Original Research Article

Chemometric Evaluation of Terpanes for Composition Allocation of Two Niger Delta Crude Oil Mixtures

O. J. Abayeh[1], L. C. Osuji[1], A. J. Bowofade[1], M. O. Onyema[1*]

[1]Petroleum and Environmental Geochemistry Research group, Department of Pure and Industrial Chemistry, University of Port Harcourt, Port Harcourt, 500001 Nigeria

*Corresponding author: M. O. Onyema
Petroleum and Environmental Geochemistry Research group, Department of Pure and Industrial Chemistry, University of Port Harcourt, Port Harcourt, 500001 Nigeria

Abstract

Chemometric methods were employed to evaluate terpanes that correspond to the variations in the compositions of two Niger Delta crude oils (samples A-10 and F-00) in their mixes (samples B-82, C-64, D-46, and E- 28). Total abundances, plots, and hierarchical cluster analysis of C_{19} to C_{35} terpanes in the six oil samples suggest that the abundances and distributions of terpanes significantly correlated and did not mix correspondingly in the oil samples. Principal component analysis (PCA) showed that the first principal component (PC1) accounted for 95.77% of the total variance of individual terpanes among the oil samples with H30, OL, and NH30 explaining the most variance and NH25a, NH25b, and TR20d explaining the least variance, respectively. Plots of nine (9) ratios, derived from the most and least variant terpanes, gave R^2 values (0.4919 - 0.9749) that indicate the ratios moderately to very strongly correspond to the compositions of the two Niger Delta crude oils in the oil samples. Ratios of H30/H30+NH25b, OL/OL+NH25b and NH30/NH30+NH25b, with the highest R^2 values, showed very strong predictability (97.33%, 97.03% and 97.49%) for estimating the compositions of the two Niger Delta crude oils in their mixes.
Keywords: Niger Delta, crude oil mixtures, plot, ratio, terpane, allocation, HCA, PCA.

INTRODUCTION

Analysis of data in order to extract relevant information used to solving problems in chemistry is termed chemometrics (Wold, 1995). Chemometric methods employed for analyses of oil hydrocarbon data include principal component analysis (PCA), hierarchical cluster analysis (HCA), partial least-squares (PLS), regression analysis, discriminant analysis (DA), ratio plots, and factor analysis (Mudge, 2002; Christensen and Tomasi, 2007; Al-Kaabi *et al.,* 2017). A diverse set of instrumental methods are used for oil hydrocarbon data gathering however, gas chromatography (GC) techniques are the most widely used (Bayona *et al.,* 2015; Ismail *et al.,* 2016). The GC coupled with a flame ionization detector (GC-FID) or mass spectrometry (GC-MS) separates and detects oil hydrocarbons present in crude oils or source rock extracts, even in low quantities (ppm and sub-ppm level), and facilitates their detailed distribution and composition (Wang *et al.,* 1999).

Chemometric methods have been used to analyse oil hydrocarbon data generated by GC to accomplish a variety of goals including correlation and/or differentiation of source rocks, crude oils and tarballs and to discriminate between adulterated and unadulterated gasoline samples as well as accurately detect the different solvents used in adulterations (Hostettler *et al.,* 2004; Mendes and Barbeira 2013; Wang *et al.,* 2018). Oil spill identification, monitoring and environmental forensics have also adopted chemometrics to resolve the source(s) of hydrocarbon pollutions in the environment and distinguish them from closely related sources, suggest mixed sources, determine weathering state and monitor pollution patterns and their sources particularly where multiple sources are present (Aboul-Kassim and Simoneit, 1995; Lavine *et al.,* 1995; Stout *et al.,* 2001; Malmquist *et al.,* 2007). Crude oil mixtures have been evaluated using chemometric methods to unmix a set of oil samples mixed in the laboratory, identify reservoir continuity and distinct compartments and back allocate the compositions of end member crude oils from different

contributing fields to commingled pipeline crude oil production (Kaufman *et al.,* 1987; Hwang, *et al.,* 2000; Zhan *et al.,* 2016; Kanshio, 2020).

Aliphatic and polycyclic aromatic hydrocarbons composition similarities between oil spilled soil samples and their source crude oil were assessed by Onyema *et al.,* (2013, 2016) using HCA and PCA. Ekpo *et al.,* (2018), using geochemical fingerprinting and multivariate statistical data showed crude oils from three reservoir units, from western offshore Niger Delta of Nigeria, originated from mixed marine/terrigeneous source rocks deposited under oxidizing condition and the existence of a single compartment with vertical and lateral reservoir continuity. Onojake *et al.,* (2015) has also used multivariate representation of molecular marker data to discriminate Niger Delta crude oils into two genetic families which differed comparatively by the relative marine inputs. In this study, we have evaluated the terpane data of two Niger Delta crude oils and their mixtures using chemometric methods. This is aimed at extracting terpanes that correspond to the composition variations of two Niger Delta crude oils in their mixes.

EXPERIMENTAL
Samples
The Niger Delta region, situated in Southern Nigeria, is an extremely prolific petroleum-producing Tertiary delta basin located between longitudes 5°- 8° E and latitudes 3° - 6° N (Reijers *et al.,* 1997; Tuttle *et al.,* 1999). Nine (9) states comprise the petroleum-producing Niger Delta region of Southern Nigeria. They are Ondo, Edo, Delta, Imo, Abia, Bayelsa, Rivers, Akwa-Ibom and Cross-River states. Two crude oil samples (one each) were obtained from onshore oil producing fields in Rivers and Delta states (samples A-10 and F-00). Four mixtures of the two crude oil samples were made at different compositions of 80:20, 60:40, 40:60 and 20:80, and labelled samples B-82, C-64, D-46, and E-28, respectively. All six oil samples were placed in pre-cleaned glass jars and stored in the refrigerator at a temperature of less 4 °C until analysis.

Crude Oil Fractionation
50 mg of each crude oil sample was weighed into a labelled centrifuge tube. Excess pentane was added to precipitate the asphaltenes. The mixtures (oil sample and pentane) were allowed to stand for three hours and then centrifuged for thirty minutes to coalesce the precipitated asphaltenes. The pentane soluble (PS) fractions were decanted and concentrated under a gentle stream of nitrogen. The PS fraction of each oil sample was transferred to the top of a glass column (30 x 1 cm) plugged with glass wool at the base and packed with activated silica (mesh 100-200). *n*-Hexane was poured in to elute the saturates, which contain the terpanes. The eluents were collected into pre-weighed vials and the solvents evapourated to dryness under a gentle stream of nitrogen at 40 °C.

Gas Chromatography-Mass Spectrometry (GC-MS) Analysis
Analyses of the saturate fraction of each oil sample was performed on an Agilent 7890A gas chromatograph (GC) system equipped with a HP-5 silica capillary column (50 m x 320 μm i.d and 0.25μm film thickness) and an Agilent 5975 mass selective detector (MSD). With the aid of a G4513A automatic liquid sampler, 1μL of the saturate fraction was injected into the GC capillary column in splitless mode. The GC oven was set to an initial temperature of 80 °C for 5 min., then ramped to 300 °C at a rate of 4 °C min^{-1} and held at this temperature for 30 min. The GC analysis was performed in selected ion mode at the mass to charge (m/z) 191 and the terpanes were identified by their relative retention times in comparison with related literature (Wang *et al.,* 2006). Quantification was acquired by area integration of each identified compound peak, which was processed by Chemstation OPEN LAB CDS software. The eluents were collected into pre-weighed vials and the solvents evapourated to dryness under a gentle stream of nitrogen at 40 °C.

Chemometric Analysis
Data from GC-MS analyses were subjected to chemometric analysis. Microsoft's Excel 2013 spreadsheet package was used for preparation of tables, calculation of ratios, and plots. Hierarchical cluster analysis (HCA) and principal component analysis (PCA) were applied to test data for correlations, sort out differences, and extract essential information which correspond to the composition variations of the two Niger Delta crude oils in their mixes using PAST software version 4.11 (Hammer, 2022).

RESULTS AND DISCUSSION
Distribution of Terpanes
GC-MS analyses of the saturate fractions at m/z 191 showed well resolved peaks for all the six oil samples. This indicates that the two Niger Delta crude oils (samples A-10 and F-00) and their compositional mixes (samples B-82, C- 64, D-46, and E-28) contain terpanes. The distributions of terpanes in the oil samples are shown in Figure 1.

Forty-three (43) terpanes were detected in all the crude oil samples which range from C_{19} tricyclic terpane (TR19) to C_{35} homohopanes (H35). Terpanes detected in the six oil samples are listed in Appendix I. Total abundance of terpanes in the oil samples increased in the order E-28 < F-00 < A-10 < C-64 < D-46 < B-82. This suggests the total abundance of terpanes did not mix correspondingly with the composition of the two Niger Delta crude oils in their mixes. Hopane (H30), oleanane (OL) and 30-norhopane (NH30) were the most abundant terpanes in all six oil samples and constitute 12.80 - 17.36%, 11.89 - 15.57% and 7.83 - 13.01% of total terpane abundances, respectively (fig. 1). From figure 1, it was also observed

that the abundances of C_{20} (TR20a-d) and C_{21} (TR21a-d) tricyclic terpanes were reduced in sample A-10, but increased in samples B-82, C-64, D-46, E-28 and F-00. Total abundance of C_{20} (TR20a-d) tricyclic terpanes in sample A-10 constituted 3.35% of total terpane abundance, but in samples B-82, C-64, D-46, E-28 and F-00 constituted 12.94%, 19.44%, 22.67%, 17.65% and 17.01%, respectively. The total abundance of C_{21} (TR21a-d) tricyclic terpanes in sample A-10 constituted 2.74% of total terpane abundance, but in samples B-82, C-64, D-46, E-28 and F-00 constituted 6.61%, 8.74%, 9.84%, 7.26% and 6.86%, respectively. These terpane distributions and abundances suggest the oil samples are generally similar and did not mix correspondingly to

permit differentiation of the compositions of the two Niger Delta crude oils in their mixes

Multivariate Correlations

In crude oil correlations with a large number of data, the use of multivariate statistical methods are beneficial (Peters *et al.,* 2005). Hierarchical cluster analysis (HCA), a multivariate statistical method, was employed for comparison of the oil samples. HCA sorts out the similarities among individual elements and the result is displayed graphically as a dendrogram which shows the similarities (correlation) among the elements (Pavon *et al.,* 2006).
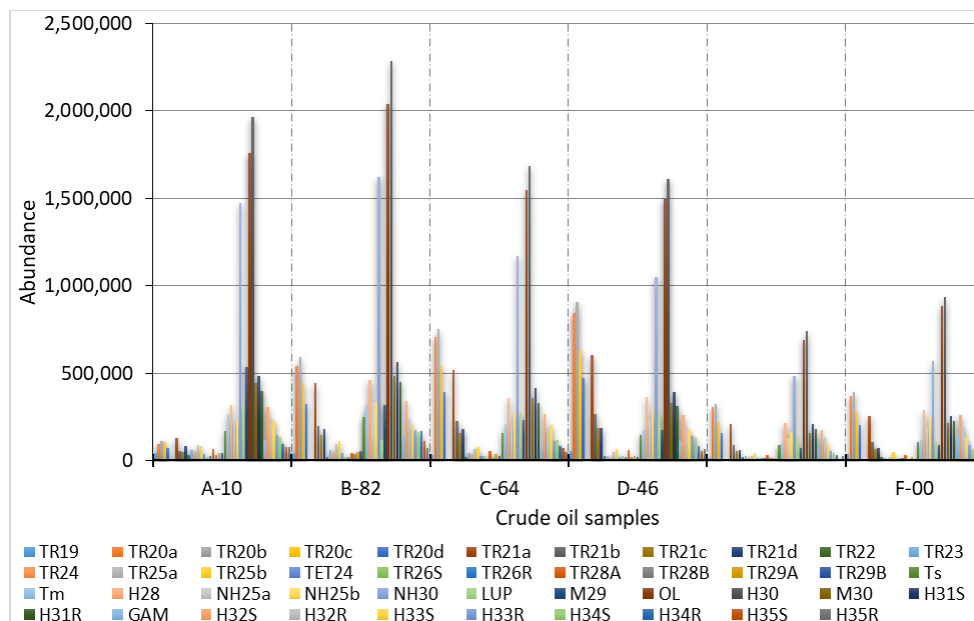

Figure 1: Distribution of terpanes in the two Niger Delta crude oils (samples A-10 and F-00) and their compositional mixtures (samples B-82, C-64, D-46, and E-28)
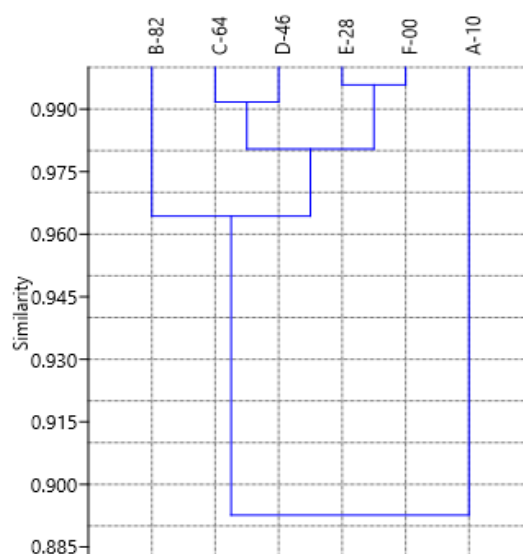

Figure 2: Dendogram of hierarchical cluster analysis (HCA) showing the similarities in the terpane (C_{19} - C_{35}) distributions among the oil samples

From the dendogram (fig. 2), HCA separated the oil samples, according to similarities in their terpane distributions, into two clusters. Cluster-1 comprise sample A-10, while cluster-2 comprise samples B-82, C-64, D-46, E- 28 and F-00 with similarities ranging from 0.9652 to 0.9958. The similarity between cluster-1 and cluster-2 samples was 0.8904. Similarity values of HCA usually range between -1 and 1. The greater the number, the stronger the relationship between the variables, while a value of zero (0) indicate that the two sets of variables are unrelated and distinct. From fig. 2, the six crude oil samples are strongly related (correlated) in the distributions of their terpanes. Samples B-80:20, C- 60:40, D-40:60 and E-20:80, the four compositional mixtures, are strongly related with sample A-10 (89.04%) and a very strongly related to

(96.52% - 99.58%) with sample F-00. The HCA result indicates all the oil samples are significantly similar in their terpane distributions to permit differentiation and consequently allocation of the compositions of the two Niger Delta crude oils in their mixes.

Principal component analysis (PCA) was used for evaluation of individual (C_{19} - C_{35}) terpanes among the six crude oil samples. PCA is a multivariate statistical method used to transform data sets with correlated variables into a new set of uncorrelated variables, and assesses each variable's contribution to the overall variance of the data (Burns *et al.,* 1997; Stella *et al.,* 2002). The result of PCA is presented as a graph plotting in fig. 3.
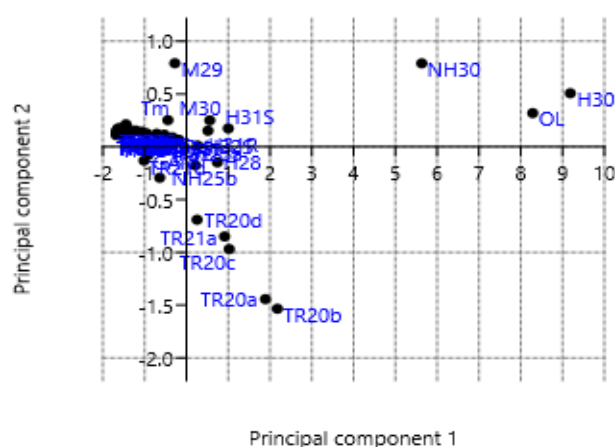
| Principal Component (PC)	Eigenvalue	% variance
1	5.74622	95.77
2	0.213635	3.5606
3	0.029959	0.49932
4	0.007305	0.12175
5	0.002468	0.04114
6	0.000415	0.0069198

Figure 3: Principal component analysis (PCA) plot showing the variance of C_{19} - C_{35} terpanes among the six oil samples

From the PCA plot, most of the terpanes were clustered close to each other; others such as TR20a, TR20b, TR20c, TR20d, TR21a and M29 were slightly separated while H30, OL and NH30 were considerably separated, outliers (fig. 3). Four principal components (PCs) accounted for 99.95% of the total variance of individual terpanes among the six crude oil samples (fig. 3). Principal component 1 (PC1) with an eigenvalue score of 5.7462 explained 95.77% of the total variance of individual terpanes among the six crude oil samples, PC2 with a score of 0.2136 explained 3.56%, PC3 with a score of 0.0300 explained 0.50% and PC4 with a score of 0.0073 explained 0.12%. The principal components (PC1-4) scores of individual terpanes are listed in Appendix II. The significance of each component is expressed by the variance (i.e.

eigenvalue) of its projections on the PCs score plots or by the proportion of the variance explained (Zumberge, 1987). From PC1 scores, which explained most of the variance (95.77%), H30, OL and NH30 had considerably high eigenvalues of 9.1860, 8.2858 and 5.6309 compared to other individual terpanes (Appendix II). This PCA results indicate that H30, OL and NH30 significantly explained the variance in the compositions of the two Niger Delta crude oils in their mixes.

Terpane Plots

Terpanes which explained the most variance among the six oil samples were plotted to justify the variance (fig. 4).
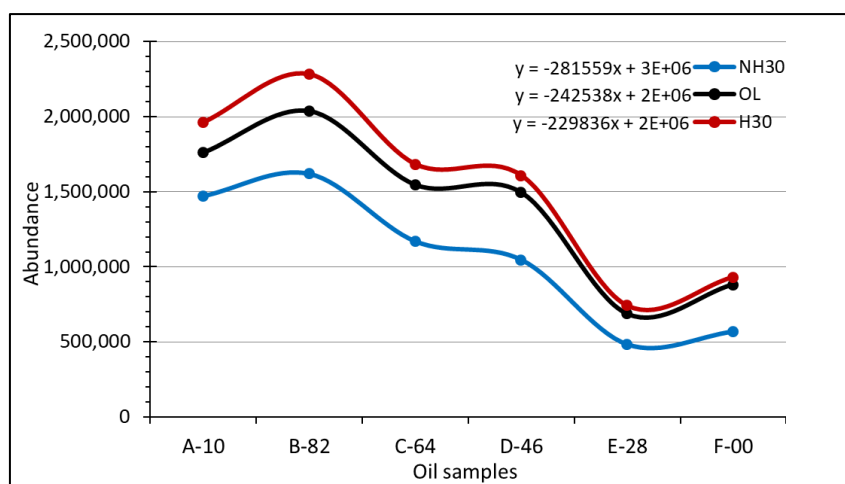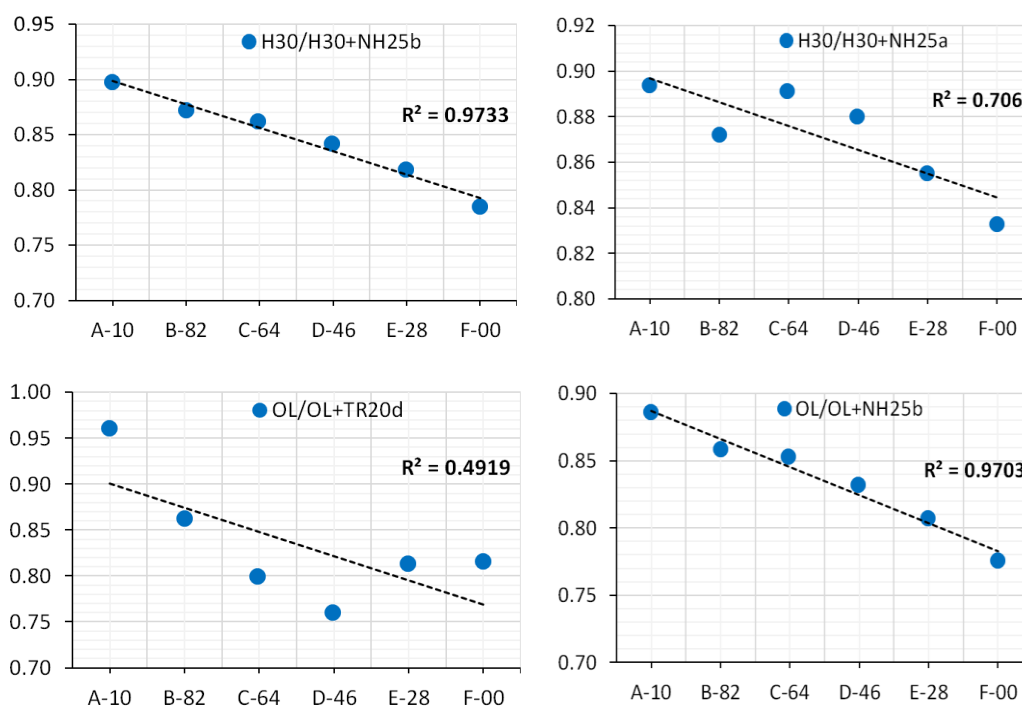
Figure 4: Plots of the abundance of hopane (H30), oleanane (OL) and 30-norhopane (NH30) showing their variation among the six crude oil samples

Figure 4 explained the variation pattern of the abundances of H30, OL and NH30 among the six crude oil samples. Generally, the abundances of H30, OL and NH30 decreased with increase in the composition of oil sample F-00 vis-à-vis increased with increase in the composition of oil sample A-10, respectively. From the linear equation of each plot ($y = mx + c$), the gradients (m) of H30, OL and NH30 were determined as -229,836, -242,538 and -281,559, respectively (fig. 4). The gradient results revealed that the abundances of H30, OL and NH30 significantly varied among the six crude oil samples. From the linear equation of each plot ($y = mx + c$), the gradients (m) determined as -229,836, -242,538 and -281,559 revealed that the

abundances of H30, OL and NH30 significantly varied among the six crude oil samples, respectively (fig. 4).

For distinguishing geochemical variations in crude oils, ratios of hydrocarbons which are most variant and least variant are used (Halpern 1995; Onyema *et al.*, 2020). From the PCA results, H30, OL, and NH30 explained the most variance among the oil samples with scores of 9.1860, 8.2858 and 5.6309, while NH25a, NH25b, and TR20d explained the least variance, with scores of -0.17614, 0.21272 and 0.25438, respectively (Apendix II). Nine (9) ratios were derived from these terpanes (H30, OL, NH30 and NH25a, NH25b, TR20d) calculated and plotted against the oil samples (fig. 5).
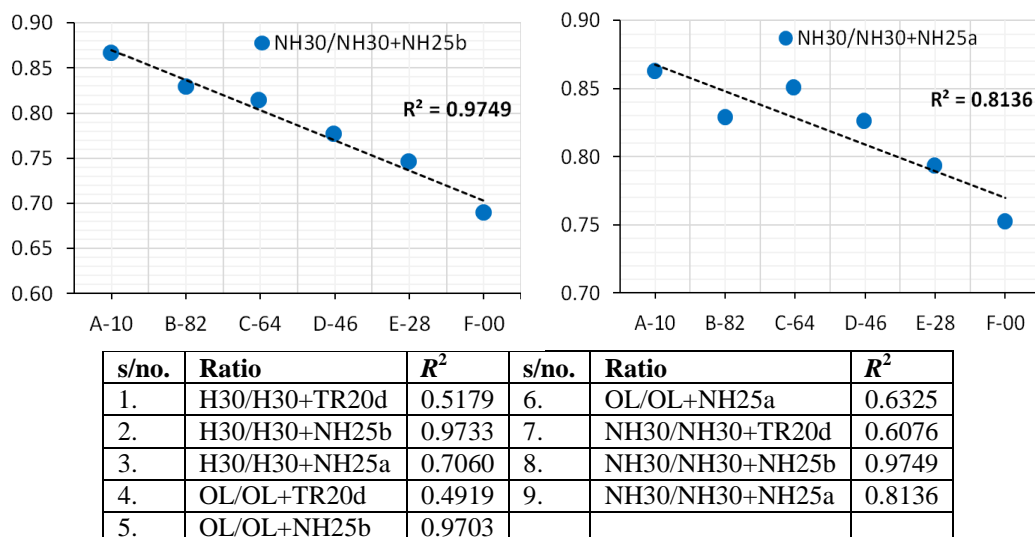
s/no.	Ratio	R^2	s/no.	Ratio	R^2
1.	H30/H30+TR20d	0.5179	6.	OL/OL+NH25a	0.6325
2.	H30/H30+NH25b	0.9733	7.	NH30/NH30+TR20d	0.6076
3.	H30/H30+NH25a	0.7060	8.	NH30/NH30+NH25b	0.9749
4.	OL/OL+TR20d	0.4919	9.	NH30/NH30+NH25a	0.8136
5.	OL/OL+NH25b	0.9703			

Figure 5: Plots of some derived terpane ratios showing their correlation (R^2) with the composition(s) of the two crude oils in the samples

From the plots of the derived terpane ratios, the line of best fit was used to evaluate for the correlation coefficient of determination (R^2). The values of R^2 normally ranges from 0 to 1; the closer the value is to 1 the stronger the relationship between the variables, while a value of 0 indicate the variables are not related (distinct). Values of R^2 for the 9 derived terpane ratios ranged from 0.4919 - 0.9749 (fig. 5). This indicate that the composition(s) of the two Niger Delta crude oils in the oil samples moderately corresponds with H30/H30+TR20d and OL/OL+TR20d; strongly corresponds with H30/H30+NH25a, OL/OL+NH25a, NH30/NH30+TR20d and NH30/NH30+NH25a; very strongly corresponds with H30/H30+NH25b, OL/OL+NH25b and NH30/NH30+NH25b ratio values.

Coefficient of determination (R^2) also gives information about the predictability of a model based on the proportion of total variation explained by the model *(Devore, 2011)*. A value of 0 indicate that the independent variables is not a predictor of the dependent variable, higher values indicate a better predictability of the dependent variable from the independent variables and a value of 1 indicates that the predictions of the dependent variable from the independent variables are exactly correct (*Draper and Smith, 1998*). From fig. 5, H30/H30+NH25b, OL/OL+NH25b and NH30/NH30+NH25b ratios had the highest R^2 values that accounted for 97.33%, 97.03% and 97.49% of the variability of the composition(s) of the two crude oils in the samples, respectively. These results indicate that H30/H30+NH25b, OL/OL+NH25b and NH30/NH30+NH25b ratios show very strong predictability for estimating the composition(s) of the two Niger Delta crude oils in their mixtures.

CONCLUSIONS
Chemometrics were used to evaluate terpanes (C_{19} to C_{35}) that correspond to the composition(s) of

two Niger Delta crude oils (samples A-10 and F-00) in their four mixtures (samples B-82, C-64, D-46, and E-28). Plots and HCA showed all six oil samples were generally similar in the terpane distributions, did not mix correspondingly and were geochemically related to permit differentiation and allocation of the composition(s) of the two Niger Delta crude oils in their mixtures. PCA result indicated that the first principal component (PC1) explained 95.77% of the total variance of individual terpanes among the oil samples with H30, OL, and NH30 explaining the most variance and NH25a, NH25b, and TR20d explaining the least variance, respectively. Nine (9) ratios were derived from the most and least variant terpanes and from their plots, R^2 values indicated that the ratio values moderately to very strongly correspond with the composition(s) of the two Niger Delta crude oils in the oil samples. R^2 results also indicated that H30/H30+NH25b, OL/OL+NH25b and NH30/NH30+NH25b ratios show very strong predictability (97.33%, 97.03% and 97.49%) for estimating the variability of the composition(s) of the two Niger Delta crude oils in their mixtures.

REFERENCES
- Aboul-Kassim, T. A. T., & Simoneit, B. R. T. (1995). Petroleum hydrocarbon fingerprinting and sediment transport assessed by molecular biomarker and multivariate statistical analyses in the Eastern Harbour of Alexandria, Egypt. *Marine Pollution Bulletin*, 30, 63-73.

- Al-Kaabi, N. S., Kristensen, M., Zouari, N., Solling, T. I., Bach, S. S., Al-Ghouti, M., & Christensen, J. H. (2017). Source identification of beached oil at Al Zubarah, Northwestern Qatar. *Journal of Petroleum Science and Engineering*, 149, 107-113.
- Bayona, J. M., Domínguez, C., & Albaigés, J. (2015). Analytical developments for oil spill fingerprinting: Review Article. *Trends in Environmental Analytical Chemistry*, 5, 26-34.
- Burns, W. A., Mankiewicz, P. J., Bence, A. E., Page, D. S., & Parker, K. R. (1997). A principal-component and least-squares method for allocating polycyclic aromatic hydrocarbons in sediment to multiple sources. *Environmental Toxicology and Chemistry*, 16, 1119-1131.
- Christensen, J. H. & Tomasi G. (2007). Practical aspects of chemometrics for oil spill fingerprinting: Review: *Journal of Chromatography A*, 1169(1-2), 1-22.
- *Devore, J. L. (2011). Probability and statistics for engineering and the sciences (8th ed.). Cengage Learning, Boston, MA.*
- *Draper, N. R., & Smith, H. (1998). Applied Regression Analysis. Wiley-Interscience.*
- Ekpo, B. O., Essien, N., Neji, P. A., & Etsenake, R. O. (2018). Geochemical fingerprinting of western offshore Niger Delta oils. *Journal of Petroleum Science and Engineering*, 160, 452-464.
- Halpern, H. I. (1995) Development and applications of light-hydrocarbon-based star diagrams. *American Association of Petroleum Geologists (AAPG) Bulletin*, 76, 801-815.
- Hammer, O. (2022). Past 4.11: Paleontological Statistics software package, version 4.11. https://www.nhm.uio.no/english/research/resources/past/ downloaded on September 15, 2022.
- Hostettler, F. D., Rosenbauer, R. J., Lorenson, T. D., & Dougherty, J. (2004). Geochemical characterization of tarballs on beaches along the California coast (Part I). *Organic Geochemistry*, 35, 725-746.
- Hwang, R. J., Baskin, D., K., & Teerman, S. C. (2000). Allocation of commingled pipeline oils to field production. *Organic Geochemistry*, 31, 1463-1474.
- Ismail, A., Toriman, M. E., Juahir, H., Kassim, A. M., Zain, S. M., Ahmad, W. K. W., Wong, K. F., Retnam, A., Zali, M. A., Mokhtar, M., & Yusri, M. A. (2016). Chemometric techniques in oil classification from oil spill fingerprinting. *Marine Pollution Bulletin*, 111(1-2), 339-346.
- Kanshio, S. (2020). A review of hydrocarbon allocation methods in the upstream oil and gas industry. *Journal of Petroleum Science and Engineering*, 184, 106590.
- Kaufman, R. L., Ahmed, A. S., & Hempkins, W. B. (October 1987). "A New Technique for the Analysis of Commingled Oils and its Application to Production Allocation Calculations," *Proc.*, Indonesian Petroleum Assn., 16th Annual Convention, Jakarta, 247.
- Lavine, B. K., Mayfield, H., Kromann, P. R., & Faruque, A. (1995). Source identification of underground fuel spills by pattern recognition analysis of high-speed gas chromatograms. *Analytical Chemistry*, 67, 3846-3852.
- Malmquist, L. M. V., Olsen, R. R.., Hansen, A. B., Andersen, O., & Christensen, J. H. (2007). Assessment of oil weathering by gas chromatography–mass spectrometry, time warping and principal component analysis. *Journal of Chromatography A*, 1164(1-2), 262-270.
- Mendes G. & Barbeira P. J. S. (2013). Detection and quantification of adulterants in gasoline using distillation curves and multivariate methods. *Fuel*, 112, 163-171.
- Mudge, S. M. (2002). Reassessment of the hydrocarbons in Prince William Sound and the Gulf of Alaska: Identifying the source using partial least-squares. *Environmental Science and Technology*, 36, 2354-2360.
- Onojake M. C., Abrakasa S., & Osuji L. C. (2015). Chemometric representation of molecular marker data of some Niger Delta crude oils. *Egyptian Journal of Petroleum*, 24, 139–143.
- Onyema, M. O., Okoroh, N. C., Okorie, I. H., & Osuji, L. C. (2020). Geochemical characterization of two Niger Delta crude oils and their mixtures II: Correlation of bulk properties and aliphatic hydrocarbons. *Science Journal of Chemistry*, 8(5), 118-123.
- Onyema, M. O., Osuji, L. C., & Isiuku, B. O. (2016). Compositional Alteration of Polycyclic Aromatic Hydrocarbons by Burning Oil Spilled Soils. *Academia Journal of Environmental Science*, 4(11), 213-219.
- Onyema, M. O., Osuji, L. C., & Ofodile, S. E. (2013). Geochemical fingerprinting of an oil-impacted site, Niger Delta: Source and weathering profile of aliphatic hydrocarbons. *Researcher*, 5(10), 16-21.
- Pavón, J. L. P., Peña, A. G., Pinto, C. G., & Cordero, B. M. (2006). Differentiation of types of crude oils in polluted soil samples by headspace-fast gas chromatography–mass spectrometry. *Journal of Chromatography A*, 1137(1), 101-109.
- Peters, K. E., Walters, C. C., & Moldowan, J. W. (2005). The biomarker guide: Biomarkers and isotopes in petroleum and earth history, 2nd ed, Cambridge University Press, Cambridge, UK.
- Reijers, T. J. A., Petters, S. W., & Nwajide, C. S. (1997). Chapter 7: The Niger delta basin. *Sedimentary Basins of the World*, 3, 151-172.
- Stella, A., Piccardo, M. T., Coradeghini, R., Redaelli, A., Lanteri, S., Armanino, C., & Valerio, F. (2002). Principal component analysis application in polycyclic aromatic hydrocarbons "mussel

watch" analyses for source identification. *Analytica Chimica Acta*, 461, 201-213.

- Stout, S. A., Uhler, A. D., & McCarthy, K. J. (2001). A strategy and methodology for defensibly correlating spilled oil to source candidates. *Environmental Forensics*, 2, 87-98.
- Tuttle, L. W. M., Charpentier, R. R., & Brownfield, E. M. (1999). Chapter A. Tertiary Niger Delta (Akata-Agbada) petroleum system (No. 719201) Niger Delta Province, Nigeria, Cameroon, and Equatorial Guinea, Africa. U.S. Geological Survey, World Energy Project Open-File Report: 99-50-H.
- Wang, Y. P., Zou, Y. R., Shi, J. T., & Shi J. (2018). Review of the chemometrics application in oil-oil and oil-source rock correlations. *Journal of Natural Gas Geoscience*, 3(4), 217-232.
- Wang, Z. D., Fingas, M., & Sigouin, L. (2000). Characterization and source identification of an unknown spilled oil using fingerprinting techniques by GC-MS and GC-FID. *LC-GC*, 10: 1058-1068.
- Wang, Z., Fingas, M., & Page, D. S. (1999). Oil spill identification - A review. *Journal of Chromatography A*, 843:369-411.
- Wold, S. (1995). Chemometrics; what do we mean with it, and what do we want from it? *Chemometrics and Intelligent Laboratory Systems*, 30(1), 109-115.
- Zhan, Z. W., Zou, Y. R., Shi, J. T., Sun, J. N., & Peng, P. (2016). Unmixing of mixed oil using chemometrics. *Organic Geochemistry*, 92, 1-15.
- Zumberge, J. E. (1987). Prediction of source rock characteristics based on terpane biomarkers in crude oils: A multivariate statistical approach. *Geochimica Cosmochim. Acta*, 51, 1625-1637.

APPENDIX I

Labels of terpane peaks detected in the mass fragmentogram (m/z 191) of the oil samples

Peak	Terpanes	Code	formula
1	C_{19} tricyclic terpane	TR19	$C_{19}H_{34}$
2	C_{20} tricyclic terpane (a)	TR20a	$C_{20}H_{36}$
3	C_{20} tricyclic terpane (b)	TR20b	$C_{20}H_{36}$
4	C_{20} tricyclic terpane (c)	TR20c	$C_{20}H_{36}$
5	C_{20} tricyclic terpane (d)	TR20d	$C_{20}H_{36}$
6	C_{21} tricyclic terpane (a)	TR21a	$C_{21}H_{38}$
7	C_{21} tricyclic terpane (b)	TR21b	$C_{21}H_{38}$
8	C_{21} tricyclic terpane (c)	TR21c	$C_{21}H_{38}$
9	C_{21} tricyclic terpane (d)	TR21d	$C_{21}H_{38}$
10	C_{21} tricyclic terpane (e)	TR21e	$C_{21}H_{38}$
11	C_{21} tricyclic terpane (f)	TR21f	$C_{21}H_{38}$
12	C_{22} tricyclic terpane	TR22	$C_{22}H_{40}$
13	C_{23} tricyclic terpane	TR23	$C_{23}H_{42}$
14	C_{24} tricyclic terpane	TR24	$C_{24}H_{44}$
15	C_{25} tricyclic terpane (a)	TR25a	$C_{25}H_{46}$
16	C_{25} tricyclic terpane (b)	TR25b	$C_{25}H_{46}$
17	C_{24} tetracyclic terpane	TET24	$C_{24}H_{42}$
18	C_{26} S tricyclic terpane (S)	TR26S	$C_{26}H_{49}$
19	C_{26} tricyclic terpane (R)	TR26R	$C_{26}H_{50}$
20	C_{28} tricyclic terpane (a)	TR28A	$C_{28}H_{52}$
21	C_{28} tricyclic terpane (b)	TR28B	$C_{28}H_{52}$
22	C_{29} tricyclic terpane (a)	TR29A	$C_{29}H_{54}$
23	C_{29} tricyclic terpane (b)	TR29B	$C_{29}H_{54}$
24	Ts: 18a (H),21β (H)-22,29,30-trisnorhopane	Ts	$C_{27}H_{46}$
25	Tm: 17a (H),21β (H)-22,29,30-trisnorhopane	Tm	$C_{27}H_{46}$
26	17a (H),18a (H),21β (H)-28,30-bisnorhopane	H28	$C_{28}H_{48}$
27	17a (H),21β (H)-25-norhopane	NH25a	$C_{29}H_{50}$
28	17β (H),21a (H)-25-norhopane	NH25b	$C_{29}H_{50}$
29	17a (H),21β (H)-30-norhopane	NH30	$C_{29}H_{50}$
30	17a (H)-diahopane (Lupane)	LUP	$C_{30}H_{52}$
31	17a (H),21β (H)-30-norhopane (normoretane)	M29	$C_{29}H_{50}$
32	18a (H) and 18β (H)-oleanane	OL	$C_{30}H_{52}$
33	17a (H),21β (H)-hopane	H30	$C_{30}H_{52}$
34	17β (H),21a (H)-hopane (moretane)	M30	$C_{30}H_{52}$
35	22*S*-17a (H),21β (H)-30-homohopane	H31S	$C_{31}H_{54}$
36	22*R*-17a (H),21β (H)-30-homohopane	H31R	$C_{31}H_{54}$

37	Gammacerane	GAM	$C_{30}H_{52}$
38	22*S*-17a (H),21β (H)-30,31-bishomohopane	H32S	$C_{32}H_{56}$
39	22*R*-17a (H),21β (H)-30,31-bishomohopane	H32R	$C_{32}H_{56}$
40	22*S*-17a (H),21β (H)-30,31,32-trishomohopane	H33S	$C_{33}H_{58}$
41	22*R*-17a (H),21β (H)-30,31,32-trishomohopane	H33R	$C_{33}H_{58}$
42	22*S*-17a (H),21β (H)-30,31,32,33-tetrakishomohopane	H34S	$C_{34}H_{60}$
43	22*R*-17a (H),21β (H)-30,31,32,33-tetrakishomohopane	H34R	$C_{34}H_{60}$
44	22S-17a (H),21β (H)-30,31,32,33,34-pentakishomohopane	H35S	$C_{35}H_{62}$
45	22*R*-17a (H),21β (H)-30,31,32,33,34-pentakishomohopane	H35R	$C_{35}H_{62}$

APPENDIX II
First four Principal components (PC1-4) scores of individual terpanes

	PC 1	PC 2	PC 3	PC 4
TR19	-1.5505	0.098595	0.019672	-0.01453
TR20a	1.8887	-1.4436	0.072988	0.043887
TR20b	2.1771	-1.535	0.08622	0.036308
TR20c	1.0239	-0.96879	0.12011	-0.02019
TR20d	0.25438	-0.69048	0.11458	-0.0269
TR21a	0.91275	-0.84966	0.15248	-0.01647
TR21b	-0.63989	-0.2922	0.13745	-0.03001
TR21c	-1.0151	-0.13044	0.14236	-0.04104
TR21d	-0.90752	-0.075795	0.16635	-0.05422
TR22	-1.6388	0.12089	0.016279	0.02167
TR23	-1.542	0.17501	0.054031	0.000707
TR24	-1.6077	0.17517	0.060493	0.003879
TR25a	-1.4479	0.21292	0.11956	-0.02324
TR25b	-1.2955	0.13114	0.006282	-0.03355
TET24	-1.6551	0.15052	0.061254	-0.01943
TR26S	-1.6764	0.11054	0.060561	0.006443
TR26R	-1.6741	0.11996	0.015543	-0.00118
TR28A	-1.4568	0.11206	0.042163	0.049529
TR28B	-1.677	0.1487	0.0616	-0.02293
TR29A	-1.596	0.14769	0.044318	-0.02574
TR29B	-1.6409	0.17309	0.07693	-0.04586
Ts	-0.71274	0.12449	-0.03402	-0.07555
Tm	-0.4442	0.25324	0.042142	-0.06615
H28	0.74058	-0.14706	-0.36704	-0.01785
NH25a	-0.17614	0.065945	-0.22637	-0.09537
NH25b	0.21272	-0.17598	-0.38252	-0.02019
NH30	5.6309	0.79111	0.32199	-0.03303
LUP	-0.33243	0.089372	0.12679	0.43221
M29	-0.27695	0.79201	0.25336	0.17674
OL	8.2858	0.31978	-0.080442	0.011039
H30	9.186	0.50705	0.02706	-0.05374
M30	0.55881	0.25107	-0.022186	-0.00897
H31S	1.0011	0.17412	-0.10812	0.006608
H31R	0.51556	0.15253	-0.1699	0.026359
GAM	-0.79066	-0.017423	-0.42043	0.0555
H32S	0.26577	0.011545	-0.44923	0.046429
H32R	-0.29051	0.059707	-0.33054	0.062029
H33S	-0.53085	0.11554	-0.087236	0.073785
H33R	-0.94933	0.12668	-0.0046818	-0.03642
H34S	-1.0689	0.1424	0.1087	-0.03491
H34R	-1.213	0.1542	0.045214	-0.1267
H35S	-1.4135	0.18351	0.078977	-0.11003
H35R	-1.4335	0.13583	0.047257	0.001079