Submission code: L69MQP7V6A

See this and other abstracts on our website: https://home.sbgf.org.br/Pages/resumos.php

Random Forest for Seismic Facies Segmentation of the Brazilian Pre-Salt strata: Reducing Labeling Effort with High Accuracy

MARIANA MEIRELES LEITE, Fernanda Costa-Moura, Jose Geraldo Fernandes, Luiz Facury de Souza, Humberto Reias (HR CONSULTING LTDA.), Mauricio Marengoni

Random Forest for Seismic Facies Segmentation of the Brazilian Pre-Salt strata: Reducing Labeling Effort with High Accuracy

Introduction

Subsurface geological investigation is primarily conducted using seismic reflection methods. However, seismic exploration in ultra-deep waters, such as the Mesozoic deposits of the Brazilian pre-salt, is still challenging due to its complex geometries and high acoustic impedance contrasts. Although recent studies have succeeded on applying machine learning techniques to improve and automatize the seismic interpretation at great depths, these approaches often demand large volumes of labeled data for training.

Method and Theory

In this work we propose a qualitative and quantitative segmentation of the pre-salt carbonate reservoirs of the Búzios Field (Santos basin) using multiple seismic attributes and a limited number of labeled horizons. Our goal is to automate this labor-intensive process through the application of a Random Forest algorithm, which is a machine learning method based on an ensemble of decision trees. It combines the outputs of multiple trees to improve classification or regression accuracy, is robust for overfitting and performs well even when using noisy or incomplete data. A single crossline section was used for training and as input to the algorithm we incorporated the top-to-the pre-salt strata (salt base) and basement surfaces, a velocity model, and the amplitude, envelope, phase, frequency, sweetness, semblance, Marfurt coherence, Sobel filter, RMS amplitude, and TECVA seismic attributes. The trained model was then tested on five additional crosslines located at a maximum of 1100 crosslines interval (20, 100, 200, 500, 1100) away from the training section, and also one inline.

Results and Conclusions

Evaluation metrics included the Jaccard index, F1-score, and accuracy. The Jaccard value for the training crossline reached 0.904, while test crosslines yielded values between 0.90 and 0.86. Seismic quality, rather than spatial distance, was identified as the primary factor influencing performance. The accuracy was 0.949 in the training section and ranged from 0.908 to 0.871, also showing no clear correlation with spatial distance from the training line. The results were also consistent in the inline with no additional labeling, yielding 0.808 and 0.894 for the Jaccard and accuracy scores, respectively. A preliminary analysis indicates that the Random Forest algorithm tends to produce mismatches in areas with low data resolution, whether due to poor signal quality or geological complexity, the same regions where an interpreter is also more prone to error. Random Forest proved to be a robust tool for seismic interpretation in complex settings, significantly reducing the manual effort required for labeling while delivering results comparable to traditional interpretations and other AI methods. However, further tests are needed to refine the methodology and improve results, particularly in low-quality seismic zones or even in other pre -salt fields.