



Descriptive statistics and probability distributions of volumetric parameters of a Nigerian heavy oil and bitumen deposit

O. O. Mosobalaje¹ · O. D. Orodu¹ · D. Ogbe²

Received: 12 July 2017 / Accepted: 4 June 2018 / Published online: 8 June 2018
© The Author(s) 2018

Abstract

The absence of geostatistical modeling of volumetric parameters of the long-discovered Nigerian heavy oil and bitumen deposits is responsible for the inconsistencies surrounding estimates of hydrocarbon-in-place contained therein. An exploratory data analysis (EDA) is a pre-cursor to such modeling. As part of EDA, this work presents the descriptive statistics and probability distributions of the volumetric parameters of a Nigerian heavy oil and bitumen deposit. Raw data from the existing works have been assembled into a database. Using basic principles, porosity have been computed, from the raw data, for several core samples retrieved from the two bituminous horizons in the deposit. The computed database has been partitioned into the two horizons, using depth-to-top and thickness data. Furthermore, this work has conducted detailed analyses and offers robust discussions on the descriptive statistics and probability distributions of the porosity, depth-to-top, and thickness databases. The statistics and distribution curves obtained are observed to exhibit good correlations with existing geologic, stratigraphic, and textural data. An hypothesis suggesting the two horizons belong to same geological population has been formulated and tested at field and well levels; with results affirming the hypothesis. The descriptive statistics and probability distributions obtained offer a significant understanding of the characteristics and features of the available data. In addition, the distributions now become prior information to which reservoir descriptions would be constrained, in the future conditional simulation stage of this work. The correlation of core data obtained here with the existing geologic, stratigraphic, and textural data would promote data integration in the characterization of this deposit.

Keywords Heavy oil and bitumen · Nigeria · Descriptive statistics · Probability distribution · Porosity · Payzone thickness

Introduction

Revenues from the oil and gas sector are the mainstay of the economy and foreign exchange of the Nigerian state. Recent reports (OPEC 2016) attributed 35% of the nation's GDP and 92% of total export revenue to the sector. However, as evident in recent reports, the scenario currently prevailing in the sector is that of stagnant reserves, declining production capacity, small Reserves-to-Production (R/P) ratio. The nation's proved crude oil reserves has been stagnant at 37 billion barrels since 2010, while its

production capability declined from 2.048 million barrels/day in 2010 to 1.748 million barrels/day in 2016 (OPEC 2011, 2016). At the current reserves and production rate, the R/P ratio is estimated to be 58 years; the nation's current reserves would be exhausted in 58 years if no additional reserves are booked. Conversely, the internal consumption demand for crude oil had increased from 259,000 barrels/day in 2010 to 407,000 barrels/day in 2016 (OPEC 2011, 2016). In the face of the foregoing disturbing trends, it is expedient that the nation should seek to grow its reserves. On the global scale, the potential for reserves growth lies more in unconventional sources of crude oil such as heavy oil, natural bitumen, and shale oil than in the conventional fields (Stark et al. 2007). In Nigeria's case, the conventional fields are becoming harder to find and occur less frequently in the matured Niger Delta oil province. Interestingly, vast deposits of heavy oil and natural bitumen have been long-discovered in the Dahomey basin (Benin basin) southwestern Nigeria. Several studies

✉ O. O. Mosobalaje
olatunde.mosobalaje@covenantuniversity.edu.ng;
olu645online@yahoo.com

¹ Department of Petroleum Engineering, Covenant University, Ota, Nigeria

² African University of Science and Technology, Abuja, Nigeria

(Adegoke et al. 1980; Conoco, 2002 cited in; Lawal 2011a; ERA, 2003 cited in Lawal 2011a; Energy Commission of Nigeria, ECN, 2003; Adebisi et al. 2005; MSMD, 2006; Meyer et al. 2007; Attanasi and Meyer 2010; MMSD, 2010 cited in Lawal 2011a; and Lawal 2011a) have attempted to estimate the volume of resource in-place in these Nigerian heavy oil and natural bitumen deposits. The values reported by these studies have been inconsistent and vary widely from 30 to 420 Billion Barrels. The inconsistency has prohibited any serious commercial investment towards the exploitation of the resources; hence, the resources has not been booked as part of the nation's reserves. Acknowledging the fact that the official value has been fixed at 43 Billion Barrels (MSMD, 2006); the possible oil reserves from Nigeria's heavy oil and natural bitumen deposits can grow the nation's reserves to more than double its current value. The undeniable need to grow Nigeria's oil reserves, in the face of reduced frequency of discovery of the conventional oilfields, and the presence of vast deposits of heavy oil and natural bitumen, therefore, provides a strong motivation for this paper.

The inconsistency and divergence that characterize the outcomes of the existing attempts at estimating the volumetrics of the deposits are attributable to the absence of spatial variability analysis in such studies. None of the studies have considered the spatial variability (heterogeneity) of key volumetric parameters of the deposits; rather, the studies have used either assumed constant or averaged values of these key parameters and have often assume homogeneity of the entire deposit belt or of identified layers in the belt. A recent study (Lawal 2011a), while reporting an attempt to improve on the estimation of the hydrocarbon potentials of the bitumen resource deposits, acknowledged its own limitations and recommended that future studies be focused on the construction of a comprehensive geo-model to accommodate all available data and, more importantly, to account for both areal and vertical variability in the distribution of key volumetric parameters. The construction of such integrated geo-model is in the realm of geostatistical modeling and stochastic simulation. However, to obtain credible results, a crucial preliminary exploratory data analysis (EDA) must be conducted on available database prior to geostatistical modeling. While EDA is not the core geostatistical modeling, it is essential to gain an integrated understanding of the data, data trends, and relationships between data variables. Such understanding makes the modeler develop intimacy with the data. Through EDA, possible sources of errors and suspicious data are detected and deleted. Exploring the descriptive statistical measures and probability distribution of the data is an integral preliminary part of EDA. Descriptive statistics and probability distribution offers a means of profiling the available data. With the objective of acquiring intimate understanding of available data, this work conducted

detailed analyses and offers discussions on various univariate descriptive statistics and probability distribution of the volumetric parameters of a part of the Nigerian heavy oil and natural bitumen deposits.

The study area: Agbabu heavy oil and bitumen field

Diverse studies have confirmed the existence of heavy oil and natural bitumen deposits in the southwestern part of Nigeria. Some of these studies (Adegoke et al. 1980; Enu 1985; Olabanji et al. 1994; Fayose 2005; MSMD, 2006; Lawal 2011a; Akinmosin et al. 2012 and; Adeyemi et al. 2013) reported both outcrop and sub-surface sections of the deposits traversing an East–West belt spanning Ogun and Ondo states in particular. Enu (1985), Olabanji et al. (1994), Fayose (2005), Lawal (2011a), and Akinmosin et al. (2012) included parts of Lagos and Edo states. Figure 1 is the map of Nigeria showing these four states and the bitumen belt across them. The exposed outcrop section of the deposits spans an East–West length of about 120–140 km, while the North–South width is about 4–6 km (Lawal 2011a and; Akinmosin et al. 2012). The Nigerian deposits are accumulated in the eastern part of the Dahomey basin; the entire coastal sedimentary basin stretches from Ghana-Ivory Coast border to western Nigeria. The outcrop exposure of the block-faulted Cambrian basement rock that underlies the basin marks its northern border, while the southern border, not distinctly defined, is thought to lie beneath the ocean floor (MSMD, 2006). The formation of the basin during the early Cretaceous time was a consequence of the block faulting that heralded the creation of the Atlantic Ocean (Adegoke et al., 1980 and Fayose 2005). Comprehensive reviews of the geology of the basin are presented in Enu (1985), Fayose (2005), MSMD (2006), Odunaike et al. (2009), Akinmosin et al. (2012), and Adeyemi et al. (2013).

The stratigraphic sequence of the deposits has been reviewed by Anukwu et al. (2014). Omatsola and Adegoke (1981, cited in Anukwu et al. 2014) identified three (3) distinct formations in the deposits: Ise formation at the bottom, Afowo formation, and Araromi formation at the top; all occurring within the lower Cretaceous-to-Paleocene sequence of the basin. The oldest formation which lies on the Cambrian basement rock is identified as Ise formation; it is continental and is made up of unconsolidated well-sorted conglomerates, coarse sands (quartz), and medium-grained sands. This formation is essentially water bearing. The age of Ise formation is Neocomian and its maximum thickness of 609 m occurs in Ise – 1 well. (Enu 1985; Fayose 2005 and MSMD, 2006). Afowo formation, which is transitional to marginally marine, lies on the Ise formation and is made up of fine-to-medium-grained sands interbedded with shales

data, the generated porosity database was partitioned into the two horizons (Horizon X and Y). At the core of this work, measures of univariate descriptive statistics and probability distribution curves of porosity, depth-to-top, and thickness data were generated for the entire study area, each horizon and each wellbore. Thereafter, the statistical measures and distribution curves obtained were correlated with geological/facies and textural data available. Figure 3 shows the X–Y coordinate of the locations of the wells drilled in the study area.

Data computations and statistical analysis methods

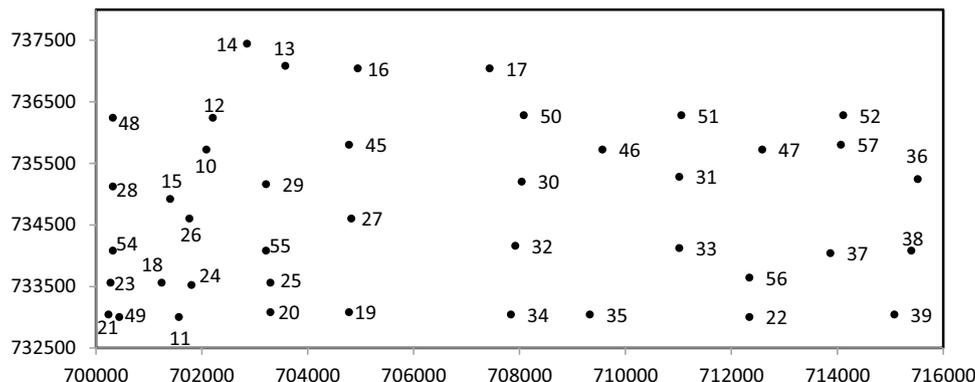
The data contained in the existing database includes core sample number, depth of sampling, wet tar (bitumen) weight percent, dry tar weight percent ($%W_o$), and water content weight percent ($%W_w$) for all core samples, as well as depth-to-top and thickness of horizons for all wells. The computations and statistical analyses done in this current work are only for core samples retrieved from Horizon X and Horizon Y strata, a total of 443 core samples. Each core sample is reported to weigh 20 g. On the basis of total core mass, $M_T = 20$ g; mass of bitumen (M_o) and mass of water (M_w) contained in each sample were computed. Mass of rock grain (M_r) was computed as the balance the total core mass. The volumes of each component of the core sample (bitumen, water, and rock grains) were obtained from the masses using their respective specific gravity values. Ola (1991) measured and reported the average specific gravity of the Agbabu tar sand rock grains (SG_r) to be 2.58. In converting bitumen metric tons to bitumen barrels, Falebita et al. (2014) estimated the conversion factor to be 6.4977 barrel/ton. According to the authors, this value is based on average specific gravity of bitumen samples obtained from Agbabu. Using this conversion factor, we computed the average bitumen specific gravity (SG_o) to be 0.9679. In justifying the use of this value, we first utilized it to compute bulk density

for each of the core samples and observed that the average bulk density for all 445 samples is 2.15; a value which we adjudged sufficiently close to 2.24 being the single value reported by Adegoke et al. (1980). We refrained from using a single constant value for bulk density, because it is our considered view that assuming that a constant value for bitumen specific gravity is more realistic than assuming a constant value for core bulk density. Core bulk density is expected to vary with fluid porosity and fluid saturation; these parameters are known to vary widely in the study area. Furthermore, the adopted value of SG_o (i.e. 0.9679) is equivalent to an API gravity of 14.69° for the Agbabu bitumen; this value falls within range of API gravity for oils classified as heavy oil/bitumen by Tedeschi (1991) and Meyer et al. (2007). Water-specific gravity (SG_w) was taken to be 1.0. Ultimately, porosity value (in decimal fractions) for each core sample was simply computed as a ratio of total fluid (bitumen and water) to total sample volume given thus:

$$\phi = \frac{\frac{0.01 \times \%W_o \times M_T}{SG_o} + \frac{0.01 \times \%W_w \times M_T}{SG_w}}{\frac{0.01 \times \%W_o \times M_T}{SG_o} + \frac{0.01 \times \%W_w \times M_T}{SG_w} + \frac{0.01(100 - \%W_o - \%W_w)M_T}{SG_r}} \quad (1)$$

A perusal of the porosity values obtained revealed that some of the values are suspiciously high; a few values were as high as 0.76. Considering the fact that the tar sand is unconsolidated and only held together the bitumen, large errors may occur in porosity measurements due to the disturbance to natural packing during handling. While it is expected that porosity values for such unconsolidated loose sands as the Agbabu tar sands would be high, maximum values expected have been pegged at 0.45 (Xuetao et al. 2017). Furthermore, a large database of porosities of the Canadian Athabasca tar sands (Clark 1957) contains values only as high as 0.52 after it has been subjected to error detection process. The Athabasca tar sand is a popular analog of the Nigerian tar sand. On the basis of the foregoing, we reckoned all porosity values greater than 0.5 as spurious and

Fig. 3 X–Y coordinates of the locations of the wells drilled in the study area. Adapted from Falebita et al. (2014)



deleted such from the database in this work. Thirty-five (35) of such data values were deleted; consequently, the statistical analyses in this work involved 408 data points spread across only 33 wells. Other wells either have none or insufficient data points, after deleting the spurious data values.

A suite of univariate descriptive statistical measures was selected for analysis on the porosity, depth-to-top, and thickness databases: mean, median, standard deviation, skewness, kurtosis, minimum value, maximum value, first quartile, and third quartile. Mean and median are measures of central tendency, i.e., measures of typical, expected, and representative data values. Since the mean is only a sample estimate of the population (true) mean, we include the 95% confidence interval of the mean in the suite of descriptive statistics reported. This interval is the range (lower and upper limit) of values that would (with a probability of 0.95) contain the true mean. Standard deviation is a measure of spread/dispersion away from the typical values. It measures the degree to which the typical values represent the measured data values (in the database) and the unmeasured data values (in the underlying population). Skewness is a measure of symmetry of the data distribution about the center (mean). The analyses in this work also included graphical representations of the probability distribution of the porosity, depth-to-top, and thickness data. These representations reveal the frequency of occurrence of data within certain ranges (classes). Such univariate probability distributions are used as prior information to which reservoir descriptions are constrained in conditional simulation stage of geostatistical modeling. In this work, three distribution curve types were used. Relative frequency (probability) histogram was the basic graphical representation employed in this work. Porosity is a continuous random variable, and hence, it is important that, in addition to histogram (which is discrete in nature), the underlying probability density function (PDF) be studied. This work used kernel density plots to study the underlying PDF. Kernel density estimation is a non-parametric method that makes no assumption about the underlying distribution (i.e., assumes none of the popular distribution functions). Many statistical applications require normality of data distribution; hence, it is expedient to study the extent to which the data distributions approximate normal (Gaussian) distribution. To do this, we have included the normal density curves as part of analysis tools in this work. To enhance the analysis, both the kernel density and normal density curves are superimposed on the histograms, for each data set plotted in this work. Further still, normal quantile–quantile (Normal Q–Q) plot analysis is also included in this work. A normal Q–Q plot is a graphical measure of the plausibility of approximating the data as being normally distributed. It is a scatter plot created by plotting the quantiles of the data against the quantiles of a theoretical normal distribution. If the data are approximately normally distributed, the scatter points will

result in approximately a straight line. To detect the presence of outliers (in spite of the 0.5 cut-off set for porosity), box plots were also generated for each porosity data set.

All the computations and statistical analyses herein outlined were done using version 3.3.3 of R (R Core Team 2017), a language and environment for statistical computing and graphics. Skewness and Kurtosis which are not included in base R are computed using an R package known as “moments” (Komsta and Novomestky 2015).

Data analysis results and discussion

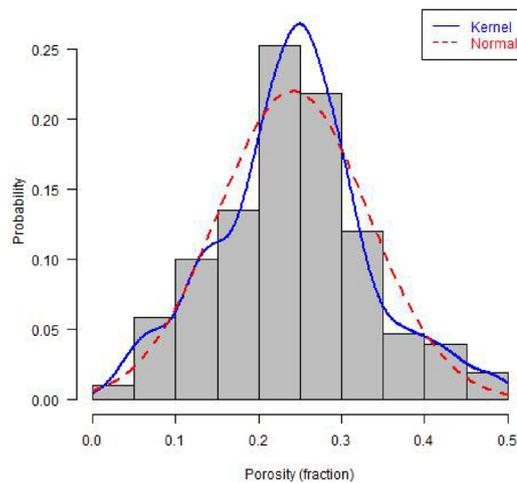
Results of analyses conducted on porosity, depth-to-top, and thickness data are herein presented and discussed. The analyses were done per the under-listed hierarchies to facilitate proper understanding of the database.

- Field (study area)-wide porosity analysis
 - Individual porosity values
 - Well-mean porosity values
- Horizon-wide porosity analysis
 - Individual porosity values
 - Mean porosity values
- Per-well porosity analysis
 - Well-payzone porosity values
 - Well-horizon porosity values
- Depth-to-top and thickness analysis

Field-wide porosity analysis

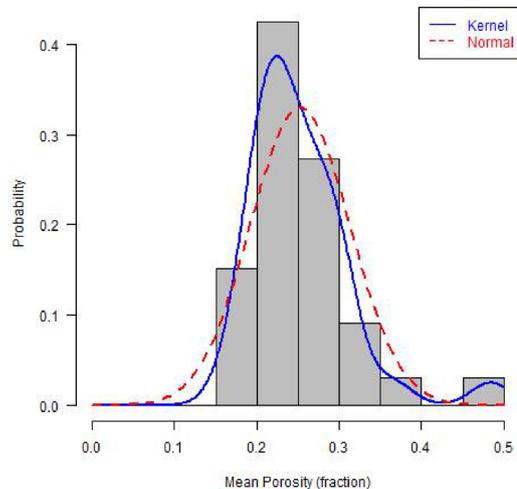
First, the data analysis (descriptive statistics and distribution curves) for the whole set of individual porosity values at all sampled points in the entire study area was conducted. Thereafter, mean porosity for each well (well-mean) was obtained and the analysis was conducted for well-mean porosity values. Figures 4, 5 present the analysis results for the individual and well-mean analyses, respectively. The mean of the data sets for individual porosity and well-mean porosity cases are 0.2415 (with a 95% confidence interval of 0.2327–0.2503) and 0.2525 (with a 95% confidence interval of 0.2310–0.2739), respectively. The implication of the 95% confidence interval is that there is 95% probability (confidence) that the error in using these sample means as estimates of the true means would not exceed 0.0088 and 0.0214, in both cases respectively. The modal class in both cases is 20–25%. Porosity values obtained for 70 representative cores in the same Agbabu area by Enu (1985),

Fig. 4 Descriptive statistics and distributions of all individual porosity values—field-wide



Data Attribute	Value
Mean	0.2415
95% Confidence Interval of Mean	0.2327 – 0.2503
Median	0.2470
Standard Deviation	0.0907
Kurtosis	3.2354
Skewness	0.1609
Minimum	0.0132
First Quartile	0.1882
Third Quartile	0.2938
Maximum	0.4964
N	408

Fig. 5 Descriptive statistics and distributions of well-mean porosity values—field-wide



Data Attribute	Value
Mean	0.2525
95% Confidence Interval of Mean	0.2310 – 0.2739
Median	0.2367
Standard Deviation	0.0605
Kurtosis	7.6695
Skewness	1.7933
Minimum	0.1642
First Quartile	0.2088
Third Quartile	0.2817
Maximum	0.4836
N	33

using textural analysis ranges from 0.24 to 0.35. A perusal of some of the descriptive statistics vis-a-vis the descriptive statistics of grain size distribution of the core samples reported in Adegoke et al. (1980) shows good correlation between porosity and grain size distributions. The positive (right) porosity distribution skewness here mirrors the positively skewed attribute of the grain size distribution in Adegoke et al. (1980), although the latter work reported some strongly coarse (negative) skewness. Both the porosity and the grain size distributions also exhibited positive kurtosis (mesokurtic-to-leptokurtic).

In general, porosity distributions are known to be nearly normally distributed, i.e., they are nearly symmetric about the mean. A fundamental probability theorem known as central limit theorem (CLT) is at the root of this behaviour of porosity. The central limit theorem establishes the fact that a variable obtained by summing up a large number of independent variables will exhibit an approximately normal distribution. Core porosity is simply a sum of the volume of

several pores in the core (Jensen et al. 2000). The histograms of both individual porosities and well-mean porosities in Figs. 4 and 5 are observed to exhibit reasonable degrees of symmetry. Furthermore, the kernel density distributions in both cases match closely with the respective normal distributions. However, the degree of normality is higher in the case of individual porosity distribution as evident in the skewness (a measure of departure from normal distribution) values of both cases. This higher degree is attributable to the large number of data points (sample size) available in the case of individual porosities (408); the CLT normal distribution approximation becomes more accurate as sample size becomes larger. In addition, the averaging in the case of well-mean porosities has reduced the variability in the data (reduced standard deviation), thereby resulting into fewer large (greater than mean) values, a situation that manifests as positive skewness (i.e., longer tail at the right of the curve).

The normal Q–Q plots of both individual porosities and well-mean porosities are presented in Figs. 6 and 7. In both

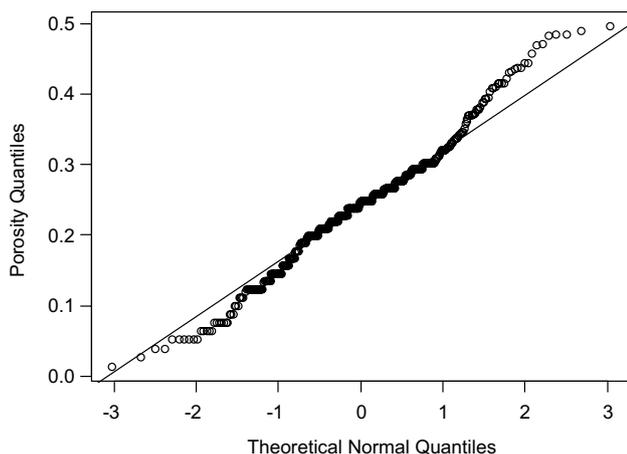


Fig. 6 Normal Quantile–quantile plot of all individual porosity values—field-wide

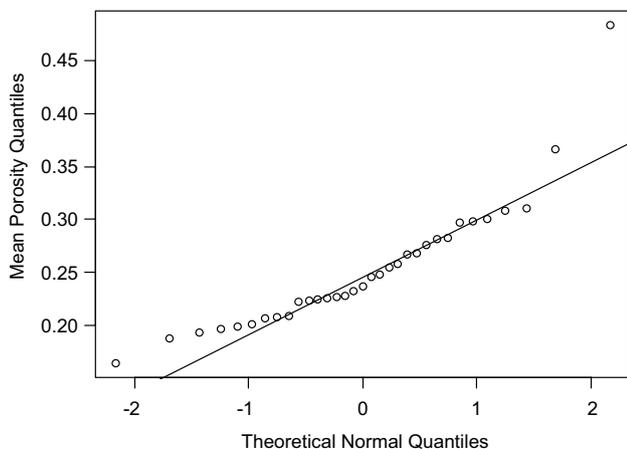


Fig. 7 Normal Quantile–Quantile plot of well-mean porosity values—field-wide

figures, most of the scatter points cluster around the fitted straight lines which, by default, pass through the first and third quartiles. Figure 6 shows that, for the individual porosities, the large values in the data set are larger than would be expected for a perfectly normal distribution. In a similar manner, the small values are also smaller than expected for a perfect normal distribution. These two deviations at the extremes of the plot are approximately equal, giving rise to low skewness in the histogram. In addition, the deviations imply a heavy tail in the distribution of the porosity data. Conversely, Fig. 7 shows that, for the well-mean porosities case, the large values are larger than expected, while the small values are not as small as would be expected for a perfectly normal distribution. This gives rise to a longer (heavy) tail at the right and a shorter (truncated) tail at the left. Consequently, the well-mean porosities data feature a more pronounced positive skewness than the individual

porosities case do. This analysis stopped short of conducting goodness of fit tests (such as Kolmogorov–Smirnov, Anderson–Darling, Shapiro–Wilk, and Lilliefors tests) of normal approximation of the data sets. Such tests are known to give false rejection of the null hypothesis (that the data are normally distributed) if data values are not independent of each other (McDonald 2014). Petrophysical data are known to be spatially correlated; hence, values are not independent of each other.

We observe that both distributions (Figs. 4, 5) are unimodal, in spite of the presence of two distinct geological horizons in the data set. If two (or more) geological horizons from statistically different geological populations (corresponding to different environments of deposition) are present in a data set, the resulting distribution will be bimodal (or multi-modal) with each mode corresponding to each population. On the basis of the foregoing, we hypothesize that both Horizons X and Y might belong to the same population, a hypothesis which we test in the next section of this paper. In spite of the removal of computed porosity values greater than 0.5 as discussed earlier, this analysis still detected a few outliers in the individual porosity data set and only one outlier in the mean porosity data set. In this analysis, outliers are data points farther than 1.5 times the interquartile range from the box. In the future spatial analysis and property estimation stages of this work; the removal of such outliers is expected to improve variogram interpretability and reduce uncertainty. The box plots for both data sets are presented in Fig. 8. The near-absence of outlier points in the well-mean porosity case is attributed to the reduced variability due to averaging. In fact, the only detected outlier is not a mean porosity; rather, it is the single individual value available for Well BH11.

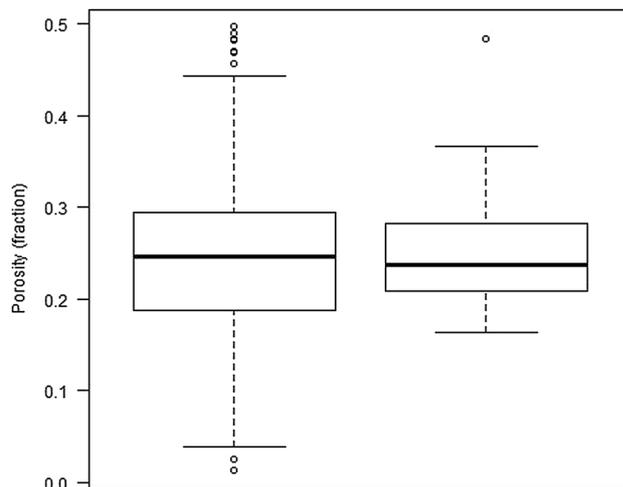


Fig. 8 Box plots of individual porosity values (left) and well-mean porosity values (right)—field-wide

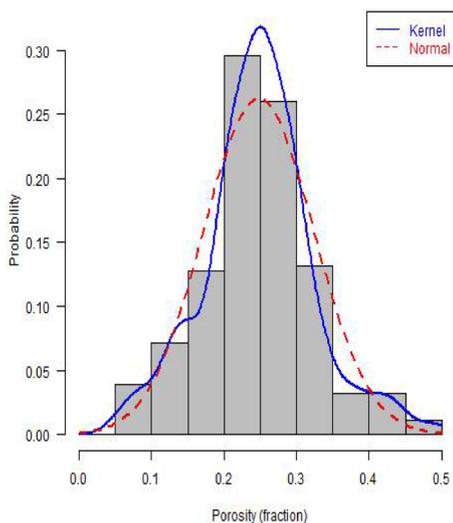
Horizon-wide porosity analysis

As done for the entire study area, data analysis was conducted for the whole set of individual porosity values at all sampled points in each of Horizons X and Y. In addition, for each of the two horizons, the means of porosity values within the horizon in each well (horizon well-mean) were obtained, and analysis conducted thereon. Figures 9, 10 present the analyses results for the individual porosity data sets, for the two horizons. The mean of the individual porosity data sets for Horizon X and Horizon Y is 0.2478 (95% confidence interval of 0.2388–0.2567) and 0.2277 (95% confidence interval of 0.2074–0.2480), respectively. The standard deviation for Horizon X data set is 0.0761, while the standard deviation for Horizon Y data set is 0.1158. The higher value of standard deviation in Horizon Y implies higher variability in Horizon Y than in Horizon X. The modal class for Horizon X is 20–25%, while that of

Horizon Y is 10–15% with class 20–25% occurring nearly as frequent as the modal class. For most of the statistics measures reported, Horizon X values seem to have more influence on the field-wide values than Horizon Y values have. This is due to the sampling size being larger in Horizon X (281) than in Horizon Y (127).

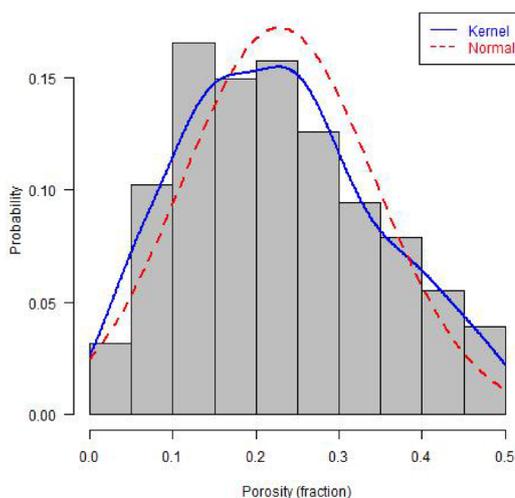
The histograms of the two individual porosity data sets (Figs. 9, 10) exhibit reasonable degrees of symmetry about the mean. However, Horizon X exhibits more symmetry (less skewness) than Horizon Y; again, this is attributable to the larger sample size in the former. In addition, kernel density distributions in both horizons match closely with the respective normal distributions. It is observed that the tails of Horizon Y histogram slope gradually away from the center, while the tails slope rapidly in the Horizon X. This observation confirms the higher variability in Horizon Y. Normal Q–Q plots of the individual porosity data sets, for the two horizons, are presented in Figs. 11 and 12.

Fig. 9 Descriptive statistics and distributions of all individual porosity values—Horizon X



Data Attribute	Value
Mean	0.2478
95% Confidence Interval of Mean	0.2388 – 0.2567
Median	0.2478
Standard Deviation	0.0761
Kurtosis	3.7943
Skewness	0.1967
Minimum	0.0516
First Quartile	0.2086
Third Quartile	0.2933
Maximum	0.4964
N	281

Fig. 10 Descriptive statistics and distributions of all individual porosity values—Horizon Y



Data Attribute	Value
Mean	0.2277
95% Confidence Interval of Mean	0.2074 – 0.2480
Median	0.2186
Standard Deviation	0.1158
Kurtosis	2.3984
Skewness	0.3310
Minimum	0.0132
First Quartile	0.1399
Third Quartile	0.3040
Maximum	0.4889
N	127

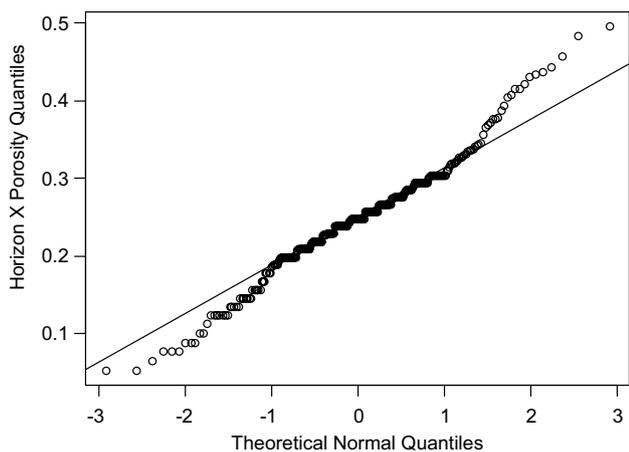


Fig. 11 Normal Quantile–quantile plot of individual porosity values—Horizon X

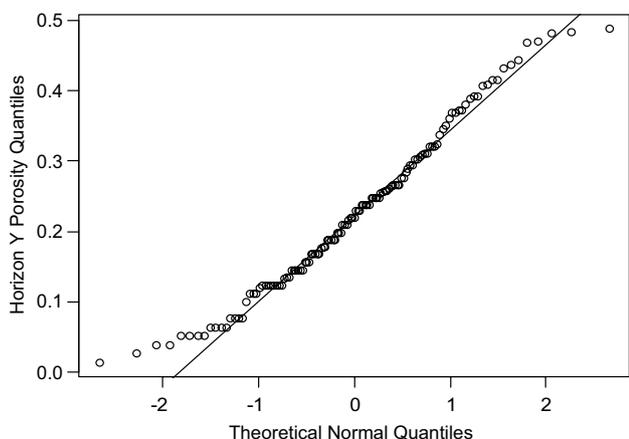


Fig. 12 Normal Quantile–quantile plot of individual porosity values—Horizon Y

In both figures, most of the scatter points cluster around the fitted straight lines. The low skewness in Horizon X data set is evident in the roughly equal but opposite deviations at the left and right extremes of its Normal Q–Q plot (Fig. 11). The short (truncated) left tail of Horizon Y histogram is evident in the deviations in the left extremes of its Normal Q–Q plot (Fig. 12) where the small values are not as small as expected of a perfect normal distribution.

Here, we test our earlier hypothesis that both Horizons X and Y statistically belong to the same geological population (i.e., the sediments in both horizons belong to the same environment of deposition). Horizon X and Horizon Y porosity data sets were the two samples tested. The two-sided two-sample *t* test, conducted using R, was set up as follows:

- Null hypothesis, H_0 : the difference between the true means of Horizon X and Horizon Y is zero; i.e., $\mu_X - \mu_Y = 0$
- Alternative hypothesis, H_a : the difference between the true means of Horizon X and Horizon Y is not zero; i.e., $\mu_X - \mu_Y \neq 0$
- Significance level: 0.05
- Variance assumption: unknown and unequal.

The output of the test is as follows:

- *t* Statistic: 1.7858
- Critical region: $t < -1.9735$; $t > 1.9735$
- *p* value: 0.0759
- 95% confidence interval: -0.0021–0.0422.

On the account of the *t* statistic not being within the critical region, and the *p* value being greater than the significance level, we fail to reject the null hypothesis. Therefore, we conclude that the statistical difference between the true means of Horizon X and Horizon Y porosity values is not significantly different from zero. In addition, the fact that the 95% confidence interval contain zero implies a 0.95 probability that the interval contain the true difference between the population means of the porosity of both horizons. With the difference of the true mean being in the region of zero, we conclude that both Horizon X and Horizon Y are from the same geological population (i.e., the same environment of deposition). Furthermore, a Q–Q plot of Horizon X quantiles plotted against Horizon Y quantiles is here shown in Fig. 13. The fact that the points form a roughly straight-line (not the line shown) implies both Horizons X and Y have somewhat similar probability distributions. However, the 45° line shown on the plot indicates that the large values of Horizon Y are larger than the corresponding large values

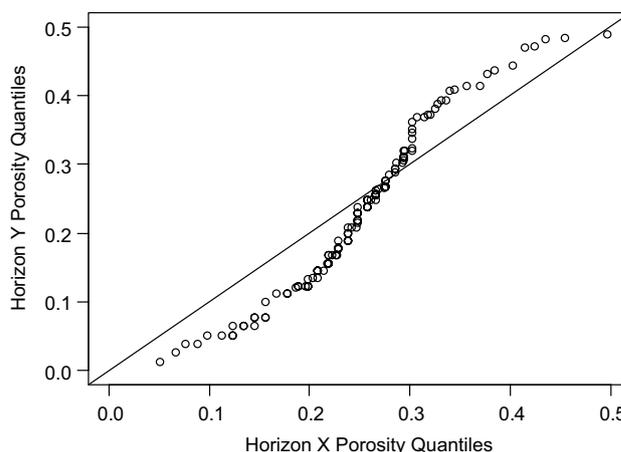


Fig. 13 Q–Q plot of Horizon X quantiles plotted against Horizon Y quantiles

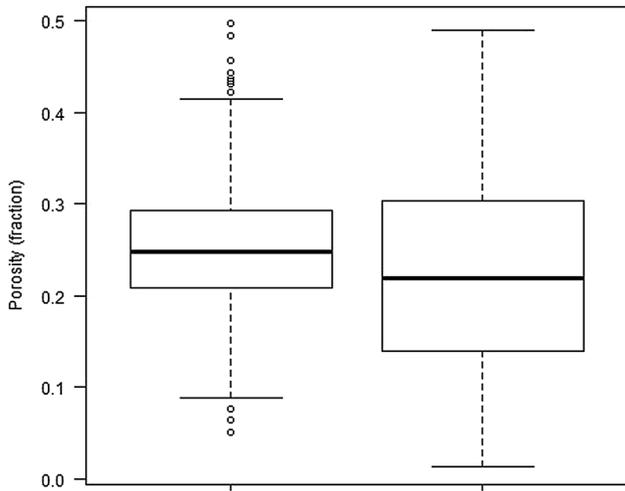


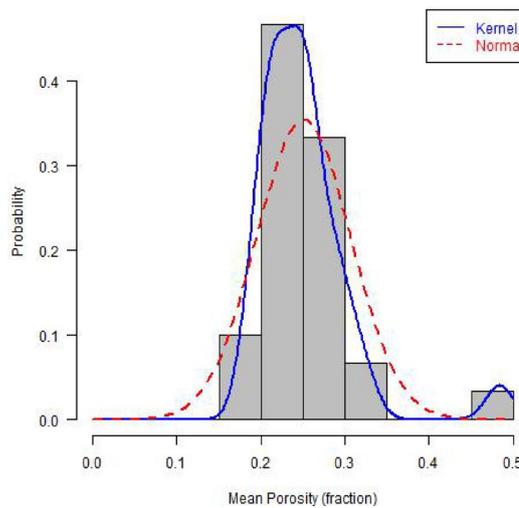
Fig. 14 Box plots of Horizon X individual porosity values (left) and Horizon Y individual porosity values (right)

of Horizon X. Similarly, the small values of Horizon Y are smaller than the corresponding small values of Horizon X. These deviations are evident in the heavier tails featured in the histogram of Horizon Y data set.

The box plots for individual porosity data sets, for Horizons X and Y, are presented in Fig. 14. A few outliers are detected in the Horizon X data set, while no outlier was detected in Horizon Y data sets. The detection of outliers in Horizon X is attributable to the smaller variability in Horizon X data set as compared to the variability in Horizon Y data set. The bigger box in Horizon Y is another indication of the higher variability in Horizon Y.

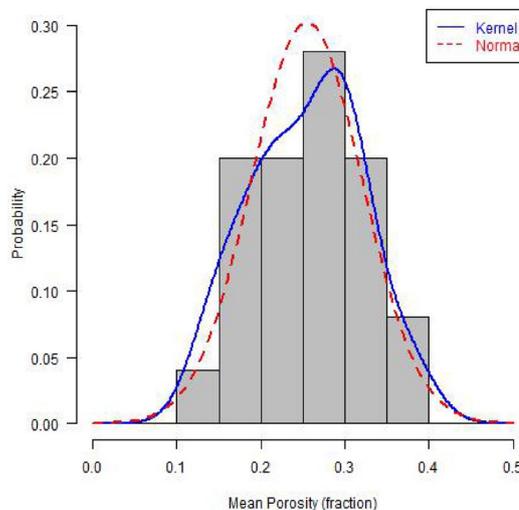
Figures 15, 16 present the analyses results for the well-mean porosity data sets, for the two horizons. A noticeable difference between the histograms of individual and well-mean porosity for Horizons X (Figs. 9, 15) on one hand and the histograms of individual and well-mean porosity for Horizons Y (Figs. 10, 16) on the other hand is the degree of variability reduction due to averaging. Normally, averaging is expected to reduce variability of a random variable data.

Fig. 15 Descriptive statistics and distributions of well-mean porosity values—Horizon X



Data Attribute	Value
Mean	0.2511
95% Confidence Interval of Mean	0.2301 - 0.2721
Median	0.2438
Standard Deviation	0.0562
Kurtosis	10.8001
Skewness	2.4151
Minimum	0.1880
First Quartile	0.2169
Third Quartile	0.2654
Maximum	0.4836
N	30

Fig. 16 Descriptive Statistics and Distributions of well-mean porosity values—Horizon Y



Data Attribute	Value
Mean	0.2547
95% Confidence Interval of Mean	0.2275 - 0.2820
Median	0.2666
Standard Deviation	0.0660
Kurtosis	2.2101
Skewness	-0.0592
Minimum	0.1358
First Quartile	0.2154
Third Quartile	0.3062
Maximum	0.3811
N	25

This reduction in variability manifests as reduced range of values in histograms of well-mean cases. The histograms of well-mean porosity for both horizons do, indeed, have reduced range when compared to their respective histogram of individual porosity counterparts. However, for Horizon X, this variability reduction is more pronounced. It is necessary to state here that the right-most bin (class 0.45–0.5) in the histogram for Horizon X well-mean porosity (Fig. 15) is not actually a mean value; it is the single individual value available for Well BH11. With the right-most bin excluded, the histogram of Horizon X well-mean porosity spans through four (4) bins compared to nine (9) bins in the histogram of Horizon X individual porosity. For Horizon Y, the histogram of well-mean porosity spans through six (6) bins compared to ten (10) bins in the histogram of Horizon Y individual porosity. Upon reviewing the geology of the two horizons, we surmise that the difference in variability reduction, after averaging, is not due (at least, directly) to any geological differences between the two horizons; rather, it is attributable to the skewness of each data set. From Figs. 9 and 10, it is observed that Horizon Y is more positively skewed than Horizon X is. The implication of this is that the difference in sum of positive deviations from mean and sum of negative deviations from mean is more pronounced in Horizon Y than that in Horizon X. Since averaging is essentially a summation process, these opposing deviations tend to cancel each other out. The more pronounced the difference in the sum of opposing deviations, the less the cancelling effect, and, consequently, the less the variability reduction upon averaging.

Per-well porosity analysis

The foregoing analyses are also performed for each of the 33 wells being considered in this work. First, the porosity data set for each well was analyzed as a whole; thereafter, the data set for each well was partitioned into Horizons X and Y (where both are penetrated), and each well-horizon data analyzed. The mean porosity for wells ranges from 0.1642 to 0.3657, excluding Well BH11 with only one data point (0.4836); while the standard deviation ranges from 0.0181 to 0.1667. The mean porosity for well-Horizon X partitions ranges from 0.1880 to 0.3244, excluding Well BH11; the corresponding standard deviation ranges from 0.0181 to 0.1261. In the case of well-Horizon Y partitions, the mean ranges from 0.1358 to 0.3811, while standard deviation ranges from 0.0058 to 0.1667. The foregoing values confirm the higher variability in Horizon Y, as earlier noted. In spite of small sample size (less than 30) in each of the wells, the porosity distributions of some of the wells and well-horizons approximate the normal distribution. Twenty-eight (28) wells exhibit skewness between -1 and $+1$, while none of the well have skewness outside the range -2 to $+2$. Of the 30 well-Horizon X partitions, 25 exhibit skewness

between -1 and $+1$; 22 of the 25 well-Horizon Y partitions have skewness in that range. Distribution and Normal Q–Q plots of some wells/well-horizons whose porosity distributions exhibit near-symmetry around their respective means are presented in Figs. 17 and 18.

Despite the earlier conclusion, based on the field-wide test of hypothesis, that both Horizon X and Horizon Y are from the same geological population; the occurrence of bi-modal histograms in some of the wells sparks fresh concern on the validity (or otherwise) of this conclusion, at individual well level. Typically, a bi-modal histogram implies the presence of two statistically distinct sub-populations in a data set. This occurrence, therefore, necessitated the implementation of test of hypothesis at such wells, to reaffirm or refute the earlier conclusion. Wells BH 10, BH 17, BH 23, and BH 50 whose histograms are bi-modal (see Fig. 19 below) were chosen for the well-level two-sided two-sample t tests. For each well, the corresponding Horizon X and Horizon Y porosity data sets were the two samples tested to determine if both horizons, at that well, belong to the same population. The tests were set up as done for the field-wide test described above. The outputs of the tests are presented in Table 1. In each of the wells tested, the t statistic is not within the corresponding critical region; also, the p value is greater than the significance level (0.05). Hence, we fail to reject the null hypothesis, and conclude that the statistical difference between the true means of Horizon X and Horizon Y porosity values, at each well, is not significantly different from zero. Furthermore, the fact that the 95% confidence intervals contain zero reinforces this conclusion. With the difference of the true mean being in the region of zero, at all the tested wells, we reaffirm our earlier conclusion that both Horizon X and Horizon Y are from the same geological population (i.e., the same environment of deposition).

From the boxplots generated for each of the wells, one or two outliers were detected in 13 of the 33 wells; Horizon X partitions accounted for most of the outliers.

Depth-to-top and thickness analysis

The foregoing analyses are also performed on the depth-to-top and thickness data, for both Horizon X and Horizon Y. Figures 20, 21 present the analysis results for depth-to-top of Horizons X and Y, respectively. Depth to the top of Horizon X ranges from 8.2ft to 210ft, with a mean of 84.6ft and standard deviation of 67.4ft. On the other hand, depth to the top of Horizon Y ranges from 62.3ft to 275.6ft, with a mean of 144.3ft and standard deviation of 63.8ft. The high values observed for the standard deviations are in agreement with the significant variations in depth of marker beds observed in the geoelectrical–geological data correlation conducted by Adegoke et al. (1980). These significant variations in depth-to-top had been stated to be an indication of local faulting

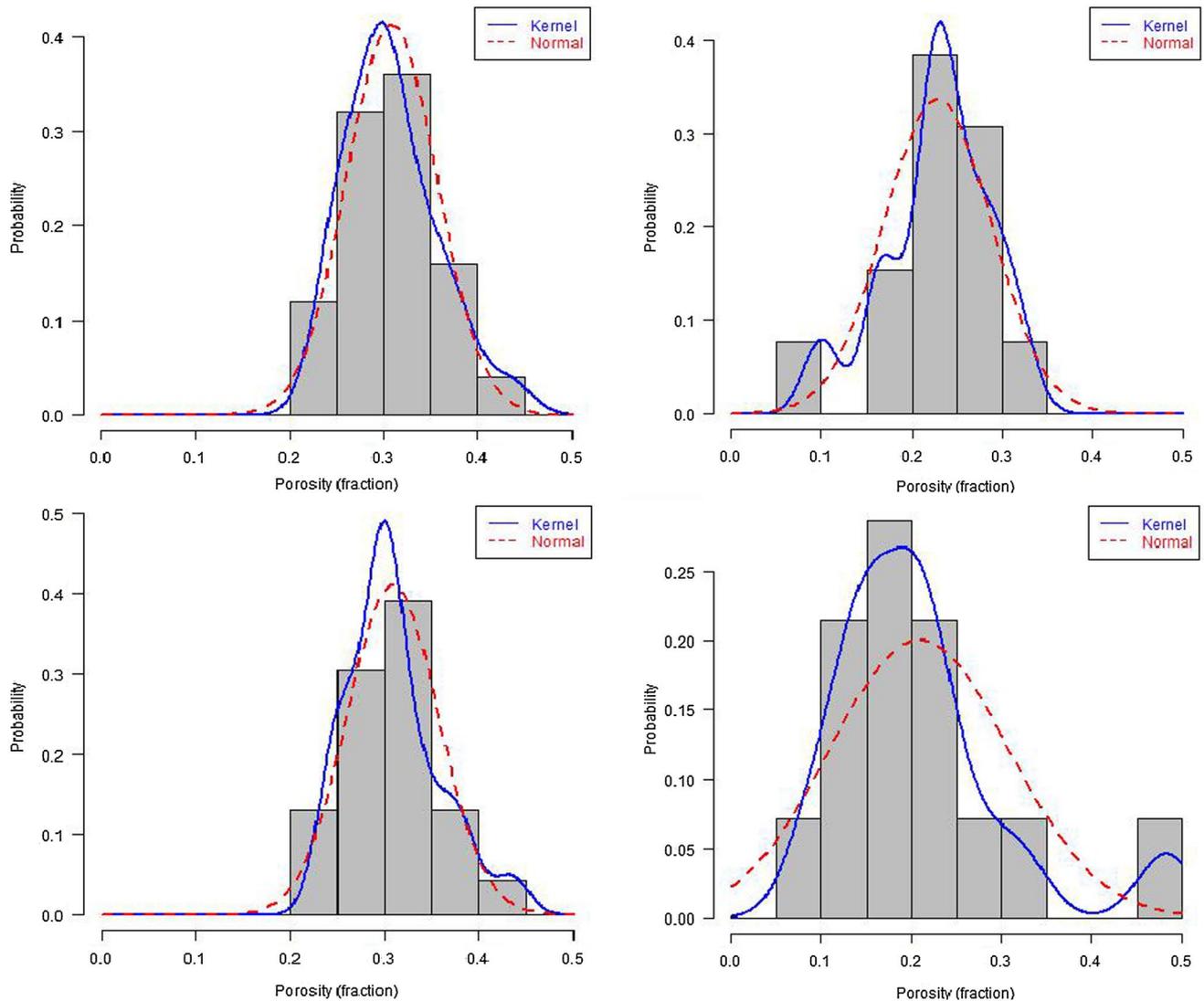


Fig. 17 Porosity values distributions for Well BH 19 (top left) and Well BH 29 (top right); Horizon X of Well BH 19 (bottom left) and Well BH 28 (bottom right)

in the deposits, considering the tectonic features of the area (Adegoke et al. (1980)).

Figures 22, 23 present the analysis results for the thickness data of Horizons X and Y, respectively. Thickness of Horizon X ranges from 16.40ft to 72.18ft, with a mean of 49.06ft and standard deviation of 14.69ft. Thickness of Horizon Y ranges from 6.56ft and 98.43ft, with a mean of 44.17ft and standard deviation of 26.00ft. Comparatively, the ranges, standard deviations, and 95% confidence intervals show that there exists more variability in the thickness of Horizon Y than in the thickness of Horizon X. This observation correlates well with the existing stratigraphic information. Adegoke et al. (1980) had reported that the thickness of Horizon Y varies widely across different regions of the study area; it (Horizon Y) even becomes irregular and thins out in the

southeastern region. Meanwhile, the thickness of Horizon X had been reported to be fairly constant Adegoke et al. (1980). The presence of multiple peaks (modes) in the histogram of thickness of Horizon Y suggests that there exists distinct ‘thickness’ sub-populations each corresponding to different regions of the study area. Such regional variation (known as non-stationarity) observed here is expected to have profound influence on the spatial data analysis which is a future stage of this work. The observed depth ranges indicates that the heavy oil and natural bitumen reserves in the study area would be recoverable by both surface mining and in-situ techniques. Deposits occurring at depths no more than 250ft can be recovered via mining while deposits that occur at depths greater than 250ft are recoverable in-situ (Attanasi and Meyer 2010).

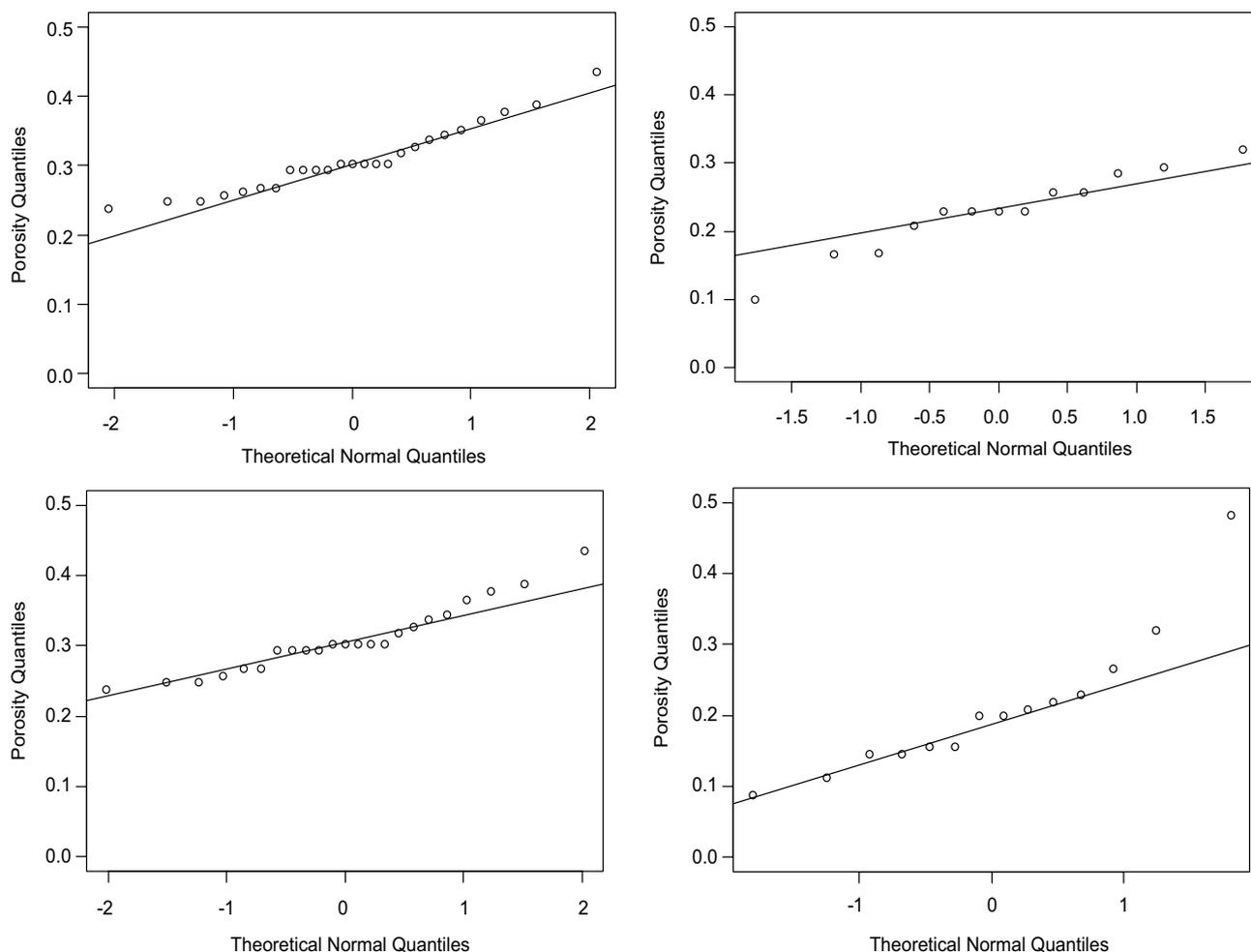


Fig. 18 Normal Q–Q plots for Well BH 19 (top left) and Well BH 29 (top right); Horizon X of Well BH 19 (bottom left) and Well BH 28 (bottom right)

Summary and conclusions

Using basic principles, this work has computed porosity values for 443 core samples retrieved from 33 wells drilled into the two bituminous sand horizons in Agbabu heavy oil and natural bitumen deposit. Spurious values have been deleted from the computed database; sound justifications for the choice of cut-off value have been provided. The computed database has been partitioned into the two horizons (Horizons X and Y), using the depth-to-top and thickness data. At its core, this work has conducted detailed analyses and offers robust discussions on various univariate descriptive statistics and probability distributions of the porosity, depth-to-top, and thickness databases. The statistics and distribution curves so obtained have been correlated with the existing geologic, stratigraphic, and textural data. In addition, a hypothesis on the likelihood of the two horizons belonging to same geological population

(i.e., the same environment of deposition) has been formulated and tested.

From the results presented in this work, the following conclusions are drawn.

- The field-wide average porosity in the study area is 0.2415, with a 95% confidence interval of 0.2327–0.2503 (i.e., a standard error of 0.0088).
- The field-wide average well-mean porosity in the study area is 0.2525, with a 95% confidence interval of 0.2310–0.2739 (i.e., a standard error of 0.0214).
- The histograms of both individual porosities and well-mean porosities exhibit good degrees of symmetry. Furthermore, the kernel density distributions in both cases match closely with the respective normal distributions; just as the Normal Q–Q plots confirmed the plausibility of normal distribution approximations.

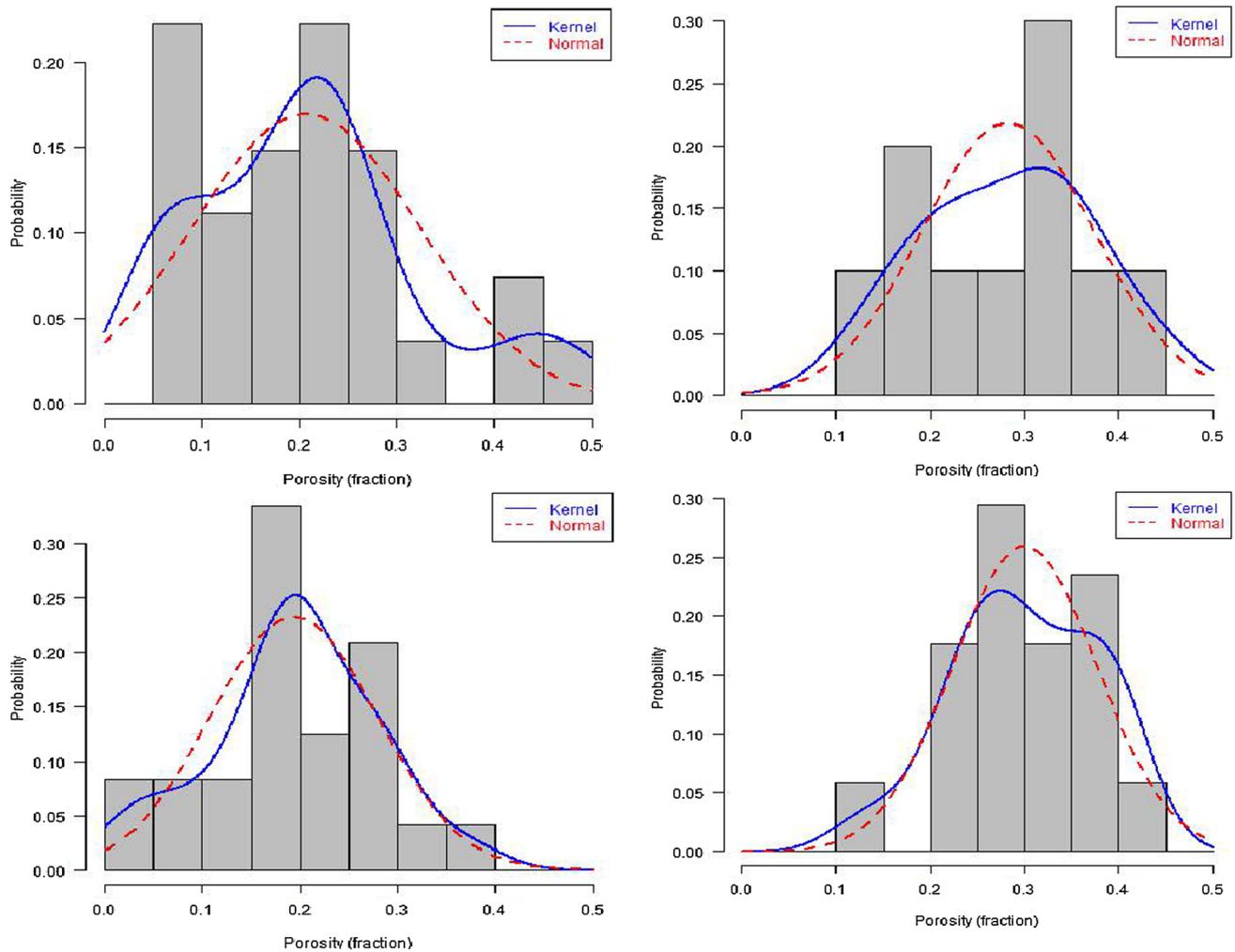


Fig. 19 Porosity values distributions for Well BH 10 (top left), Well BH 17 (top right), Well BH 23 (bottom left), and Well BH 50 (bottom right)

Table 1 Outputs of well-level two-sided two-sample t tests

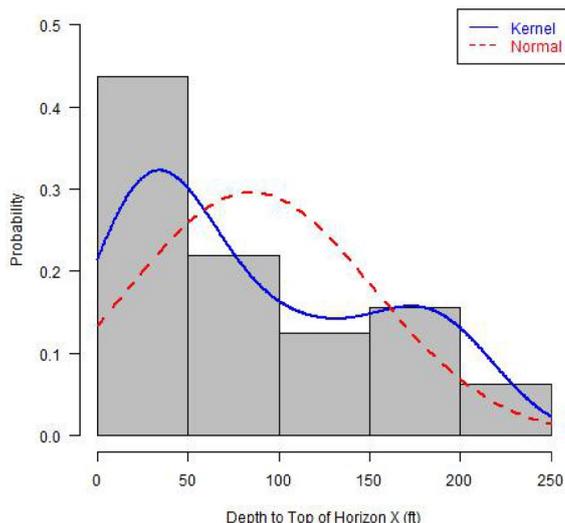
Well name	t Statistic	Critical region	p value	95% Confidence interval
BH 10	1.9916	$t < -2.0806$; $t > 2.0806$	0.0597	-0.0036 to 0.1650
BH 17	-0.8434	$t < -2.4642$; $t > 2.4642$	0.4322	-0.2044 to 0.1002
BH 23	1.8835	$t < -2.1326$; $t > 2.1326$	0.0793	-0.0089 to 0.1429
BH 50	-0.6746	$t < -2.2984$; $t > 2.2984$	0.5186	-0.1189 to 0.0650

- Horizon X average porosity in the study area is 0.2478, with a 95% confidence interval of 0.2388–0.2537 (i.e., a standard error of 0.0089) while Horizon Y average porosity in the study area is 0.2277, with a 95% confidence interval of 0.2074–0.2480 (i.e., a standard error of 0.0203).
- The degree of data variability is higher in Horizon Y than in Horizon X.
- The statistical difference between the true means of Horizon X and Horizon Y porosity values is not significantly different from zero; hence, both horizons are

from the same geological population (i.e., the same environment of deposition). In addition, the Q–Q plot of Horizon X quantiles plotted against Horizon Y quantiles indicates that both horizons have somewhat similar probability distributions.

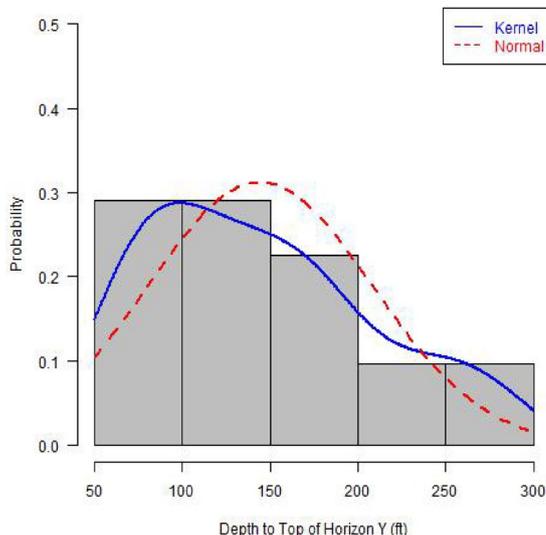
- The mean porosity for each well range from 0.1642 to 0.3657; the mean porosity for the Horizon X partitions in each well ranges from 0.1880 to 0.3244; and the mean porosity for the Horizon Y partitions in each well ranges from 0.1358 to 0.3811.

Fig. 20 Descriptive statistics and distributions of depths to top of Horizon X



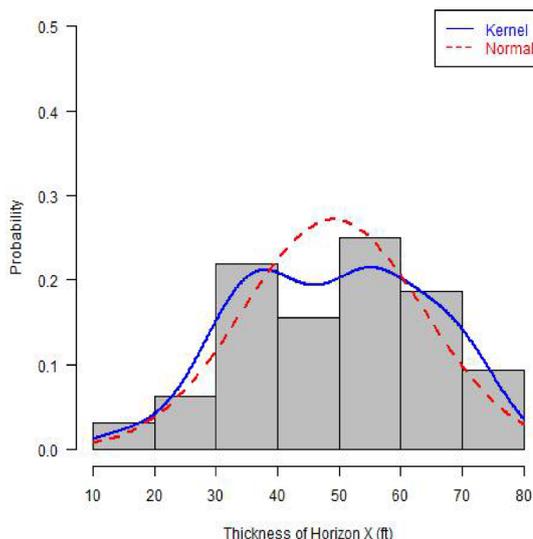
Data Attribute	Value
Mean	84.64ft
95% Confidence Interval of Mean	60.35 – 108.92ft
Median	56.59ft
Standard Deviation	67.36ft
Kurtosis	1.8352
Skewness	0.5558
Minimum	8.2021ft
First Quartile	26.65683ft
Third Quartile	143.1266ft
Maximum	209.9738ft
N	32

Fig. 21 Descriptive statistics and distributions of depths to top of Horizon Y



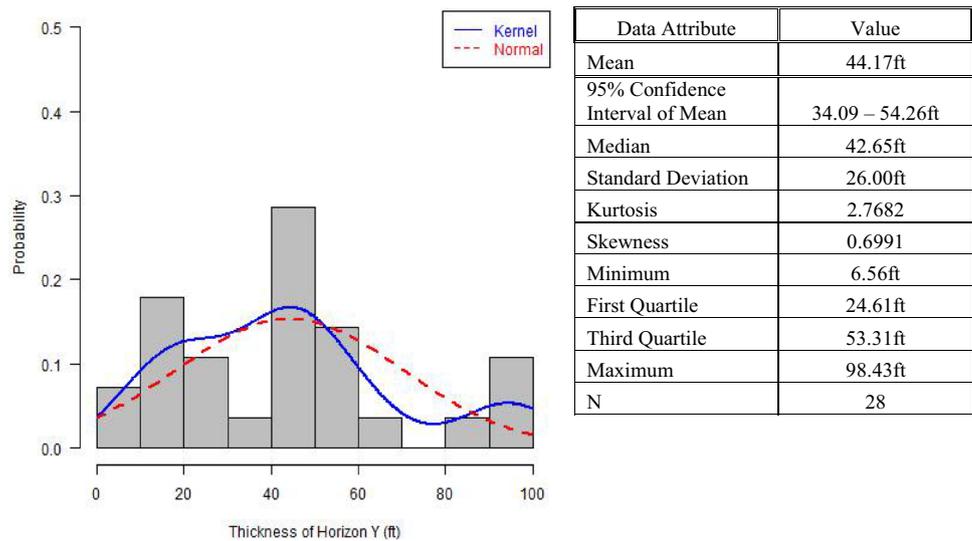
Data Attribute	Value
Mean	144.25ft
95% Confidence Interval of Mean	120.84 – 167.67ft
Median	131.23ft
Standard Deviation	63.84ft
Kurtosis	2.3173
Skewness	0.5893
Minimum	62.34ft
First Quartile	83.66ft
Third Quartile	176.35ft
Maximum	275.59ft
N	31

Fig. 22 Descriptive statistics and distributions of thickness of Horizon X



Data Attribute	Value
Mean	49.06ft
95% Confidence Interval of Mean	43.76 – 54.35ft
Median	51.67ft
Standard Deviation	14.69ft
Kurtosis	2.1178
Skewness	-0.1189
Minimum	16.40ft
First Quartile	37.32ft
Third Quartile	60.70ft
Maximum	72.18ft
N	32

Fig. 23 Descriptive statistics and distributions of thickness of Horizon Y



- The porosity distributions of some of the wells and well-horizons approximate the normal distribution.
- The statistical difference between the true means of Horizon X and Horizon Y porosity values, at each well, is not significantly different from zero. Hence, at each well, both horizons are from the same geological population (i.e., the same environment of deposition).
- Depth to the top of Horizon X ranges from 8.2 to 210 ft, with a mean of 84.6 ft and standard deviation of 67.4 ft. Depth to the top of Horizon Y ranges from 62.3 to 275.6ft, with a mean of 144.3ft and standard deviation of 63.8ft.
- Thickness of Horizon X ranges from 16.40 to 72.18 ft, with a mean of 49.06 ft and standard deviation of 14.69 ft. Thickness of Horizon Y ranges from 6.56 and 98.43 ft, with a mean of 44.17ft and standard deviation of 26.00ft.
- There exists more variability in the thickness of Horizon Y than in the thickness of Horizon X.
- The histogram of thickness of Horizon Y exhibits multiple peaks (modes), suggestive of a mixture of ‘thickness’ sub-populations.

Acknowledgements The authors acknowledge Dr. A. A. Adepelumi for facilitating access to previous report of the Geological Consultancy Unit of Obafemi Awolowo University. Dr. Christopher Thron’s workshop on R as well as insightful discussions with Dr. K. A. Lawal were very helpful.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adebiyi FM, Bello OO, Sonibare JA, Macaulay SRA (2005) Determination of SARA constituents of southwestern Nigerian Tar sands and their physical properties. *Eng J Univ Qatar* 18:29–38
- Adegoke OS, Ako BD, Enu EI, Afonja AA, Ajayi TR, Emofurieta WO, Soremekun OA (1980) Geotechnical investigations of the ondo state bituminous sands—volume I: geology and reserves estimate. Geological Consultancy Unit, Department of Geology, University of Ife, Nigeria (Unpublished)
- Adeyemi GO, Akinmosin AA, Aladesanmi AO, Badmus GO (2013) Geophysical and sedimentological characterization of a Tar sand rich area in South-western Nigeria. *J Environ Earth Sci* 3(14):71–82
- Akinmosin AA, Osinowo OO, Osei MM (2012) Geophysical and sedimentological studies for reservoir characterization of some tar sands deposits in southwest Nigeria. *Mater Geoenviron* 59(4):413–427
- Anukwu G, Odunaike K, Fasunwon O (2014) Oil sands exploration using 2-D electrical imaging technique. *J Nat Sci Res* 4(4):68–73
- Attanasi ED, Meyer RF (2010) Natural Bitumen and Extra-Heavy Oil. In: Trinnaman J, Clarke A (eds) 2010 Survey of energy resources. World Energy Council, London, pp 123–150
- Clark KA (1957) Bulk densities, porosities and liquid saturations of good grade Athabasca oil sands. Research Council of Alberta. http://ags.aer.ca/document/INF/INF_022.PDF. Accessed 20 Jan 2017
- Energy Commission of Nigeria, ECN (2003) National energy policy. The Presidency, Federal Government of Nigeria. http://www.energy.gov.ng/index.php?option=com_docman&task=doc_download&gid=42&Itemid=49. Accessed 28 Mar 2016
- Enu EI (1985) Textural characteristics of the Nigerian Tar sands. *Sed Geol* 44:65–81
- Falebita DE, Oyebanjo OM, Ajayi TR (2014) A geostatistical review of the Bitumen reserves of the upper cretaceous afowo formation, Agbabu area, Ondo State, Eastern Dahomey Basin, Nigeria. *Pet Coal* 56(5):572–581
- Fayose EA (2005) Bitumen in Ondo State: Prospects and Promises. (lecture transcript)
- Jensen JL, Lake IW, Corbett PWM, Goggin DJ (2000) Statistics for petroleum engineers and geoscientists, 2nd edn. Elsevier, Amsterdam

- Komsta L, Novomestky F (2015) Moments: moments, cumulants, skewness, kurtosis and related tests. R package version 0.14. <https://CRAN.R-project.org/package=moments>. Accessed 27 Mar 2017
- Lawal KA (2011) Reconciling and Improving the Volumetrics of Nigerian Heavy Oil and Bitumen resources. In: SPE (Society of Petroleum Engineers), In: Nigerian annual international conference and exhibition (Paper 150802). Abuja, Nigeria. 30 July–3 August, 2007
- McDonald JH (2014) Handbook of biological statistics, 3rd edn. Sparky House Publishing, Baltimore
- Meyer RF, Attanasi ED, Freeman PA (2007) Heavy oil and natural bitumen resources in geological basins of the world (pdf) United States Geological Survey (Open-File Report 2007–1084). <http://pubs.usgs.gov/of/2007/1084/OF2007-1084v1.pdf>. Accessed 21 Mar 2016
- Milos C (2015) Bitumen in Nigeria—weighing the true costs of extraction. https://ng.boell.org/sites/default/files/bitumen_in_nigeria.pdf. Accessed 1 Feb 2017
- Ministry of Solid Mineral Development, MSMD (2006) Technical overview: Nigeria's bitumen belt and development potential. (pdf) Abuja: MSMD. <http://www.mmsd.gov.ng/privatisation/docs/bitumen%20overview.pdf> Accessed 16 Jan 2011
- Nigeria and its States (n.d.). <http://file.scirp.org/Html/4-9401351.files/image002.jpg>
- Odunaiké RK, Ijeoma GC, Edigbe RO, Babatope AH, 2009. Oil sands exploration in Ijebu-Imushin using magnetic and electrical resistivity methods. In: SAGA, 11th SAGA biennial technical meeting and exhibition. Swaziland, pp 16–18 Sept
- Ola SA (1991) Geotechnical properties and behaviour of Nigerian tar sand. Eng Geol 30(1991):325–336
- Olabanji SO, Haque AM, Fazinic S, Cherubini R, Moschini G (1994) PIGE-PIXE Analysis of Nigerian tar sands. J Radioanal Nuclear Chem 177(2):243–252
- Organization of Petroleum Exporting Countries, OPEC (2011) Annual Statistical Bulletin (pdf). OPEC. http://www.opec.org/opec_web/static_files_project/media/downloads/publications/ASB_2010_2011.pdf. Accessed 28 Apr 2017
- Organization of Petroleum Exporting Countries, OPEC (2015) Annual Statistical Bulletin (pdf). OPEC. http://www.opec.org/opec_web/static_files_project/media/downloads/publications/ASB2015.pdf Accessed 28 Apr 2017
- Organization of Petroleum Exporting Countries, OPEC (2016) Annual Statistical Bulletin (pdf). OPEC. http://www.opec.org/opec_web/static_files_project/media/downloads/publications/ASB2016.pdf Accessed 28 Apr 2017
- R Core Team (2017) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org/>. Accessed 15 Mar 2017
- Stark PH, Chew K, Fryklund B (2007) The role of unconventional hydrocarbon resources in shaping the energy future. In: SPE (Society of Petroleum Engineers), international petroleum technology conference (paper 11806). Dubai, United Arab Emirate, 4–6 December, 2007
- Tedeschi M (1991) Reserves and Production of heavy crude oil and natural bitumen. In: WPC (World Petroleum Congress), 13th world petroleum congress. (Paper 24201). Buenos Aires, Argentina. 20–25 October, 1991
- Xuetao H, Shuyong H, Fayang J, Su H (2017) Physics of petroleum reservoirs. Springer, Berlin

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.