

CLUSTER ANALYSIS AS A TOOL FOR SAMPLING AND DECISION SUPPORT IN A GEOCHEMICAL INVESTIGATION OF POTIGUAR BASIN OILS, BRAZIL

Erica T. MORAIS¹ and Henrique L. B. PENTEADO¹

1. Petrobras/CENPES/PDEXP/GEOQ-PROMOB, Rua Horácio Macedo, 950, Cid. Universitária, Rio de Janeiro, CEP 21941-915, Brazil

The Potiguar Basin lies in the northeastern part of Brazil and covers an area of approximately 60,000 km², with its onshore portion occupying 40% of total area. The Potiguar Basin is a rift basin that evolved to a passive margin after the break-up of the African and South American continents in the Early Cretaceous. During more than 30 years of exploration activities in the basin, various studies were performed aiming at the identification of different oil families and their spatial relationship.

Oil correlation studies in the onshore part of the Potiguar Basin based on gas chromatography (GC), GC coupled to mass spectrometry (GC-MS) and carbon isotopic data allowed the identification of three basic oil types: oils generated in the continental sequence (Neocomian and Aptian lacustrine freshwater shales), oils generated in the transitional sequence (Aptian marine evaporitic shales and marls), and oils derived from the mixture of the two previous types (Trindade *et al.*, 1992; Mello *et al.*, 1993).

The developments in analytical methods have established Organic Geochemistry as one of the most effective tools to characterize petroleum in terms of origin, thermal evolution and biodegradation. These developments allowed an increasing number of organic compounds to be identified and quantified. This capability, in turn, has allowed scientists to attack ever more complex problems, but the larger number of samples that must be analysed and the number of constituents that can be measured per sample give rise to data sets of enormous size and complexity. The main objective of this study is to present one methodology of sampling in which the data set recovered from a larger one is representative of distinct groups of oils based on geochemical parameters throughout the basin. The sampling of a representative subset will allow statistical analyses to be performed more effectively in a relatively smaller number of samples. This subset of samples will be used in the future for further classification techniques to investigate new grouping patterns.

The first step of the method consists in running a cluster analysis on the original data set. Cluster analysis is one of the multivariate statistical methods that measure the similarity between objects. This technique determines the structural characteristics of a data set by dividing the data into related groups. Cluster analysis is an exploratory data analysis procedure, hence it is usually applied to data sets for which there is no a priori knowledge

concerning the class membership of the samples. The steps followed in this methodology are presented in figure 1. It was chosen for decision support because samples of the Potiguar Basin oil types are not well distributed in geographic and frequency terms, thus preventing the use of random sampling techniques. Furthermore, the a priori recognized classes (lacustrine, marine and mixture) might have subdivisions which were not previously identified and must be considered for a proper representative sampling. The groups recognized by cluster analysis using GC, GC-MS and carbon isotopic data were plotted on a map to allow for a visualization of the area coverage of each group, including geographic superposition of oil groups. Random sampling was then performed within each group to obtain a smaller data set on which more sophisticated statistical methods will be run for sample classification. At the end, the consistency of the grouping from the cluster analysis with that of the previously recognized oil types was checked.

The use of this methodology allowed a successful representative sampling of different types of oils. This technique can be advantageous in planning future strategy and decision support in oil studies involving a large number of samples and geochemical parameters.

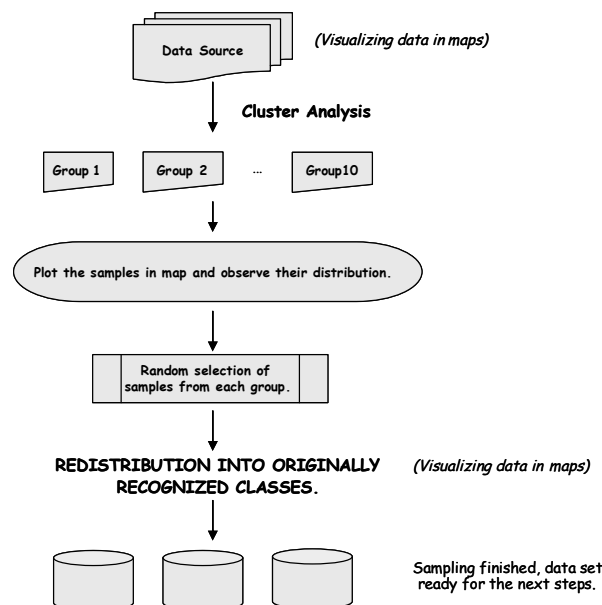


Figure 1. Flow diagram of steps followed by this methodology.

REFERENCES

- Trindade L.A.F., Brassell S.C. and Santos Neto E.V (1992) Petroleum migration and mixing in the Potiguar Basin, Brazil. *AAPG Bulletin*, **76** (12), 1903-1924.
- Mello M.R, Koutsoukos E.A.M, Santos Neto E.V. and Telles Jr. A.C.S. (1993) Geochemical and micropaleontological characterization of lacustrine and marine hypersaline environments from Brazilian sedimentary basins. In: Katz B.J & Pratt L.M. eds. Source rocks in sequence stratigraphic framework. *AAPG Studies in Geology*, **37**, 17-34.