LITHOFACIES CLASSIFICATION USING BAYES THEOREM METHOD: CASE STUDY WESTERN DESERT, EGYPT

Abdelrahman Moataz Mohamed Gomaa
PhD candidate, University of Miskolc, Institute of Geophysics and Geoinformatics
3515 Miskolc, Miskolc-Egyetemváros, e-mail: mh.geophysicist@sci.asu.edu.eg

Abstract

This paper shows the availability of using the Bayesian classification method to predict class membership probabilities in one of the deep tight reservoirs in Western Desert, Egypt. The workflow of our project that using the Bayesian method used the deterministic petrophysical results of three training wells to train the data and extract the classifiers. The classified data were modeled using Gaussian distribution for each lithofacies. The used wells were acquired from a deep Jurassic gas reservoir in the Western Desert of Egypt. The fitting between actual and modeled data has been reached by minimizing the L2 norm. Besides, a cross-validation process was used for validating the resulted classifiers. Finally, the Bayesian classification method can predict the GWC with an accuracy of 4 m. To avoid probability interference caused by the compacted shale more data should be added to the initial model.

Keywords: Bayesian Classification, lithofacies prediction, formation evaluation

1. Introduction

The main objective of this paper is to classify the different rock lithofacies using the Bayesian Classification method and correlate the predicted results to the petrophysical laws based on calculated results at a complex gas field. Four wells contain Quadro compo logs were used in this study. In this research, the input data used for training wells were derived from Abdel-Moktader A. El-Sayed, et al, (2017), where it shows that the Jurassic sandstone reservoir which called (Lower Safa member) is considered a heterogeneous tight reservoir due to the effect of diagenetic processes.

The gas field is located in the Northern part of the Western Desert of Egypt, it is far from Cairo by approximately 450 km in the WNW direction. The field is situated in the Western flank of the Matruh basin and it is considered the main Jurassic gas and condensate asset in the northern part of the Western Desert of Egypt, about 50 kilometers south of the Mediterranean coast. Figure 1. shows the location of the study area.

Several studies implied that the northern part of Egypt is more complex than the southern part. So, they divided Egypt according to structure stability into four structural divisions, which are the Hinge Zone, Unstable Shelf in the north, Stable Shelf, and the Nubian/Arabian Cratons in the south (Said, 1962). The exploration activity has been largely concerned with the northern area (approximately north of 28° N latitude) that known as the Unstable Shelf.

Figure 1. *Location of the study area.*

The stratigraphic column shows that the Western Desert consists of four transgression sequences each sequence is alternative beds from Carbonate and clastic. The most important Jurassic reservoir is Khatatba Formation where several researchers divided into four members: Zahra, Upper Safa, Kabrite, and Lower Safa. Figure 2 shows the stratigraphic column of the Western Desert. Ghanima et al. (2015) said that Upper Safa is considered as a source rock of Lower Safa gas reservoir, while Kabrite is a seal rock, Lower Safa reservoir composed of sandstone and shale intercalations.

2. Basic Concepts

As with any classification technique, this paper used the cross-validation process and started using the deterministic petrophysical analyses for training the data and extract the classifiers then applied into the target.

2.1 Petrophysical Evaluation

The petrophysical analyses were carried out analytically and graphically for the training wells by using different equations and standard charts. They include the following steps:

2.1.1   *Zonation*

The first step after data editing is to detect the depths of interests' zones by identifying the response of the wireline logs like the gamma-ray log (GR), compressional sonic log, and density log for lithology identification. The zonation step may be using the formation picks information from the well site engineer.

2.1.2   *Shale Volume Calculation*

Shale volume can be considered as the most important petrophysical parameters because it is the backbone of all the further petrophysical calculations. Methods of shale volume calculation are single

indicator methods such as gamma-ray log method and/or double indicator methods such as Neutron – Density logs method.
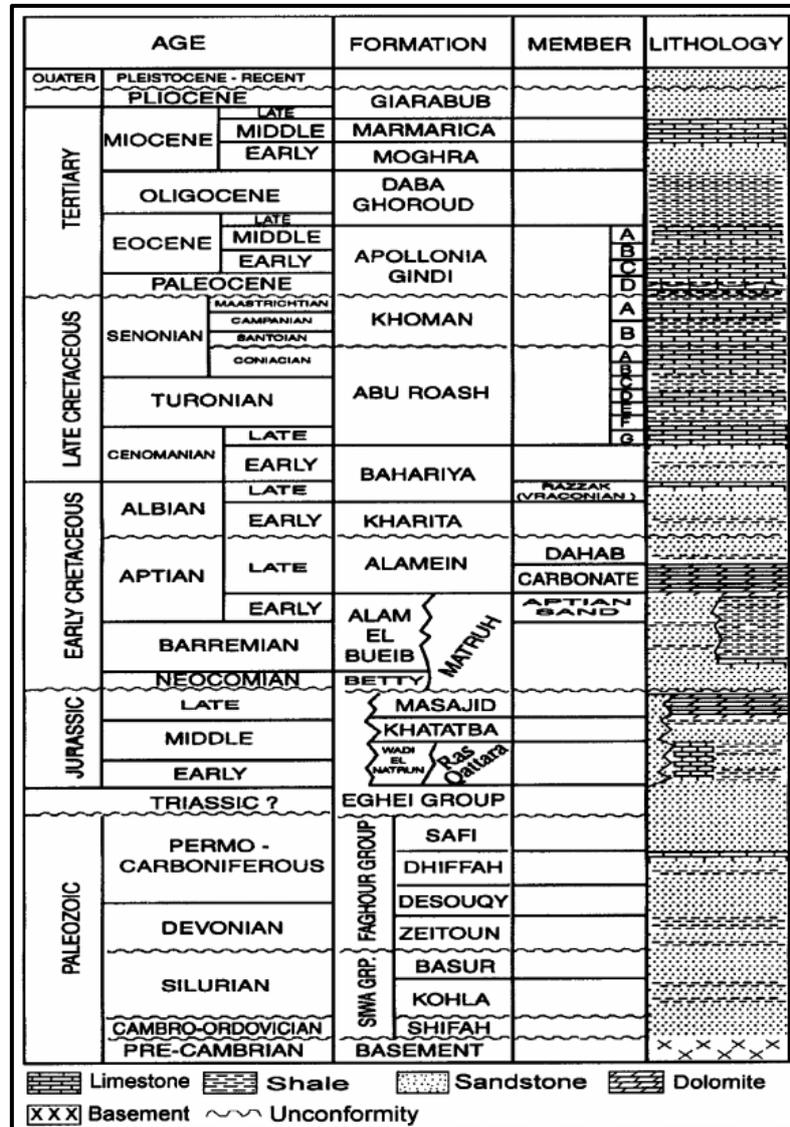
AGE			FORMATION	MEMBER	LITHOLOGY
QUATER	PLEISTOCENE - RECENT				
	PLIOCENE		GIARABUB		
TERTIARY	MIOCENE	LATE	MARMARICA		
		MIDDLE			
		EARLY	MOGHRA		
	OLIGOCENE		DABA GHOROUD		
	EOCENE	LATE	APOLLONIA GINDI	A	
		MIDDLE		B	
		EARLY		C	
	PALEOCENE			D	
LATE CRETACEOUS	SENONIAN	MAASTRICHTIAN	KHOMAN	A	
		CAMPANIAN			
		SANTOIAN		B	
		CONIACIAN			
	TURONIAN		ABU ROASH	A	
				B	
				C	
				D	
				E	
				F	
	CENOMANIAN	LATE	BAHARIYA	G	
		EARLY			
EARLY CRETACEOUS	ALBIAN	LATE		RAZZAK (VRACONIAN)	
		EARLY	KHARITA		
	APTIAN	LATE	ALAMEIN	DAHAB	
				CARBONATE	
		EARLY	ALAM EL BUEIB	APTIAN SAND	
	BARREMIAN				
	NEOCOMIAN		BETTY		
JURASSIC	LATE		MASAJID		
	MIDDLE		KHATATBA		
	EARLY		WADI EL NATRUN		
	TRIASSIC ?		EGHEI GROUP		
PALEOZOIC	PERMO - CARBONIFEROUS		SAFI		
			DHIFFAH		
			DESOUQY		
	DEVONIAN		ZEITOUN		
	SILURIAN		BASUR		
			KOHLA		
	CAMBRO-ORDOVICIAN		SHIFAH		
	PRE-CAMBRIAN		BASEMENT		

Limestone ▦ Shale ▤ Sandstone ▦ Dolomite ▦
XXX Basement ∼∼ Unconformity

Figure 2. Lithostratigraphic column of the North-Western Desert (El Nady, 2008).

2.1.3   Single indicator method (The Gamma-Ray Log)

The analysis of the intensity of the gamma-ray log is the most common method used for shale volume calculation using the following equation:

$$(V_{sh})_{GR} = \frac{(GR_{log} - GR_{min})}{(GR_{max} - GR_{min})}, \tag{1}$$

where, (GR_{log}) is the reading in the zone of interest, (GR_{min}) is the minimum reading in clean sands or carbonates, (GR_{max}) is the maximum reading in nearby 100% shale, and (V_{sh})$_{GR}$ volume of shale.

2.1.4 Double Indicator method (Neutron and Density Logs (ØN and ØD))

The decreasing pattern in the density log and the increasing pattern in the neutron log may be a result of shale presence. The volume of shale can be calculated using the following equation:

$$V_{sh} \leq \frac{\emptyset_N - \emptyset_D}{\emptyset_{Nsh} - \emptyset_{Dsh}}, \tag{2}$$

where $\emptyset_N$ is the porosity from the neutron log, $\emptyset_D$ is the porosity from the density log, and $\emptyset_{Nsh}$ is the neutron porosity in front of a shale bed, and $\emptyset_{Dsh}$ is the density porosity in front of a shale bed.

2.1.5 Effective Porosity Calculation

The relationship between porosity and density can be used for effective porosity calculations by using the following equation:

$$\emptyset = \frac{\rho_{mat} - \rho_{log}}{\rho_{mat} - \rho_f} - V_{sh} * \frac{\rho_{mat} - \rho_{sh}}{\rho_{mat} - \rho_f}, \tag{3}$$

where ρ_{mat} is the density of the matrix, ρ_{log} is the reading value of density log, ρ_f is the density value of fluid, ρ_{sh} is the reading value of density log at a thick representative shale zone.

2.1.6 Water Saturation Calculation

There are many methods to estimate the water saturation in the reservoir, this study used mono porosity cross plot of Pickett (Pickett, 1972) for Archie equation, where Archie equation depends on the resistivity of formation water (R_w), saturation exponent (n), tortuosity (a) and cementation exponent (m).

$$S_w = \left(\frac{a \, R_w}{\emptyset^m \, R_t} \right)^{\frac{1}{n}}. \tag{4}$$

2.1.7 Initial Gas in Place

Reserve estimation and Initial gas in place calculations have been carried out by the following equation:

$$G = 43560 * \emptyset * (1 - Sw) * H * A, \tag{5}$$

where, G = initial gas in place, SCF, $\emptyset$ = effective porosity, fraction, S_w = initial water saturation, fraction, H = Productive interval thickness, ft, and A = drainage area, acres.

 The lithological description of the interesting intervals within the wells can be detected graphically using neutron-density cross plot, M-N cross plot, and MID cross plot.

2.2 Bayesian Classification

The Bayesian classification method is one of the classification techniques that deal with difficult formulating problems (Jiawei Han, et. al., (2012)). For example, identifying diagnoses for patients who start with a series of certain symptoms, determining customers' creditworthiness according to their

credit history and demographics. In each of these problems, the computing of a label for each analyzed part of data that depends on the properties of the data is the challenge (Berger, J. 2006).

According to (Wang, S. and Campbell, B. (2013)) and (Robbins, H. (1956)), Bayesian classification is based on the Bayes theorem that needs a strong independence assumption between the different attributes. Abstractly, naive Bayes is a conditional probability model.

From the conditional probability theory, we can define the probability of event *A* to occur in the condition of *B* is happening as:

$$P(A \mid B) = \frac{P(A \cap B)}{(P(B))}, \tag{6}$$

in the same way, we can have

$$P(B \mid A) = \frac{P(A \cap B)}{P(A))}, \tag{6}$$

from a previous couple of information, we can say that

$$P(A \mid B) \times P(B) = P(B \mid A) \times P(A), \tag{7}$$

hence,

$$P(A \mid B) = (P(B \mid A) \times \frac{P(A)}{P(B)}, \tag{8}$$

by redefining equation (9) for borehole evaluation,

$$P(Sand|GR) = \frac{P(GR|Sand) \times P(Sand)}{P(GR)}, \tag{9}$$

In real-world applications, we may need to incorporate not only one attribute as information to update our guesses, but we may also need more than one attribute. Naïve Bayesian classification theory assumes that the posterior probability knowing N of information can be calculated by multiplication of Bayesian probabilities of the attributes series as:

$$P(SD|I_1, I_2, \dots, I_N) = P(SD) \times P(I_1|SD) \times P(I_2|SD) \times \dots \times P(I_N|SD). \tag{10}$$

This equation is well known as the Bayesian classification rule. The posterior probability of event A to occur with knowing B as a fact is the product of the prior guess of *A* and the conditional probability of B with A happening divided by the probability of *B*. Naïve Bayesian classification theory assumes that the posterior probability knowing N of information can be calculated by multiplication of Bayesian probabilities of the attribute's series. Then, for each attribute a probability distribution curve will be constructed based on each lithofacies, figure 3. shows the probability distribution curve of the different attributes based on lithofacies distribution. The data modeled using the Gaussian equation.

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \tag{11}$$

where *x* is the attribute value, μ is the mean, and σ is the standard deviation
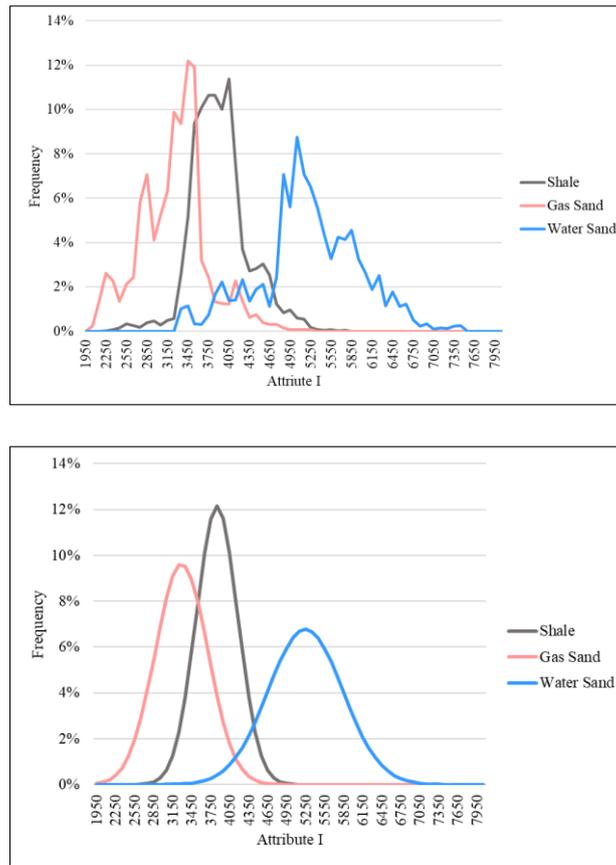
Figure 3. *The actual and modeled probability curve.*

The inversion problem has been solved by minimizing the L_2 -norm of the difference between the measured and modeled log data in front of each lithofacies as a percentage of misfit calculated (L_2-norm) using the following equation:

$$L_2 = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - E_n)^2},$$ (12)

where the x_i data is the modeled and the E_n is the log measured data, and the letter *i* refer to an *i*th log value of each attribute.

3. Results and Discussion

3.1 Deterministic Petrophysical Evaluation

Wireline log analysis gives information about the reservoir lithology description, petrophysical parameters calculation, and lithofacies prediction. Figures 4-7. show that the reservoir is a thick sandstone layer with shale interbeds, where sandstone layers have low Gamma Ray (GR) and low neutron log response.

According to the graphical and analytical calculations, the lithology of the interested member could be divided into three units (A-B), C, and D. Where units from old to young consist of pure sandstone, overlying by shale bed and lately covered by shaly sand layer. It is noted from log responses and neutron density cross plot that, the sand units containing gas saturation have a lower density than the sand of water saturation, figure 4. shows the gas effect on the neutron –density cross plot. The reservoir interval shows a high percentage of porosity, except unit C, and show that the water saturation increases at depth (TVDss) of 3870 m. Figure 5. shows that the reservoir interval has a low intensity of gamma-ray log that reflects the sandstone log response, while figure 6. shows the neutron-density cross plot and the different rock units A, B, C, and D, where the units A and B are composed of sandstone rocks with gas presence, unit C is a shale bed, and unit D is composed of sandstone rocks with water presence. Figure 7. shows the final interpreted logs of the reservoir interval (composite log) of one of the training wells.
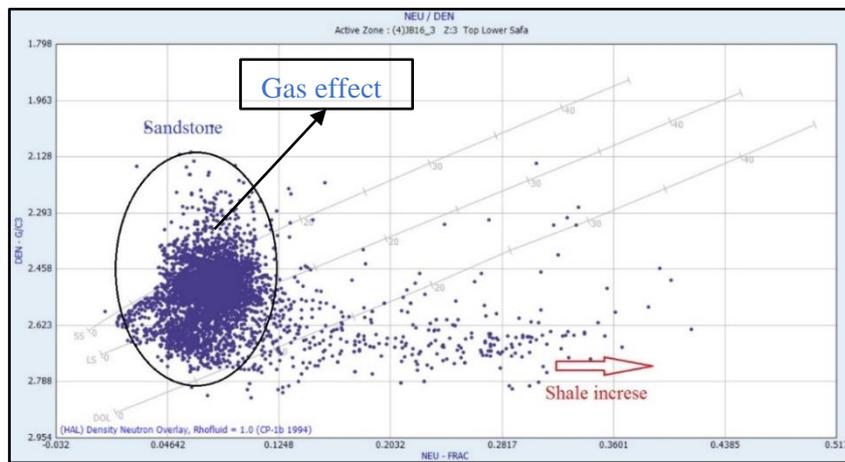
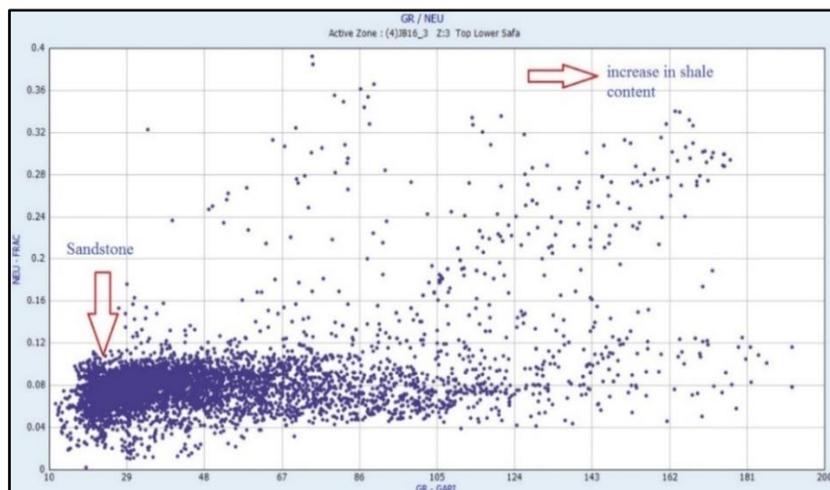**Figure 4.** *Lithology identification with neutron - density cross plot of the reservoir interval.*

Figure 5. *Lithology identification with neutron porosity- GR cross plot of the reservoir interval.*
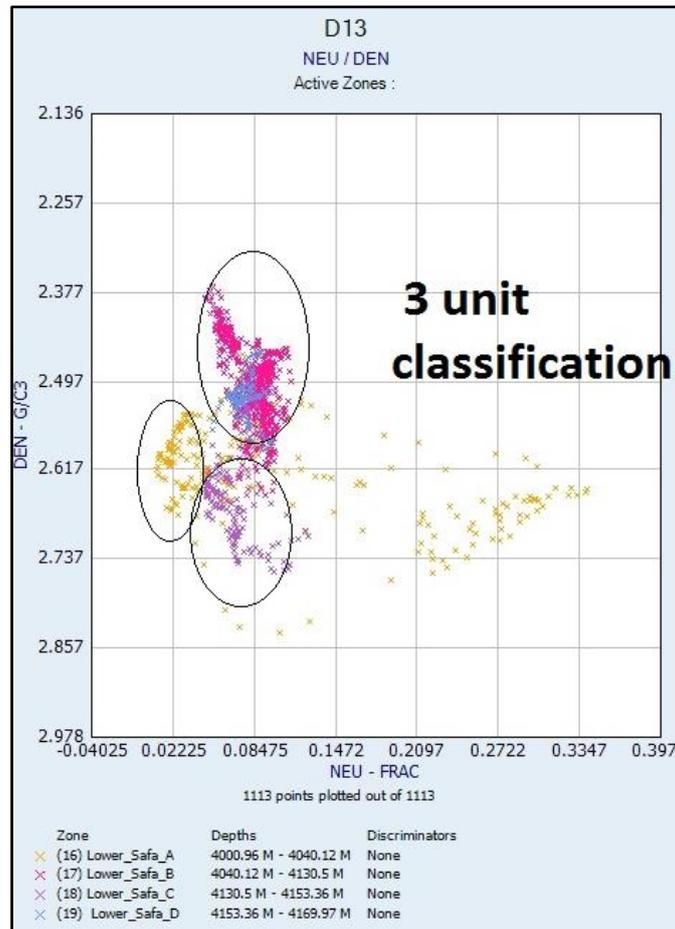
Figure 6. *Neutron-density cross plot with units A, B, C, and D.*

3.2 Bayesian classification

The Bayesian classification technique was carried out using the cross-validation method by using three training wells and one target well. Training wells were divided into three main lithofacies or attributes (gas-sand, water-sand, shale) depending on the previous deterministic evaluation. Thus, constructed a distribution curve of the filtered data of each attribute.

According to the Bayesian classification theory, the knowledge of the distributions of feature vectors and the prior probabilities of the classes are the basis of obtaining the optimal classification decision. Consequently, each lithofacies was modeled by using the normal Gaussian distribution equation with a mean μ and standard deviation σ, so that the modeled Gaussian distribution curve was constructed to fit the measured data. For sake of the comparability, the Gaussian distribution curve needs to be scaled using a scaled factor for each attribute. Determining the scaled factor relied on the try and error method to obtain the best fit between the two types of data. Graphical fitting of the distribution curves for actual and modeled data for GR, neutron, and density logs for gas-sand lithofacies (attribute) figure 8.
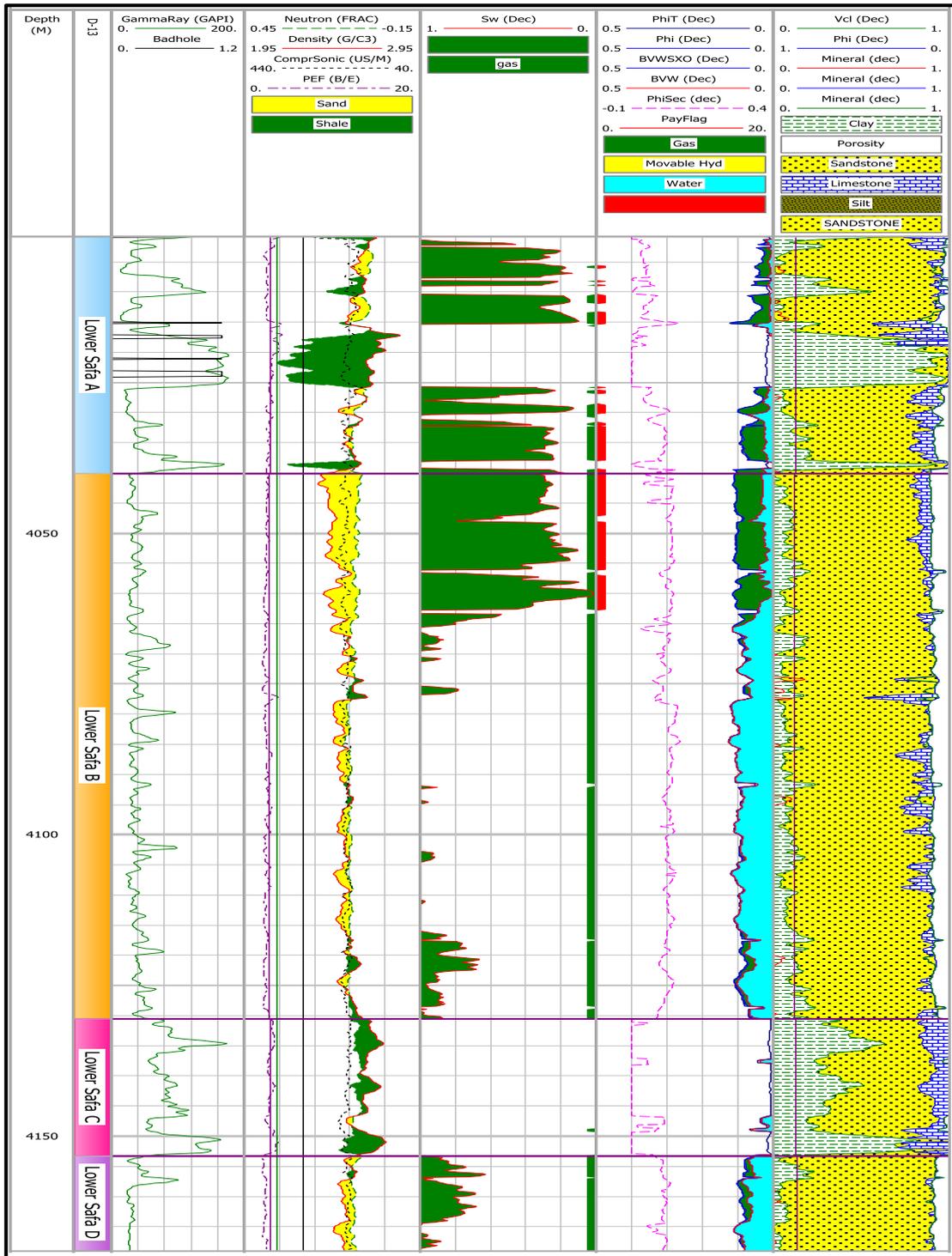
Figure 7. *Litho-saturation cross plot of the reservoir interval.*
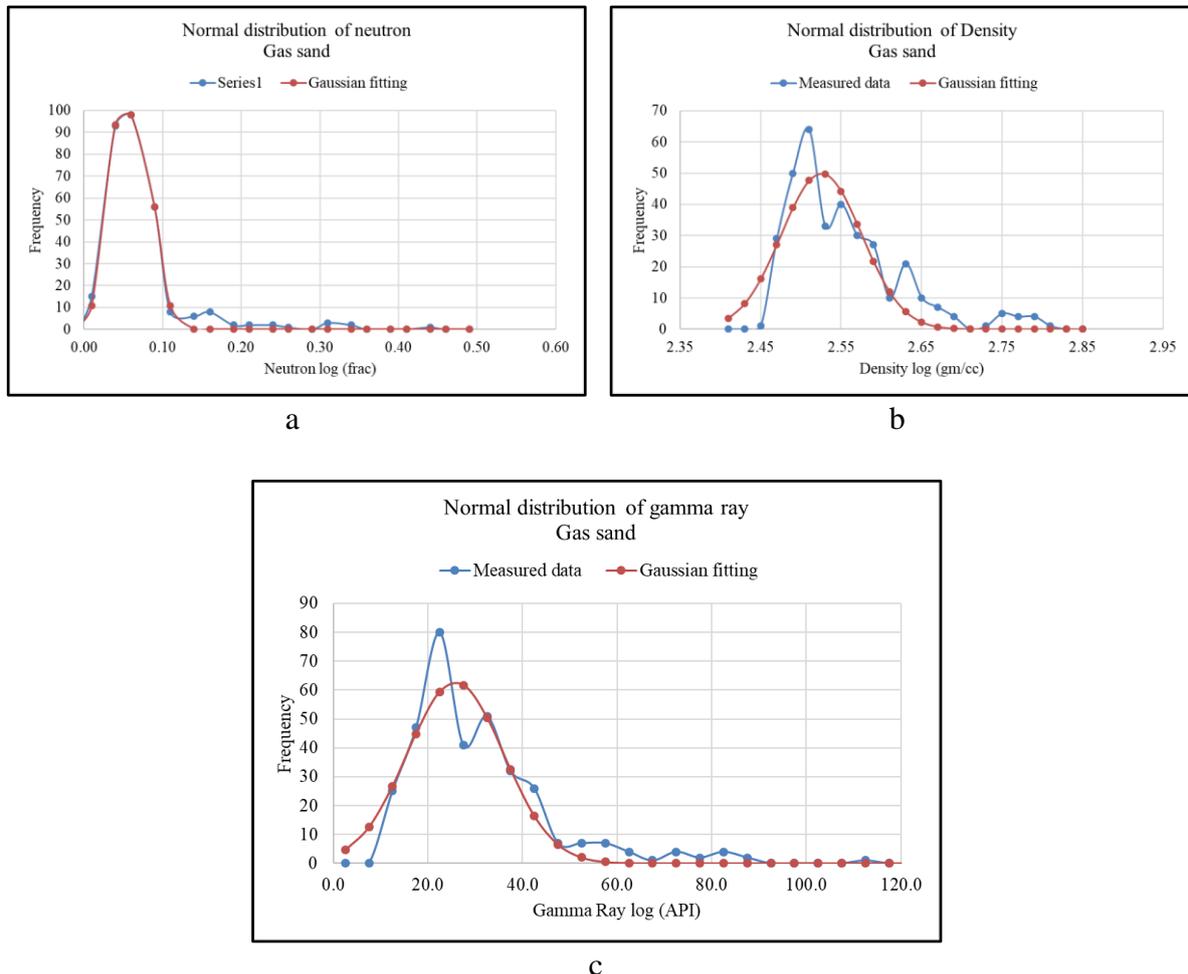
a

b

c

Figure 8. *The distribution curves for actual and modeled data for GR (b), neutron (a), and density (c) logs for gas-sand lithofacies (attribute).*

To predict the lithofacies of the density log, P (Den|SDg) P (SDg) is evaluated for each lithofacies class, where SDg is the gas-sand lithofacies class. The classifier predicts that the lithofacies of the desired log value are one of the three lithofacies classes if and only if

$$P \ (Den \ |SDg) \ P \ (SDg) > P \ (Den \ |SDw) \ P \ (SDw),$$

where SDw is the water sand lithofacies class. In other words, the predicted attribute label is the attribute i for which P (Den |SDg) P (SDg) is the maximum. Figures 9. and Figure 10. show the applied classifier results on the training wells.
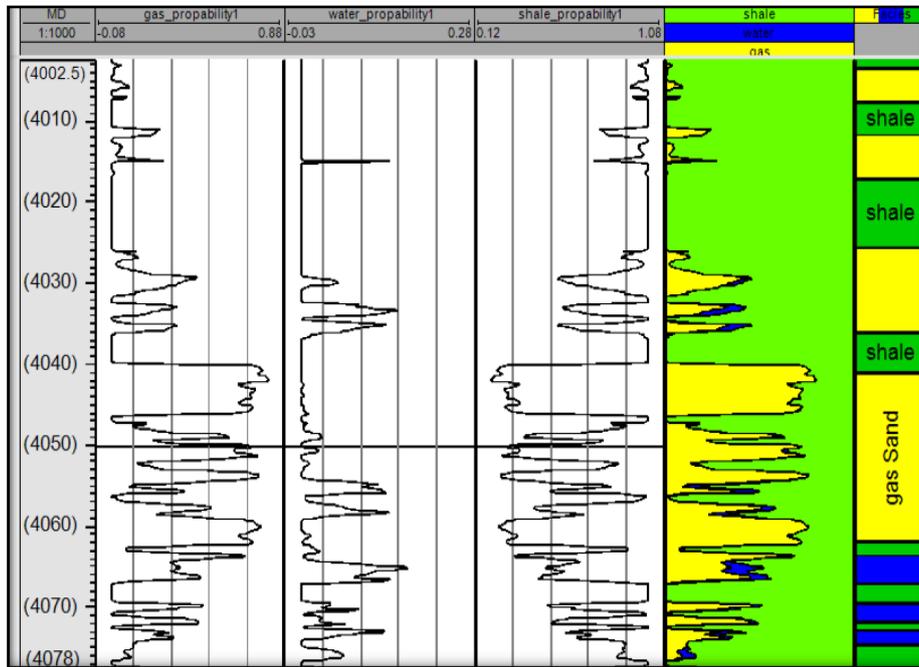
Figure 9. *Bayesian Classification results for training well no.1.*
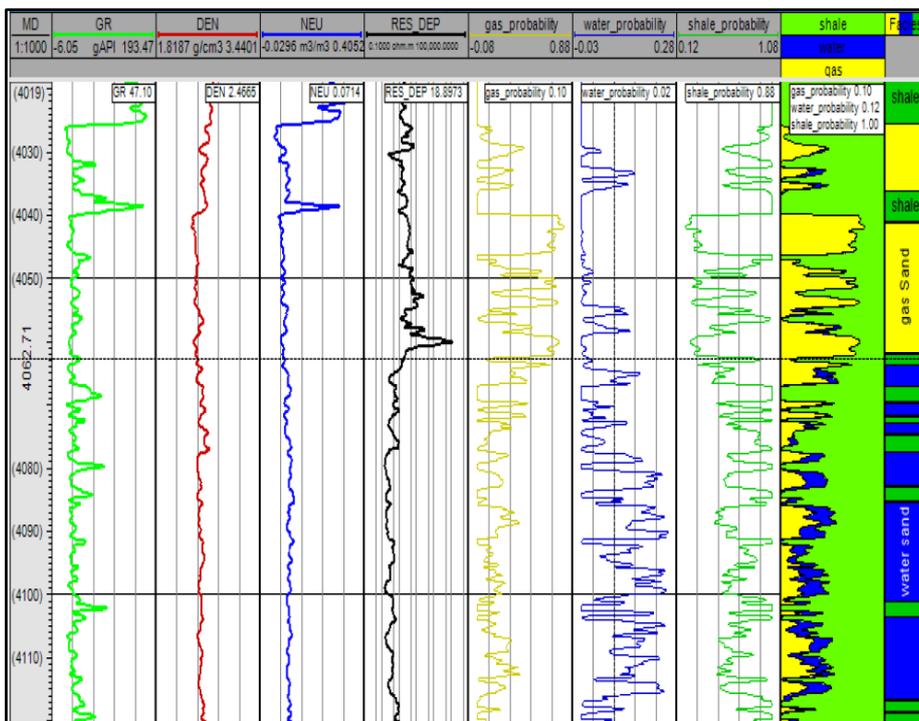

Figure 10. *Bayesian classification results for training well II.*

Figure 11. shows the difference between the facies log of the deterministic evaluation and Bayesian classifiers of the Bayesian classifiers on the training well *II*. The classifiers could distinguish between different lithofacies but with certain accuracy where there are some interferences with water-sand and gas- sand lithofacies which may be resulted from the fact that this reservoir is a deep tight reservoir that causes a change in the behavior of the log in front of the sand reservoir interval. Besides that, the overburden weight of the overlying layers forms a very compacted shale that causes an increase in shale density that affects the reading of the density log and shows no noticeable change in the reading of density log in front of the compacted shale and sand intervals. Furthermore, the classifiers could predict the presence of the gas-water contact (GWC), where it is located at a true vertical depth (TVD_{ss}) of 3873 m and the actual depth is 3869 m.

Figure 12. shows the results of using the Bayesian classifiers on the target well with reasonable accuracy for facies prediction but still, the effect of the compacted shales can be seen in certain areas.
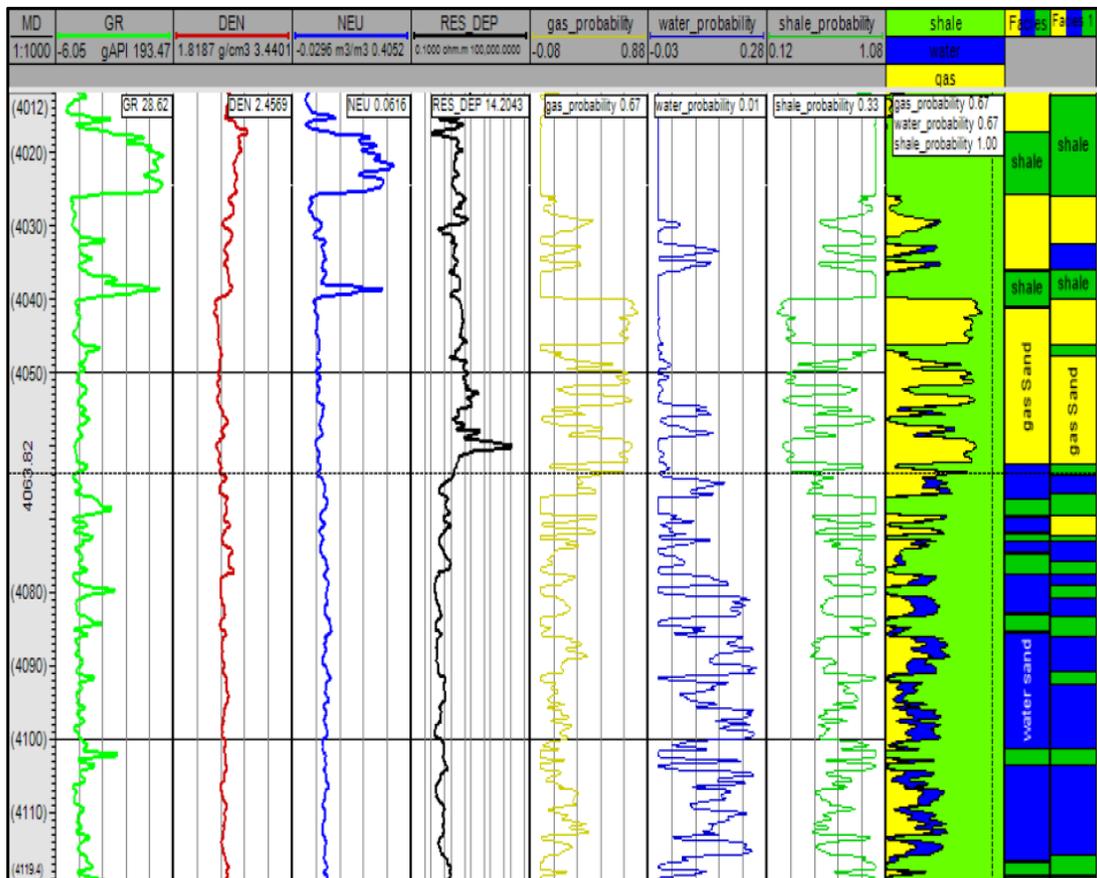

Figure 11. *The differences between facies logs of deterministic valuation (left facies log) and Bayesian classifiers (right facies log).*
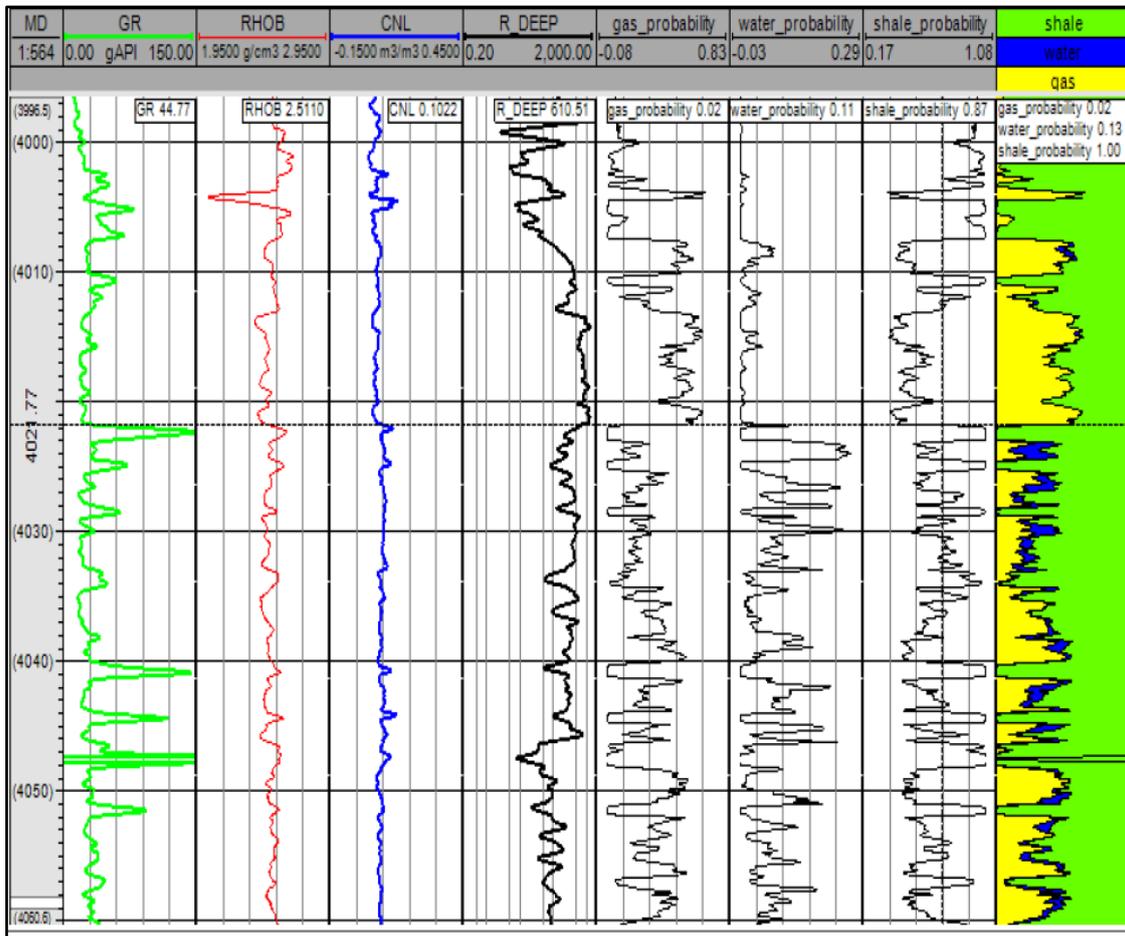
Figure 12. *Bayesian Classification results for target well.*

4. Conclusion

This paper shed the light on the rule of the Bayesian classification method in fluid and lithofacies prediction of a deep tight reservoir. Moreover, the extracted classifiers reduce the workload of humans to discriminate between different lithofacies concerning different well-log data. The Bayesian classifiers could predict the GWC with an accuracy of 4 m. The effect of the compacted shales is clearly shown in the results of a classification in front of the shaly areas in the shape of pseudo-water probability values. Therefore, the classification procedure needs more classification to initial lithofacies of the initial model to avoid probability interference.

The deviation from the normal distribution in certain good log data could be avoided by increasing the number of input data, which could enhance the accuracy of the classifiers and provide the initial model of the classifiers with more data in front of various rock qualities. The results provide that using the Bayesian classification can be an aid in the evaluation process for deep tight reservoirs and for predicting lithofacies change and by increasing the input data the initial model will be more robust.

5. References

[1] Abdel-Moktader A. El-Sayed, et al, (2017): *Building a Reservoir Static Model of the Lower Safa Member at El-Obaiyed Field, North Western Desert, Egypt,* Egyptian Geophysical Society journal (EGS),

[2] M. M. El Nady (2008): *Biomarkers Assessment of Crude Oils and Extracts from Jurassic-Cretaceous Rocks, North Qattara Depression, North Western Desert, Egypt,* Petroleum Science and Technology, 26:9, 1063-1082, **https://doi.org/10.1080/10916460701208496**

[3] Ghanima, Ahmed et. al., (2015): *An Integrated Petrophysical, Geochemical and Operational Approach to Identify Unconventional Resource, Case Study, Obaiyed Field, Western Desert-Egypt*, **https://doi.org/10.2118/175744-MS**

[4] Han, J., Kamber, M., & Pei, J. (2012): *Data mining: Concepts and techniques, third edition* (3rd ed.). Waltham, Mass.: Morgan Kaufmann Publishers. ISBN 978-0-12-381479-1.

[5] Said, R. (1962): *The geology of Egypt*, Elsevier Publishing Co., Amsterdam, New York, 337 p.

[6] Pickett, G. (1972): *Practical formation evaluation*, Golden Colorado.

[7] Berger, J. (2006): The case for objective Bayesian analysis. Bayesian Anal. 1, 385−402 (electronic). **https://doi.org/10.1214/06-BA115**

[8] Robbins, H. (1956): An empirical Bayes approach to statistics. In Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954−1955, Vol. I, 57−163. University of California Press, Berkeley and Los Angeles.

[9] Wang, S., and Campbell, B. (2013): Mr. Bayes goes to Washington. Science 339, 758−759. **https://doi.org/10.1126/science.1232290**