. This procedure deals with mathematical science that can be applied to the task of interpreting formation lithology and reservoir properties in case of the absence of data.

2. Materials and methods

The task of interpreting well logging data on the Bazhenov Formation interval implies determining specific layers in the section, their lithology, a possible association in the group (lithotypes) and predicting petrophysical and geochemical properties through statistical relationships. These relationships between data can be obtained separately for each lithotype at the stage of lithotype interpretation using well logging. It deals with the fact that relationships between data, for example, between the porosity and well logging porosity methods (neutron and sonic logging), represent a cloud of points, the regression equation of which cannot characterize the entire section of the formation. This is primarily due to the extremely heterogeneous composition of the Bazhenov Formation characterized by different proportions of mineral components. Therefore, the interpretation of lithotypes is an important stage of well logging data interpretation, during which it is possible to predict the petrophysical properties and determine promising development zones on the Bazhenov Formation interval.

The first important step was to match core data with well logging diagrams. Core data matching was made on the basis of measurements taken in other areas: Salym. In this area, core was matched with depths on the basis of studying the relationship of radioactive elements such as uranium in the core samples and measuring natural radioactivity in a well, because uranium is the most significant contributor to the radioactivity of rocks in the Bazhenov Formation [2] The second step of Bazhenov Formation interpretation was using accessible log curves. Curves of gamma, neutron, acoustic, electric logs (both conventional logs at different depths of investigation and microresistivity logs) are required to determine specific layers in the formation section. Caliper is also an essential tool for Bazhenov Formation section interpretation, especially when there is a need to distinguish between dense and brittle sections confined to specific lithotypes. Caliper is sensitive to shale intervals due to their ability to demonstrate flaggyness. The methods of GR, LL, NGL and DT allowed clearly determining dense rocks associated with potential reservoir intervals. Also, these layers were determined on the basis of microresistivity logs that indicate the invasion of drilling mud filtrate into the formation. This may be due to both the natural and artificial fracturing of rock (caused by drilling) on these intervals. It is impossible to calculate and predict petrophysical properties at this step due to the absence of relationships between petrophysical values on core and well logging parameters for the entire section of the formation where the relations are a "cloud" of points without an explicit dependence. It is convenient to use relationships obtained from the interpreted groups of rocks – lithotypes (Figure 1, 2).

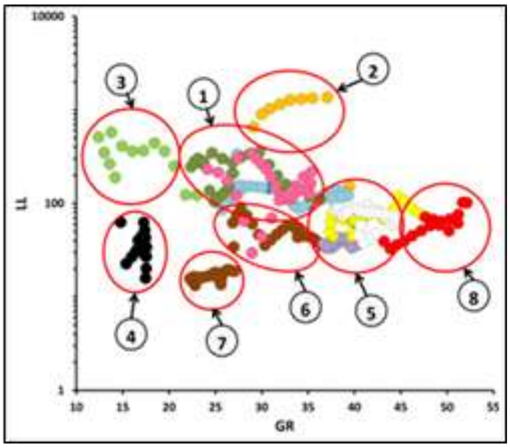

Figure 1. GR-LL plot for lithotype identification.

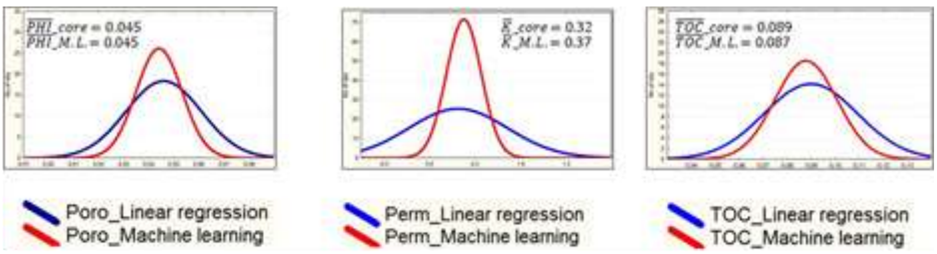

Figure 2. Different algorithm results.

Determining lithotypes at the stage of well logging analysis allows studying the interpreted layers and finding specific features on the basis of which these layers can be grouped. Only three logs of the whole complex of well log curves are essential to determine lithotypes: GR, LL and NGL. The method of specific area identification on GR-LL and GR-NGL was used to interpret the lithotypes. In this case, the NGL units should be converted into NPHI because it is inappropriate to use conventional units for neutron logs due to the different calibration of logging devices. The lithotypes are determined most clearly on the GR-LL plot, but there are areas on this plot where it is impossible to identify any lithotypes. In this case, the GR-NPHI plot should be used. The plot allows making several conclusions. As we can see from the GR-LL plot, there is a small decrease in resistivity vs. the increase in GR values. The upper part of the Bazhenov Formation is characterized by higher values of gamma ray logging and lower values of resistivity logging and vice versa for the lower part of the section. It means the increase in the carbonate content in the lower part of the formation. Higher carbonate content testifies to the greater effective thickness of the lithotypes being a potential reservoir. With lithotype interpretation, it was possible to determine such lithotypes as: 1) Siliceous argillite with low TOC; 2) Siliceous rocks (silicite); 3) Shaly-carboniferous siliceous rock; 4) Transition zone-2 (TZ-2); 5) Interbedding shaly-kerogen and siliceous rocks; 6) Shaly-carboniferous siliceous rocks; 7) Transition zone-1 (TZ-1); 8) Siliceous rocks with high TOC (radiolarite)

Core material was interpreted with the help of the data obtained during laboratory analysis which was conducted in the Field R study area. Using the results of lithological, mineralogical, geochemical and petrophysical studies is an important part of lithotype interpretation on the basis of core, and they are further used as a training sample for machine learning algorithms. In addition, core is important, as it is "hard data" characterizing a formation section. The lithotypes were identified on the basis of core using two sets of data obtained during mineralogical and geochemical laboratory analyses. The mineralogical core analysis is focused on the percentage of accessory and authigenic minerals within the Bazhenov Formation core interval. These minerals include quartz, calcite, dolomite, chlorite, illite, kaolinite, plagioclase, muscovite, alternating-layer mineral (ALM), feldspar, siderite and pyrite. Variations of mineral values that best reflect different lithologic characteristics are observed only for 4 components, such as quartz, dolomite, calcite and argillite, which are a mixture of clay minerals (illite, chlorite, kaolinite) and mica (muscovite and ALM). Core can be divided into 4 lithotypes that correspond to the well log data curves. The lithotypes interpreted on the basis of core (silicites, argillite, carbonates, transition zone) were also interpreted with the help of a geochemical analysis and mineralogical analysis. The parameters obtained from the pyrolysis of core samples, such as S1, S2, TOC, HI, OI, were used as a foundation for lithotype interpretation. It is interesting to note that parameter S2 repeats the TOC curve, which indicates the high oil-sourcing potential of the Bazhenov Formation in the area under study.

One of the well logging data interpretation problems is finding a relationship between the parameters of physical fields measured in a well during well logging and reservoir properties determined during a laboratory analysis of core. Also, it would be convenient to forecast geochemical properties, one of which being TOC, in order to determine possible future oil- and gas-bearing areas within the Bazhenov Formation

It is impossible to obtain a reliable relationship between porosity and standard types of logs measuring a physical field being responsible for the same porosity (neutron and sonic logs) on the whole interval of the Bazhenov Formation. It is noteworthy that porosity log measurements reflect the porosity of rock less effectively due to the specificity of the Bazhenov Formation. Deep resistivity data (LL) were used to find relationships between core porosity and a geophysical field. The absence of deep resistivity in other wells of the field R area section can be compensated for by the recalculation of induction log values to deep resistivity log values. The most reliable dependence is derived on the basis of rock with porosity values obtained via the thermogravimetric analysis. Due to the lack of essential data, petrophysical and geochemical relationships can be obtained only for two lithotypes: 1 and 3. The correlation coefficient for porosity and lateral log resistivity is equal to 0.78. The permeability values measured via the thermogravimetric method also correspond to the thermogravimetric porosity correlation coefficient being equal to 0.75. TOC has a good correlation only with the total gamma activity due to the uranium nature of organic content with the correlation coefficient being equal to 0.8.

Applying the theory of machine learning to the task of interpreting lithotypes and recognizing them in other wells is the second global stage of the study. The whole learning process is based on core data, since core is non-changeable evidence that directly characterizes the Bazhenov Formation section in the area under study. A training sample can be created on the basis of the lithotypes identified from core and it can be used to determine lithotypes in other wells via a reference well. It represents a transfer of the research scale from core analysis to the reference well and to the entire field. In general, the problem of machine learning is divided into many types of problems, but there are two specific objectives the algorithms of which were used in this study. These objectives comprise supervised learning and unsupervised learning. Automatic classification deals with raw data, a so-called training set. The training set is a finite set of objects or events selected from the general population of precedents to solve machine learning problems. The basis of the training sample is formed by three well logs selected as the key ones in lithotype interpretation. A class can be assigned to each lithotype. Thus, the training set is a set of classes corresponding to the lithotypes that can be identified throughout the Bazhenov Formation section.

Clustering with the k-means algorithm is the most common method used to solve unsupervised learning problems. The method is based on the fact that all raw data are divided into several disjoint groups, whereas objects from the same group have similar characteristics. The characteristics of objects from different groups differ from each other. Groups where objects have similar characteristics and differ from other objects in other groups are called clusters. Data clustering was performed by dividing clusters into 4 groups, as four lithotypes were interpreted from core. To control clustering quality, average distances in each cluster should have approximately small equal values. It testifies to high-quality clustering and the absence of a residual point. If a high average value is obtained, it means that one of the groups is characterized by unsatisfactory clustering quality. Clustering results are compared with manual core interpretation data. If the ratio of truly and falsely interpreted clusters is 50/50, it means unsatisfactory clustering.

Due to the fact that clustering shows the low convergence of the results and low quality interpretation, supervised algorithms should be used. Before the application of supervised methods, we should describe the specificity connected with the lack of data. Core was sampled only from the half-interval of the Bazhenov Formation and this core allowed characterizing only 4 lithotypes. Therefore it was decided to extend the lithotypes that were not characterized by core to the lithotypes that were interpreted from core. As a result, the Bazhenov Formation is characterized by only 4 lithotypes. It is convenient to use machine learning algorithms. These lithotypes are described in Table 3. Potential reservoirs in this case are lithotype 2 and lithotype 3. Before applying the algorithms of machine learning, the training set was divided into test and training sets to select more stable algorithms and an optimal setting for them. Train data are a range of values (in case of well R4 – GR, LL, NGL) with an assigned class. Test data are a range of values used as raw data without any classes assigned, but actually a specific class is determined for a test set. It is used to compare learning quality with training data. Quality control was performed with the ROC analysis where the quality of classification can be numerically estimated with the AUC (area under curve) parameters. If AUC is equal to 1, then the algorithm is considered high-quality. If this parameter is equal to 0.5, then the classifier does not work properly. The most stable algorithms can be selected during the test/training procedure. They are kNN, RF and NB. The CT and SVM algorithms show unsatisfactory results, as they are very sensitive to different sets of data. For supervised learning, about 50-60% of the data from the entire training set can be sufficient for reliable learning. Also, more sensitive parameters having a high influence on the area under curve were selected for each algorithm. When the algorithms are set up, machine learning can be used for the reference well and then for other wells of the field.

The kNN algorithms show the best machine learning results in terms of percentage (74.1% of the truly classified parameters). Also, they ensure more reliable visual interpretation. Therefore, the kNN algorithms should be used to apply machine learning in other wells. Other wells should be adjusted to the reference well in order to apply the machine learning procedure. Further, the kNN algorithm should be used. For lithotypes 2 and 3, it is possible to create a thickness map and determine a potential zone for further drilling. Using the algorithms of machine learning, it is also possible to distribute petrophysical and geochemical properties. A porosity and permeability map was constructed for lithotypes 2 and 3. A TOC map can also be obtained. It is possible to perform quality control for machine learning of petrophysical and geochemical property prediction by comparing values obtained via machine learning with values obtained with the help of the linear regression prediction of properties. Linear regression relationships, as it has already been mentioned, were obtained only for the interval of lithotype 3. Figure 2 shows that the machine learning procedure makes a reliable prediction of properties, as the average values for linear regression prediction and machine learning are almost the same.

3. Conclusion

The following results were obtained during the study:

1. Main lithotypes were interpreted on the basis of well logging and core data. Eight lithotypes were obtained from well logging data, whereas core data allowed obtaining only 4 lithotypes. The remaining lithotype intervals that were not characterized by core were extended to the interpreted core lithotypes.
2. A machine learning procedure was applied. First, optimal settings for the algorithms were selected. Second, the authors determined the most stable algorithm showing good automatic interpretation results.
3. The automatic interpretation of the Bazhenov Formation was performed for other wells. Lithotypes with the highest development potential were identified. A potential reservoir map was created. Promising zones for drilling were selected.
4. A machine learning procedure was used to predict petrophysical and geochemical properties. Porosity, permeability maps for the potential reservoir and a TOC map for all fields were created. Promising zones were selected.
5. The analysis of petrophysical and geochemical data prediction was made. Machine learning techniques allowed obtaining reliable results.

References

[1] Alekseev A D 2014 The Bazhenov Formation: In Search of Big Shale Oil in Upper Salym *Rogtecmagazine*
[2] Pavlova M A, Kamkina A D and Suhorukova K V 2012 *Geology-Geophysical Model of the Bazhenov Formation* Report (Novosibirsk)
[3] Bukhanov N, Belozerov V B, Demyanov V V and Evdokimov A V 2015 Classification of Bazhenov Formation Using Seismic Traces, EAGE/SPE Joint Workshop *Exploration of Shale Oil Resources and Reserves* (Moscow)
[4] Rotatu A and Rukavishnikov V S 2015 Description and Comparison of Bazhenov Reservoir Models *EAGE Conf.*(Tyumen)