An Application of Clustering Analysis over 3D Seismic Data for Direct Hydrocarbon Indicators Identification

Matheus R. S. Barbosa[1,2] and Alexsandro G. Cerqueira[1,2], [1]INCT-GP/UFBA, [2]GAIA/UFBA

Abstract

Understanding the tectonostratigraphic evolution of sedimentary basins, the study of depositional environments, and the characterization of direct hydrocarbon indicators (DHI) anomalies are only possible through interpreters' performance on geophysical data. In 3D seismic surveys, interpreting a large amount of data without intelligent algorithms is a real challenge and has become too slow or infeasible. Hence, the use of clustering analysis algorithms on 3D seismic data increases recurrently and has become even more necessary to offer a faster and more reliable interpretation. In this sense, this paper presents an application of the Gaussian Mixture Model (GMM) algorithm that aims to identify DHI anomalies and generate its probability volume from seismic attributes cubes. The study was carried out in the open seismic dataset F3 Block located at the offshore region of the North Sea in the Netherlands. Although it was possible to use the seismic cube in its entirety, in this approach, the cluster analysis is performed in zones of interest limited by two 3D seismic horizons, using the following seismic attributes as input data: RMS amplitude, reflection strength, sweetness, coherence, acoustic impedance, and spectral decomposition (25 and 45 Hz). Once null data and outliers have been removed, the Principal Component Analysis (PCA) algorithm performs a linear transformation that projects the multidimensional dataset over the most meaningful orthogonal basis to re-express it, ensuring an optimal representation of the original data through a smaller number of variables. The four principal components explained the variability of 92.8% of the original information. The intra-cluster variance, which measures the grouping quality based on the distance between cluster members and their centroids, alongside the elbow method, defined the metric to estimate the optimal number of clusters that best models data. The Elbow Method suggested three as the optimal cluster number for the GMM algorithm. Although the validation criterion is a helpful technique for selecting the value of K, geological characteristics should also be considered. Therefore, our analysis demonstrated that the number of groups equal to four was able to isolate the DHI anomaly and showed a high correlation to the geomorphological setup. The Gaussian Mixture Model algorithm accurately delimited at least five promising regions as direct hydrocarbon indicators anomalies. Besides that, it was possible to estimate a probability cube showing the responsibility of each sample belonging to the cluster associated with DHI anomalies, revealing the zones that were classified already as anomalies linked with hydrocarbons and the domain with an intermediary to low probability to be a DHI. In addition, the algorithm also highlighted some areas with high probabilities of DHI occurrence close to the faulted areas. Furthermore, the interpreter's knowledge combined with petrophysical information and lithological observation derived from well-logs played a part in the interpretation of the Hydrocarbon Indicator anomaly and clusters unrelated to the DHI anomaly.