Mathematical analysis of synthetic bulk density logs around the world- Part I- the use of linear regression

Mayra D. L. Carrasquilla, Caique P. Carvalho, Marcus. Danilo F. B. Costa, Carolina B. da Silva, José J. S. de Figueiredo, Igor J. S. Souza, Carlos E. Amanajas, Celso Rafael Lima, João Rafael S. Silveira and Lucas Rautino (CPGf-Federal University of Pará)

This paper was prepared for presentation during the 17th International Congress of the Brazilian Geophysical Society, held in Rio de Janeiro, Brazil, 16-19 August 2021.

Abstract

The density rock is a property generally used to identify rock types and, in geological studies, being a physical property affected by geological time, pressure, and temperature. The bulk density is a crucial petrophysical property in hydrocarbon exploration, geology, and several other fields within geophysics. This attribute is helpful in compositional, lithological, and stratigraphic studies. This article focuses on the evaluation using linear methods for calculating bulk density proposed by some authors in the universal literature. In the first part of this work, empirical models were solved numerically using least squares regression. In this case, the density was estimated from V_{sh} and V_p logs in a model and from V_p only for another model. The regression parameters (coefficients) for these were analyzed from aspects mathematical, geophysical, tectonic, geological, and geochronological. It was performed the validity of these mathematical models on data from oilfield wells in various regions of the planet, differently from the area in which the models were generated. The results show that t e error between the estimated density and the actual density increases when we extrapolate the empirical equation from one field to another. In the same field, the mean squared error (MSE) is less than 2 % fit between real and synthetic logs for both models. This error can increases to 8 % (for NPRA field-Alaska) when the age and composition of rock differ significantly from the geology of the well-log of reference (Araki field).

Introduction

It is well known that rock density (bulk) is an important petrophysics parameter in the hydrocarbon oil exploration. In the seismic context, the density with compressional velocity is used to obtain the impedance and consequently the seismic reflectivity.

The lack of density logs in some wells is quite common to find, either by omission of measurement or by the conditions in which the well and/or the tool presents that will make it difficult to take the density measurement of the rocks. To find a mathematical model that allows the reproduction of the density log based on other logs measured in the well, in order to be a support for the interpretation and correlations of the existing rocks and fluids present in the well, is also of great help for geological, geophysical studies and for the oil industry in exploratory fields.

The density log tends to contain a large percentage of error in its measurement due to the sensitivity of the tool and the general condition of the well's walls. The Vertical resolution is 18 inches ($45,72$ cm) for the density measurement (Serra and Serra, 2004), which is a deficit for correct density measurement, it is also affected by the filtration of the drilling mud, variation of the diameter of the well, crumbling and roughness of the well's walls (Figure 1). It is useful to propose a mathematical method that allows the validation of data and generate a quality control of them. The implementation of this model allows the correction of problematic logs, cleaning and adjusting these from noise interference and erroneous data measured in the well, in order to guarantee the quality and veracity of the studies directly done in the well from the evaluation of the information obtained in these.

There are mathematical models for obtaining bulk density logs proposed by different authors, in which using logs such as P-wave velocity (V_p) and shale volume (V_{sh}), it is possible to calculate a synthetic bulk density log. Sometimes the results of one of the models are more efficient than the other depending on the case and the behavior of the data in the study area.

For the realization of this study, a large number of data was taken from wells belonging to the oil fields of Taranaki−New Zealand (Science, 2012), Teapot Dome−USA (DNPOSR, 1976; Energy-Department, 2011), The National Petroleum Reserve in Alaska−NPRA (Houseknecht and Schenk, 2001; Saltus et al., 2002), the Niger delta basin (Reijers, 2011), Campos Basin−Brazil (Guardado, 1990) and Viking Graben field in the North Sea−Norway, which have age ranges among the lower-Mississipian to later-pliocene/holocene. Evaluating the different mathematical approaches to obtain bulk density is intended to analyze the influence of geological, tectonic and geophysical aspects of each field on the values of the resulting parameters and at the same time determine the effect of applying the parameters of one field in others fields with different characteristics.

Theoretical Background

The Least Squares method

The theory of least squares is a numerical analysis method where, by providing a set of variables (both dependent

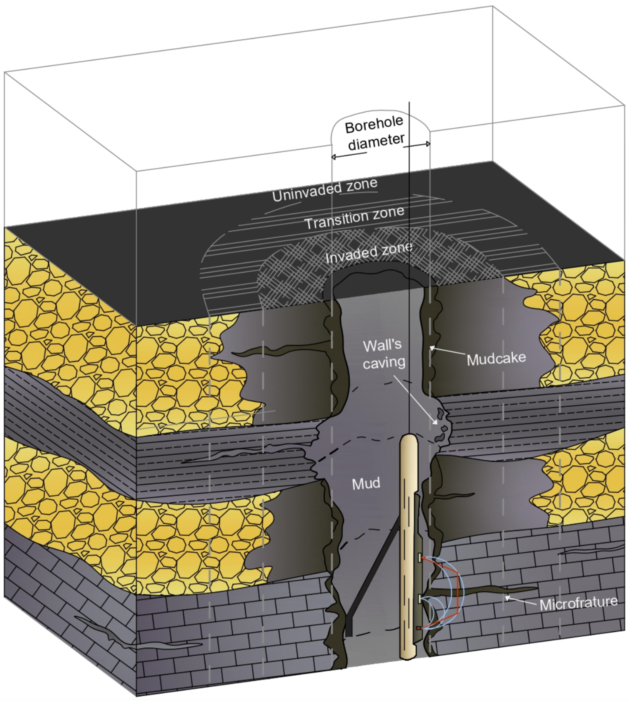

Figure 1: Degradation of the formation during and after drilling. Overpressured mud is indicated to be invading porous and permeable sand formations with the formation of a mudcake. The mud circulation also causes borehole washout in the shale zones (Ellis and Singer, 2007).

Figure 2: World map showing the general location of the six oil fields studied in this work, which are located in different places around the world, giving a global approach to the research.

and independent) and a group of functions, the aim is to determine a continuous function -within those provided- that shows the greatest adjustment with the data provided by the criterion of minimum square error (Leon et al., 2006). When an attempt is made to find a set of coefficients c_i[1] where $f(x)$ interpolates with the set $\{(x_j, y_j)\}_{j=1}^{n}$ - being (x_j, y_j) a set of points in the Cartesian plane - then this is fulfilled:

$$\sum_{i=1}^{m} c_i f_i(x_j) = y_j, \tag{1}$$

[1]where i=1,2,3,...,N

Empirical models

Having knowledge of the equation for the calculation of bulk density proposed by Gardner et al. (1974):

$$\rho_b = k V_p^b, \tag{2}$$

where ρ_b is the bulk density, V_p is P-wave velocity and k and b are constants.

As well as the modified Birch's Model (Birch, 1960; Kowallis et al., 1984; Tosaya and Nur, 1982; Castagna et al., 1985; Han et al., 1986; Oloruntobi and Butt, 2019):

$$\rho_b = QV_p + ZV_{sh} + P, \tag{3}$$

where ρ_b is the bulk density, V_p is P-wave velocity, V_{sh} is the shale volume and Q, Z and P are constants.

Method

The workflow of our case study is divided into five (5) steps. Its description is pictured in Figure 3. According to step one (1), we started selecting well-log data set from five (5) regions spread in different locations on the planet. In other words, the well-logs data is from Taranaki (New Zealand), Niger Delta (Nigeria), Alaska National Oil Reservoir (USA), Teapot Dome (USA), Viking Graben (the North Sea - Norway), and Campos Basin (Brazil) oil fields (see Figure 2). We also performed - in step(1) - a pre-processing of well-logs. Here, we removed the random noise elimination (despiking), mud infiltration correction on the density log (based on the caliper log), and shale volume estimation (from gamma-ray index (IGR) log).

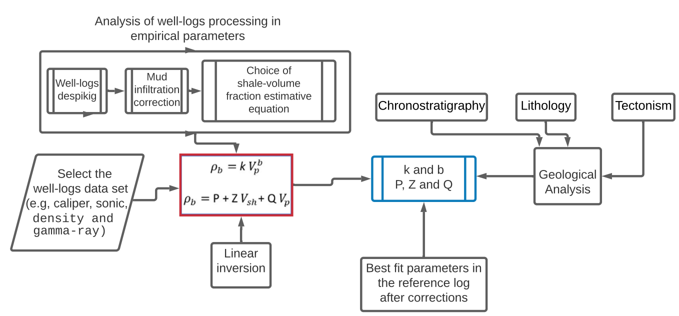

Figure 3: Flow-chart showing the work methodology applied in the first part of this study.

Based on the factors that affect the speed of sound waves and vary the results of the natural radiation emission record of the rock component elements, three geographically and geochronologically different oil fields are selected for the execution of the study, doing a previous investigation of its geological and tectonic aspects. After performing tests and data analysis, representative wells of each field were selected according to the results obtained.

To verify the influence of the different characteristics of each field in the estimation of the parameters (k, b, Q, Z and P) with their respective correlation applying and validating the results obtained from one field to another, as well as the variations thereof being evaluated in different wells within the same oil field, thus observing the behavior of the resulting bulk density synthetic logs.

Using the Matlab programming language, the data is loaded (logs ρ_b, V_{sh}, V_p) and corrected by applying an

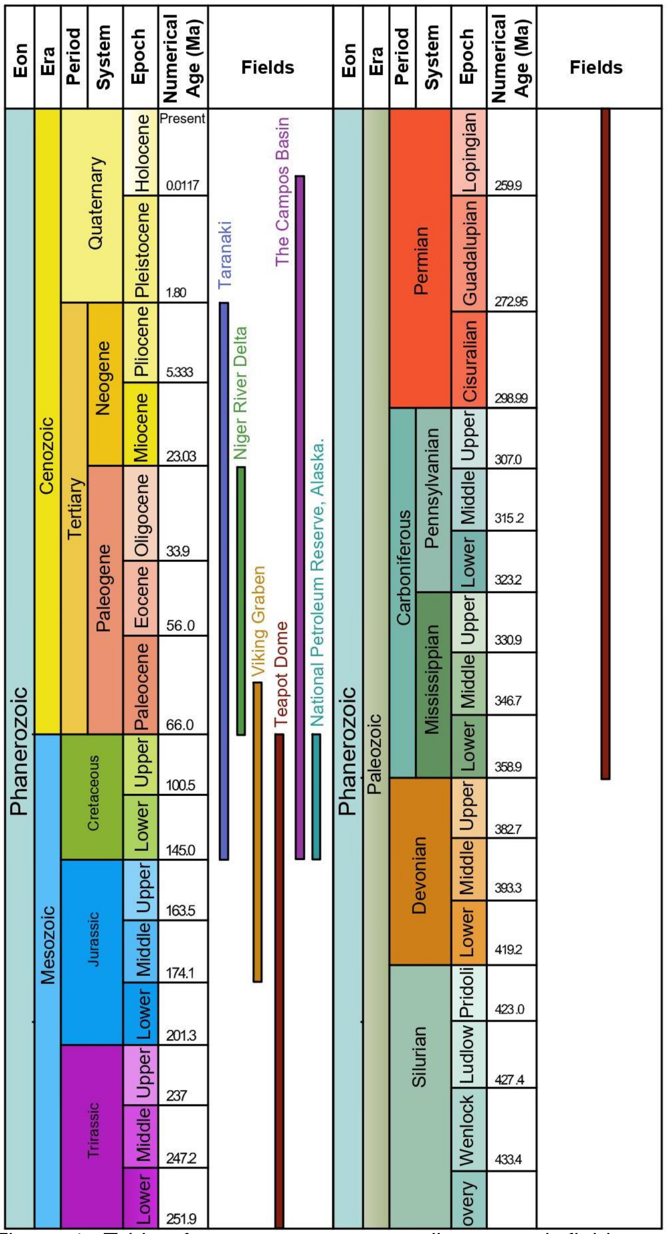

Figure 4: Table of age range corresponding to each field involved in the study

encoded function called Despike -which is responsible for cleaning the noise present in the data measured in well- and a mud filtrate correction using the caliper log to remove the effect of wall's caving, roughness and variation of the diameter of the well; thereby preventing the reproduction of noise and erroneous data by parts of the parameters calculated later (de Macedo et al., 2020).

When despike filtering was applied on the V_{shale}, ρ_b and V_p logs data it was generated a smoothing of the points with outliers in the data.

The mud filtrate correction on density log is based on the caliper log measure of the variation in the borehole diameter, which shows that when the well's diameter is increasing, the well is passing through an unconsolidated formation or walls' caving. The mud filtrate correction was made based on the paper from de Macedo et al. (2020).

Using the programming language, a code is constructed that allows the calculation of the different parameters showed in equations 2 and 3, for each one of the fields, selecting the representative parameters of each field to be applied later on a well selected as a base (Ariki Well / Taranaki oil field) due to its wide geological and seismic information, besides a great geological ages variability along its stratigraphic column; thus observing the behavior, viability and reproduction of the parameters of each field when applied in fields of equal and different geological and lithological ages and affecting factors (tectonism, magmmatism and external agents). For this last item, the parameters proposed in the study carried out in Nigeria by Oloruntobi and Butt (2019) are also applied using the linear equation model proposed by the same for obtaining a synthetic bulk density log.

Results

When the Despike correction is applied on the sonic register, gamma-ray and density log, for the traditional Gardner equation, the parameters show an inversely proportional behavior between them, where the variations of k and b after the correction are not greater than 1.5-2.0% of the parameters' value without correction, which indicates little loss of data but a great effect on the results' variation. For eq. 3 it is generally observed, when the Despike correction is applied, growth in the value of the parameters Q and Z and decrease in the value of P, which is justified based on the proportional and equitable behavior of the parameters to guarantee the linearity of the equation.

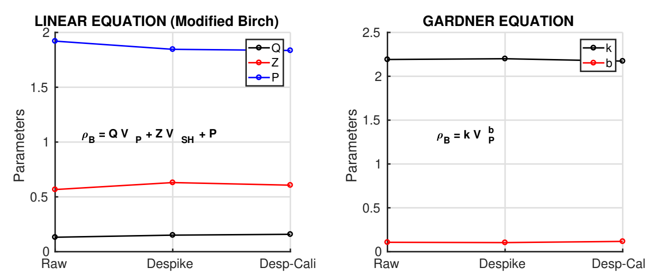

Figure 5: Variation graph of the values of the parameters "k, b" for traditional Gardner and "Q,Z,P" for the linear equation, applying the despike correction and mud filtrate correction to the density to the records and logs worked within equations in question, taking for this case the data from well ID 5002320012000 of NRPA as a sample.

The mud filtrate correction applied to the density log in traditional Gardner model, as well as the despike correction, shows an inversely proportional behavior in the variations of the values for the parameters k and b. For the modified Birch's model, the correction of the density selected to decrease the value of the parameter Q with respect to the result thereof with the application of Despike. The values of Z and P increased concerning their initial values (raw data) Figure 5.

In Figure 6 the parameters are calculated after all the corrections are applied on the reference well, it is possible to denote that these parameters make a better adjustment when applied to calculate the synthetic density log in a different well by avoiding the reproduction of noises present

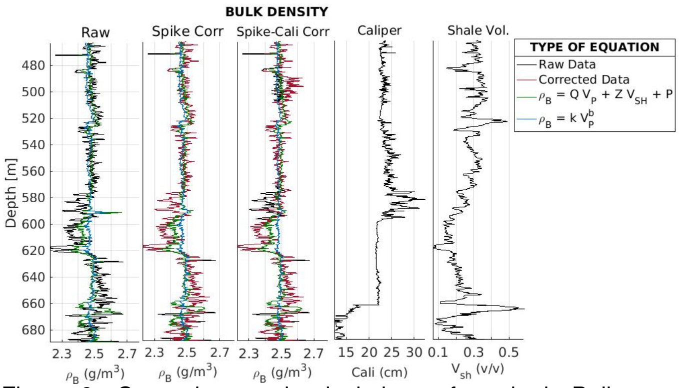

Figure 6: Corrections and calculations of synthetic Bulk Densities on the NPRA well identified as 500232001200. the black curve corresponds to the original bulk density data of the well. The red curve shows the result of the corrected data. The blue curve is the synthetic log obtained using the Gardner's traditional model. The green curve corresponds to the synthetic log determined with the modified Birch's model proposed by Birch (1960), Han et al. (1986) and Oloruntobi and Butt (2019).

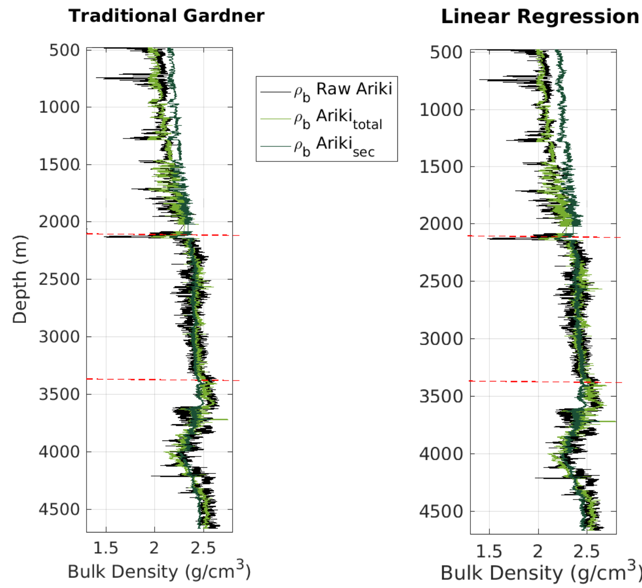

Figure 7: Comparison of calculated parameters of the Ariki well (Taranaki oil field) using the entire log and a behaved segment of it (demarcated on the graph between red dashed lines), being applied to reproduce the synthetic log of the source data. It can be seen how the parameters do not reproduce noises when they are found using prior corrections (Despike and mud filtrate correction).

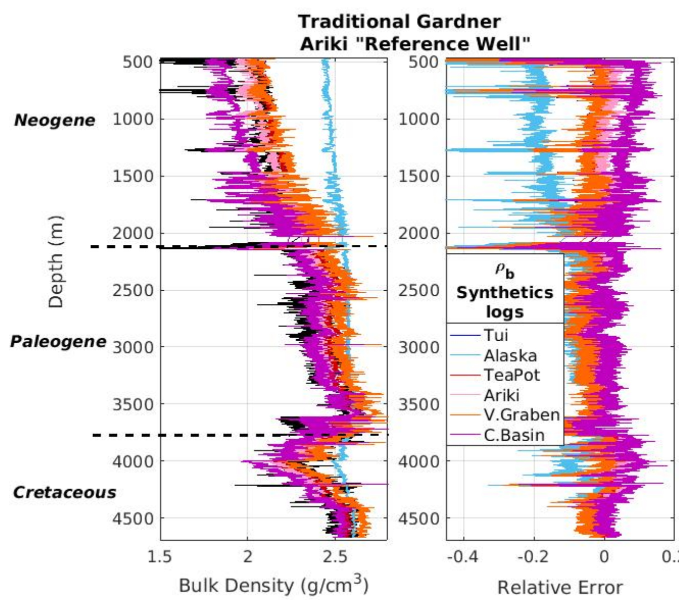

Figure 8: Synthetic bulk density logs of the Ariki well (taranaki basin oil field) using calculated parameters of the NPR Alaska, Teapot Dome, Campos Basin, Viking Graben and Taranaki basin - New Zaland (Tui and Ariki) oil fields in the traditional Gardner equation. The graph is divided into geological time sections corresponding to the Ariki well, delimited by dashed lines based on the interpretation of the OTS-5 seismic line.

in the original data. In other words, an intrinsic correction is generated.

When performing the calculations, the variability and veracity of the parameters found for the same well is confirmed using the entire log and also regions with greater data content within it. On Figure 7 can be seen the reproduction of the synthetic density log of the Ariki well, where the resulting log with the parameters found using the total of the log has a higher reproduction, less limitation of the amplitude of the synthetic register compared to the parameters resulting from the calculation based on a behaved section (in this case the one corresponding to the region present between 2100 and 3400 m deep) of the original record. Numerically, when calculating with a behaved section, the parameter k in traditional Gardner has an increase, generating a decrease in the b value, since these two parameters act inversely proportional in this case, which explains the limitation of the amplitude in the synthetic record reproduced since this parameter regulates this characteristic. The parameters for the linear equation show a relative increase in Z and P and a decrease in Q, which is the parameter -among the three existing ones- that regulates the amplitude mainly.

The traditional Gardner equation for the calculation of bulk density shows a restricted reproduction, where the adjustment is somewhat limited. Even so, it shows variations between one field and another, despite only one variable (V_p) who generates the different notable changes along the well under test, as shown in Figure 8, where the variations are denoted of bulk density when tested in a well with parameters of fields of different ages and the same age based on data from the Ariki well of the Taranaki basin field. The smallest error in reproduction is given by the synthetic log reproduced with the parameters calculated using the data from the same base well (Ariki), showing the greatest adjustment.

The modified Birch's equation shows a more adaptable behavior -greater sensitivity- to the original data, giving a greater adjustment with respect to the original bulk density data, being appreciably less limiting present for the reproduction of the synthetic density logs, as can be seen in Figure 9.

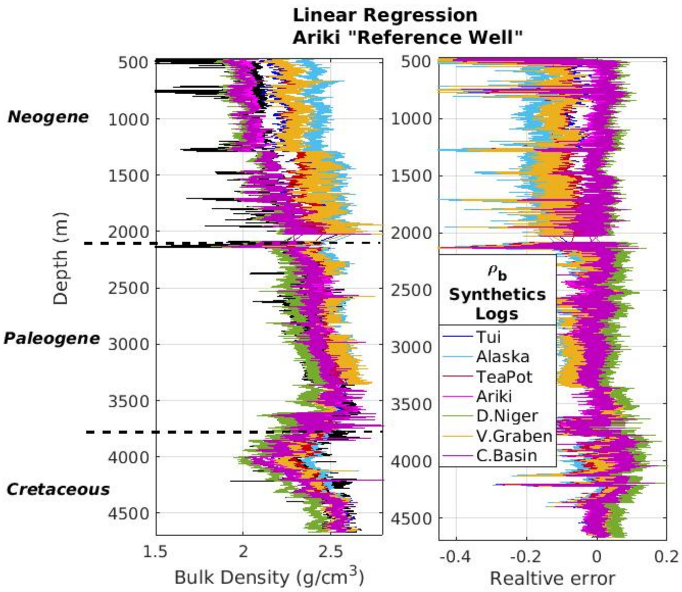

Figure 9: Synthetic bulk density logs of the Ariki well (Taranaki basin oil field) using calculated parameters of the NPR Alaska, Teapot Dome, Campos Basin, Viking Graben and Taranaki basin - New Zealand (Tui and Ariki) oil fields in the modified Birch's equation. The graph is divided into geological time sections corresponding to the Ariki well, delimited by dashed lines based on the interpretation of the OTS-5 seismic line

Tabela 1: Table of mean squared errors of each bulk density log calculated with the parameters of different fields and applied to the Ariki well, using the Gardner's traditional equation and the linear regression eq. 3 proposed by Oloruntobi and Butt (2019)

Synthetic Logs	(MSE) Gardner's traditional equation $\rho_b = kV_p^b$	(MSE) Linear Regression $\rho_b = QV_p + ZV_{sh} + P$
Ariki	0.0068	0.0073
Tui	0.0087	0.0270
NPRA	0.0782	0.0636
Teapot Dome	0.0085	0.0238
Viking Graben	0.0166	0.0355
Campos Basin	0.0150	0.0088
Delta Niger	-	0.0192

the information provided by the V_p is not enough to generate parameters that reproduce and adjust in a greater percentage of geological variations (which will be represented in the density log) present in the wells, as it is also possible to observe in Figure 8, the reproduction does not generate a sufficiently detailed adjustment to allow such an approximation of the reproduced density synthetic log using the Teapot Dome parameters (lower-carboniferous to the upper-cretacic) or the Campos Basin parameters (Holocene/Pleistocene to Jurassic) of the Ariki well to the original density log from Ariki.

The inclusion of V_{sh}, in the case of the linear equation proposed by Oloruntobi and Butt (2019), as a variable within the mathematical equation generates more dispersion in the results, allowing to obtain better detail with respect to the area studied compared to traditional Gardner. For that reason, eq. 3 It shows greater sensitivity of the parameters -before lithological, geological age and other variation- when a reproduction of a synthetic density log is generated as can be seen in Figure 9. Synthetic logs showed little impact when calculated using a V_{sh} under the influence of the different methods of obtaining GI, being the results obtained with the Clavier model, generally giving intermediate values between the set of data resulting from the parameters at be evaluated by the four models mentioned above, taking this equation proposed by Clavier et al. (1971) as the most suitable for a GI calculation, without generating distinctions or limitations according to the age of the rocks present in the wells tested.

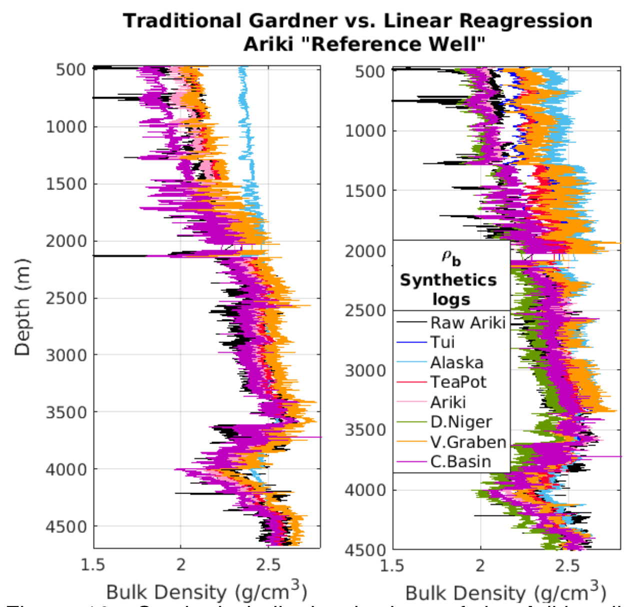

Figure 10: Synthetic bulk density logs of the Ariki well (taranaki basin oil field) using calculated parameters of the NPR Alaska, Teapot Dome, Taranaki basin - New Zealand (Tui and Ariki) and Niger Delta in the traditional Gardner equation and modified Birch's model.

Reproduction tests of the synthetic density logs using the reproduction parameters based on well data from the different fields, which are exemplified in Figure 10, shows that the different characteristics of each field create, by applying its parameters on the same base well, a synthetic density log that in some examples -such as Alaska vs. Ariki or Viking Graben vs. Ariki- the curves approach and/or overlap sections where they have equality and similar age.

However, the implicit distribution in the structure of a linear equation generates an adjustment deficit, which causes greater adjustment and evidence of the geological and geological age particularities of each field, where it can be observed as the curves created based on the parameters calculated for each field give greater adjustment and reproduction, overlapping the original curve and the other synthetic curves -generated with the parameters of the other fields- in the sections where the mimes have equity in age.

The causes and the different limitations that occur during the reproduction of a synthetic density curve, based on the parameters of a different field, are attributed to the differences between the geological and geochronological characteristics in each field both in general terms and local; since the properties of a rock can change in a matter of a few centimeters or meters, which is characteristic of anisotropic media.

Throughout the study, it was possible to verify that, by calculating the parameters on a well reference field and then using these parameters in another well of the same field, these managed to occasionally adjust better to the original density log of the test well (although not totally) compared to the adjustment using parameters from wells outside the field.

Conclusions

1) The application of previous corrections in the parameters increases the reliability of the results when applied for the calculation of the bulk density logs, and in the same manner, increases their adjustment generating a more efficient reproduction of the synthetic logs. It is necessary to perform the calculation of the parameters taking all the data from the well logs with the highest accuracy and confidence in the results when applying the parameters, in order to avoid generating geological and geochronological limitations and also to get a better adjustment, reproduction and sample of characteristics of each field, either by applying the parameters on wells of the same base field that they were obtained or in fields of different age and characteristics. Furthermore, the total data collection will be very useful for studies of correlation and geological age effects between the different logs of a well.

2) The use of parameters calculated with wells external to the reference field, generates an error greater than 1% when applied to wells external to the field of the well taken as a basis for its own calculation, either using traditional Gardner's equation or the linear model, which exceeds the permitted limit in margin of error establishes an admissible adjustment of the parameters. Occasionally errors of less than 5% are present, however, the density estimation of the parameters outside the test field shows unsatisfactory reproductions.

3) The traditional model for obtaining bulk density proposed by Gardner generates some parameters with a rather limited adjustment and with a higher percentage of error compared to the results obtained using the parameters calculated with the linear model, due to the fact that its calculation only depends on one variable (V_p) which, although it shows useful information about the media it passes through and tries to reproduce these characteristics in the synthetic log, the data provided are not sufficient to give a good approximation and adjustment to the original bulk density log of the well under test.

4) The results obtained when generating a synthetic apparent density log by applying the linear parameters proposed in the study carried out by Oloruntobi and Butt (2019), do not show an appropriate or reliable adjustment, giving errors greater than 1% , which corroborates that the parameters obtained in one field when applied to another different field, may generate a high percentage of error in the synthetic reproduction and low reliability in the results.

Acknowledgment

The authors also would like to thank CAPES, INCT-GP and CNPq from Brazil and the graduate program at Federal University of Pará for the financial support in this research.

Bibliography

Birch, F., 1960, The velocity of compressional waves in rocks to 10 kilobars: 1.: Journal of Geophysical Research, **65**, 1083–1102.

Castagna, J. P., M. L. Batzle, and R. L. Eastwood, 1985, Relationships between compressional-wave and shear-wave velocities in clastic silicate rocks: geophysics, **50**, 571–581.

Clavier, C., W. Hoyle, D. Meunier, et al., 1971, Quantitative interpretation of thermal neutron decay time logs: part i. fundamentals and techniques: Journal of Petroleum Technology, **23**, 743–755.

de Macedo, I. A., J. J. S. de Figueiredo, and M. C. de Sousa, 2020, Density log correction for borehole effects and its impact on well-to-seismic tie: Application on a north sea data set: Interpretation, **8**, T43–T53.

DNPOSR, 1976, Teapot dome field, naval pet 3: Environmental impact statement: Department of Naval Petroleum and Oil Shale Reserve. 3.

Ellis, D. V., and J. M. Singer, 2007, Well logging for earth scientists: Springer, **692**.

Energy-Department, 2011, Characterization of cement from a well at teapot dome oil field: Implications for geological sequestration: National Energy Technology Laboratory (U.S.).

Gardner, G., L. Gardner, and A. Gregory, 1974, Formation velocity and density—the diagnostic basics for stratigraphic traps: Geophysics, **39**, 770–780.

Guardado, L. R., 1990, Petroleum geology of the campos basin, brazil, a model for a producing atlantic type basin: Divergent/passive margin basins.

Han, D.-h., A. Nur, and D. Morgan, 1986, Effects of porosity and clay content on wave velocities in sandstones: Geophysics, **51**, 2093–2107.

Houseknecht, D., and C. Schenk, 2001, *in* DEPOSITIONAL SEQUENCES AND FACIES IN THE TOROK FORMATION, NATIONAL PETROLEUM RESERVE – ALASKA (NPRA): 179–199.

Kowallis, B. J., L. Jones, and H. Wang, 1984, Velocity-porosity-clay content systematics of poorly consolidated sandstones: Journal of Geophysical Research: Solid Earth, **89**, 10355–10364.

Leon, S. J., I. Bica, and T. Hohn, 2006, Linear algebra with applications: Pearson Prentice Hall Upper Saddle River (NJ, USA).

Oloruntobi, O., and S. Butt, 2019, The new formation bulk density predictions for siliciclastic rocks: Journal of Petroleum Science and Engineering, **180**, 526–537.

Reijers, T., 2011, Stratigraphy and sedimentology of the niger delta: 133–162.

Saltus, R., T. Hudson, J. Phillips, C. Kulander, J. Dumoulin, and C. Potter, 2002, Basement geology of the national petroleum reserve alaska (npra), northern alaska.

Science, G., 2012, Taranaki basin.

Serra, O., and L. Serra, 2004, Well logging: data acquisition and applications: Serralog Méry Corbon, France.

Tosaya, C., and A. Nur, 1982, Effects of diagenesis and clays on compressional velocities in rocks: Geophysical

Research Letters, **9**, 5–8.