

Technical Note

# Pearson Correlation Coefficient Applied to Petroleum System Characterization: The Case Study of Potiguar and Recôncavo Basins, Brazil

Érica T. Morais <sup>1</sup>, Gabriel A. Barberes <sup>2,\*</sup> , Igor Viegas A. F. Souza <sup>1</sup>, Fabiano G. Leal <sup>1</sup>, Jarbas V. P. Guzzo <sup>1</sup> and André L. D. Spigolon <sup>1</sup>

<sup>1</sup> Petrobras/CENPES, Geochemistry Division, Rio de Janeiro 21941-915, Brazil; ericat@petrobras.com.br (É.T.M.); igorviegas@petrobras.com.br (I.V.A.F.S.); fabianoleal@petrobras.com.br (F.G.L.); guzzo@petrobras.com.br (J.V.P.G.); andrespigolon@petrobras.com.br (A.L.D.S.)

<sup>2</sup> Center of Geoscience, Faculty of Science and Technology, University of Coimbra, Largo Marquês de Pombal, 3000-272 Coimbra, Portugal

\* Correspondence: gabriel.barberes@gmail.com

**Abstract:** This study applied the Pearson correlation coefficient and principal component analysis as tools for unsupervised qualitative petroleum system evaluation techniques. A total of 252 oil samples (32 features per sample) representative of two Brazilian sedimentary basins (Recôncavo and Potiguar) were used to classify them according to their respective degrees of maturation and origin. The large initial set of variables comprises data on  $\delta^{13}\text{C}$  composition, saturate, aromatic, polar compound fractions, and the techniques reduced biomarkers to the most important variables, maintaining the global pattern of variance. The results were efficient in discriminating different petroleum systems from lacustrine, marine, and mixing sources, as observed in the studied accumulations from the Lower Cretaceous sediments of the Recôncavo and Potiguar basins. The methodology proved to be very useful to better characterize the petroleum systems. This methodology can be applied to analyze a large amount of oil samples, using simple software and spending relatively less time.

**Keywords:** Pearson correlation coefficient; principal component analysis; petroleum system



**Citation:** Morais, É.T.; Barberes, G.A.; Souza, I.V.A.F.; Leal, F.G.; Guzzo, J.V.P.; Spigolon, A.L.D. Pearson Correlation Coefficient Applied to Petroleum System Characterization: The Case Study of Potiguar and Recôncavo Basins, Brazil. *Geosciences* **2023**, *13*, 282. <https://doi.org/10.3390/geosciences13090282>

Academic Editors: Jesus Martinez-Frias and Sebastian Lüning

Received: 26 June 2023

Revised: 6 September 2023

Accepted: 14 September 2023

Published: 18 September 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Chemometric methods such as the principal component analysis (PCA) and Pearson correlation coefficient (PCC) provide useful tools for more extensive and large-scale analyses of petroleum geochemistry data (e.g., [1,2]). The primary objective of these statistical analyses is to reduce the dimensionality of the initially obtained information to a few important components that best explain the variation in the data [2,3].

The data are processed and graphically displayed in chemometric exploratory data analysis (EDA). According to several authors, one of the most common techniques for EDA is the PCA (e.g., [2]). On the other hand, PCC, a dimensionless measure of the mutual association between a pair of variables, is one of the most used statistical quantitative methods in geosciences (e.g., [4–7]).

Particularly on Pearson correlation, there are not many references in the scientific literature about applying this method in petroleum geochemistry analyses regarding oil parameters. One of the few was published the last year by Wang et al. [8], although the method was already used to analyze other organic geochemical variables. Some examples are given below.

Rasheed et al. [9] established the PCC of various light hydrocarbon components to help assess the hydrocarbon prospects of the Kerala–Konkan offshore basin (India). El-Nady et al. [10] carried out cluster analysis and PCC between different parameters (TOC, S1, S2, Tmax, HI, OI, and PI) to (1) quantify the relationship between petroleum

potential and maturity and (2) analyze diagrams of HI, QI, and I versus maturity, in the Ras Gharib oilfield (Central Gulf of Suez, Egypt). El-Nady and Lotfy [11] applied linear regression analysis and PCC in samples from the Safir oilfield (Western Desert of Egypt) to investigate the relation between several parameters (S1, S2, S1 + S2, HI, OI, BI, PI, and TOC) of petroleum potentiality and to investigate the impact of changes in the Tmax and Ro%. Skupio and Barberes [7] established an easy and rapid way to indicate hydrocarbon saturated intervals in the borehole profile before well log interpretation in the Pomerania region (Poland), using radioactive isotopes ( $^{40}\text{K}$ ,  $^{238}\text{U}$ , and  $^{232}\text{Th}$ ) via thorium normalization and the PCC method. Wang et al. [8] applied PCC to understand the cause of the unusually high abundance of rearranged hopanes from crude oil obtained from the Southwest Depression of the Tarim Basin (China).

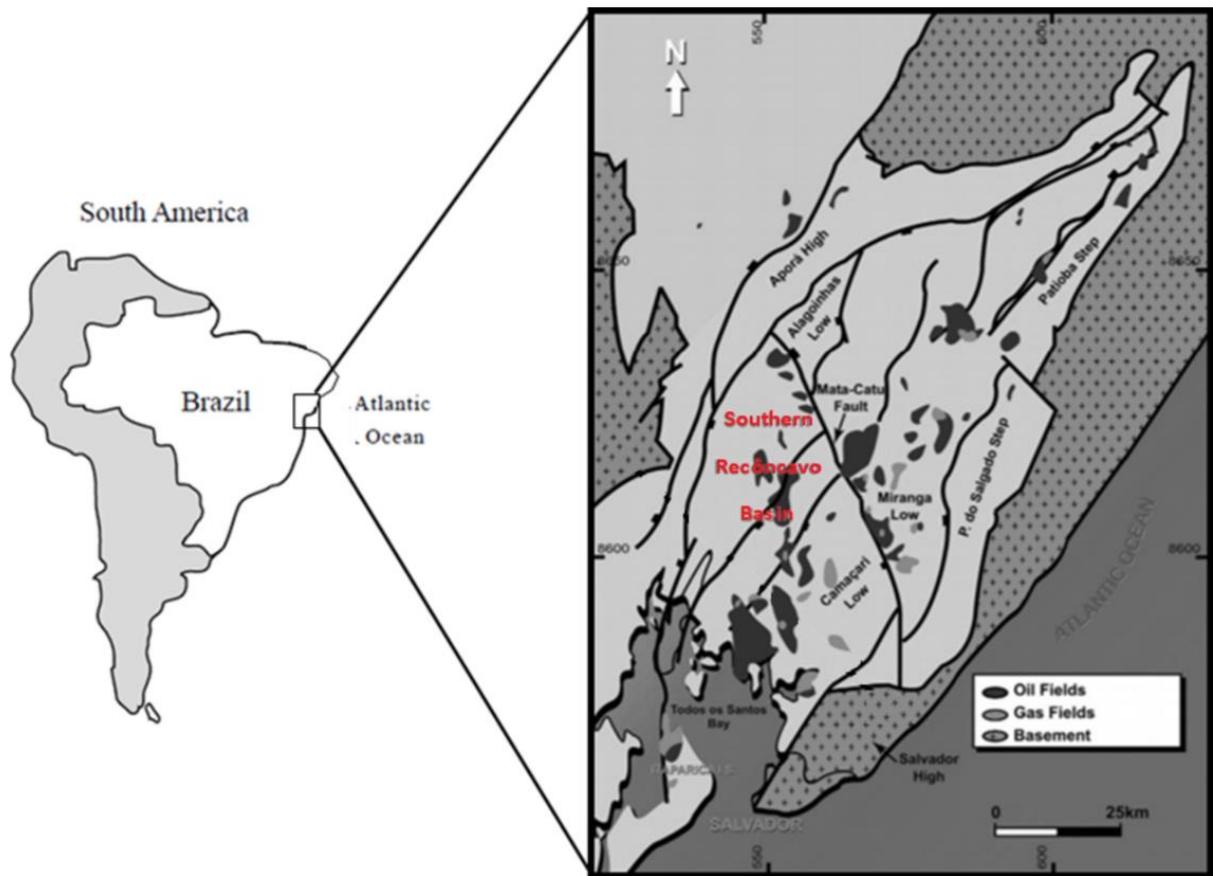
On the other hand, the principal component is a known method for analyses of petroleum geochemistry data. Serra et al. [12] applied PCA to separate 3000 oil samples from several Brazilian sedimentary basins to classify the paleoenvironmental conditions. According to the authors, previous studies with biomarkers and isotopes (e.g., decision trees) have shown that it is possible to classify oil samples from sedimentary basins, hitting percentages bigger than 95%.

In order to assist the petroleum system characterization, this paper applies a qualitative unsupervised discrimination (in terms of thermal evolution and source) technique in geochemical data obtained from the onshore part of two different Atlantic basins: Potiguar and Recôncavo. The main goal of this work is to compare the results obtained by the PCC technique with the traditional classification methods presented by Aguiar [13] for the Recôncavo Basin and the traditional coupled with artificial intelligence (AI) methods applied by Morais [14] for the Potiguar Basin.

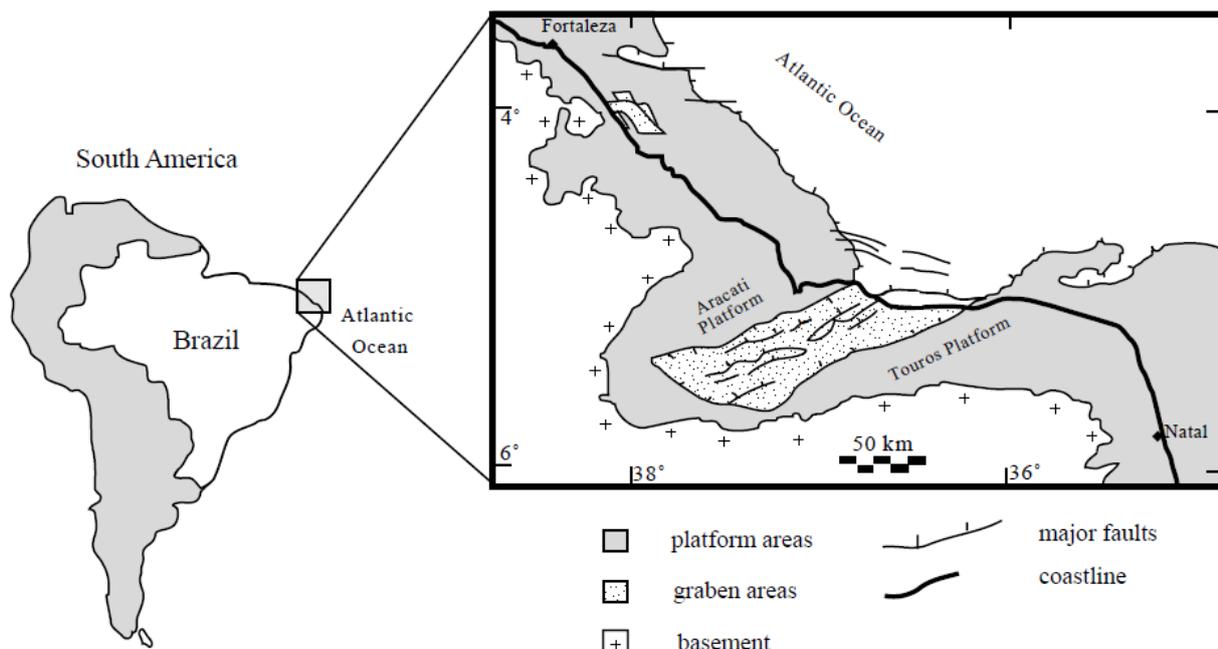
#### *Geological Framework and Petroleum Systems*

The Recôncavo Basin is part of the Recôncavo–Tucano–Jatobá rift system (Figure 1), which is an aborted intracontinental rift formed during the Gondwana break-up at the beginning process aperture of the South Atlantic in the Late Jurassic to Early Cretaceous periods [15,16]. The basin occupies an onshore area of approximately 11,500 km<sup>2</sup> [17]. The sedimentation in the Recôncavo Basin can be divided into two main sequences: the pre-rift and rift sections. The pre-rift section records subaerial sedimentation, which was gradually drowned, culminating with lacustrine deposits (Aliança, Sergi, Itaparica, and Candeias formations). The rift sequence begins with the deposition of a thick shaly sequence associated with turbidite deposits. This sequence evolves to deltaic sedimentation (Ilhas Group) and extensive fluvial sediments of the Massacará Group at the end of the rift phase [18]. The main source rock intervals are associated with the Candeias Fm. [19] and are represented by dark shales from Tauá Mb. and calciferous shales from Gomo Mb. [20]. The original reconstituted organic carbon map for the source rock interval shows values around 4% in the main depocenters of the Reconcavo Basin [21]. In terms of petroleum accumulation, the Reconcavo Basin can be subdivided into three major systems: pre-rift, rift-Candeias, and rift-Ilhas, with 15 exploratory plays, all associated with lacustrine oils generated in different thermal maturity degrees [13,18].

The Potiguar Basin is part of the northeast Brazilian rift system that includes an area of approximately 48,000 km<sup>2</sup>, of which 21,500 km<sup>2</sup> is onshore and 26,500 km<sup>2</sup> is offshore (Figure 2) [23,24]. The tectono-sedimentary evolution is related to three main stages: rift, transitional, and drift. The petroleum habitat is distinguished according to structural style and reservoir-source rocks association [25]. Two petroleum systems were recognized in the basin, Pendência (!) and Alagamar-Açu (!), which are linked to lacustrine and marine evaporitic source rocks, respectively [26]. The amount and the good preservation of the organic matter from both Pendência and Alagamar formations characterize them as oil-prone source-rock [27,28].



**Figure 1.** Location of Recôncavo Basin with structural framework and hydrocarbon fields. Modified from Magnavita et al. [22].



**Figure 2.** Location map and schematic tectonic framework of the Potiguar Basin, northeastern Brazil. Adapted from Santos Neto and Hayes [26].

The hydrocarbon accumulations that occur inland are located along the northeast–south–west trends, associated with offshore kitchens, and in the southern part of the basin within the grabbens [29,30].

## 2. Materials and Methods

Mathematically, the PCC is a linear association between two variables or a set of variables, where the coefficient is calculated by the ratio between the covariance of dependent and independent variables and the square root of the variance for these same variables [31]. It can be equivalently defined by:

$$r_{xy} = \frac{\sum(x_i - \bar{x})\sum(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2}\sqrt{\sum(y_i - \bar{y})^2}} \quad (1)$$

where  $\bar{x} = \frac{1}{n} \sum_{i=1}^N x_i$  denotes the mean of  $x$  and  $\bar{y} = \frac{1}{n} \sum_{i=1}^N y_i$  denotes the mean of  $y$ . The coefficient  $r_{xy}$  ranges from  $-1$  to  $1$ , and it is invariant to linear transformations of either variable [32]. A correlation of  $+1$  indicates a perfect direct relationship between two variables; a correlation of  $-1$  indicates that one variable changes inversely in relation to the other. Zero indicates the lack of any linear relationship at all.

It is important to note that according to Reimann et al. [33], the Pearson method requires a careful study of the univariate distribution of each of the variables prior to estimating the correlation coefficient.

The PCA is a distance-based ordination technique used primarily to display patterns in multivariate data. It aims to display the relative positions of data points in fewer dimensions while retaining as much information as possible and exploring relationships between dependent variables [32]. The technique is based on calculating the covariance matrix eigenvectors and eigenvalues. The term “eigenvalues” is linked to the variance of the data matrix, and it is an important index for determining the number of factors to retain in the analysis. The simplest method (rule) refers to the use of factors with eigenvalues greater than 1 (also known as the Kaiser criterion) [34].

Once eigenvectors are calculated from the covariance matrix, they are ordered by the highest to the lowest eigenvalue. The principal components are defined by score and loading values from the covariance matrix variation eigenvector, with the highest eigenvalue called the “first principal component” (PC1) of the dataset. In other words, the PC1 is a linear combination of original variables that captures the maximum variance in the dataset [31,35].

For this work, the PCA was calculated using PAST [36] and the PCC using Microsoft Excel™ software, respectively.

The analyses of all samples were carried out at the Petrobras Research Center (CEN-PES) and submitted to the same analytical procedures: liquid chromatography (LC), gas chromatography (GC), gas chromatography coupled to mass spectrometry (GC-MS), and mass spectrometry for stable carbon isotopes (MS), and which are better detailed in Morais [14]. All samples were analyzed simultaneously, following the timing of exploration and production activities.

The first studied case presents the petroleum samples from the Potiguar Basin, previously classified using traditional methods coupled with artificial intelligence into three classes (lacustrine, marine evaporitic, and mixed) [14]. For this case study, 198 oil samples (32 features per sample) were analyzed, giving a dataset of  $198 \times 32$ . The significant variables used for PCC were defined by PCA (after normalization), following the Kaiser criterion (Table 1).

**Table 1.** Respective eigenvalues and variance were obtained for each principal component. Applying the Kaiser criterion, the principal components are able to explain 81.6% of the data variance.

PC	Eigenvalue%	Variance
1	11.6	29.1
2	8.7	21.9
3	4.3	10.8
4	2.3	5.8
5	1.8	4.6
6	1.3	3.3
7	1.2	3.2
8	1.1	2.9

The second case study was based on petroleum samples from the Recôncavo Basin, which were previously classified, using classical methods, in only one class (lacustrine) [13]. It presents a dataset of size 54 (oil samples)  $\times$  32 (features). Two rounds of PCC were run to reach the objectives. The large initial set of variables was reduced via PCC to the most important variables, with correlation values higher than 0.6 or lower than  $-0.6$ .

The initial set of 32 features for both studied cases were the following: carbon isotope ( $\delta^{13}\text{C}$ ); C<sub>24</sub>:tetracyclics/C<sub>26</sub>:tricyclic terpane (TET24/26TRI); 19/23TRI; 21/23TRI; 24/25TRI; 26/25TRI; 26/28TRI; pristane/phytane (PRIS/PHY); PRI/nC17; phytane/nC18; hopane/sterane (HOP/STER); tricyclic/hopane (TRIC/HOP); C<sub>29</sub>:norhopane/C<sub>30</sub>:hopane (H29/H30); C<sub>29</sub>BB/C<sub>29</sub>; gammacerane index (GAM/H30); C<sub>35</sub>:pentakishomohopane/C<sub>34</sub>:tetrahomohopane (H35/H34); saturates fraction (%); aromatic fraction (%); polar fraction (%); % 27 a $\beta\beta$ R&S (218); % 28 a $\beta\beta$ R&S (218); % 29 a $\beta\beta$ R&S (218); diasterane index (DIA/C27AA); S/S+R hopanes; C<sub>23</sub>/C<sub>24</sub> tricyclic terpanes (23/24 TRI); C<sub>30</sub> tetracyclic polyprenoid (TPP (m/z 259)); 18 $\alpha$ -trisorneohopane/17 $\alpha$ -trisorhopane (TS/(TS+TM)); NORNEO/H29; DIAH/H30; and 17/(17+C27); CPI-1.

### 3. Results and Discussion

#### 3.1. Potiguar Basin

The large initial collection of 32 variables (including, basically, total carbon isotope, biomarkers, saturate, aromatic, and polar compounds fractions) was reduced via the PCA to 9 high-weighted variables: carbon isotope ( $\delta^{13}\text{C}$ ); gammacerane index (GAM/H30); C<sub>29</sub>:norhopane/C<sub>30</sub>:hopane (H29/H30); C<sub>35</sub>:pentakishomohopane/C<sub>34</sub>:tetrahomohopane (H35/H34); C<sub>24</sub>:tetracyclics/C<sub>26</sub>:tricyclic terpane (TET24/26TRI); pristane/phytane (PRIS/PHY); hopane/sterane (HOP/STER); polar fraction (% NSO); and saturates fraction (% saturates). From the subset of variables selected by the method, it was possible to observe that those are the ones classically applied to identify oil sources [2]. Figure 3 shows an example of the gas chromatogram, fragmentograms ( $m/z$  191 and  $m/z$  217), and total carbon isotopic composition for an oil typical for each class (from Potiguar Basin) defined by Morais [14].

In general, the parameters pointed out by the unsupervised discrimination are quite similar to the previous geochemical assessments. This fact is linked to the special discriminant characteristics imprinted via the paleoenvironmental conditions of the source rocks in the basin. The Alagamar Fm., for example, presents very low PRIS/PHY, HOP/STER ratios, and high GAM/H30, pointing out a marine evaporitic under anoxic conditions, which contrasts with the values measured in lacustrine samples that suggest an oxic environment. The ratios H29/H30 and H35/H34 do not present relevant variation, showing a homogeneous behavior for the analyzed interval, with values lower than 1, that matches molecular geochemistry from marine environments and anoxic conditions [2]. The percentage of the NSO is an important variable, as the oils derived from Alagamar Fm., recovered from reservoirs of the Açu Fm., are normally biodegraded. The biodegradation process was not observed or is insignificant in the reservoirs in the rift session (Pendência Fm.). However,

the PCC identified one sample with a lower correlation than those from the same oil class. After its technical supervised evaluation, we could observe that this sample is biodegraded.

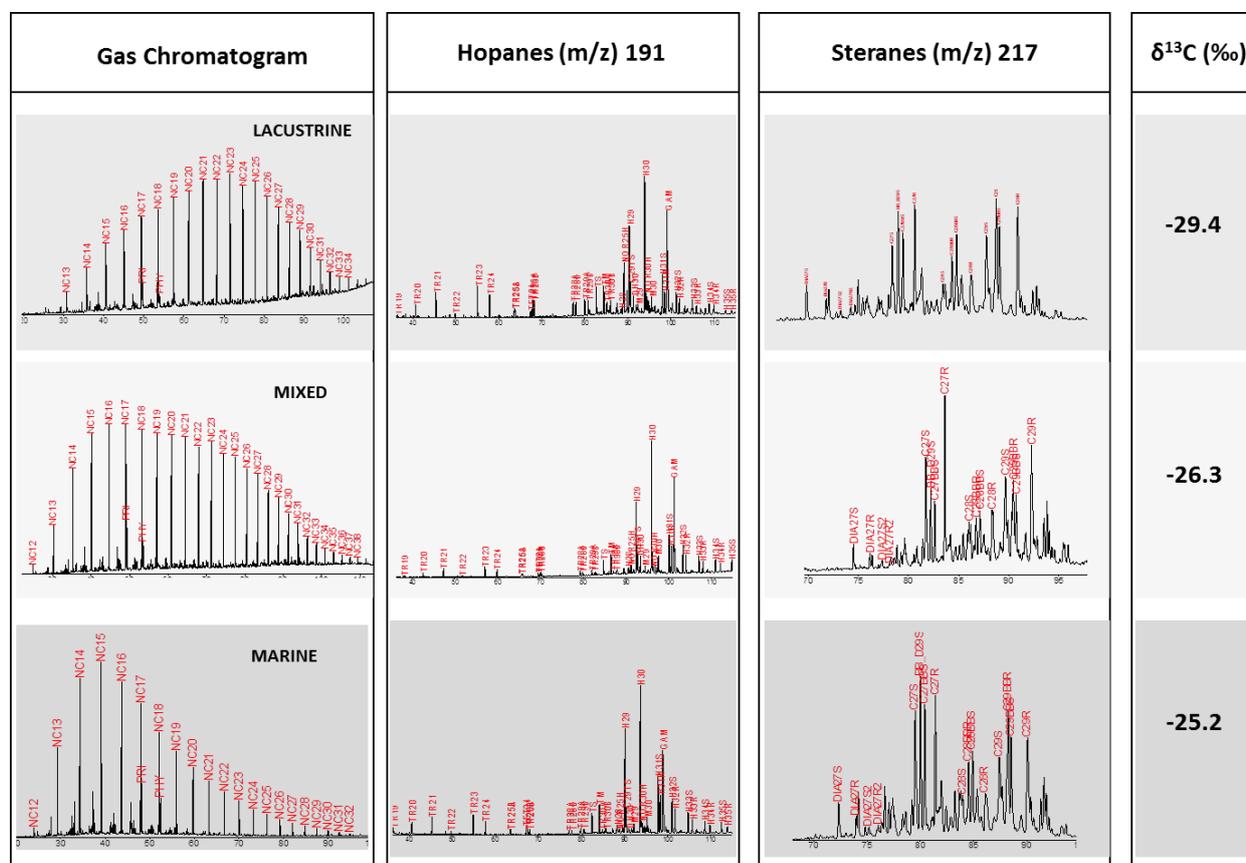


Figure 3. Summary of geochemical parameters for the three oil families analyzed in this study [14].

The similarities and differences between marine evaporitic and lacustrine samples are clear, and this is even more evident when we analyze the PCC results between these two systems (Table 2). The results from mixed oils presented an interesting variability, probably caused by the classification subjectivity. In other words, if the sample contains a bigger contribution from the marine evaporitic system (>50%), the result becomes closer to a “pure marine evaporitic sample”, and the same logic goes for the lacustrine system contribution (see Supplementary Data: File S1). It is important to highlight that the proposed method was not able to distinguish the four types of lacustrine oil samples (A, B, Siliciclastic A, and Siliciclastic B) classified by Morais [14], so all of them are classified as lacustrine.

Table 2. Average PCC results between the three different classes. The lacustrine system is the largest contributor to oils classified as mixed.

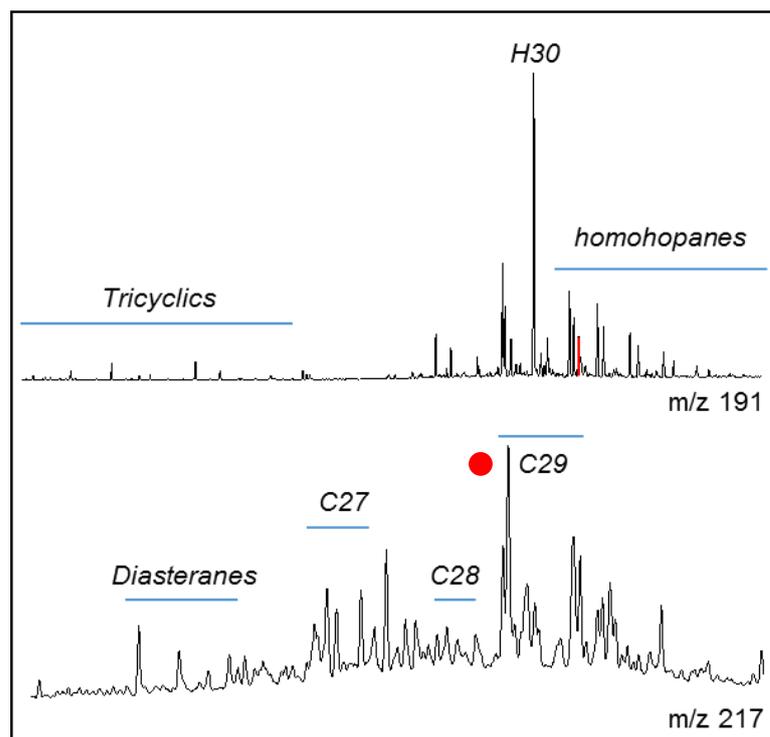
Classes	Lacustrine	Mixed	Marine
Lacustrine	1	0.93	0.75
Mixed	0.93	0.95	0.82
Marine	0.75	0.82	1

### 3.2. Recôncavo Basin

Since this case study is only one genetic class (lacustrine), the approach applied was different from the Potiguar basin. Two rounds of PCC were run, and the large initial set of variables was reduced via PCC to the most important variables GAM/H30, 21/23 tricyclics (21/23TRI), PRI/PHY, and tricyclics/hopane (TRI/HOP), with correlation

values higher than 0.6 and lower than  $-0.6$ . The first step was applied to the relation between the biomarkers (and the ratios between them) and other geochemical parameters ( $\delta^{13}\text{C}$ , saturate, aromatic, and polar compounds fractions). Then, these selected variables were used to quantify the degree of relation between the samples (second round of PCC). According to the data, the samples with correlation values higher than 0.8 were considered highly correlated. It was possible to identify three oil groups and three oil samples that do not present any correlation (see Supplementary Data: File S2).

Group A is characterized by high values of HOP/STER, TPP/(TPP + diasterane), significant presence of methylsteranes, low values of GAM/H30, tricyclic/hopanes (TRIC/HOP), and diasterane/ $\text{C}_{27}\alpha\alpha$ , associated with  $\delta^{13}\text{C}$  values  $< -28.15\text{‰}$  (Figure 4). Many authors have interpreted these features as indications of freshwater lacustrine sources [3,17,37].

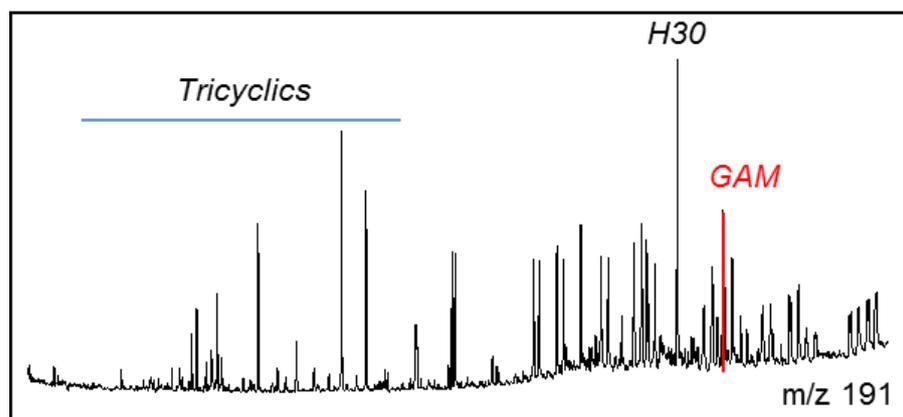


**Figure 4.** Group A terpanes and steranes' distributions. The red peak in the terpanes' distribution is the gammacerane compound. The red dot is the  $\alpha\alpha\alpha\text{RC}_{29}$  sterane.

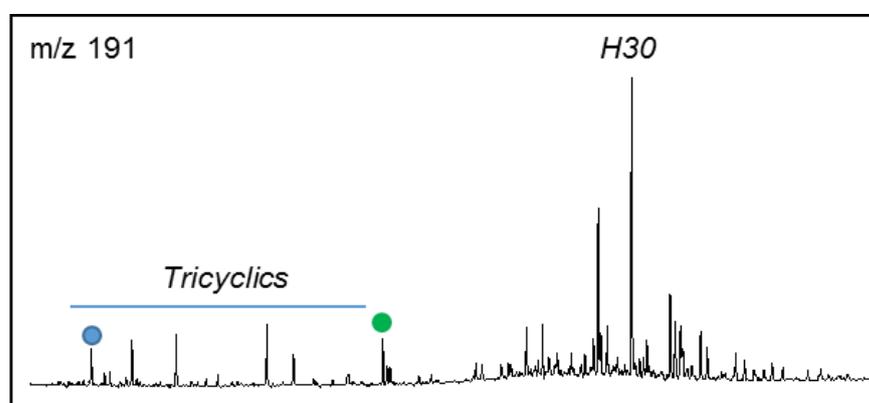
Group B (Figure 5) has a similar origin to group A but is characterized by high values of GAM/H30, 21/23TRI, and TRIC/HOP that were interpreted as indicative of high thermal evolution, as proposed by Aguiar [13]. This group of oils is from the Cexis Field, which is located near Camaçari Low. This interpretation is in accordance with the migration model proposed by Penteado [38] and Coutinho [21].

Group C has only two oil samples, from the Cexis and Massapê fields, and is characterized by extremely high TRIC/HOP values and cracked biomarkers. These features suggest that Group C was generated under extremely high thermal conditions.

Three oil samples do not present a correlation with any samples (File S2). The first sample presents high GAM/H30, associated with low TRIC/HOP. The second sample is characterized by high TRIC/HOP, associated with low GAM/H30, and the third sample shows high diasterane/ $\text{C}_{27}\alpha\alpha$ , associated with low TRIC/HOP and significant values of  $\text{C}_{19}$  tricyclic and  $\text{C}_{24}$  tetracyclic/ $\text{C}_{26}$  tricyclic ratio (Figure 6). These differences were interpreted as changes in paleoenvironmental conditions of the source rock in a specific basin compartment (e.g., water stratification and river inputs). The statistic tool did not highlight the third oil sample with no correlation, probably due to the set of sample characteristics.



**Figure 5.** Terpanes' distribution of Group B. The gammacerane peak is highlighted in red.



**Figure 6.** Terpanes' distribution of the sample presents no correlation with other oils. Blue dot: C<sub>19</sub> tricyclic; green dot: C<sub>24</sub> tetracyclic).

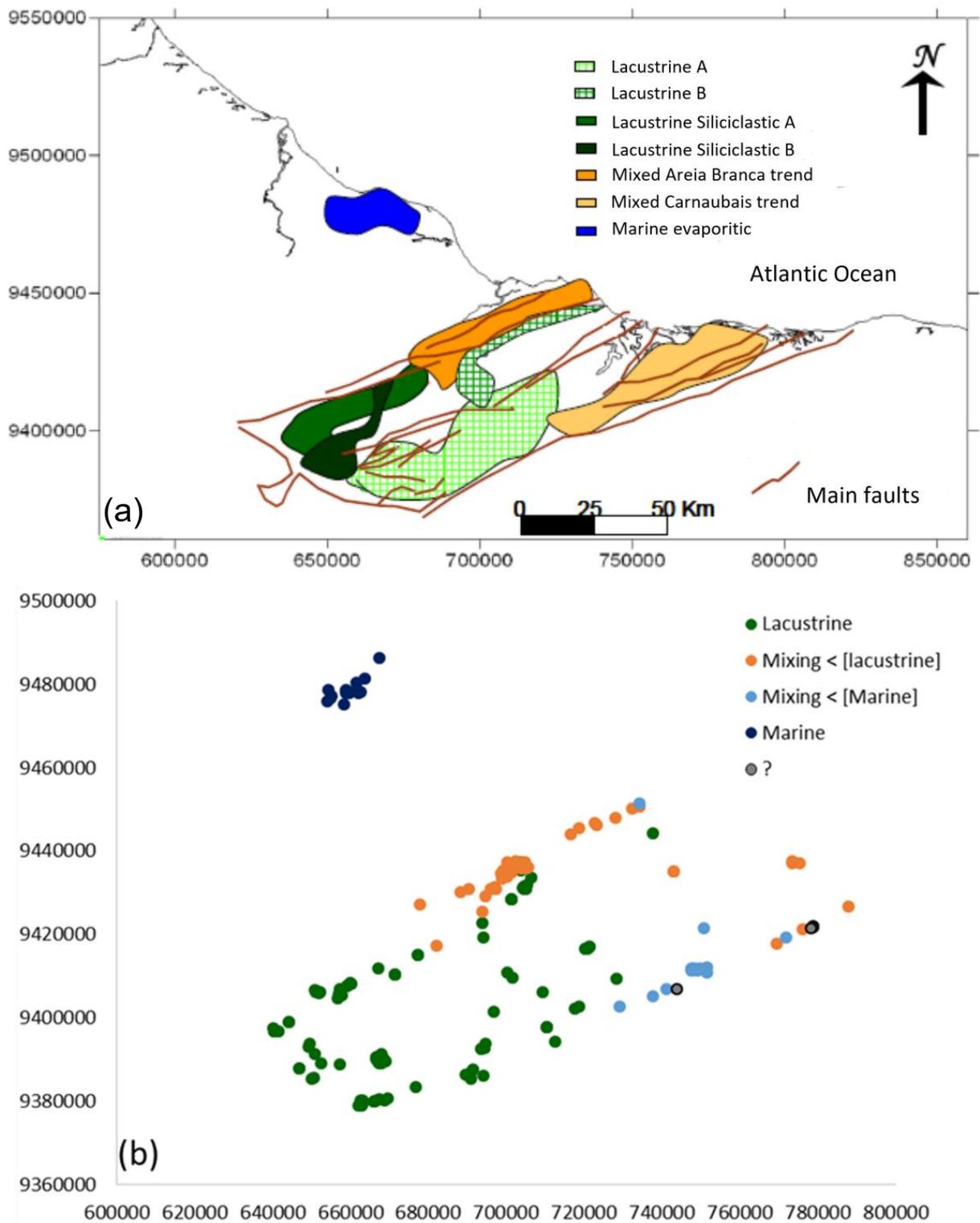
It seems that the Cexis area was filled with three different oil compositions, pointing out a complex migration and filling history that should be evaluated. Additionally, the classical parameters used to estimate maturation as  $T_s/(T_s + T_m)$ ,  $\alpha\beta\beta/(\alpha\beta\beta + \alpha\alpha\alpha)$ , and  $20S/(20S + 20R)$ , do not appear to be useful due to bitumen contamination that occurred during the migration or reservoir overprint (Figure 4—anomalous  $\alpha\alpha\alpha$  R C<sub>29</sub> sterane for oils) as pointed out by Hughes and Dzou [39].

### 3.3. Exploratory Implications

The origin, maturation, and preservation diagnoses are based on parameters of known empirical significance obtained from the petroleum biomarker distributions. These analyses allow genetic correlations between different oil samples and the consequent recognition of general migration patterns, thus reducing the exploratory risks.

By quantifying the degree of correlation between the variables, the PCC method helped the authors of the previously mentioned works understand how the different analyzed parameters interact with each other and highlight the implication of these results for an integrated petroleum system analysis.

In this work, we obtained an excellent correlation between the previously published classes for the Potiguar and Recôncavo oil samples and the PCC classification. The maps of Figure 7 compare the classification obtained using clustering and machine learning techniques, as proposed by Morais [14], with the combinations between PCA and PCC used in this study. The technique also provided an alternative to refine classes into subclasses when only one major class is present, for example, in the Recôncavo case study.



**Figure 7.** (a) Map showing the good correlation between the classes proposed by Morais [14] using AI methods. (b) The classes resulting from the classification using PCA coupled with PCC (this study). Observe the good correlation between the previous and current classes proposed in the previous study.

#### 4. Conclusions

The potential use and application of the Pearson correlation coefficient for petroleum system analysis have been poorly explored.

The combination of PCA and PCC proved efficient for the unsupervised discrimination of Pendência and Alagamar-Açu petroleum systems from the Potiguar Basin. The interaction between the different oil parameters (carbon isotope, biomarkers, saturate, aromatic, and polar compounds fractions) can be an important indicator for paleoenvironmental characterization and maturation analysis.

At the Recôncavo Basin, it was possible to identify three oil groups (different degrees of maturation) and three oil samples that do not present any correlation, probably due to the set of sample characteristics (representing a specific depocenter, for example). The PCC was very efficient for the unsupervised qualitative maturation evaluation of oil samples generated by Gomo and Tauá members.

These methodologies can be applied to analyze large amounts of oil samples, using simple software and spending less time than the classical approach. As said before, the correlation coefficients between geochemical parameters are very useful in characterizing petroleum systems.

**Supplementary Materials:** The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/geosciences13090282/s1>, Table S1 (Correlation matrix (pearson) of Potiguar basin data) and Table S2 (Correlation matrix (pearson) of Recôncavo basin data).

**Author Contributions:** É.T.M.: wrote the paper, conceived and designed the statistical experiments, and analyzed the data; G.A.B.: wrote the paper, conceived and designed the statistical experiments, statistical experiments, and analyzed the data; I.V.A.F.S.: wrote the paper, conceived and designed the statistical experiments, and analyzed the data; F.G.L.: wrote the paper, conceived and designed the statistical experiments, statistical experiments; J.V.P.G.: wrote the paper and analyzed the data; A.L.D.S.: wrote the paper and analyzed the data. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The correlation matrices are available as Supplementary Materials. The geochemical data used as input for statistical calculations are available in Morais (2007).

**Acknowledgments:** The authors are thankful to Petrobras for the opportunity to publish this paper.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Christensen, J.H.; Tomasi, G.; Hansen, A. Chemical fingerprinting of petroleum biomarkers using time warping and PCA. *Environ. Sci. Technol.* **2005**, *39*, 255–260. [[CrossRef](#)] [[PubMed](#)]
2. Peters, K.E.; Walters, C.C.; Moldowan, J.M. *The Biomarker Guide: Biomarkers and Isotopes in Petroleum Systems and Earth History*; Cambridge University Press: Cambridge, UK, 2005; Volume 1. [[CrossRef](#)]
3. Schiefelbein, C.F.; Zumberge, J.E.; Cameron, N.C.; Brown, S.W. Geochemical comparison of crude oil along the South Atlantic margins. In *Petroleum Systems of South Atlantic Margins*; AAPG Memoir 73; Mello, M.R., Katz, B.J., Eds.; The American Association of Petroleum Geologists: Tulsa, OK, USA, 2000; pp. 15–26.
4. Mudelsee, M. Estimating Pearson's correlation coefficient with bootstrap confidence interval from serially dependent time series. *Math. Geol.* **2003**, *35*, 651–665. [[CrossRef](#)]
5. Landim, P.M.B.; Bandeira de Mello, C.S.; Serra, A.C.S.; Macedo, A.B. Metodologia Estatística Multivariada em Geoquímica. In *Prospecção Geoquímica—Depósitos Minerai s Metálicos, Não-Metálicos, Óleo e Gás*; Licht, O.A.B., Bandeira de Mello, C.S., Silva, C.R., Eds.; CPRM: Brasília, Brazil, 2007; pp. 619–660. ISBN 9788574990576.
6. Howarth, R.J. *Dictionary of Mathematical Geosciences with Historical Notes*; Springer: Berlin/Heidelberg, Germany, 2017; 893p. [[CrossRef](#)]
7. Skupio, R.; Barberes, G.A. Spectromic gamma radiation of shale cores applied to sweet spot discrimination in Eastern Pomerania, Poland. *Acta Geophys.* **2017**, *65*, 1219–1227. [[CrossRef](#)]

8. Wang, Y.P.; Luo, T.; Gao, Y.; Yu, S.; Wang, S.; Zou, Y.R. Pearson correlation analysis of factors controlling the high abundance of rearranged hopanes in crude oils from the Southwest Depression of the Tarim Basin, China. *Geochem. J.* **2020**, *54*, 105–115. [[CrossRef](#)]
9. Rasheed, M.A.; Lakshmi, M.; Srinivasa Rao, P.L.; Kalpana, M.S.; Patil, D.J.; Dayal, A.M. Evaluation of Petroleum Prospects using Geo-microbial prospecting method from Seabed Sediment Samples of Gulf of Mannar, Kerala-Konkan offshore Basin, India. *Int. J. Curr. Microbiol. App. Sci.* **2014**, *3*, 281–295.
10. El-Nady, M.M.; Ramadan, F.S.; Hammad, M.M.; Lotfy, N.M. Evaluation of organic matters, hydrocarbon potential and thermal maturity of source rocks based on geochemical and statistical methods: Case study of source rocks in Ras Gharib oilfield, central Gulf of Suez, Egypt. *Egypt. J. Pet.* **2015**, *24*, 203–211. [[CrossRef](#)]
11. El-Nady, M.M.; Lotfy, N.M. Multivariate geochemical and statistical methods applied to assessment of organic matter potentiality and its correlation with hydrocarbon maturity parameters (Case study: Safir-1x well, North Western Desert, Egypt). *Egypt. J. Pet.* **2016**, *25*, 555–563. [[CrossRef](#)]
12. Serra, A.C.S.; Rangel, M.D.; Bandeira de Mello, C.S.; Bedregal, R.P.; Evsukoff, A.G.; Espíndola, R.P. Application of artificial intelligence techniques for oil origin classification. In *Book of Abstracts 10th Alago Congress on Organic Geochemistry*; ALAGO: Barcelona, Spain, 2006; pp. 298–299.
13. Aguiar, H.G.M. Análise do Comportamento de Parâmetros Físicos em Função da Composição Global e Maturação de Óleos Lacustres da Bacia do Recôncavo. Master's Thesis, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2005.
14. Morais, E.T. Aplicações de Técnicas de Inteligência Artificial para Classificação Genética de Amostras de Óleo da Porção Terrestre, Bacia Potiguar, Brasil. Master's Thesis, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2007.
15. Szatmari, P.; Milani, E.; Lana, M.; Conceição, J.; Lobo, A. How South Atlantic rifting affects Brazilian oil reserves distribution. *Oil Gas J.* **1985**, *83*, 107–113.
16. Szatmari, P.; Françolin, J.B.L.; Zanotto, O.; Wolff, S. Evolução tectônica da margem equatorial brasileira. *Rev. Bras. Geociências* **1987**, *17*, 180–188. [[CrossRef](#)]
17. Caixeta, J.M.; Bueno, G.V.; Magnavita, L.P.; Feijó, F.J. Bacias do Recôncavo, Tucano e Jatobá. *Bol. Geociências Petrobras* **1994**, *8*, 163–172.
18. Santos, C.F.; Cupertino, J.A.; Braga, J.A.E. Síntese sobre a geologia das bacias do Recôncavo, Tucano e Jatobá. In *Origem e Evolução das Bacias Sedimentares*; Raja Gabaglia, G.P., Milani, E.J., Eds.; Petrobras: Rio de Janeiro, Brazil, 1990; pp. 235–266.
19. Gaglianone, P.C.; Trindade, L.A.F. Caracterização Geoquímica dos óleos da Bacia do Recôncavo. *Geochim. Bras.* **1988**, *2*, 15–39.
20. Braga, J.A.E.; Zabalaga, J.C.; Oliveira, J.J.; Aguiar, G.A.; Silva, O.B.; Mato, L.F.; Daniel, L.M.F.; Magnavita, L.P.; Bruhn, C.H.L. Recôncavo Basin, Brazil: A prolific intracontinental rift basin. In *Interior Rift Basins*; AAPG Memoir, 59; Figueiredo, A.M.F., Ed.; The American Association of Petroleum Geologists: Tulsa, OK, USA, 1994; pp. 157–203. [[CrossRef](#)]
21. Coutinho, L.F.C. Análise do balanço material do petróleo em uma região em fase de exploração madura—Bacia do Recôncavo, Brasil. Ph.D. Thesis, COPPE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil, 2008.
22. Magnavita, L.P.; Szatmari, P.; Cupertino, J.A.; Destro, N.; Roberts, D. The Recôncavo basin. In *Regional Geology and Tectonics. Phanerozoic Rift Systems and Sedimentary Basins*; Roberts, D.G., Bally, A.W., Eds.; Elsevier: Amsterdam, The Netherlands, 2012; p. 1200, ISBN 978-0444563569.
23. Matos, R.M.D. Tectonic evolution of the equatorial South Atlantic. In *Atlantic Rifts and Continental Margins*, AGU Geophysical Monograph; Mohriak, W.U., Talwani, M., Eds.; American Geophysical Union: Washington, DC, USA, 2000; Volume 115, pp. 331–354. [[CrossRef](#)]
24. Pessoa Neto, O.C.; Soares, U.M.; Silva, J.G.F.; Roesner, E.H.; Florencio, C.P.; Souza, C.A.V. Bacia Potiguar. *Bol. Geociências Petrobras* **2007**, *15*, 357–369.
25. Bertani, R.T.; Costa, I.G.; Matos, R.M.D. Evolução tectono-sedimentar, estilo estrutural e habitat do petróleo na Bacia Potiguar. In *Origem e Evolução de Bacias Sedimentares*, 2nd ed.; Raja Gabaglia, G.P., Milani, E.J., Eds.; Petrobras: Rio de Janeiro, Brazil, 1991; pp. 291–310.
26. Santos Neto, E.V.; Hayes, J.M. Use of hydrogen and carbon isotopes characterizing oils from the Potiguar Basin (onshore), Northeastern Brazil. *AAPG Bull.* **1999**, *83*, 496–518. [[CrossRef](#)]
27. Spigolon, A.L.D. Palinofácies e faciologia orgânica da Formação Alagamar (Aptiano da Bacia Potiguar, Brasil): Paleoambiente e maturação térmica. Master's Thesis, Instituto de Geociências, Universidade de Brasília, Brasília, Brazil, 2003.
28. Spigolon, A.L.D.; Santos Neto, E.V. Geoquímica Orgânica do Eocretáceo da Bacia Potiguar: Implicações paleoambientais e paleoclimáticas. In *Livro de Resumos do X Congresso Brasileiro de Geoquímica*; Sociedade Brasileira de Geoquímica: Porto de Galinhas, Brazil, 2005.
29. Santos Neto, E.V.; Mello, M.R.; Rodrigues, R. Caracterização geoquímica dos óleos da Bacia Potiguar. *An. Congr. Bras. Geol.* **1990**, *1*, 974–985.
30. Trindade, L.A.F.; Santos Neto, E.V.; Brassell, S.C. Petroleum migration and mixing in the Potiguar basin, Brazil. *AAPG Bull.* **1992**, *76*, 1903–1924.
31. Marcucci, M.; Sharma, S. Applied Multivariate Techniques. *Technometrics* **1997**, *39*, 100–101. [[CrossRef](#)]
32. Davis, J.C. *Statistics and Data Analysis in Geology*. John Wiley & Sons: Hoboken, NJ, USA, 1973; 550p, ISBN 978-0471172758.
33. Reimann, C.; Filzmoser, P.; Garrett, R.G.; Dutter, R. *Statistical Data Analysis Explained—Applied Environmental Statistics with R*; John Wiley & Sons: Hoboken, NJ, USA, 2008; 362p. [[CrossRef](#)]

34. Piedmont, R.L. Eigenvalues. In *Encyclopedia of Quality of Life and Well-Being Research*; Michalos, A.C., Ed.; Springer: Dordrecht, Germany, 2014. [[CrossRef](#)]
35. Hammer, Ø. *PAST: Paleontological Statistics. Version 3.20. Reference Manual*; University of Oslo: Oslo, Norway, 2018; 262p.
36. Hammer, Ø.; Harper, D.A.T.; Ryan, P.D. PAST: Paleontological Statistics Software Package for Education and Data Analysis. *Palaeontol. Electron.* **2001**, *4*, 9.
37. Mello, M.R.; Koutsoukos, E.A.M.; Mohriak, W.; Bacocoli, G. Selected petroleum systems in Brazil. In *The Petroleum System—From Source to Trap*; AAPG Memoir 60; Magoon, L.B., Down, W.G., Eds.; The American Association of Petroleum Geologists: Tulsa, OK, USA, 1994; pp. 499–512. [[CrossRef](#)]
38. Penteadó, H.L.B. Modélisation Compositionnelle 2D de la Genèse, Expulsion et Migration du Pétrole dans le Compartiment sud du Bassin de Recôncavo, Brésil. Ph.D. Thesis, De L'Université Pierre et Marie Curie (Paris VI), Paris, France, 1999.
39. Hughes, W.B.; Dzou, L.I.P. Reservoir overprinting of crude oils. *Org. Geochem.* **1995**, *23*, 905–914. [[CrossRef](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.