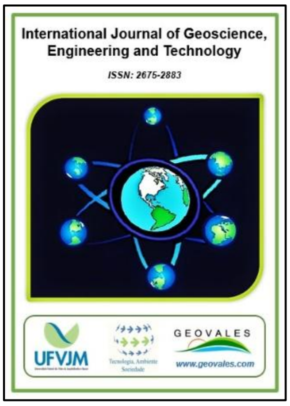
International Journal of Geoscience, Engineering and Technology
ISSN: 2675-2883
https://doi.org/10.70597/ijget.v9i1.533
Volume 9 – Nº 1 – Year: 2024 – pp.1-20



Received:
February 5, 2024

Accepted:
April 17, 2024

Published:
April 30, 2024

Well Log Data Statistical Processing for Unbiased Qualitative and Quantitative Analyses: Case Study from the Gulf of Guinea

André Noukpo Djoï [1]

[1] Independent Researcher, Cotonou, Benin.

Email address
djoi.andrew@gmail.com (André N. Djoï) – Corresponding author.

Abstract

Over oil and gas fields exploration and development phases, one of the main challenges of geoscientists and petroleum engineers is the petrophysical characterization of potential or discovered fields or reservoirs. Well logs data play key roles in wells stratigraphic column establishment and the computation of reservoir formations petrophysical parameters. Due to the conditions and the environment of well log data acquisition, they undergo some technical processing. For gamma ray log data for instance, although the technical data processing, the representative minimum and maximum values of recorded GR are required for unbiased qualitative and quantitative analyses. This study aims to propose statistical techniques for gamma ray logs data processing that will contribute to the reduction of biases related to their qualitative and quantitative analyses. A case study has been performed on a Gulf of Guinea's offshore well gamma ray log data. The results show that the difference between the maximum and minimum values for the semi-processed data is almost twice the one of the processed data, what will lead to the underestimation of formations shale volumes and therefore to the overestimation of reservoirs effective porosity and flow performance. Moreover, the baselines (shaly sand, sandy shale and shale baselines) obtained from the semi-processed data are respectively located more rightward to those from the processed data. The main consequence is that the semi-processed data analysis has hidden the shaliness of formations comparatively to the processed data analysis. A comparative analysis shows that the semi-processed data analysis has globally underestimated the thickness of thicker formations and underestimate the shale volumes of thicker formations and those for which the estimated thicknesses from both analyses are the same or close to each other. In summary, the statistical processing of gamma ray log data prevents from the underestimation of thicker formations thicknesses and formations shale volumes. The main practical advantage is that it will prevent geologists, petro-physicists and reservoir engineers from the overestimation of oil or gas reservoirs effective porosity and flow performance and therefore from the overestimation of oil or gas initially in place and reserves.

Keywords: Well log, Gamma ray log data, Statistical processing, Gulf of Guinea.

1. Introduction

Over oil and gas fields exploration and development phases, geoscientists and petroleum engineers have to characterize the potential or discovered fields or reservoirs. One of the main reservoir characterizations is the petrophysical characterization of reservoir formations through direct and indirect methods (Djoï, Nwosu and Ikiensikimama, 2022). The indirect methods use information recorded on the reservoir from surface or by lowering tools in wellbores. Well logs data play key role in the reservoir characterization through indirect techniques. The common petrophysical parameters of reservoir formations computed from well log data are, but not limited to, shale volume, porosity, permeability, fluids (water, oil and gas) saturation, reservoirs net pay thickness, fluids contacts (WOC and GOC), the skin effect and reservoir pressure (Ekwere, 2004). Qualitative analysis can also be performed with well log data to identify reservoirs and determine their lithologies as well as those of the overburden formations thereof (Djoï, 2012). After recording well logs information, some technical processing is performed by the logging engineers to eliminate errors related to measurement environments, measurement conditions (environmental, borehole geometry, formation temperature, etc.) and tools response limitations (Schlumberger, 2004). Apart from technical processing, statistical processing is required for some parameters log data unbiased analysis. Indeed, for instance, for formations shale volume estimation with spontaneous potential (SP) or gamma ray (GR) logs, one needs to determine the spontaneous potential and gamma ray minimum and maximum values (HLS Asia Limited, 2007; Szabó, 2011). Their misevaluation can lead, among others, to an overestimation of reservoir formations shale volume and underestimation of formation effective porosities. Moreover, misevaluations of shale volumes render biased the oil and gas reservoirs identification (Djoï, 2015). Indeed, when formations shale volumes are overestimated, the reservoir layers considered as high shaly formations would be rejected wrongly. The experience of well logs data analysis has shown that for shale volume computation with spontaneous potential or gamma ray for instance, statistical processing of spontaneous potential and gamma logs data is required for a good qualitative and quantitative analysis.

This study aims to propose statistical techniques for gamma ray (GR) logs data processing that will contribute to the reduction of biases related to their qualitative and quantitative analyses.

A case study will be performed on gamma ray logs data of wells located in the Gulf of Guinea for wells stratigraphic columns determination and reservoir formations shale volume computations. This case study will show the impact of statistical processing on the data analyses results.

1.1. Logging and Gamma Ray Log Data Acquisition and Processing

Log in the oil industry means a recording against depth of any of the characteristics of the underground rock formations traversed by a measuring apparatus in the well-bore (Oberto, 1984). The underground rock formations characteristics measurement can be performed while drilling wells or after cessation or the end of the drilling operations. Base on that, two wide types of loggings exist: logging while drilling (LWD) also called measurement while drilling (MWD) and wireline logging. The while-drilling logs are recorded during the drilling operations. They are mainly used for drilling control parameters (rotation rate, weight on bit, drilling-rate, mud-loss, torque, etc.) measurement (Maget, 1990 and Chapelier, 2004). Wireline logging is performed after an interruption (or the termination) of drilling activity, and is thus distinguished from drilling-logs (of such things as drilling-rate, mud-loss, torque, etc.) and mud-logs (drilling mud salinity, pH, mud-weight, etc.) obtained during drilling operations (HLS Asia Limited, 2007).

Logging can be carried out in open or cased holes. According to Oberto (1984), well-logs are important (useful) for the following reasons:

- core data are difficult or expensive to obtain for some parameters (such as natural gamma-ray radiation, neutron hydrogen index, sonic transit time, bulk density, etc.) measurements;
- well-log data allow an analysis of a larger volume of formation than the core, especially when we consider that core measurements are themselves made on thin sections or plugs of material;
- well-log data offer the possibility of making a computer analysis of quantitative data more than core data;
- well-log information is continuous and permanent.

The first well-log, a measurement of electrical resistivity, devised by Marcel and Conrad Schlumberger, was run in September 1927 in Pechelbronn, France (Schlumberger, 2004 and Bateman, 2020). Log measurements are made using a measuring sonde (with electronic cartridge) lowered on a cable from a winch, which is mounted on a logging truck or offshore unit (Oberto, 1984). The truck and unit are laboratories containing the recording equipment (optical and tape), control panels, and perhaps a computer.

Tens of logs are developed and used. There are, among others, spontaneous potential logs, electric logs, induction logs, resistivity logs, caliper logs, gamma ray logs, neutron logs, density logs, acoustic logs, directional surveys and nuclear magnetism logs (Oberto, 1984 and Schlumberger, 2009). This study will focus on gamma ray logs.

1.1.1. Gamma Ray Log Data Acquisition

The gamma ray (GR) log is a continuous recording of the intensity of the natural radiations emanating from the formations penetrated by the borehole versus depth (Oberto, 1984 and Bassiouni, 1994). The sources of natural radioactivity are the isotopes of potassium (^{40}K), thorium (^{232}Th) and uranium (^{238}U) contained in the formation minerals (Bassiouni, 1994). There are two types of natural gamma ray (NG) log. One, the total or standard GR log, measures only the total radioactivity. The other, the NGS (Natural Gamma Ray Spectrometry) log, measures the total radioactivity and the concentrations of potassium, thorium, and uranium producing the radioactivity (Baron, Cariou and Thorion, 1989 and Schlumberger, 2004). Asquith and Krygowski (2004) added that the spectral gamma ray log records not only the number of gamma ray emitted but by the formation but the energy of each, and processes that information into the curves representative of the amount of potassium, thorium and uranium present in the formation. Bateman (2020) highlighted that the first standard and spectral gamma ray log tools were developed and run in 1938 and 1969 respectively.

As far as GR is concerned, it is measured with a radiation counter placed in a sonde (Baron, Cariou and Thorion, 1989). The measurement is not stable, knowing that the emission phenomenon is variable over time. Only an average intensity is measured over a long time. Based on that, GR log is a very long operation and the recording presents statistical fluctuations. According to Oberto (1984) the GAPI unit used corresponds to microgram equivalent of uranium per tons (µg Raeq/t): the conversion relationship is 16.5 GAPI = 1 µg Ra-eq/t). For Maget (1990) and Bassiouni (1994), one API gamma ray unit is defined as 1/200 of the difference in log deflection between the two lower concrete zones of low and high radiation in the calibration unit. All gamma ray tools are calibrated to API standards record gamma ray radiation in the same unit of measurement.

For Djoï, Nwosu and Ikiensikimama (2022), the intensity of radiations from potassium are much more important than thorium and uranium and is characteristic of clay presence in formations. Thorium are present in two main types of sedimentary rocks (salt or anhydrite and limestone) while sandstone are wealthy in uranium (Djoï, 2012). As a result, high values of GR indicate the presence of shale and lower values characterize formations with less clay content. Indeed, GR value is related to the amount of radioactive items in the formation.

1.1.2. Gamma Ray Log Data Processing

As highlighted by Oberto (1984), although we would like logs to be direct measurements of the formation, log responses are invariably affected by the presence of the well-bore (presence of casing and cement), certain near-hole phenomena associated with the drilling of the well (presence of drilling mud invasion and mud cake), the logging procedure (speed) and the geometry of the logging tool itself. Operational problems may be posed by temperature and pressure in the well. Rabaute (2009) added that log measurements are carried out in difficult or at least less optimal conditions that it is always necessary to correct errors related to the logging environments and procedure. He states two types or errors: systematic error due to the tool and analytic error related to environment, standard uses, calibration and filtering. This error correction, called technical log data processing in this study, is referred to as data correction.

Raw log data processing can be performed on at least three levels: downhole in the tool, uphole the truck, and at a central computing center (Schlumberger, 2004). Where the processing is done depends on where the desired results can most efficiently be produced, where the extracted information is first needed, where the background expertise exists, or where technological considerations dictate. Whenever it seems desirable, the logging tool is designed so that the data are processed downhole and the processed signal is transmitted to the surface.

The logs data processing is assured by the data acquisition company before data be provided to the client.

The common log technical data processing set is the depth calibration, the filtering and the required corrections regarding the type of log. For GR logs, the main corrections are borehole (hole size) and mud weight corrections for open and cased holes, potassium correction for open and cased holes and environmental corrections. The equations (or formulae) and chart for corrected GR determination are available in log data acquisition companies' manuals for the different corrections. The following is a brief development of GR logs corrections for

Schlumberger. The principle consists of determining the correction factor and the corrected gamma ray (GR_{cor}) is determined from Equation (1).

$$GR_{cor} = CF * GR_{mes} \quad (1)$$

With GR_{mes} the GR measured and CF the correction factor.

Schlumberger developed methods for GR correction for borehole and mud weight. Some borehole and mud weight correction techniques use an intermediate parameter t. Equation (2) helps to calculate t for open holes while t for cased holes is computed with Equation (3). The correction factor is function of t and is determined with the charts designed for that purpose, for different tool diameters. Figures (1) and (2) show the charts for some tool diameters.

$$t = \frac{W_m}{8.345}\left[\frac{2.54(d_h)}{2} - \frac{2.54(d_{sonde})}{2}\right] \quad (2)$$

With W_m the mud weight in lbm/gal, d_h the diameter of wellbore in inch and d_{sonde} the outer diameter (OD) of the tool in inch.

$$t = \frac{2.54}{2}\left[\frac{W_m}{8.345}(d_{IDcsg} - d_{sonde}) + \rho_{csg}(d_{ODcsg} - d_{IDcsg}) + \rho_{cement}(d_h - d_{ODcsg})\right] \quad (3)$$

With W_m the mud weight in lbm/gal, d_h the diameter of wellbore in inch and d_{sonde} the outer diameter (OD) of the tool, d_{IDcsg} and d_{ODcsg} the inner and outer diameter of the casing in inch, ρ_{cement} the density of the cement in g/cm^3 and ρ_{csg} the density of the casing in g/cm^3.

Figures (3) and (4) show the charts for borehole correction factors determination for some bit sizes and tool diameters respectively.

The chart of Figure (5) helps in computing the correction factor for potassium (K) correction for open holes for 6.75-in logging tool.

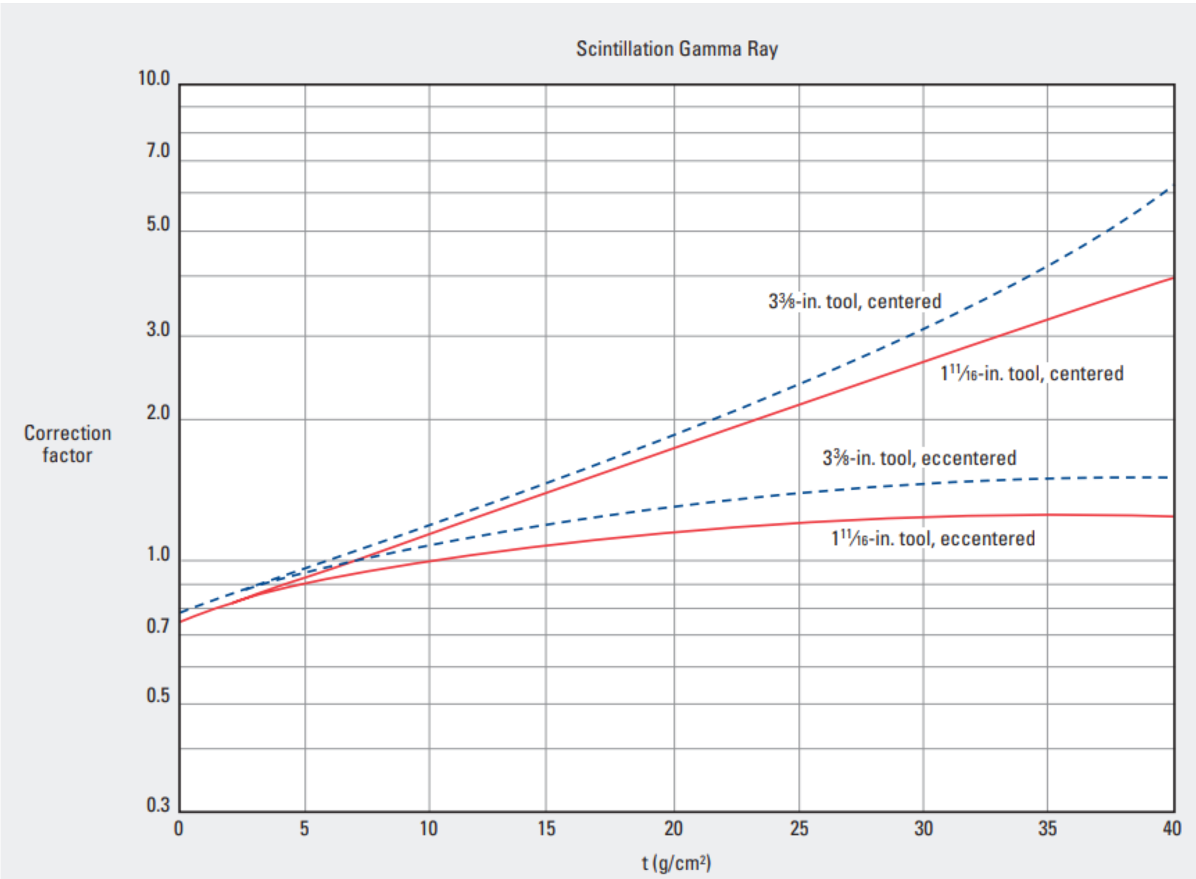


Figure 1 – GR borehole and mud weight correction chart for open holes for four tool diameters (Schlumberger, 2009).

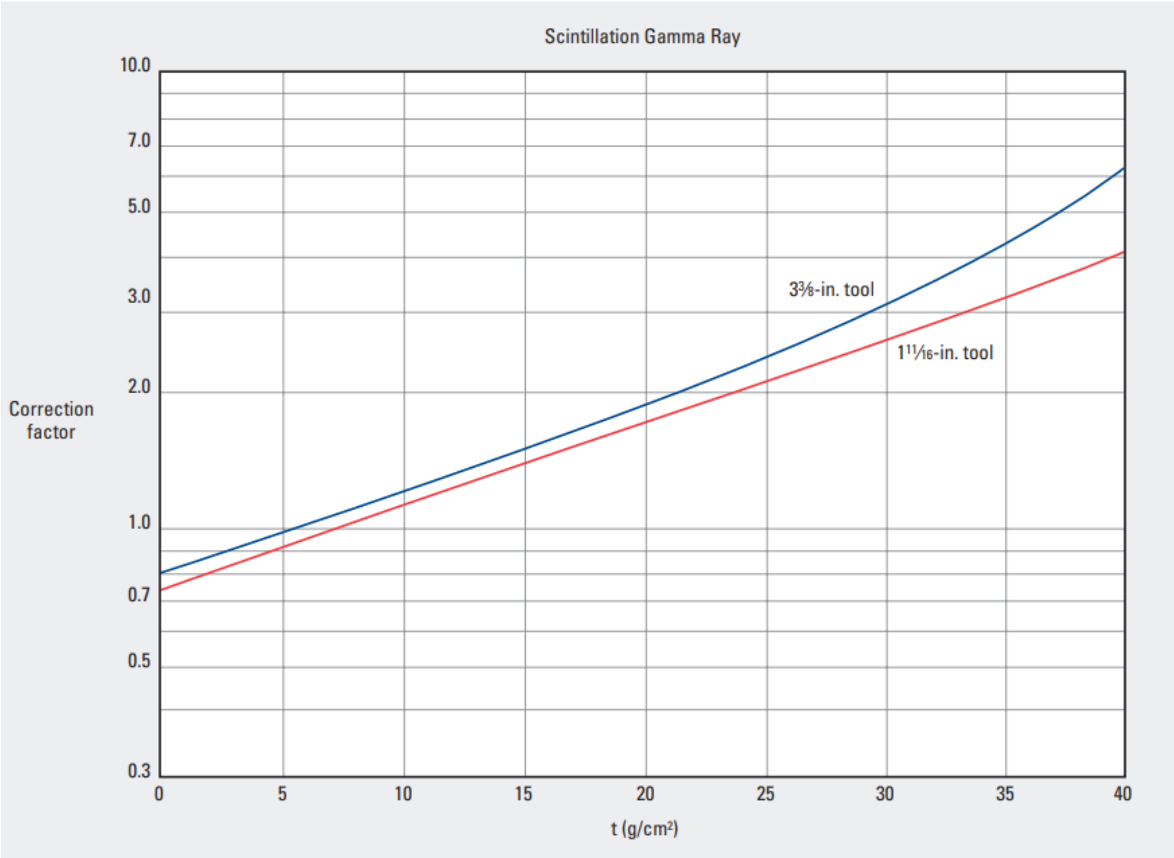


Figure 2 – GR borehole and mud weight correction chart for cased holes for two tool diameters (Schlumberger, 2009).

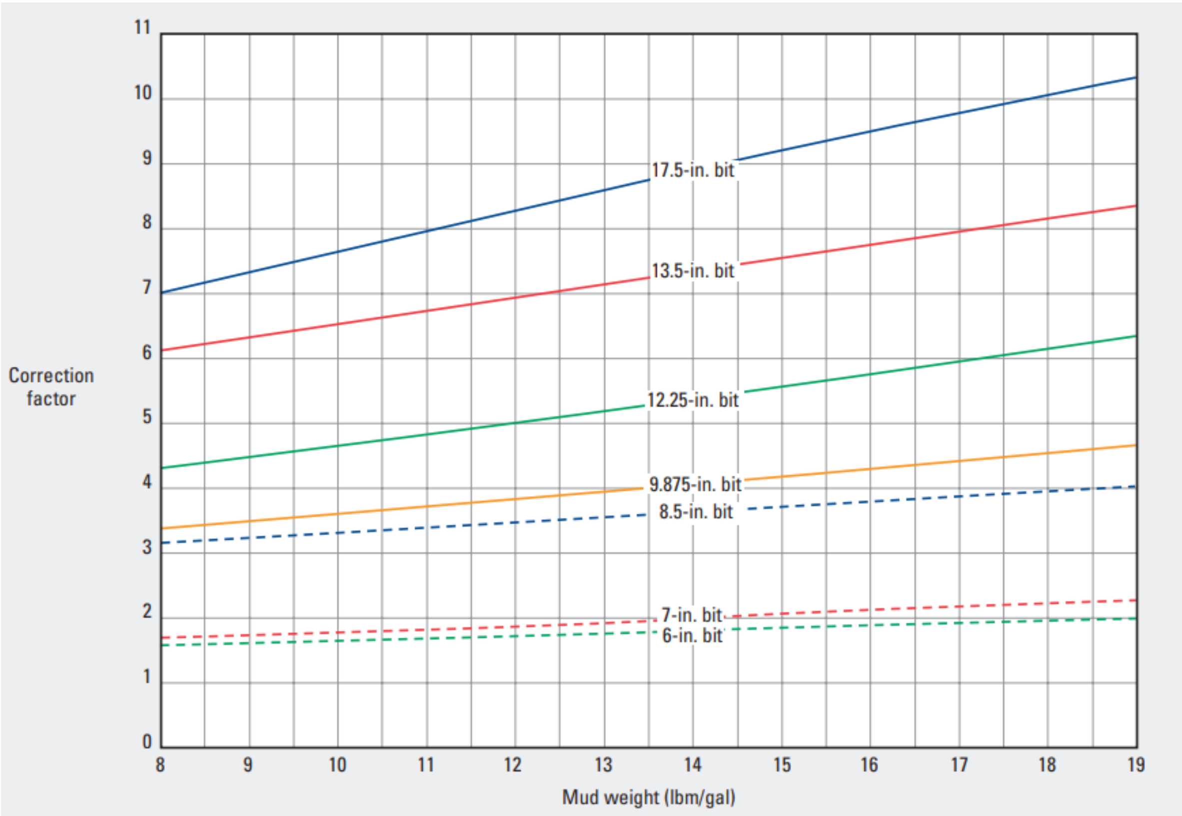


Figure 3 – GR borehole and mud weight correction chart for open holes for seven tool diameters (Schlumberger, 2009).

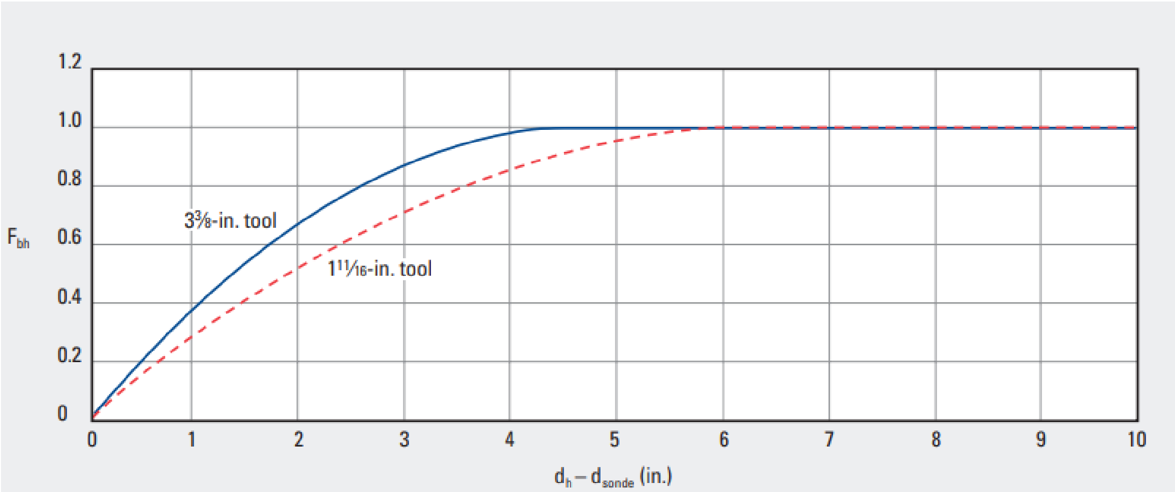


Figure 4 – GR borehole correction chart for open holes for two tool diameters, F_{bh} is the correction factor (Schlumberger, 2009).

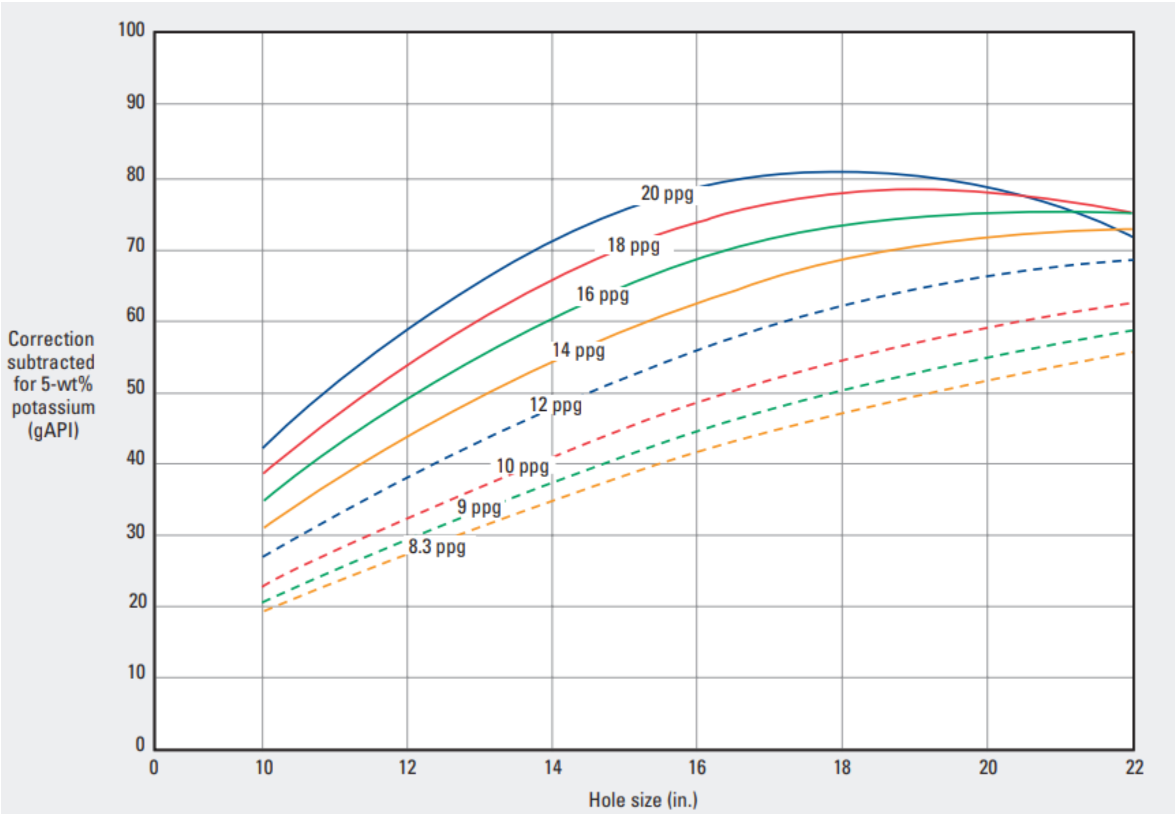


Figure 5 – GR potassium correction chart for open holes for 6.75-in tool for mud weight between 8.3 and 20 ppg (Schlumberger, 2009).

Bateman (2020) spotlighted that using Schlumberger GR data correction methods, other things being equal, the magnitude of the corrections:

- changing hole size from 8.5 inch to 9.625 inch produces an observed GR reduction of about 20%;
- changing mud weight from 8 to 12 lb/gal produces a reduction of about 40%;
- changing from open hole to cased hole with 7v casing produces a reduction of 60% or so.

After these raw data corrections, the log data users need to perform some statistical processing for a good interpretation.

1.2. Importance of Gamma Ray Log Data Statistical Processing

For the qualitative and quantitative gamma ray log data analysis, one needs to know the minimum and maximum values of GR recorded over a well. The minimum GR value characterizes clean sandstones or milestones while the maximum value constitutes the DNA of shales (Oberto, 1984; Kamel and Mabrouk, 2003; Schlumberger, 2004 and Djoï, 2012). In most cases, the minimum and maximum GR values used are single values and do not really represents formations. Since single minimum and maximum GR values of the GR measured dataset are odds and their will bias qualitative and quantitative interpretations. Therefore, the representative minimum and maximum values must be determined in the ways they really represent formations (clean sandstones/limestones and shales).

Moreover, gamma ray log dataset may contain outliers that must be detected and treated before any analysis.

The gamma ray data outliers' detection and treatment as well as its representative minimum and maximum values determination can be done through statistical data processing. That is why log data statistical processing is required for good GR log data analysis.

1.3. Statistics for Gamma Ray Logs Data Processing

The statistical approach to be proposed by this study for gamma ray log data processing requires the understanding of some statistical aspects. This section will address the main mathematical and statistical tools needed. An eye on the well logs data acquisition process, makes notice that gamma ray log is a statistical variable. Regarding the random nature of formations responses to logging tools and the recording points choice, gamma ray data acquisition must be considered as random sampling (Djoï, Nwosu and Ikiensikimama, 2022). A well gamma ray data is therefore a gathering of random samples of penetrated formations GR. As a result, GR log data is a random sample of penetrated formations GR random variable. Statistical processing of GR log data will bring out some statistical items such as outliers, histogram, mode, modal class, etc.

1.3.1. Outliers Detection and Treatment

Iglewicz and Hoaglin (1993) define an outlier as an observation which deviates so much from other observations as to arouse suspicions that it was generated by a different mechanism. They are much smaller or much larger than the vast majority of the observations (Cuisineau and Chartier, 2010). In order words, an outlier is an observation that appears to deviate markedly from other members of the sample in which it occurs (Cuisineau and Chartier, 2010). In case of GR log, the outliers are the measured values that are much smaller or much larger than the vast majority of the observations. Mathematically, an outlier is an observation that is less than the lower fence or greater than the upper fence of the dataset (Iglewicz and Hoaglin, 1993).

Iglewicz and Hoaglin (1993) and Djoï, Nwosu and Ikiensikimama (2023) define datasets lower (L) and upper (U) fences with Equations (4) and (5):

$$L = Q1 - 1.5\,IQR = Q1 - 1.5\,(Q3 - Q1) \quad (4)$$

$$U = Q3 + 1.5\,IQR = Q3 + 1.5\,(Q3 - Q1) \quad (5)$$

With Q1 the lower quartile, Q3 the upper quartile and IQR the interquartile range.

For Illowsky and Dean (2021), a random variable sample interquartile interval or interquartile range, noted IQR, is referred to as the difference between the upper quartile (Q_3) and the lower quartile (Q_1). That is, (Equation (6)):

$$IQR = Q3 - Q1 \quad (6)$$

The quartiles of a numerical random variable X sample $\{x_1, x_2, \dots, x_n\}$ are the three attributes that divide the sample into four populations of same size (Djoï, Nwosu and Ikiensikimama, 2023). Their are noted Q_k $(1 \le k \le 3)$. Q1, Q2 and Q3 are respectively called first, second and third quartile.

There are several ways for outlier detection in a data set such as boxplot and z-score methods. The one this study proposes is data visualization through boxplot.

A numerical data set boxplot (also called Box-and-whiskiers diagram or plot) is a plot which depicts a summary of the lower or first quartile (Q1), median or second quartile (Q2), upper or third quartile (Q3) and the fences of the dataset (Williamson, Sawaryn and Morrison, 2006 and Aguinis, Ryan and Harry, 2013). A box plot may also indicate which observations, if any, might be considered as outliers.

Keeping the outliers in a dataset affects negatively the analysis results. A rule of thumb for outliers' treatment is as follows.

Outliers are either removed or corrected (replaced) by the sample mode or the center of its modal class (Djoï, Nwosu and Ikiensikimama, 2023).

1.3.2. Other Statistical Tools

A numerical dataset histogram, a sample mode or modal class are other statistical tools required for gamma ray log data statistical processing.

Histograms are a common visual representation of a quantitative variable. Histograms summarize the data using rectangles to display either frequencies or proportions as normalized frequencies (Joseph, 2007). Making a histogram, consists of (a) dividing the range of data into bins of equal width (usually, but not always); (b) counting the number of observations in each class and (c) drawing the histogram rectangles representing frequencies or percentages by area. For Joseph, (2007), a histogram helps in determining overall pattern of the sample and its deviation from the pattern.

A statistical variable or sample mode is the most frequent value. There can be more than one mode in a data set as long as those values have the same frequency, and that frequency is the highest (Ekwere, 2004; Illowsky and Dean, 2021). The mode of a qualitative random variable X sample $\{x_1, x_2, \dots, x_n\}$ is the attribute x_j of that has the highest probability (Equation (7)).

$$Mode\ (X) = x_j, with\ P(X = x_j) = Max_{1\leq\ i\ \leq n}\ \{P(X = x_i)\} \quad (7)$$

The modal class of a quantitative random variable X sample $\{x_1, x_2, \dots, x_n\}$ arranged in m classes $\{C_1, C_2, \dots, C_m\}$ is the class C_j that has the highest probability (Equation (8)).

$$Modal\ Class\ (X) = C_j, with\ P(X \in C_j\ ) = Max_{1\leq\ i\ \leq m}\ \{P(X \in C_i\} \quad (8)$$

2. Materials and Methods

2.1. Materials

The main materials used for the case study performed in this study are GR log data, Microsoft Excel and R Studio. Microsoft Excel and R Studio have been used for GR log data visualization, manipulation and processing. R studio has also served for GR log data qualitative and quantitative analyses.

2.2. Methods

2.2.1. Gamma Ray Log Data Statistical Processing

As underlined in the section on the importance of gamma ray log data statistical processing, for good analyses:

- the potential gamma ray data outliers' must be detected and treated; and
- representative gamma ray minimum and maximum values must be determined.

In addition to that, odd values of the data set must be detected and treated.

This study proposes an integrated technique for these tasks. It sets down methods for (1) odd values detection and treatment, (2) outlier detection and treatment and (3) representative minimum and maximum values determination.

(1) Odd Values Detection and Treatment

In well log data, some recorded values must be considered as odd for some reasons. This study focuses on the values that are not in the ranges of the parameters of interest. For gamma ray log data for instance, one knows that the values must be positive and less than or equal to 150 GAPI (on 150 GAPI scale) and 200 GAPI (on 200 GAPI scale). So, any value out of these ranges (negative or greater than the cutoffs) are odd values.

At this stage of the data processing, odd values must be detected and removed from the dataset.

(2) Outliers Detection and Treatment

The approach for outliers' detection will be based on data visualization with boxplot. The algorithm proposed for outliers' detection and treatment is as follows.

Algorithm 1: Algorithm for Outliers Detection and Treatment

Step 1: Plot the boxplot of the log dataset. If there are outliers in the data set, then move to steps 2 and 3.

Step 2: 2.a- Plot the histogram of the log dataset with class intervals of 5 to 10% of the data range.

2.b- Identify the log dataset modal class and compute the modal class center.

Step 3: Replace the log dataset outliers by the modal class center.

(1) Representative Minimum and Maximum Values Determination

While interpreting gamma ray log data, the minimum and maximum values must be determined in the ways they really represent formations (clean sandstones/limestones and shales). Indeed, the minimum GR value characterizes clean sandstones or milestones while the maximum value presents shales (Oberto, 1984; Kamel and Mabrouk, 2003; Schlumberger, 2004 and Djoï, 2012). In most cases, the minimum and maximum GR values used are single values and do not really represents formations. This study sets a comprehensive technique for gamma ray representative minimum and maximum values determination. The proposed algorithm is as follows.

Algorithm 2: Algorithm for Representative Minimum and Maximum Determination

Step 1: Determine the number N of times the minimum and maximum must be each counted in the log dataset to be able to characterize the sandstones (or limestones) and shales respectively and compute the corresponding frequency F. This number must correspond to at least 5 to 10 ppt (party per thousand, that is 0.5 to 1%) of the dataset length.

Step 2: Plot the histogram of the log dataset with class intervals that allow to have leftward classes with frequencies F_i less than F.

Step 3: Identify the class C_p that satisfies $\sum_{i=1}^{p} F_i = F$. The upper limit U_p of C_p is the representative minimum researched.

Step 4: Plot the histogram of the log dataset with class intervals that allow to have rightward classes with frequencies F_i less than F.

Step 5: Identify the class C_q that satisfy $\sum_{i=q}^{m} F_i = F$. The lower limit L_q of C_q is the representative maximum researched.

2.2.2. Evaluation of Well Log Data Statistical Processing Impact

Well log data statistical processing impact assessment will consist of evaluating how the statistical processing of gamma ray data affect positively the results of data analyses. For that, one will proceed to the qualitative and quantitative analysis of unprocessed and processed gamma ray data and identify the main improvements or differences obtained in the results. It is therefore necessary to address the methodology to be used for gamma ray log data analyses.

As for most of the well-logs, two types of analyses are usually performed on GR log data: qualitative and quantitative analyses. This section addresses the actual gamma ray qualitative and quantitative analyses.

A. Gamma Ray Log Data Qualitative Analysis

As stated by Schlumberger (2004), the main purposes of GR and NGS qualitative analysis are the following:

- differentiate potentially porous and permeable reservoir rocks (sandstone, limestone, dolomite) from nonpermeable clays and shales;
- define bed boundaries;
- tie cased hole to openhole logs;
- give a qualitative indication of shaliness;
- monitor radioactive tracers;
- aid in lithology (mineral) identification;
- in the case of the NGS log, detect and evaluate deposits of radioactive minerals;
- in the case of the NGS log, define the concentrations of potassium, thorium, and uranium;
- in the cases of the NGS log, monitor multiple isotope tracers.

As one can see, all the above converge to wells stratigraphic column determination. One of the objectives of the current study being to propose a gamma ray data statistical processing technique for unbiased qualitative analysis thereof, this subsection will focus on the gamma ray qualitative analysis that will be positively impacted by the technique to be proposed: well stratigraphic column determination. A well stratigraphic column describes the vertical location of rocks penetrated by the well; it shows the sequence of sedimentary rocks, with the oldest rocks on the bottom and the youngest on the top (Wikkipedia, 2024).

The GR is usually used to identify boundaries, primarily shale units from other lower radioactivity formations (limestones, sandstones and dolomites) and to quantify shale volume. In fractured formations, an increase in the gamma ray reading without concurrently higher formation shaliness can be observed (Karacan, 2009).

According to Djoï (2012), total gamma ray signature can be used to identify clean sandstones and limestones, shaly sandstones and limestones, sandy shales and shales. Since GR is related to the shale volume of the formations, the shale volume (Vsh) signature is:

- less than 25% for clean sandstones and limestones;
- between 25 and 50% for shaly sandstones and limestones;
- between 50 and 75% for sandy shales;
- between 75 and 100% for shales.

Contrarily to Djoï (2012), Kamel and Mabrouk (2003) set other Vsh cutoff for rock differentiation as follows: rocks can be considered as clean if Vsh < 10%, shaly if Vsh ranged from 10 to 33% and if the Vsh is more than 33%, it is considered to be shale. 0% and 100% of Vsh correspond respectively to the representative minimum and maximum values of GR.

No matter the shale volume cutoff used, the corresponding GR values to shale volume are determined on the basis of GR scale used for the log data presentation (150 or 200 GAPI scale).

As stated by Oberto (1984) and Schlumberger (2004), natural gamma ray spectrometry (NGS) can:

(a) differentiate between shales and potassium salts; these last minerals having a much higher potassium content than the clay minerals, and no thorium content since thorium is insoluble and can be considered as an indicator of detrital origin. So, in front of potassium evaporites, the Th curve will be flat and near zero while the K curve will show a high percentage of potassium and a shape generally very similar to that of the total gamma ray, at least if at the same time the uranium curve is flat and near zero (showing little organic material in the rock).

(b) Recognize the potassium evaporite mineral, through its potassium content, if this mineral form a sufficiently thick bed compared to the NGS vertical resolution. If this is not the case, a combination of the NGS with other logs is necessary for a complete and accurate mineralogy determination in evaporite series.

Since the case study will be performed on total GR log data, the following is the algorithm to use for well stratigraphic column determination.

Algorithm 3: Algorithm for formation identification

Step 1: Determine the baselines.

It consists of determines sand, shaly sand, sandy shales and shales baselines. They are referred to as follow:

- sand baseline is the vertical line whose GR value (noted GR0) corresponds to 0% shale volume in the formation and therefore to the minimum value of GR.
- shaly sand baseline is the vertical line whose GR value (noted GR25) corresponds to 25% shale volume in the formation.
- sandy shale first and second baselines are the vertical lines whose GR values (noted GR50 and GR75) correspond respectively to 50 and 75% shale volumes.
- shale baseline is the vertical line whose GR value (noted GR100) corresponds to 100% shale volume in the formation and therefore to the maximum value of GR.

The GR cutoffs GR0, GR25, GR50, GR75 and GR100 are computed from Equations (9) to (13).

$$GR0 = GR_{min} = min(GR) \tag{9}$$

$$GR100 = GR_{max} = max(GR) \tag{10}$$

$$GR25 = \frac{3\ GR_{min} + GR_{max}}{4} \tag{11}$$

$$GR50 = \frac{GR_{min} + GR_{max}}{2} \tag{12}$$

$$GR75 = \frac{GR_{min} + 3\ GR_{max}}{4} \tag{13}$$

Step 2*:* Identify the tops and bases of different layers as follows.

- clean sandstones or limestones are formations whose GR curves are deflected on the left side of the shaly sand baseline. Their tops and bases correspond to the cross points of GR curve and the shaly sand baseline respectively;
- shaly sandstones (or shaly limestones) are formations whose GR curves are deflected on the right side of the shaly sand baseline (or on the left side of the sandy shale first baseline) and located between the shaly sand and sandy shale first baselines. Their tops and bases

correspond to the cross points of GR curve and the shaly sand (or the sandy shale first) baseline respectively.

- sandy shales are formations whose GR curves are deflected on the right side of the sandy shale first baseline (or on the left side of the sandy shale second baseline) and located between the sandy shale first and second baselines. Their tops and bases correspond to the cross points of GR curve and the sandy shale first (or sandy shale second) baseline respectively.
- shales are formations whose GR curves are deflected on the right side of the sandy shale second baseline and located between the sandy shale second baseline and shale baseline. Their tops and bases correspond to the cross points of GR curve and the sandy shale second baseline respectively.

B. Gamma Ray Log Data Quantitative Analysis

The determination of reservoir quality in terms of petrophysical parameters, lithology identification, porosity, type and distribution of reservoir fluids, formation permeability and anticipated water cut estimates, is mainly based on the evaluation of shale volume (Vsh) (Kamel and Mabrouk, 2003). Gamma ray log is a shale volume indicator and can then be used for shale volume computation. Oberto (1984), Maget (1990), Hamada (1996), Karacan (2009), Szabó (2011), Fadiya, Alao and Adetuwo (2018), Mohammed (2021), Kamayoul, Ehirim and Ikiensikimama (2021) and Djoï, Nwosu and Ikiensikimama (2022) underlined in their studies that gamma ray shale volume must be computed from the gamma ray shale index models. The gamma ray shale index is computed from Equation (14) (Hamada, 1996 and Mohammed, 2021).

$$(I_{sh})_{GR} = \frac{GR - GR_{min}}{GR_{max} - GR_{min}} \tag{14}$$

With GR the total gamma ray read, GR_{min} the minimum of total GR and GR_{max} the maximum of total GR. GR_{min} and GR_{max} are respectively the GR readings in the clean formations (clean sandstone and limestone) and the pure shale (Karacan O. C., 2009).

Oberto, S. (1984) stated that natural gamma ray spectrometry can help for better shale volume index computation stressing that potassium shale volume index, $(I_{sh})_K$, and thorium shale volume index, $(I_{sh})_{Th}$, will serve as better shale indicators than the total gamma ray shale volume index $(I_{sh})_{GR}$ and uranium shale volume index, $(I_{sh})_U$, since the general random associativity of uranium with shale has been eliminated. $(I_{sh})_K$ and $(I_{sh})_{Th}$ are computed as follows (Equations (15) and (16)).

$$(I_{sh})_K = \frac{K - K_{min}}{K_{max} - K_{min}} \tag{15}$$

$$(I_{sh})_{GR} = \frac{Th - Th_{min}}{Th_{max} - Th_{min}} \tag{16}$$

With K and Th the respective potassium and gamma thorium radiations read, K_{min}, K_{max}, Th_{min} and Th_{max} the minimum and maximum of potassium and thorium gamma radiations respectively.

Several GR shale volume index models exist for formation shale volume computation: the linear and non-linear models (Fadiya, Alao and Adetuwo, 2018).

For Oberto (1984), Maget (1990), Hamada (1996), Karacan (2009), Szabó (2011), Fadiya, Alao and Adetuwo (2018) Mohammed (2021), Kamayoul, Ehirim and Ikiensikimama (2021) and Djoï, Nwosu and Ikiensikimama (2022), the linear model states that the shale volume V_{sh} is equal to the shale volume index I_{sh}. Equation (17) is the linear equation for total gamma ray shale volume computation.

Similar equations are used for potassium, thorium and uranium shale volume calculation.

$$V_{sh} = \frac{GR - GR_{min}}{GR_{max} - GR_{min}} \tag{17}$$

The non-linear method of Larionov of Equation (18) is used for shale volume estimation for tertiary and younger rocks (Kamayoul, Ehirim and Ikiensikimama, 2021). Equation (19) helps in estimating the shale volume for older rocks (Szabó, 2011 and Mohammed, 2021).

$$V_{sh} = 0.083 * \left(2^{(3.7 * I_{sh})} - 1\right) \tag{18}$$

$$V_{sh} = 0.083 * \left(2^{(2 * I_{sh})} - 1\right) \tag{19}$$

Mohammed (2021) and Djoï, Nwosu and Ikiensikimama (2022) have shown that two other non-linear models have been developed by Steiber in 1970 and Clavier in 1971. Steiber and Clavier's models are respectively given by Equations (20) and (21).

$$V_{sh} = \frac{I_{sh}}{3 - 2 * I_{sh}} \quad (20)$$

$$V_{sh} = 1.7 - \sqrt{3.38 - (V_{sh} - 0.7)^2} \quad (21)$$

2.3. Data Case Study

The case study has been performed on a Gulf of Guinea's offshore well gamma ray log data. It has consisted of:

- carrying out the statistical processing of the well gamma ray log data in accordance with the approach set in the methodology section;
- performing the qualitative and quantitative analysis of the statistically unprocessed well gamma ray log data;
- performing the qualitative and quantitative analysis of the statistically processed well gamma ray log data;
- comparing the results obtained from the unprocessed and processed data.

The raw data have undergone odds values detection and treatment. The gamma ray log data analysis done are the stratigraphic column determination and formations shale volume computation for a specific interval, using the methods defined in the methodology section.

3. Results and Discussion

3.1. Gamma Ray Data Statistical Processing

The statistical processing performed on the gamma ray log data encompasses the:

- odd values detection and treatment in accordance with the methodology set above;
- outliers detection and treatment using algorithm 1;
- representative minimum and maximum values determination using algorithm 2.

The results are as follows.

As shown in first boxplot of Figure (6), the data contains recorded values greater 150 GAPI whereas the 150 GAPI scale is used. After the removal of these odd values, it has been noticed up to 207 outliers (see boxplot 2 of Figure (6)). Replacing the outliers by the center of the dataset modal class, the boxplot 3 of the same figure is obtained. The semi-processed dataset is the one gotten after removing the odd values. Figure (7) shows the profiles of unprocessed, semi-processed and processed GR log data.

The minimum and maximum values of semi-processed data noted are respectively 9.2508 and 149.3187 GAPI while those of processed data are respectively 12.4 and 107 GAPI. The difference between the maximum and minimum values is 140.0679 GAPI for the semi-processed data and 99.5563GAPI for the processed data. One can notice the greatness of that difference of semi-processed data comparatively to processed data. As a result, for a same underground formation penetrated by the well, the semi-processed data analysis will underestimate the formation shale volume. If for instance, that formation is located in the oil or gas reservoir pay zone, then its effective porosity and flow performance will be overestimated.

The gamma ray cutoffs (GR25, GR50 and GR75) obtained from the semi-processed data are respectively 44.27, 79.28 and 114.30 GAPI while those from the processed data are 36.05, 59.70 and 83.35 GAPI. The first remark is that the baselines (shaly sand, sandy shale and shale baselines) obtained from the semi-processed data analysis are respectively located more rightward to those from the processed data analysis. The main consequence is that the semi-processed data analysis will hide the shaliness of formations. That is, in same conditions of thickness, the semi-processed data analysis results will show:

- a shaly sandstone as a sandstone;
- a sandy shale as a shaly sandstone or sandstone;
- a shale as a sandy shale or shaly sandstone or sandstone.

3.2. Well Stratigraphic Column Determination

The qualitative analysis of the gamma ray data has been performed on the semi-processed and processed data over an interval of 78.5 meters of thickness with 2,875 mMSL and 2,953.5 mMSL as top and base depths respectively. The aim is to show how the log data analysis results are affected when the statistical data processing is not carried out. The outcomes of these analyses are presented as follows.

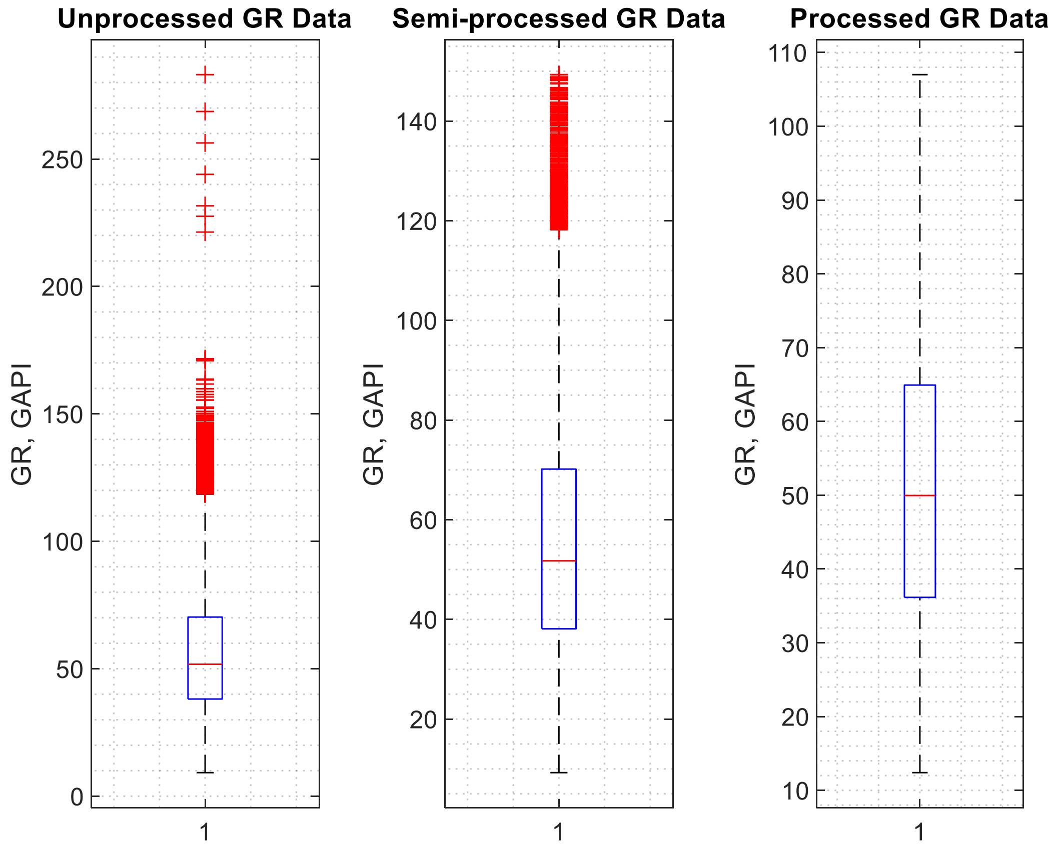


Figure 6 – Boxplots of unprocessed, semi-processed and processed gamma ray data.

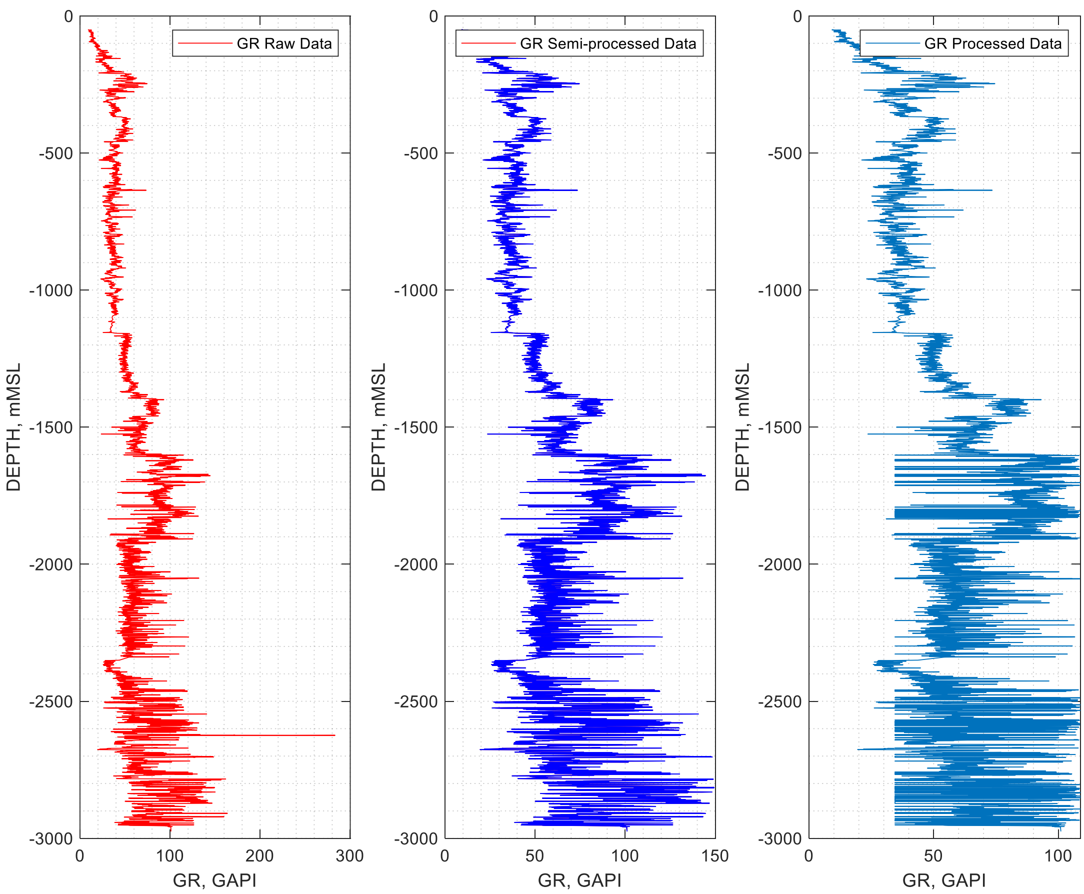


Figure 7 - Profile of unprocessed, semi-processed and processed gamma ray log data.

The predefined condition is that only layers with thickness greater than 1 meter have being considered over the qualitative analysis process.

Figure (8) shows the different layers identified from the analyses of semi-processed and processed GR log data. Twenty-one layers has been obtained from the semi-processed data with lithologies going from shaly sandstone to shale. The shaly sandstones and sandy shales are predominant with only two shale layers. The formations thicknesses range from 1.14 to 8.84 meters. Table (1) summarizes the base and top depths, lithologies and thicknesses of formations.

As far as the processed data analysis results are concerned, as summarized by Table (2), the well has penetrated twenty layers. These formations are either sandstones, or shaly sandstones, or sandy shales, or shales. Seven shales and six sandy shales have been noted, the remainder being sandstones or shaly sandstones. The thinnest formation is 0.84-meter thick while the thickest one is 10.21.

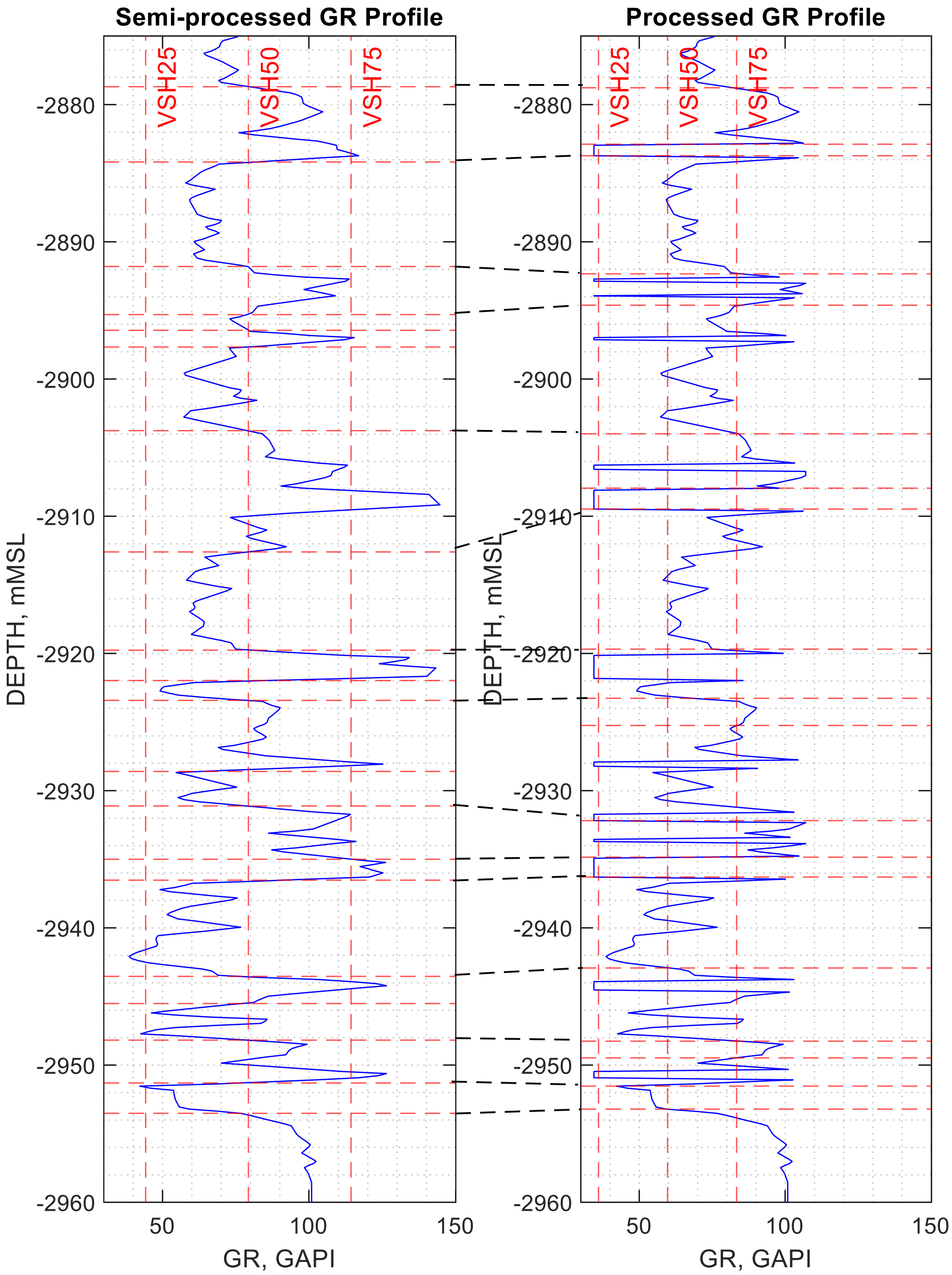


Figure 8 – Layers flow from data analysis. Left: Semi-processed data analysis results. Right: Processed data analysis results. Red vertical dashed lines are the baselines. Red horizontal dashed lines are the layers tops and bases. Black dashed lines are the tops and bases of sets of layers.

Table 1 – Results of semi-processed data qualitative analysis.

N°	Layer	Top depth (mMSL)	Base depth (mMSL)	Nature	Thickness (m)
1	A	2,875	2,878.7	Shaly sandstone	3.66
2	B	2,878.7	2,884.2	Sandy shale	5.49
3	C	2,884.2	2,891.8	Shaly sandstone	7.62
4	D	2,891.8	2,895.3	Sandy shale	3.51
5	E	2,895.3	2,896.4	Shaly sandstone	1.14
6	F	2,896.4	2,897.7	Sandy shale	1.22
7	G	2,897.7	2,903.8	Shaly sandstone	6.10
8	H	2,903.8	2,912.6	Sandy shale	8.84
9	I	2,912.6	2,919.8	Shaly sandstone	7.16
10	J	2,919.8	2,922	Shale	2.21
11	K	2,922	2,923.4	Shaly sandstone	1.45
12	L	2,923.4	2,928.6	Sandy shale	5.18
13	M	2,928.6	2,931.1	Shaly sandstone	2.51
14	N	2,931.1	2,935	Sandy shale	3.89
15	O	2,935	2,936.5	Shale	1.52
16	P	2,936.5	2,943.5	Shaly sandstone	7.01
17	Q	2,943.5	2,945.5	Sandy shale	1.98
18	R	2,945.5	2,948.2	Shaly sandstone	2.67
19	S	2,948.2	2,951.3	Sandy shale	3.12
20	T	2,951.3	2,953.5	Shaly sandstone	2.21
21	U	2,953.5	2,960.1	Sandy shale	6.55

3.3. Formation Shale Volume Computation

The results of formations shale volumes computation are summarized in Table (3). From the semi-processed data analysis, it can be seen that formations shale volumes range from 33.5 to 82.21%. The shaly sandstones A, E, G are highly shaly while the sandy shales D, F, H, L, N, Q and U are highly sandy. The smallest formation shale volume computed from the processed data is 23.34% while the highest is 87.69%. Only two shaly sandstones, O’ et S’, are highly shaly when the sandy shales D’, P’ and R’ contain great proportions of sands.

Table 2 – Results of processed data qualitative analysis.

N°	Layer	Top depth (mMSL)	Base depth (mMSL)	Nature	Thickness (m)
1	A’	2,875	2,878.8	Sandy shale	3.73
2	B’	2,878.8	2,882.9	Shale	4.11
3	C’	2,882.9	2,883.7	Sandstone	0.84
4	D’	2,883.7	2,892.3	Sandy shale	8.61
5	E’	2,892.3	2,894.6	Shale	2.29
6	F’	2,894.6	2,904	Sandy shale	9.37
7	G’	2,904	2,907.9	Shale	3.96
8	H’	2,907.9	2,909.5	Sandstone	1.52
9	I’	2,909.5	2,919.7	Sandy shale	10.21
10	J’	2,919.7	2,923.3	Shaly sandstone	3.58
11	K’	2,923.3	2,925.2	Shale	1.98
12	L’	2,925.2	2,932.2	Sandy shale	6.93
13	M’	2,932.2	2,934.8	shale	2.67
14	N’	2,934.8	2,936.3	Sandstone	1.45
15	O’	2,936.3	2,942.9	Shaly sandstone	6.63
16	P’	2,942.9	2,948.3	Sandy shale	5.33
17	Q’	2,948.3	2,949.5	Shale	1.22
18	R’	2,949.5	2,951.5	Sandy shale	2.06
19	S’	2,951.5	2,953.2	Shaly sandstone	1.68
20	T’	2,953.2	2,960.1	Shale	6.86

3.4. Comparative Analysis of Results

Since only similar formations can be compared and the shaliness of layers depends on the number of recorded GR values and therefore on their thicknesses, the correlation of both stratigraphic columns has helped to identify sixteen similar sets of layers (Figure (8)). The characteristics of these layers are those of Table (4). Figures (9) and (10) show the thicknesses and shale volumes of the sets of formations. It is noted that the semi-processed data analysis has globally underestimated the thickness of thicker formations and underestimate the shale volumes of thicker formations and those for which the estimated thicknesses from both analyses are the same or close to each other. The semi-processed data analysis underestimation of thicker formations thicknesses is due to the fact that the baselines (shaly sand, sandy shale and shale baselines) from the semi-processed data analysis are respectively located more rightward to those from processed data analysis. That is, the gamma ray cutoffs (GR25, GR50 and GR75) obtained from the semi-processed data are respectively higher than those from the processed data. In the same time, the semi-processed data analysis underestimation of shale volumes of thicker formations and formations with same estimated thicknesses from both analyses or close to each other is explained by the fact that the difference between the maximum and minimum values of the semi-processed data is greater than the one of the processed data.

Table 3 – Formations shale volumes from the semi-processed and processed data analyses.

	Semi-processed data analysis results			Processed data analysis results		
N°	**Layer**	**Nature**	**Shale Volume (%)**	**Layer**	**Nature**	**Shale Volume (%)**
1	A	Shaly sandstone	43.94	A'	Sandy shale	62.11
2	B	Sandy shale	62.68	B'	Shale	87.09
3	C	Shaly sandstone	39.18	C'	Sandstone	23.34
4	D	Sandy shale	61.31	D'	Sandy shale	56.00
5	E	Shaly sandstone	47.84	E'	Shale	77.86
6	F	Sandy shale	63.08	F'	Sandy shale	62.83
7	G	Shaly sandstone	42.35	G'	Shale	78.53
8	H	Sandy shale	60.97	H'	Sandstone	23.34
9	I	Shaly sandstone	39.92	I'	Sandy shale	61.73
10	J	Shale	78.70	J'	Shaly sandstone	34.70
11	K	Shaly sandstone	33.82	K'	Shale	78.82
12	L	Sandy shale	55.35	L'	Sandy shale	60.00
13	M	Shaly sandstone	39.50	M'	shale	83.22
14	N	Sandy shale	66.21	N'	Sandstone	23.34
15	O	Shale	80.21	O'	Shaly sandstone	45.27
16	P	Shaly sandstone	33.56	P'	Sandy shale	55.91
17	Q	Sandy shale	65.06	Q'	Shale	85.18
18	R	Shaly sandstone	38.25	R'	Sandy shale	57.41
19	S	Sandy shale	61.5	S'	Shaly sandstone	43.12
20	T	Shaly sandstone	33.91	T'	Shale	87.69
21	U	Sandy shale	62.45	-	-	-

Table 4 – Shale volumes and thickness of sets of formations.

	Semi-processed data analysis results				Processed data analysis results			
N°	**Set of layers**	**Nature**	**VSH (%)**	**Thickness (m)**	**Set of layers**	**Nature**	**VSH (%)**	**Thickness (m)**
N°	**Set of layers**	**Nature**	**VSH (%)**	**Thickness (m)**	**Set of layers**	**Nature**	**VSH (%)**	**Thickness (m)**
1	S1 = {A}	Shaly sandstone	43.94	3.66	S'1 = {A'}	Sandy shale	62.11	3.73
2	S2 = {B}	Sandy shale	62.68	5.49	S'2 = {B', C'}	Shale and sandstone	55.21	4.95
3	S3 = {C}	Shaly sandstone	39.18	7.62	S'3 = {D'}	Sandy shale	56	8.61
4	S4 = {D}	Sandy shale	61.31	3.51	S'4 = {E'}	Shale	77.86	2.29
5	S5 = {E, F, G}	Shaly sandstones intercalated by sandy shale	53.10	8.46	S'5 = {F'}	Sandy shale	62.83	9.37
6	S6 = {H}	Sandy shale	60.97	8.84	S'6 = {G', H'}	Shale and sandstone	50.94	5.49
7	S7 = {I}	Shaly sandstone	39.92	7.16	S'7 = {I'}	Sandy shale	61.73	10.21
8	S8 = {J, K}	Shale and shaly sandstone	56.26	3.66	S'8 = {J'}	Shaly sandstone	34.70	3.58
9	S9 = {L, M}	Sandy shale and shaly sandstone	47.42	7.70	S'9 = {K', L'}	Shale and sandy shale	69.41	8.92
10	S10 = {N}	Sandy shale	66.21	3.89	S'10 = {M'}	shale	83.22	2.66
11	S11 = {O}	Shale	80.21	1.52	S'11 = {N'}	Sandstone	23.34	1.44
12	S12 = {P}	Shaly sandstone	33.56	7.01	S'12 = {O'}	Shaly sandstone	45.27	6.63
13	S13 = {Q, R}	Sandy shale and shaly sandstone	51.66	4.65	S'13 = {P'}	Sandy shale	55.91	5.33
14	S14 = {S}	Sandy shale	61.51	3.12	S'14 = {Q', R'}	Shale and sandy shale	71.29	3.28
15	S15 = {T}	Shaly sandstone	33.91	2.21	S'15 = {S'}	Shaly sandstone	43.11	1.68
16	S16 = {U}	Sandy shale	62.45	6.55	S'16 = {T'}	Shale	87.69	6.88

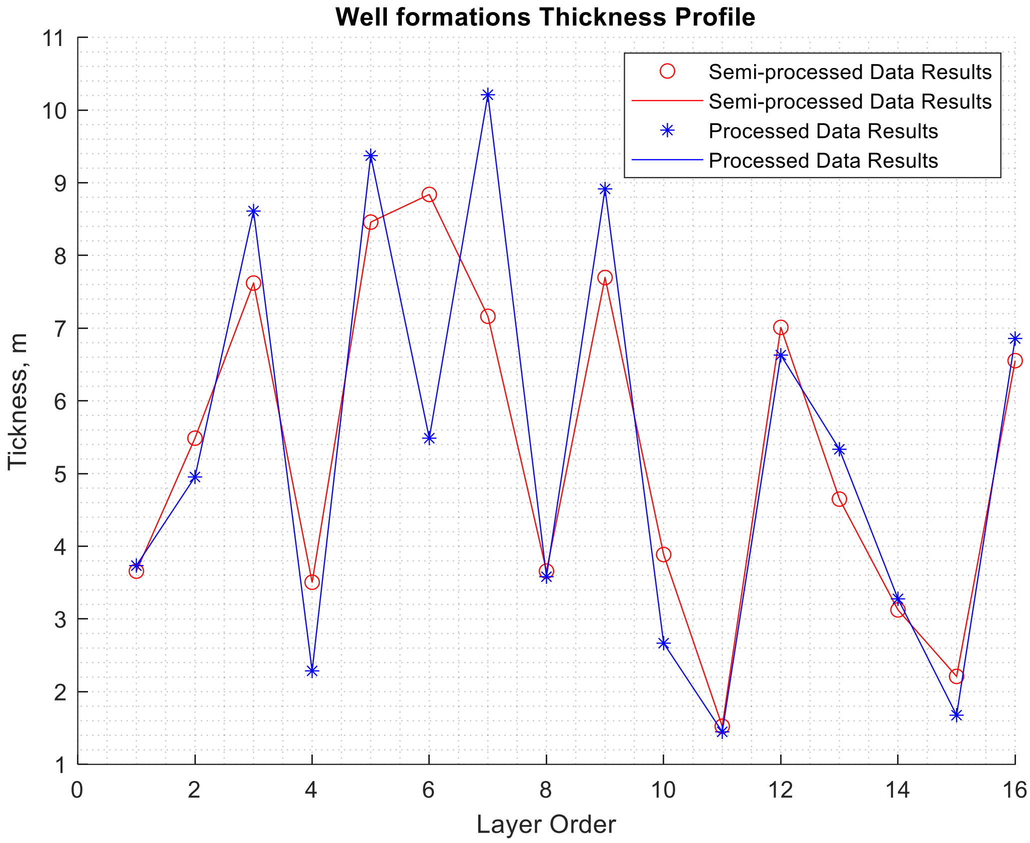


Figure 9 – Well formations thickness profile. Order 1 for the set of layers {A} and {A'}, order 2 for the sets of layers {B} and {B', C'}, …, order 16 for sets of layers {U} and {T'}.

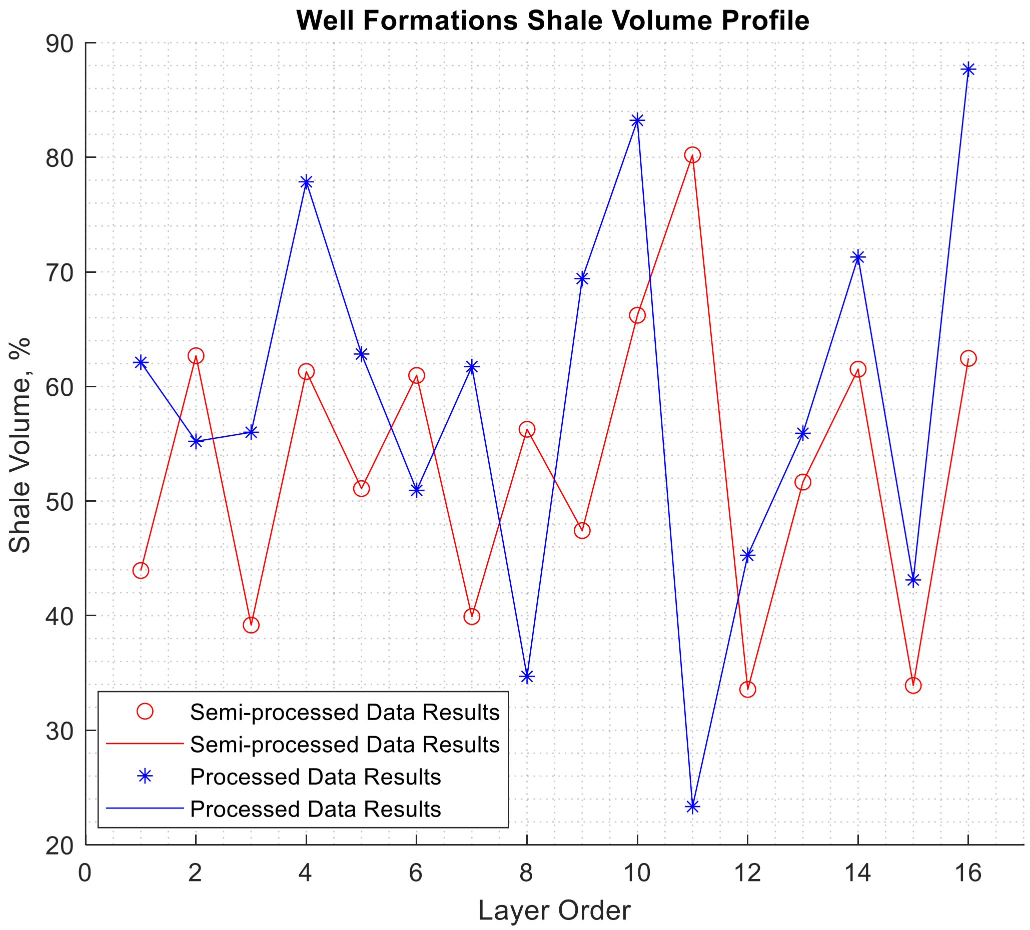


Figure 10 – Well formations shale volume profile. Figure – Well formations thickness profile. Order 1 for the sets of layers {A} and {A'}, order 2 for sets of layers {B} and {B', C'}, …, order 16 for sets of layers {U} and {T'}.

3.5. Impact of Gamma Ray Data Statistical Processing on the Analysis Results

As it can be noticed through the results of the case study, the statistical processing of gamma ray log data prevents from:

- the underestimation of thicker formations thicknesses;
- the underestimation of formations shale volumes.

The main practical advantage is that it will prevent geologists, petro-physicists and reservoir engineers from the overestimation of oil or gas reservoirs effective porosity and flow performance and therefore from the overestimation of oil or gas initially in place and reserves.

4. Conclusion

Over oil and gas fields exploration and development phases, one of the main challenges of geoscientists and petroleum engineers is the petrophysical characterization of potential or discovered fields or reservoirs. Well logs data are most used for that purpose. They play key roles in wells stratigraphic column establishment and the computation of reservoir formations petrophysical parameters such as shale volume, porosity, permeability, fluids (water, oil and gas) saturation, reservoirs net pay thickness, fluids contacts (WOC and GOC), the skin effect and reservoir pressure.

Due to the conditions and the environment of well log data acquisition, they undergo some technical processing. For gamma ray log data for instance, the minimum and maximum values of recorded GR are required for qualitative and quantitative analyses. These GR minimum and maximum values must be determined in the ways they are representative, that is, the really characterize the clean sandstones/limestones and shales respectively. Indeed, although the technical gamma log data processing, some data analysis results do not reflect the reality. The statistical data processing is important for unbiased analyses.

This study aims to propose statistical techniques for gamma ray logs data processing that will contribute to the reduction of biases related to their qualitative and quantitative analyses. A case study has been performed on a Gulf of Guinea's offshore well gamma ray log data. It has consisted of semi-statistical processing and whole statistical processing of the gamma ray log data, well stratigraphic columns establishment and formations shale volumes computation from the semi-processed and processed data.

The results show that the difference between the maximum and minimum values for the semi-processed data is almost twice the one of the processed data, what will lead to the underestimation of formations shale volumes and therefore to the overestimation of reservoirs effective porosity and flow performance. Moreover, the baselines (shaly sand, sandy shale and shale baselines) obtained from the semi-processed data are respectively located more rightward to those from the processed data. The main consequence is that the semi-processed data analysis has hidden the shaliness of formations comparatively to the processed data analysis. Indeed, the semi-processed data analysis revealed twenty-one layers showing lithologies going from shaly sandstone to shale with the predominance of shaly sandstones and sandy shales and counting only two shales. The processed data analysis has, on contrary, led to a twenty-layer column of sandstones, shaly sandstones, sandy shales and shales with six sandy shales and seven shales. A comparative analysis shows that the semi-processed data analysis has globally underestimated the thickness of thicker formations and underestimate the shale volumes of thicker formations and those for which the estimated thicknesses from both analyses are the same or close to each other.

In summary, it is noticed, through the results of the case study, that the statistical processing of gamma ray log data prevents from the underestimation of thicker formations thicknesses and formations shale volumes. The main practical advantage is that it will prevent geologists, petro-physicists and reservoir engineers from the overestimation of oil or gas reservoirs effective porosity and flow performance and therefore from the overestimation of oil or gas initially in place and reserves.

We recommend:

- to use of the statistical gamma ray log data processing technique developed in this study for gamma ray log and other similar log data processing;
- to identify other statistical data processing required for other types of log data and propose the processing approaches.

References

Aguinis, H., Ryan, K.G. and Harry, J., 2013. *Best-Practice Recommendations for Defining, Identifying, and Handling Outliers*. Organizational Research Methods 16(2), pp. 270-301. https://doi.org/10.1177/1094428112470848.

Asquith, G. and Krygowski, D., 2004. *Basic well log analysis*, *Second edition.* AAPG, Methods in Exploration Series, No. 16. American Association of Petroleum Geologists, Tlsa, Oklahoma. https://doi.org/10.1306/Mth16823

Baron, J., Cariou, J. and Thorion, R., 1989. *Les diagraphies nucléaires développées par les LPC Principes physiques, mise en œuvre, domaines d'application*. Liaison Labo. P. et Ch – 164, nov-dec 1989 – Ref – 3446. French.

Bateman, R.M., 2020. *Formation evaluation with pre-digital well logs*. Elsevier Inc., Radarweg 29, PO Box 211, 1000 AE Amsterdam, Netherlands. https://doi.org/10.1016/C2018-0-03023-6

Bassiouni, Z., 1994. *Theory, Measurement and Interpretation of Well Logs*. SPE Text Book Series Vol. 4, Society of Petroleum Engineer, Texas, USA. https://doi.org/10.2118/9781555630560

HLS Asia Limited, 2007. Basic log interpretation. Log Interpretation Seminar/ Workshop (14th – 16th May 2007, New Delhi). HLS Asia Limited, 2007.

Chapelier, D., 2004. *Diagraphie Pétrole. Support de cours. Cours online de géophysique : Cours de Diagraphies Différées - Option Réservoirs*. Université de Lausanne, Institut Français de Pétrole. French.

Cuisineau, D., and Chartier, S., 2010. *Outliers detection and treatment: a review*. International Journal of Psychological Research, 3(1), pp. 59-68.

Djoï, N.A., Nwosu, I.J. and Ikiensikimam,a S.S., 2023. *Stochastic Activity-Based Time and Cost (S-ATC) Model for Oil Wells: Case Study of Niger Delta Onshore, Nigeria*. International Journal of Engineering Trends and Technology, 71(5), pp. 49-69. https://doi.org/10.14445/22315381/IJETT-V71I5P206

Djoï, N.A., Nwosu, I.J. and Ikiensikimama S.S., 2022. *Hydrocarbon Reservoir Petrophysical Characterization with Statistical Simulations: A Case Study from the Gulf of Guinea*. International Journal of Scientific Research and Engineering Development, 5(2), pp. 298-308.

Djoï, N.A., 2015. *Détermination de l'intervalle de perforation du puits X.* Rapport de travaux. Société Béninoise des hydrocarbures, Ministère de l'Energie, Bénin. French.

Djoï, N.A., 2012. *Détermination des Paramètres Pétrophysiques de Réservoir à Partir des Diagraphies Différées*. Rapport de stage de production. Ecole Supérieure des Mines et de Géologie. Ecole Supérieure des Mines et de Géologie, Institut National Polytechnique Félis Houphouët -Boigny, Côte d'Ivoire. French.

Ekwere, J.P. 2004. *Petrophysics*. Department of Petroleum and Geosystem Engineering. University of Texas at Austin, USA 78712.

Fadiya, S.L., Alao, O.A. and Adetuwo, A.M., 2018. *A Comparative Study of the Methods of Determining Shale Volume in Radioactive Reservoirs of "AMA" Field, Niger Delta, Nigeria*. International Journal for Research in Emerging Science and Technology, 5(4), pp. 1-39.

Hamada, G.M., 1996. An Integrated Approach to Determine Shale Volume and Hydrocarbon Potential in Shaly Sand. In: *Society of Core Analysts Conference*, SCA. Number 9641.

Iglewicz, B. and Hoaglin, D.C., 1993. *How to Detect and Handle Outlier.* American Society for Quality Control, Statistics Division, vol. 16, 1993.

Illowsky, B. and Dean, S., 2021. *Introductory Statistics, 4th Edition*. College of Lake County, penStax College Publisher, July 2021.

Joseph, C.W., 2015. *An Introduction to the Science of Statistics: From Theory to Implementation*. Preliminary Edition, June 2015.

Kamayoul, V.M., Ehirim, C.N. and Ikiensikimama, S.S., 2021. *Estimating Volume of Shale in a Clastic Niger Delta Reservoir from Well Logs: A*

Comparative Study. International Journal of Geosciences, 12(10), pp. 949-959. https://doi.org/10.4236/ijg.2021.1210049

Kamel, M.H. and Mabrouk, W.M., 2003. *Estimation of shale volume using a combination of the three porosity logs*. Journal of Petroleum Science and Engineering 40(3-4), pp. 145 – 157. https://doi.org/10.1016/S0920-4105(03)00120-7

Karacan, O.C., 2009. *Reservoir rock properties of coal measure strata of the Lower Monongahela Group, Greene County (Southwestern Pennsylvania), from methane control and production perspectives*. CDC/NIOSH Pittsburgh Research Laboratory, Disaster Prevention and Response Branch, 626 Cochrans Mill Road, PO Box 18070, Pittsburgh PA, 15236, United States.

Maget, Ph., 1990. *Les diagraphies en forage: Initiation aux méthodes et aux interprétations, Application en géothermie*. Rapport du B.R.G.M 85 SGN 034 SPG. Service de Géothermie, A.F.M.E - B.R.G.M. French.

Mohammed, O.A.A., 2021. *Machine learning based shale volume prediction from the Norwegian North Se*a. MSc Thesis, Faculty of Science and Technology, University of Stavanger.

Szabó, N.P., 2011. *Shale volume estimation based on the factor analysis of well-logging data*. Acta Geophysica, 59, Pp. 935-953. https://doi.org/10.2478/s11600-011-0034-0

Oberto, S., 1984. *Fundamentals of well-log interpretation: The acquisition of logging data*. Developments in Petroleum Science, 15A. Elsevier Science Publishers B.V., 1984, Amsterdam.

Rabaute, A., 2009. *Obtenir une représentation en continu de la lithologie et de la minéralogie. Exemples d'application du traitement statistique de données de diagraphie aux structures sédimentaires en régime de convergence de plaques (Leg ODP 134, 156 et 160).* Thèse de doctorat. Université de Montpellier II Sciences et Techniques du Languedoc. French.

Schlumberger, 2004. *Gamma ray correction for hole size and mud weight Charts*. 225 Schlumberger Drive, Sugar Land, Texas 77478.

Schlumberger, 2009. *Log Interpretation Principles/Applications and Well Log Interpretation Charts*. 225 Schlumberger Drive, Sugar Land, Texas 77478.

Wikipedia, the free encyclopedia, 2024. *Stratigraphic column*. Available at: <https://en.wikipedia.org/wiki/Stratigraphic_column>. [Accessed 25 Jan 2024].

Williamson, H.S., Sawaryn, S.J. and Morrison, J.W., 2006. *Monte Carlo Techniques Applied to Well Forecasting: Some Pitfalls*. SPE Drill & Compl 21(03), pp. 216-227. Society of Petroleum Engineers (SPE), SPE Annual Technical Conference and Exhibition. https://doi.org/10.2118/89984-PA