NEW METHOD OF SEISMOFACIES IDENTIFICATION USING MACHINE LEARNING OVER SEISMIC ATTRIBUTES APPLIED TO SALT STRATIFICATIONS

Flavio Costa de Mesquita [ID] [1]*, Marco Antonio Cetale Santos [ID] [1],
Alexandre Rodrigo Maul [ID] [1,2] and Alex Laier Bordignon [ID] [3]

ABSTRACT. Salt Evaporites may be grouped into three major seismic facies: Halite, High Velocity Salts (formed by anhydrite and gypsum) and Low Velocity Salts (formed mainly by carnallite, sylvite and tachyhydrite). However, the tachyhydrite mineral presents the biggest technical and financial concerns as it causes fluid loss during the salt drilling operation as observed in the Santos Basin, Brazilian Offshore, due to its high solubility. The correct facies identification allows adjusting the drilling parameters accordingly to the estimated salt type, minimizing operational problems and economic losses. In this article, we propose the use of selected seismic attributes from 2D sections combined with some of the most recent machine learning algorithms such as Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), to identify stratifications within the salt layer and their corresponding salt type.

Keywords: machine learning, salt drilling, risk reduction, seismic attributes.

RESUMO. Evaporitos salinos podem ser agrupados em três fácies sísmicas principais: Halita, Sais de Alta Velocidade (formados por anidrita e gipsita) e Sais de Baixa Velocidade (formados principalmente por carnalita, silvita e taquidrita). No entanto, o mineral taquidrita apresenta as maiores preocupações técnicas e financeiras, pois causa perda de fluidos durante a operação de perfuração do sal, conforme observado na Bacia de Santos, offshore brasileiro, devido à sua alta solubilidade. A correta identificação de fácies permite ajustar os parâmetros de perfuração de acordo com o tipo de sal estimado, minimizando problemas operacionais e perdas econômicas. Neste artigo, propomos o uso de atributos sísmicos selecionados de seções 2D combinados com alguns dos mais recentes algoritmos de aprendizado de máquina, como Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) e Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN), para identificar estratificações dentro da camada de sal e o tipo de sal correspondente.

Palavras-chave: aprendizado de máquina, perfuração no sal, redução de risco, atributos sísmicos.

Corresponding author: Flavio Costa de Mesquita

[1] Universidade Federal Fluminense – UFF, Seismic Inversion and Imaging Group, Av. Gen. Milton Tavares de Souza s/nº Gragoatá - Campus da Praia Vermelha 24220-900 Niterói, RJ, Brazil – E-mails: flaviomesquita@id.uff.br, marcocetale@id.uff.br
[2] Petrobras - Reservoir Geophysics, Av. República do Chile, 330 / 9º andar 20031170, Rio de Janeiro, RJ, Brazil – E-mail: alexandre.maul@petrobras.com.br
[3] Universidade Federal Fluminense - UFF - Instituto de Matemática e Estatística, Rua Professor Marcos Waldemar de Freitas Reis, s/n, Blocos G e H, 24210-201 Campus do Gragoatá, São Domingos, Niterói, RJ, Brazil – E-mail: alexb@id.uff.br

INTRODUCTION

The main problems associated with drilling wells for oil and gas industry are largely related to the disturbances of balance between rock stresses and pore pressure outside the hole, the wellbore mud pressure and chemical composition that fill the well. Whenever this balance is disturbed the wellbore problems occur.

Drilling operation activities in salt zones have gained importance in Brazil due to the discovery in 2005 of large oil and gas reserves in the pre-salt province (ANP, 2013). High operating costs, associated with deep water drilling, place additional emphasis on reducing required drilling time, without quality losses.

One of the main causes of drilling problems in the Santos Basin, Brazilian Offshore, is the fluid circulation loss due to mineral absorption (Lomba et al., 2013). The continuity loss of whole mud to a formation probably is the most common and undoubting the most expensive drilling well problem. Depending on its severity, lost circulation can lead to:

- Increasing in costs for drilling mud and associated materials;
- Formation damage and decreasing of productivity;
- Wellbore fluid level drops, increasing the chance of stuck pipe, borehole instability, and kicks;
- Loss of formation evaluation data, since the information normally obtained from drilled cuttings and mud returns may be unobtainable.

One of the methods to avoid lost circulation is the correct use of mud types for each section of the well (Lomba et al., 2013):

"The drilling of evaporites with high mobility with synthetic based muds may result in wellbore collapse, high torques, difficult reaming, stuck pipe, deviations, casing collapse and, eventually, loss of the well. In some cases, stress concentration, after drilling the salt, may cause stuck bit, mainly during the connections, demanding the injection of fresh water pills for its liberation. In some wells, the frequent use of those pills results in such an irregular section as if it had been drilled with a water based fluid, thus rendering the mentioned problems more critical."

The objective of this article is to present the results after modeling the salt section in Santos Basin to identify its different types of salt and stratifications, using a selected and innovative machine learning approach, with a seismic attribute combination. This model may support the decision in changing the drilling parameters before entering a Low Velocity Salt portion that has high solubility, therefore preventing problems related with lost circulation.

Seismic attributes are extracted from the seismic data and were introduced as a part of the seismic interpretation in the early seventies. Since that time, many new seismic attributes derived from the amplitude response, and the referred application to lead the interpretation of geologic structure, stratigraphy, and rock/pore fluid properties were reported (Chopra & Marfurt, 2005). Commonly, the combination of attributes or a multi-attribute analysis is carried out to gauge more overall information than what is possible with only one single attribute. A crucial problem in a multi-attribute analysis is the selection and the number of seismic attributes to be considered. Kalkomey (1997) shows that the probability of observing a spurious correlation increases as the number of control points decreases and also as the number of seismic attributes being used increases.

Artificial neural networks (ANN) are capable of capturing linear and non-linear relationships among attributes (Hagan et al., 2002). The choice between supervised and unsupervised ANNs in geoscience is dependent on the analysis of each case (Sayago et al., 2012; Bhattacharya et al., 2016).

Machine Learning has been used in the seismic scale to predict geological structures, stratigraphy, rock and fluid properties, usually through seismic interpretation and inversion

(Iturrarán-Viveros et al., 2021). To achieve it, fully convolutional deep networks have been used in the area of fault interpretation (Long et al., 2015), and it has also been used 3D convolutional neural networks (Waldeland et al., 2018), and deep encoder-decoder networks for stratigraphic interpretation (Badrinarayanan et al., 2015). These mentioned techniques classify a 3D Post-Stack data set based on 3D sub-cubes or 2D sections, and they require a relatively low number of labels (classes). Interpretation in seismic images has long used texture attributes to better identify and highlight areas of interest. These attributes can be seen as feature maps in seismic texture. In the salt case, we note that the texture is quite chaotic, where the surrounding seismic is more "striped". However, the Santos Basin is an example of an exception where its salt layers are highly stratified.

The Santos Basin General Settings

The domain of the Santos Basin field-reservoirs is located in the central portion of the basin, approximately 180 km off the coast of Rio de Janeiro City at a water depth, in average, around 1,900 m. The reservoirs are situated between 5,000 and 6,000 m below sea level and under a layer of salt, named as Ariri Formation, which can ranges from a few hundred meters to over 2,000 m (Mohriak et al., 2012).

There are some points of attention regarding the application of Machine Learning to predict the geological structures of the Santos Basin. They are based on the available dataset that is composed of wells containing samples of the three types of salt we are interested to identify: LVS, Halite and HVS. It is worth highlighting that there are different types of saline evaporite minerals within the salt section in the Santos and Campos basins, and the research carried out in log analysis shows that not all these types of minerals are seismically detectable by the amplitude seismic attribute (Gobatto et al., 2016). Moreover, the salt section is composed of many more different types of evaporates than these three simplified classes. The

variations in this portion and the related thicknesses depend on the complexity of the hydrocarbon field and the corresponding saline structuring. Oliveira et al. (2015) observed an inverse correlation between the thickness of the salt section and the salt velocity behaviors.

Given the problem we stated before, we propose the use of a multi-attribute technique to handle several available attributes to better identify the salt mineral occurrences.

MATERIALS AND METHODS

In this section, we individually explain the processing steps that we use to model the types of salts in this article. In order to simplify, we group the proposed workflow into three main steps (Fig. 1).

The solution proposed and described by Mesquita et al. (2019) is computationally demanding and it is based on non-intuitive parameters. The workflow here proposed is comparable to applying Self Organizing Maps (SOM) such as presented by Kohonen (2001), in which a non-supervised network performs a reduction of dimensionality and the data clustering. It results in a bi-dimensional map representing the input data that were grouped by the intrinsic similarity. However, our new proposed approach differs from SOM once it offers better control of each step. It is composed of two classification steps: the first one is unsupervised and the second one supervised, as demonstrated by Moqbel & Wang (2011). Therefore, our workflow is an attempt to improve the approach proposed by Moqbel & Wang (2011) by using in our case the raw seismic trace instead of working with pixels. That decision makes our algorithm faster and capable to return even better results. For this, we use a novel dimensionality reduction technique, UMAP, followed by an unsupervised "state of the art" clusterization algorithm, HDBSCAN, to model a set of seismic facies we are interested in. Once we have the calibrated labels, we can perform the classification of a Post-Stack Depth Migration (PSDM) data.
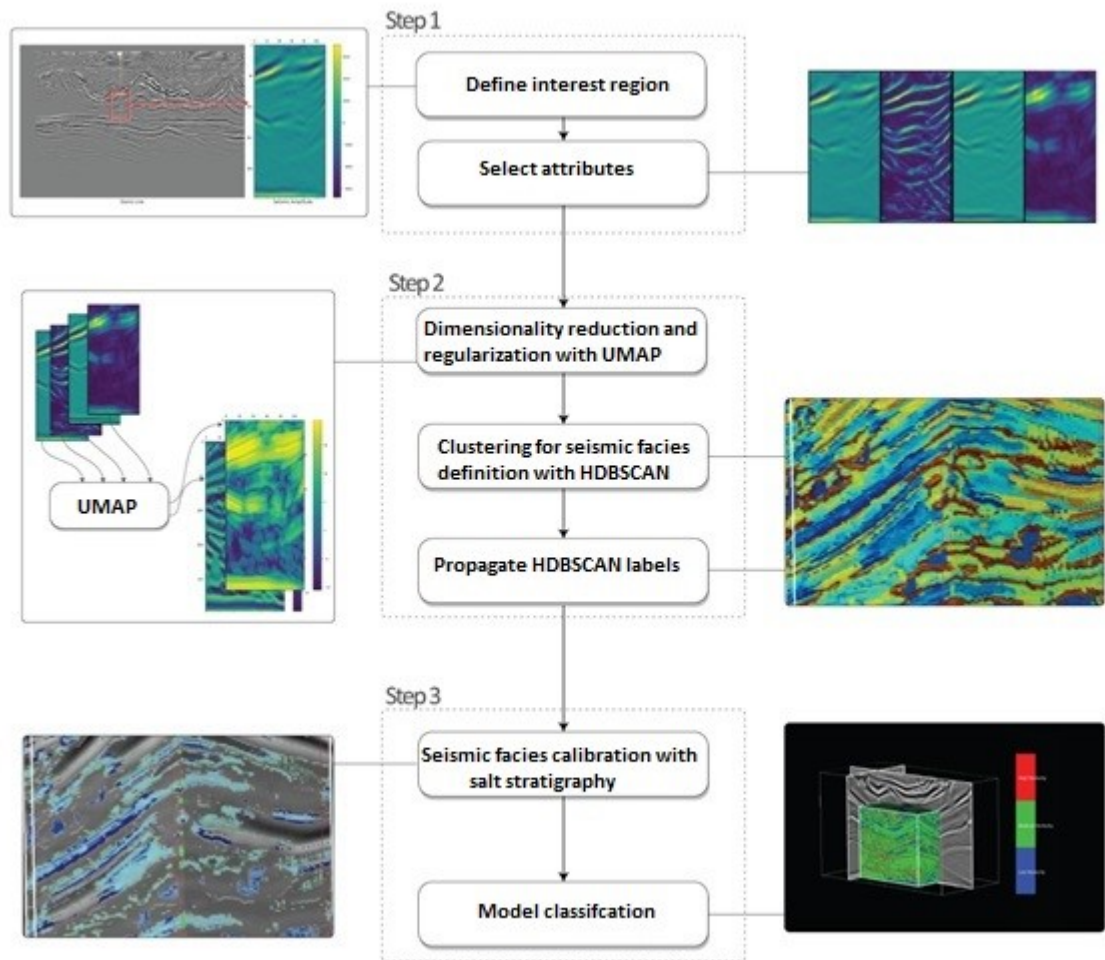
Figure 1 - Workflow for seismic facies classification using the approach we explore in this article.

WORKFLOW

Step 1 in Figure 1:

- Defining the region of interest

The input data were extracted from a region of interest defined by the top and the base of the salt (Fig. 2). In this region, we have information about the salt velocity in one well, and we select the seismic line crossing this well (Fig. 3, left part). In this seismic line, we select a rectangular window containing 120 traces, and each trace presents 250 samples, with the well in the middle as depicted in Figure 2.

For this region of interest, we select four seismic attributes, as shown in Figure 3. We end this step with four matrices containing all the data we use as input for the next step. We combine these matrices as a tensor, called I, and we refer it to a single attribute value as $I[h, w, c]$ where h is the row of the matrix, w is the column and $c \in \{1,2,3,4\}$ that represents the attribute.

In addition, we denote $I[h, w]$ as a four dimensional vector composed by the attribute samples, that is denoted in Equation 1:

$$I[h, w] =$$
$$= (\, I[h, w, 1], I[h, w, 2], I[h, w, 3], I[h, w, 4]) \qquad (1)$$

The input data consist of pieces of volumetric attributes extracted from a PSDM seismic type. From these volumetric attributes, we extracted inlines on the position of each well that shows occurrences of these salts, in special the LVS (Fig. 4).
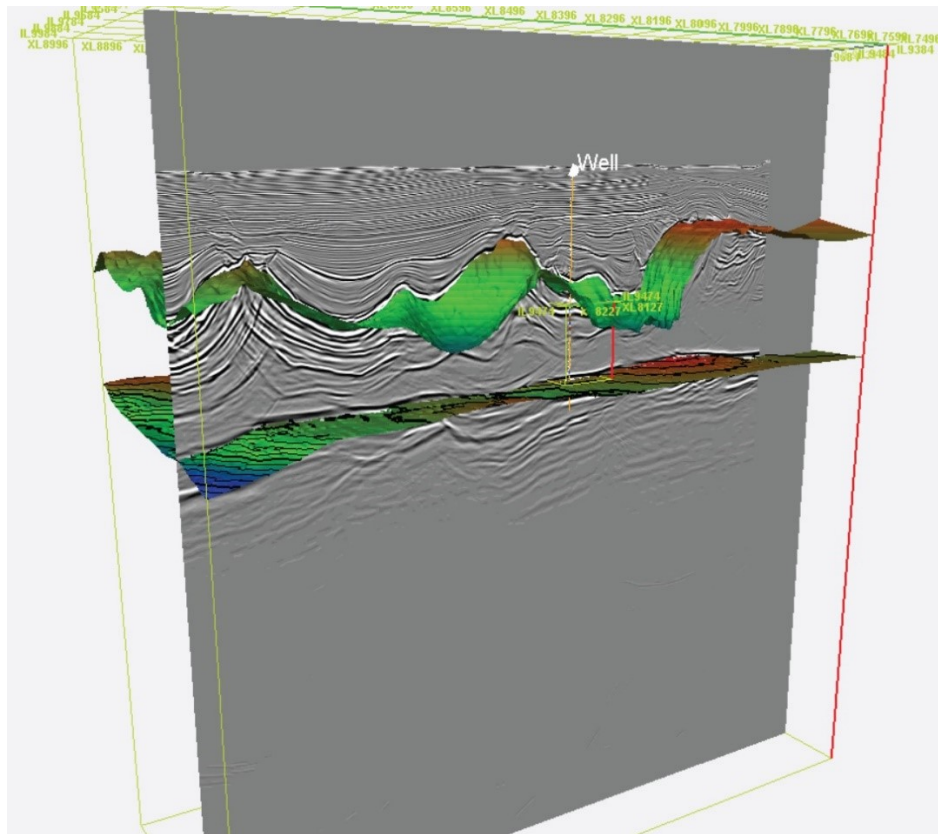
Figure 2 - Seismic line from the volume seismic with the mapped salt top and base horizons. The red, green and blue colors are associated with the surface depth positions – the red color represents the shallow portions, the green the intermediate portions and blue the deepest portion, for each of the horizons. In the middle, one of the wells used is depicted as the yellow line.
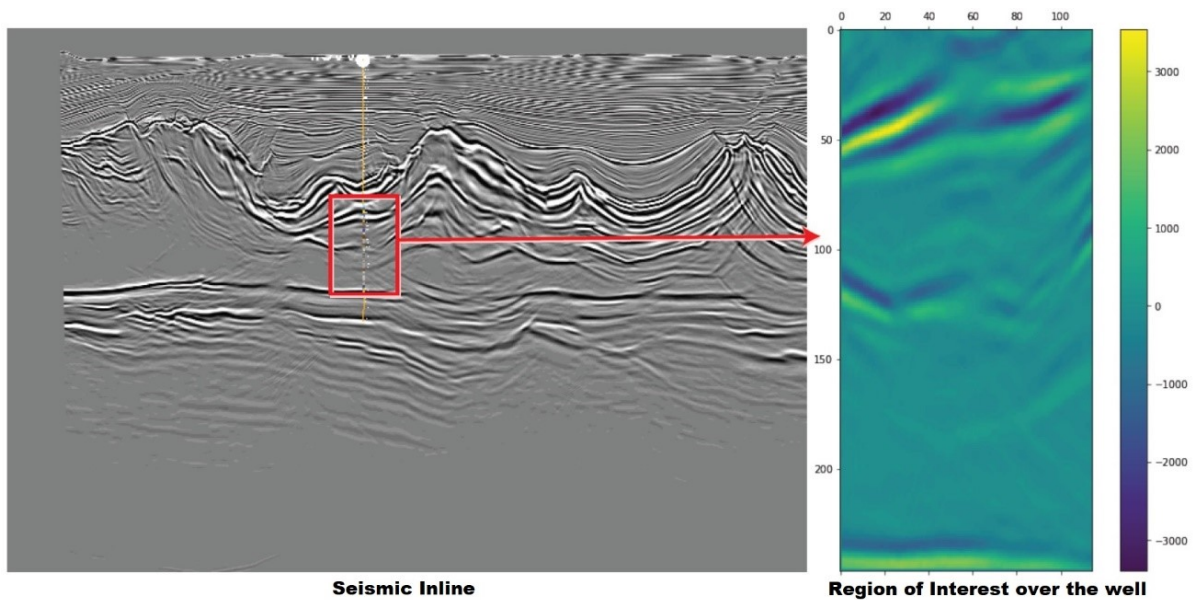
Figure 3 - Data slice from top to bottom of the salt layer on a 50 trace radius from the well location on the inline.
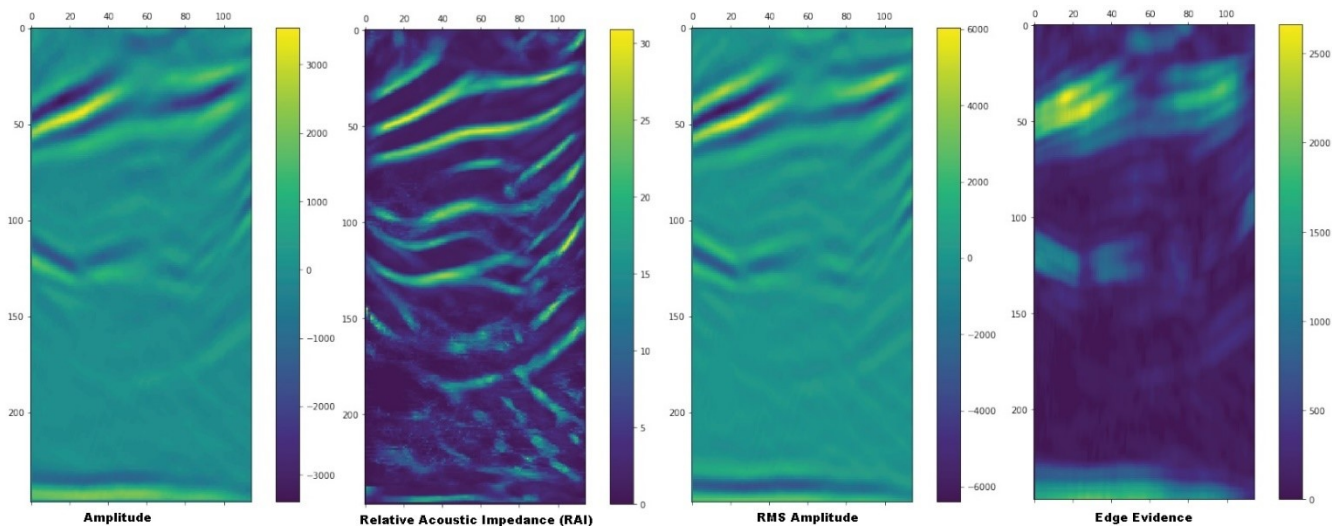
Figure 4 - Four attributes used in this seismic facies classification.

- **Choosing the Seismic Attributes**

Based on previous knowledge and experiences we select eight attributes that best highlighted the LVS occurrences. In order to know the reasonable amount of components to use, we perform a variance analysis on the data. According to the graph presented in Figure 5, four components are responsible for 97% of the dataset variance. Removing the remaining components is important to avoid redundancy and possible model overfittings, such as components that had high correlation or that do not highlight the salt reflections we are interested in when matching to the well log as proposed by Hampson et al. (2001).

In a salt evaporitic deposition sequence, to distinguish different salt minerals using only the amplitude attribute is difficult and ambiguous due to the lack of seismic resolution once some salt layers are thinner than the ones the method is capable to solve. Both anhydrite and carnallite show a seismic response from positive to negative peaks and vice-versa respectively, although having a phase shift of 180 degrees. For instance, the lack of resolution may indicate a tuning effect when a lateral lobe of a carnallite

response overprints the effective response (main lobe) of an anhydrite response (Fig. 6).

Due to this mentioned seismic behavior / weakness, the best solution to define thin salt layers is the seismic inversion usage, which generates the absolute acoustic impedance. However, as it is not the focus of this work, we generate the Relative Acoustic Impedance (RAI) volumetric attribute instead of the absolute acoustic impedance and we use it as the main attribute to conduct our research.

Knowing that we only need four attributes, we tested some of the available ones within a commercial software, and we built a correlation matrix, which can be seen in Table 1. From it, we may assume that the best attributes to handle our machine learning approach are first derivative, Edge evidence, RMS Amplitude and the RAI itself. However, in comparison to the well log outputs, when we use the method as proposed by Hampson et al. (2001), we observe that the first derivative does not reflect the salt layers we are interested in. Although it has a good ranking (low correlation with RAI), we discharge it. However, as we need four seismic attributes to run our algorithm, we use the Amplitude attribute instead.
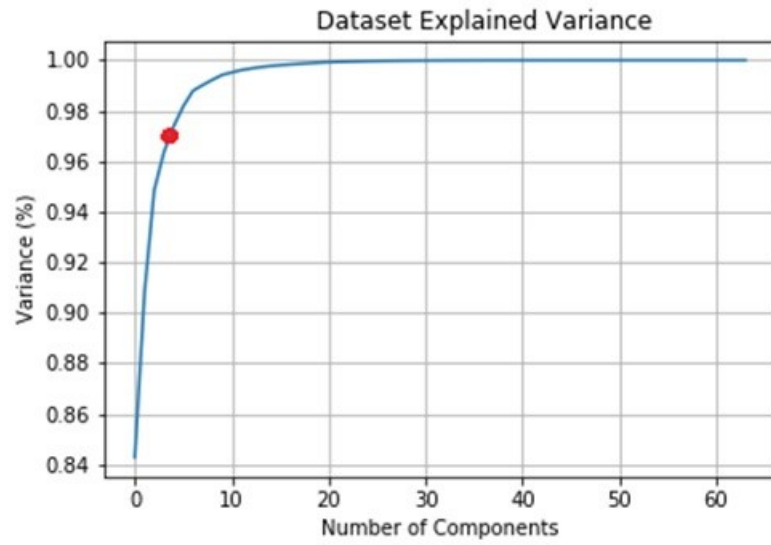
Figure 5 - Dataset explained variance chart. The first four components have 97% of the explained variance.
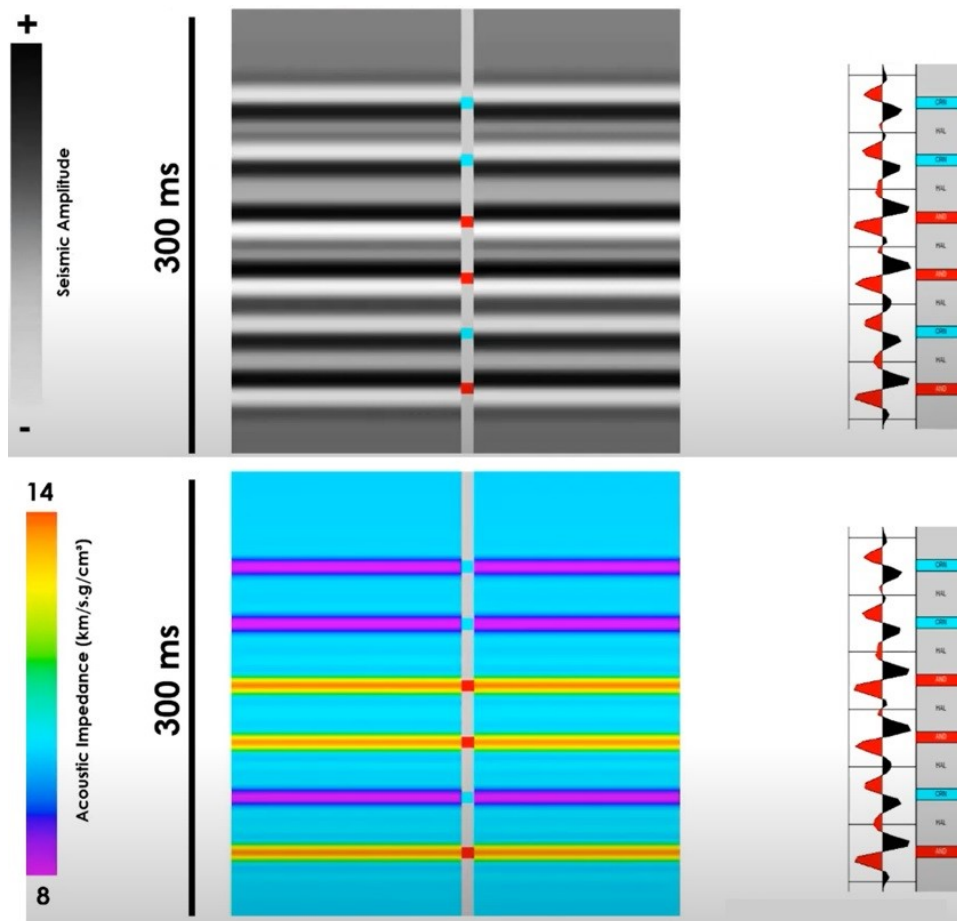
Figure 6 - Comparison of the seismic response of the amplitude attribute with the acoustic impedance attribute. Teixeira et al. (2020).

Table 1 - Correlation among the tested attributes.

	Amplitude	RMS Amplitude	Relative Acoustic Impedance	First Derivative	Instantaneous Frequency	TECVA	Edge Evidence	Amplitude Contrast
Amplitude	1.000	0.674	0.585	0.149	0.149	0.122	0.023	0.175
RMS Amplitude	0.674	1.000	0.584	0.118	0.197	0.080	0.026	0.173
Relative Acoustic Impedance	0.585	0.584	1.000	0.208	0.988	0.812	0.016	0.865
First Derivative	0.149	0.118	0.208	1.000	0.109	0.037	0.018	0.018
Instantaneous Frequency	0.149	0.197	0.968	0.109	1.000	0.991	0.056	0.004
TECVA	0.122	0.080	0.812	0.037	0.991	1.000	0.482	0.482
Edge Evidence	0.023	0.026	0.016	0.018	0.056	0.482	1.000	0.034
Amplitude Contrast	0.175	0.173	0.865	0.018	0.004	0.482	0.034	1.000

Next we describe few details about the chosen attributes:

- Amplitude: The seismic standard reflectivity amplitude is an attribute related to the physical properties of the subsurface as a function of the reflections at the different acoustic interfaces;

- Root Mean Square (RMS) Amplitude: It is a statistical measurement of the magnitude of variation in amplitude throughout a dataset. Generally, higher acoustic impedance variations are associated with variations within stacked lithology and result in higher RMS values. It is computed in a sliding tapered window of N samples as the square root of the sum of all the trace values *x* squared where *w* and *n* are the window values as presented in the Equation 2.

$$x_{rms} = \sqrt{\frac{1}{N} \sum_{n=1}^{N} w_n x_n^2} \qquad (2)$$

- Relative Acoustic Impedance (RAI): It is the relative product of density and seismic velocity, which varies among different rock layers and is commonly symbolized as Z. It is also considered as a stratigraphic method. This attribute shows apparent acoustic contrast. It indicates the sequence boundaries, the unconformity surfaces and the discontinuities. It can also support the quantification of porosity or fluid content in the reservoir. Besides, it helps to define the density contrast encountered at the interface between two distinct lithologies. It is computed by integrating the traces, then passing the result through a High-pass filter to reduce the potentially introduced low-frequency noise (Connolly, 1999). In our case, we use a frequency cutoff of 10 Hz because we note this is the dominant frequency of the wavelet inside the salt layer. It can be expressed as Equation 3 below:

$$\langle Z_n \rangle = \langle Z_0 \rangle exp\left(2 \sum_{j=0}^{n} R_j \Delta t \right) \qquad (3)$$

Where $\langle Z \rangle$ indicates the average impedance over a layer j, R is the reflectivity, j is the time sample varying from 0 to n, and t is time.

- Edge Evidence: It is a statistical edge enhancement method used to delineate fault and any geological body borders, such as salt bodies, within seismic data. It is a structural method. The algorithm is related to the Radon transform and Hough transform, but it uses an integral to detect edges within an image, and it is limited to a user-defined window. The Edge Evidence attribute works by searching locally in all directions for line segments where the values on the line differ significantly from the surrounding values (Aqrawi & Boe, 2011). This same method can be used for any arbitrary number of attributes and it can be used to assist in mapping other desirable lithologies.

Step 2 in Figure 1:

- Dimensionality reduction and regularization with UMAP

With the input data I from the previous step, we define a set of samples, where each sample $I[h,w]$ is a vector having four dimensions. Then, we transform these samples using the Uniform Manifold Approximation and Projection for Dimension Reduction (UMAP) (McInnes et al., 2020) and we cluster the transform samples applying the Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) (McInnes et al., 2017).

We construct the UMAP technique from a theoretical framework based on Riemannian geometry and algebraic topology. Dimensionality reduction seeks to produce a lower dimensional representation of higher dimensional data, preserving the relevant features. Dimensionality reduction algorithms tend to fall into two categories: those that seek to preserve the distance structures within the data and those that favor the preservation of local distances over global distances. The UMAP falls into the second described category.

The theoretical foundations for UMAP are largely based on manifold theory and topological data analysis. At a high level, UMAP uses local manifold approximations and patches together with their local fuzzy simplicial set representations to construct a topological representation of the high dimensional data. Given some low dimensional representation of the data, it uses a similar process to build an equivalent topological representation.

Thus, UMAP optimizes the layout of the data representation in the low dimensional space to minimize the cross-entropy between the two topological representations. The theoretical description of the algorithm works in terms of fuzzy simplicial sets. Computationally this is only tractable for the one skeleton, which can be described as a weighted graph. From a practical computational perspective this means that UMAP can ultimately be described in terms of construction and operations on weighted graphs. In particular, that statement situates UMAP in the class of k-neighbor based graph learning algorithms such as Laplacian Eigenmaps, Isomap and t-SNE (van Der Maaten & Hinton, 2008).

Therefore, UMAP is used mainly as a dimensionality reduction method, but its formulation also has a regularization effect. In this work, our main interest is in the regulatory effect. See Figure 7.

We can use the UMAP to transform groups of points that are very dense into groups with more uniform density points. The minimal distance between the points in the embedding can be controlled by the minimal distance hyperparameter. This has a positive effect because it facilitates the task of clustering algorithms that work towards expanding sample selection, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN) and HDBSCAN. Comparing the result of Figure 8 (b) to Figure 8 (c), we can see the effect of this regularization after employing the UMAP, once the seismic facies are better defined.
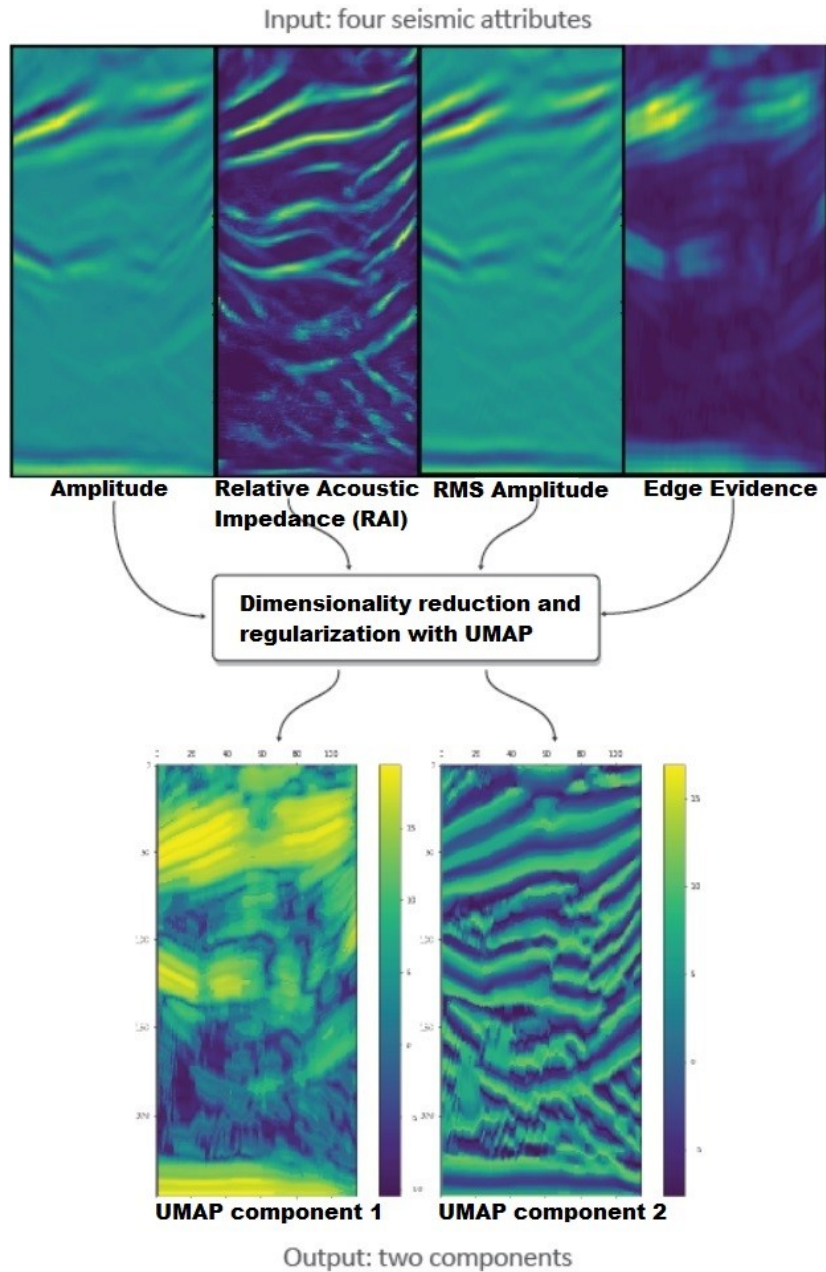
Figure 7 - Dimensionality reduction and regularization using UMAP technique. Four seismic attributes and the UMAP were used to create an embedded space with two components. The two new components are the inputs for the clustering algorithm.
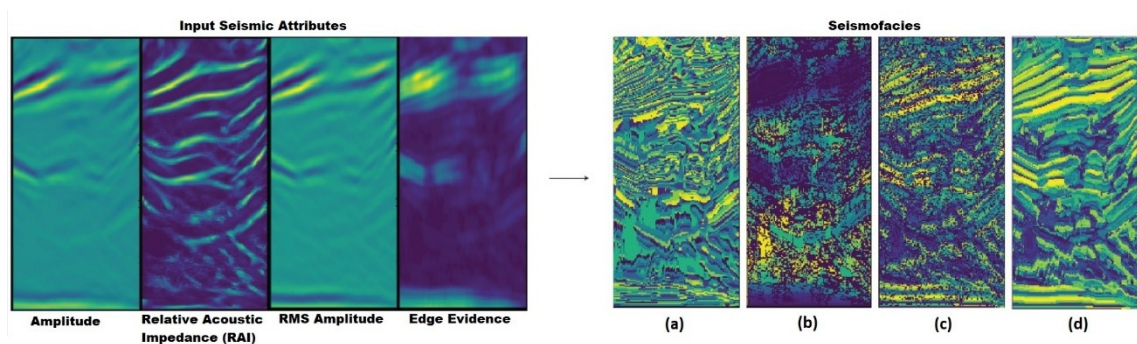
Figure 8 - Input seismic attributes and seismic facies found by different methods: (a) K-means, (b) HDBSCAN, (c) UMAP + HDBSCAN, (d) UMAP + Modified HDBSCAN.

The integrating procedure of UMAP with a hierarchical clustering algorithm provides the best evidence of existing local and global structures of data. Here we use HDBSCAN as presented by McInnes et al. (2017). It performs DBSCAN (Campello et al., 2013) over varying epsilon values, and it integrates the result to find a new clustering that gives the best stability over epsilon, which is the parameter that specifies how close points should be to each other to be considered a part of a cluster.

This allows HDBSCAN to find clusters of varying densities, rather than DBSCAN, and it is the most robust parameter for selection. In Figure 8 we compare the result of the proposed clustering method, UMAP + HDBSCAN (d) against the k-means algorithm (a) (MacQueen, 1967), with the original HDBSCAN (b) without the UMAP step and HDBSCAN with UMAP (c).

The UMAP has two main hyper-parameters: the number of neighbors that balances local versus global structure in the data, and the minimal distance that controls how tightly UMAP is allowed to pack the referred points together. We set the number of neighbors to 35 and minimal distance to 0.01, and we select these values based on a grid search and visual inspection of the facies distribution. In Figure 7 we depict the inputs and outputs generated by the UMAP algorithm with the selected hyper-parameters.

- **Clustering for seismic facies definition with HDBSCAN**

The HDBSCAN was used with the default parameters, except for the metric parameter. We chose the Mahalanobis metric, once this one is useful because it is scale invariant. In this way, it is more robust to the possible unscaled output of the UMAP transformation. We present the seismic facies obtained with this methodology in Figure 8 (c).

One warning to HDBSCAN usage is related to its formulation, once it tries to separate the clusters from the background data, causing misdetection of the points that are among the clusters, possibly making it classify them as noise. We can see this effect in Figure 8 (c), where many points were classified as belonging to the noise class (the darker blue points). To fix the problem, we modified the HDBSCAN algorithm. For each point classified as noise we look at for the nearest classified point, and we transfer its referred label in order to have the right one. We present the result of this modification in Figure 8 (d).

- **Propagate the HDBSCAN labels**

After the application of the proposed clustering method (UMAP + Modified HDBSCAN), we ended with a seismic facies map of our region of interest. Later, we want to expand this seismic facies classification to all the seismic data we are dealing with. To make this expansion properly we directly use the model, transforming the new samples applying the UMAP and using the condensed tree that was created by the HDBSCAN method to predict the new labels. The classification for new data points took approximately half a millisecond per sample in an Intel Core I5-9600KF CPU. This is little on the slow side, but the process can be directly done using parallel computing. See Figure 9 (c) where we propagate the seismic facies information to the seismic cube.

Step 3 in Figure 1:

- Seismic facies calibration with salt stratigraphy

In this step, we want to map the seismic facies from **Step 2** with the salt stratigraphy. This is a problem with incomplete information because the seismic is not perfect and contains several sources of uncertainty. Considering the example we present in Figure 10 (a), the well data show that there are LVSs (green points) in the middle of the picture. However, the seismic data are not capable to display any reflection over the well trace, but when the left and right parts of the well are observed, the referred reflections are present. There are several possible correct mappings from the seismic to the salt layers. In this sense, to create the mapping we use interpretation, which facilitates the process of creating two main sources of information as it is proposed: knowledge about the salt distribution and well logs.
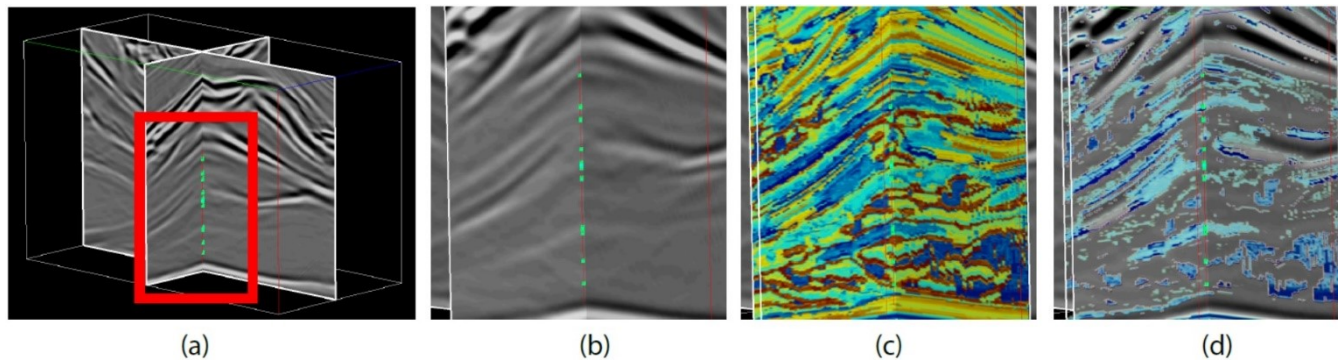
Figure 9 - Input seismic attributes and seismic facies found by different methods: (a) K-means, (b) HDBSCAN, (c) UMAP + HDBSCAN, (d) UMAP + Modified HDBSCAN.

Gao (2007) uses the method of calibrating the clusters using the control of the well, giving rise to what is called a "*posteriori* supervision". There the author uses wells along with his own knowledge of the depositional environment in which he was working on to calibrate the found clusters. Based on this procedure, we use the same process. Here the same wells that we are using as a reference serve to calibrate the seismic facies. For this, another processing step has to be performed before, which is the seismic-well tie, to place all the information in the same domain and perform the data calibration.

We want to map the seismic facies to three different stratigraphic groups: LVS, Halite (Background) and HVS, as presented in Table 2.

To create the maps we load the seismic facies obtained in **Step 2** into a commercial interpretation software, where we can visualize the well data and the seismic facies data, as shown in Figure 10. The green points over the well trace indicate the regions with the records of LVS occurrences, and we select the seismic facies associated with those regions. We present the resulting process of this selection in Figure 9 (d) and with more detail in Figure 10 (b). The facies selection is achieved in an easier way because we employ the HDBSCAN algorithm. Usual clusterization algorithms generate labels that do not have a specific order as shown in Figure 8 (a). Differently, the HDBSCAN generates ordered labels of seismic facies, and those labels are

integer numbers. This, in general words, means that facies with similar attribute characteristics have closer labels. We can observe the structure results in Figures 7(c) and (d). We use the same process to select the facies associated with HVSs and the remaining facies are associated with the Halite, our background facies. The final result is a 2D section obtained from the 3D cube, with the salt classification based on our premises (Figs. 11 and 12).
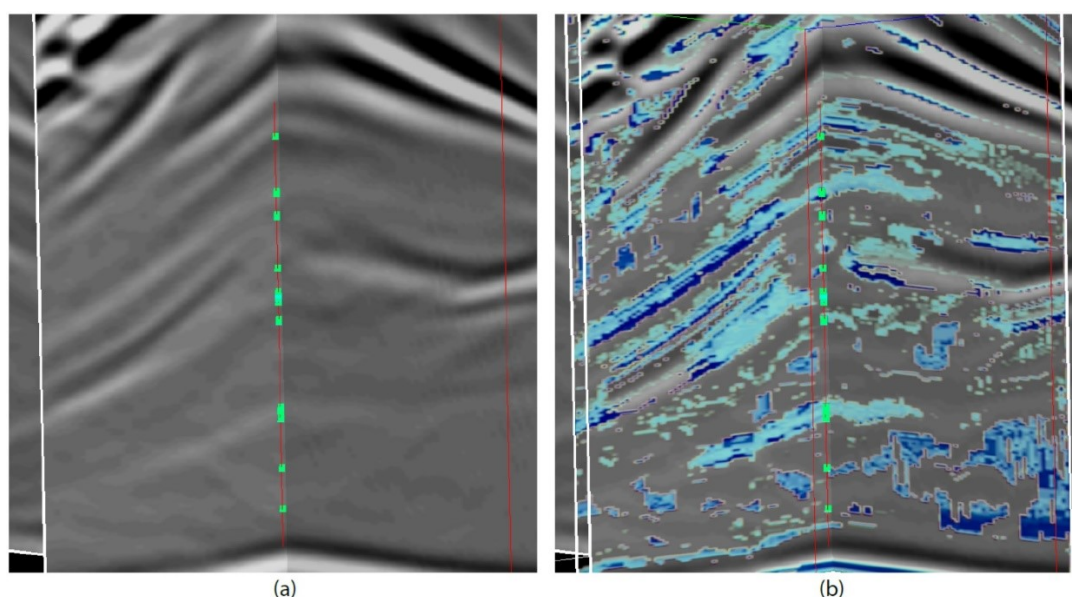
- **Model classification**

Once we classify the data, we can then assign values accordingly to our prior knowledge of the lithology and using the wells as reference. We use the theoretical values of the salt layers according to Table 2. It presents the mineral grouping as proposed by Maul et al. (2018), indicating the chemical formulas of each mineral as well as the related main acoustic properties.

RESULTS, ANALYSIS AND DISCUSSIONS

As we can see in Figure 11, and with more detail in Figure 12, the Amplitude inline was classified and values were assigned to each class that was found using our method and that we are interested. There, the green color represents the background lithology – Halite; the dark blue color represents the LVSs; and the yellow color represents the HVSs.

Table 2 - Stratigraphic groups; their mineral composition and respective properties; and average values, covering 182 drilled wells in Santos Basin, Offshore Brazil. Adapted from Maul et al. (2018).

Stratigraphic Group	Mineral	Composition	Density (g/cm³)	Interval Velocity (m/s)
LVS	Tachyhydrite	$CaMg_2Cl_6.12(H_2O)$	1.57	3,300
	Carnallite	$KMgCl_3.6(H_2O)$	1.66	3,910
	Sylvite	KCl	1.86	3,910
Background	Halite	$NaCl$	2.10	4,550
HVS	Gypsum	$CaSO_4.2(H_2O)$	2.35	5,810
	Anhydrite	$CaSO_4$	2.98	6,100

Figure 10 - (a) 3D seismic data and well data. The green points at the well path (red and green lines) indicate LVS occurrences; (b) LVS selected facies.

In this case, we did the classification for the whole inline, disregarding the post-salt and pre-salt portions. The same methodology can be used changing only the model where each different lithology is an identified class and where the input seismic attributes for each of these depositional environments (siliciclastic and carbonatic) vary from the ones we used in this work. In addition, the slice data chosen must match the environment in order to properly evaluate.

As we can see in Figure 13, we can perform the same approach considering 3D volumes, only taking into account the time consuming when performing it. We could assign any value to each of the found classes, such as impedance values. Besides, we assume that the method is versatile

and that can be further explored for many other applications, regarding seismic interpretation.

The whole process is very accessible to anyone with minimum programming skills. This workflow was implemented using Python 3.7 and ScikitLearn package (Pedregosa et al., 2011), plus Petrel interpretation software, but any other tools (for coding and interpretation) could be used.

We classified and identified some Tachyhydrite mineral (the main issue for the well drilling programs in the salt section). However, as most of them are thinner than few meters, in some case presenting centimetric thickness, which is far from and below the concept of the seismic resolution (Widess, 1973), our approach was not capable to properly separate them. Even though, the LVSs can be identified as a package.
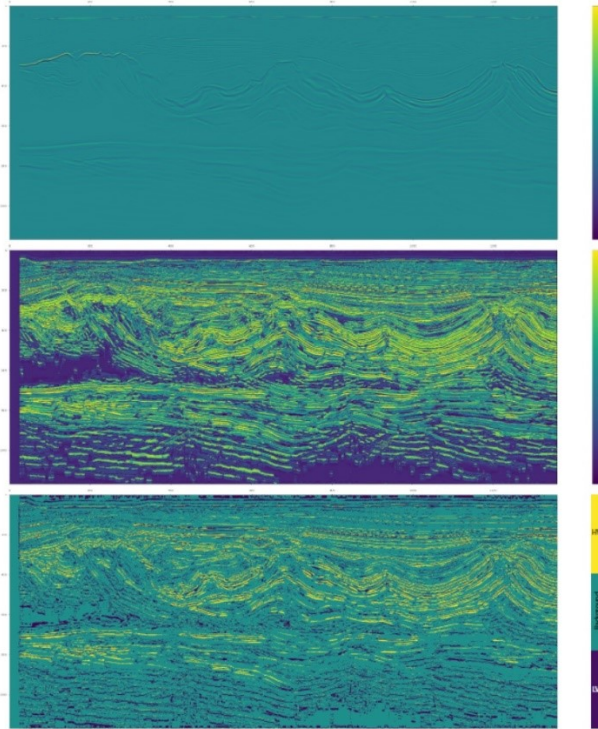
Figure 11 - Method applied to an entire seismic line: top, original seismic amplitudes; middle, seismic facies detected by the proposed UMAP + Modified HDBSCAN algorithm; bottom, the identified and calibrated salt lithology.
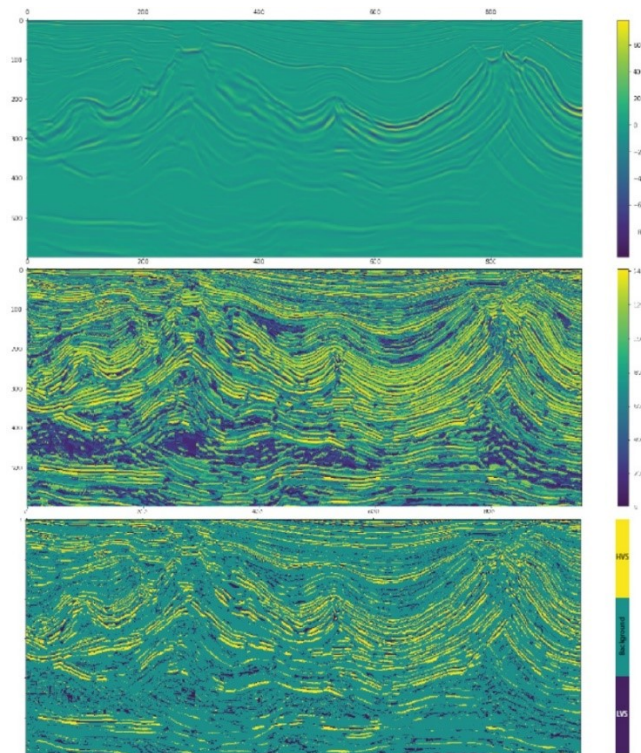
Figure 12 - Detail of Figure 11: top, original seismic amplitudes; middle, seismic facies detected by the proposed UMAP + Modified HDBSCAN algorithm; bottom, the identified and calibrated salt lithology.
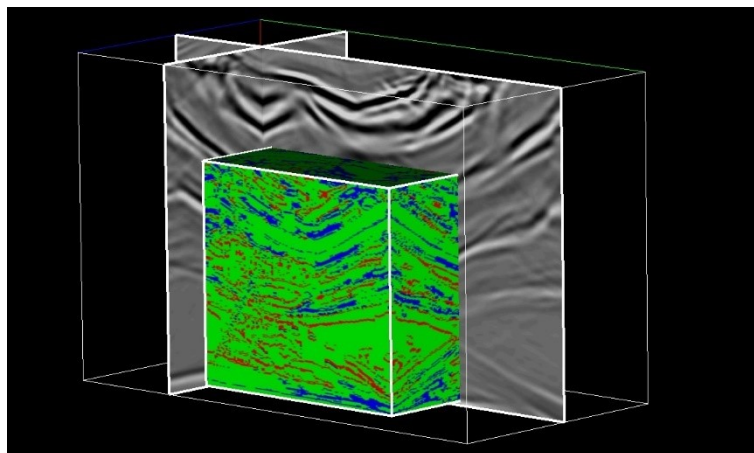
Figure 13 - 3D seismic volume classified and loaded to a commercial interpretation software. In blue color it is illustrated the LVS occurrences; in green, the halite (background) representation; and in red, the LVS presences.

CONCLUSIONS

Our proposed machine learning workflow combined with the right seismic attributes can lead to the identification of the desirable salt heterogeneities; in our case the existing stratification observed in a well and from seismic images. It helps to generate accurate velocity models for seismic processing in portions where the salt stratification is present, collaborating to the generation of reliable seismic images, especially in structural complex regions.

The same approach can be used for geomechanical issues, however the resolution matter is a subject that demands more mitigations. Using our approach we easily differentiate the three groups of salt, LVS, Halite and HVS, as described in literature. Unfortunately, our algorithm is not capable to properly split the LVS group, identifying and categorizing all the occurrence of Tachyhydrite yet. Although sometimes this full separation was not possible, the results can still be used as an aid to the drilling engineers to plan and to adapt the drilling parameters before and in the moment the well is been drilled. It helps to avoid fluid losses when entering the LVS layers that lead to a lot of time and financial issues as well as to assure safety in the drilling operations.

We can add many pre-processing steps as noise filtering and Q factor normalization to recover the attenuation effects. However, even without these steps, the method shows its strength as a much better alternative to the classic PCA + K-means normally used. It is comparable to, or better than, the Self Organizing Maps which by itself does not solve the problem without a pre-conditioning PCA plus further steps.

Besides, we introduce here the concept of using the UMAP as a regularization step to the data, diminishing the need for normalization, filtering, PCA and many other steps normally used in traditional methods. Additionally, we implement a modification to the HDBSCAN technique, allowing a hierarchical clusterization method, to work together with the distribution of varying densities, which produces much better results than K-means or KNN.

It all starts from a detailed data selection, commencing from the wells and ending with data slice and seismic inlines from the seismic cube, and then applying dimensionality reduction and clusterization followed by a calibration step. Finally, we concluded our running approach delivering a salt categorized seismic volume. We then assigned values to the categorized data according to the class behaviors, the interval velocity property in our case, allowing to build a reliable and geological 3D velocity model for the salt section. Summarizing, we start from an unsupervised classification technique, we generate the model and we use the wells and the salt proportions as a base to map the seismic facies to the salt lithology. We finish with a model showing the accurate distribution of the three salt stratification groups we were interested in: LVS, Halite and HVS.

In a future work, we plan to use autoencoders for the detection of rare events, since Tachyhydrite is an uncommon but problematic salt. We judge this technique promising and worthy to be continued.

ACKNOWLEDGEMENTS

REFERENCES

ANP. 2013. Bacia de Santos – Sumário Geológico e Área em oferta. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, Rio de Janeiro, RJ, Brazil.

AQRAWI A & BOE T. 2011. Improved fault segmentation using dip guided and modified Sobel filter. SEG Technical Program Expanded Abstracts, 999–1003. doi: 10.1190/1.3628241.

BADRINARAYANAN V, KENDALL A & CIPOLLA R. 2015. SegNet: A deep convolutional encoder-decoder architecture for image segmentation, arXiv preprint. URL: http://arxiv.org/abs/1511.00561.

BHATTACHARYA S, CARR TR & PAL M. 2016. Comparison of supervised and unsupervised approaches for mudstone lithofacies classification: Case studies from the Bakken and Mahantango-Marcellus Shale, USA. Journal of Natural Gas Science and Engineering, 33: 1119–1133.

CAMPELLO RJ, MOULAVI D & SANDER J. 2013. Density-based clustering based on hierarchical density estimates. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) 7819 LNAI, 160–172. doi: 10.1007/978-3-642-37456-2_14.

CHOPRA S & MARFURT KJ 2005. Seismic Attributes - A historical perspective. Geophysics, 70(5): 3–28. doi:10.1190/1.2098670.

CONNOLLY P. 1999. Elastic Impedance. The Leading Edge, 18: 438–452. doi: 10.1190/1.1438307.

GAO D. 2007. Application of three-dimensional seismic texture analysis with special reference to deep-marine facies discrimination and interpretation: An example from offshore Angola, West Africa. AAPG Bulletin, 91: 1665–1683, 2007. ISSN 1860-0980.

GOBATTO F, MAUL A, FALCÃO L, TEIXEIRA L, BOECHAT JB, GONZÁLEZ M & GONZÁLEZ G. 2016. Refining velocity model within the salt section in Santos Basin: an innovative workflow to include the existing stratification and its considerations. In: 87th Annual Meeting. Dallas, TX, USA. SEG – Society of Exploration Geophysicist. doi: 10.1190/segam2016-13685489.1.

HAGAN M, DEMUTH H & BEALE M. 2002. Neural Network Design. Number v. 10 in Neural network design, Campus Pub. Service, University of Colorado Bookstore. URL: https://books.google.com.br/books?id=bUNJAAAACAAJ.

HAMPSON DP, SCHUELKE JS & QUIREIN JA 2001. Use of multiattribute transforms to predict log properties from seismic data. Geophysics, 66: 220–236. doi: 10.1190/1.1444899

ITURRARÁN-VIVEROS U, MUÑOZ-GARCÍA A & CASTILLO-REYES O. 2021. Machine Learning as a Seismic Prior Velocity Model Building Method for Full-Waveform Inversion: A Case Study from Colombia. Pure Appl. Geophysics, 178: 423–448, 2021. doi:10.1007/s00024-021-02655-9

KALKOMEY CT 1997. Potential risks when using seismic attributes as predictors of reservoir properties. The Leading Edge 16, 247–251. doi: 10.1190/1.1437610

KOHONEN T. 2001. Self Organizing Maps, Ed. Springer, New York, 3rd ed., 521 pp. ISBN-10: 3540679219

LOMBA RFT, PESSANHA RR, CARDOSO JR, WF, LOMBA B, FOL-STA M, GONCALVES JT & TEIXEIRA GT. 2013. Lessons Learned in Drilling Pre-Salt Wells with Water Based Muds. In: Offshore Technology Conference Brazil, Rio de Janeiro, RJ, Brazil. doi: 10.4043/24355-MS.

LONG J, SHELHAMER E & DARRELL T. 2015. Fully convolutional networks for semantic segmentation, In: Proceedings of the IEEE conference on computer vision and pattern recognition, Boston, MA, USA, 3431–3440. URL: http://arxiv.org/abs/1411.4038.

MACQUEEN J. 1967. Some methods for classification and analysis of multivariate observations. In:

Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, v.1: Statistics, University of California Press, Berkeley, Calif, 281–297. URL:https://projecteuclid.org/euclid.bsmsp/1200512992

MAUL AR, SANTOS MAC & SILVA CG. 2018. Few Considerations, Warnings and Benefits for the E&P Industry When Incorporating Stratifications Inside Salt Sections. Brazilian Journal of Geophysics, 36(4): 461-477. doi:10.22564/rbgf.v36i4.1981.

McINNES L, HEALY J & ASTELS S. 2017. hdbscan: Hierarchical density based clustering. The Journal of Open Source Software, 2(11): 205. doi: 10.21105/joss.00205.

MCINNES L, HEALY J & MELVILLE J. 2020. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction, arXiv e-print: https://arxiv.org/pdf/1802.03426.pdf.

MESQUITA F, CETALE SANTOS M, MAUL A & BORDIGNON A. 2019. Identifying salt stratifications performing the machine learning approach over seismic attributes. In: 16th International Congress of the Brazilian Geophysical Society & EXPOGEF, Rio de Janeiro, RJ, Brazil, SBGf.

MOHRIAK WU, SZATMARI P & ANJOS S. 2012. Salt: geology and tectonics of selected Brazilian basins in their global context. Geological Society, London, Special Publications 363, 131–158, doi: 10.1144/SP363.7

MOQBEL A & WANG Y. 2011. Convolutional neural networks for automated seismic interpretation. Journal of Geophysics and Engineering, 8: 592.

OLIVEIRA LC, FERNANDES LF, MAUL AR, ROSSETO JA, GONZÁLEZ M & GONZÁLEZ G. 2015. Geological Velocity Approach in Order to Obtain a Detailed Velocity Model for the Evaporitic Section at Santos Basin, 1374–1377. In: 14th International Congress of the Brazilian Geophysical Society, Rio de Janeiro - RJ, Brazil, SBGf. doi: 10.1190/sbgf2015-273.

PEDREGOSA F, VAROQUAUX G, GRAMFORT A, MICHEL V, THIRION B, GRISEL O, BLONDEL M, PRETTENHOFER P, WEISS R, DUBOURG V, VANDERPLAS J, PASSOS A, COURNAPEAU D, BRUCHER M, PERROT M & DUCHESNAY E. 2011. Scikit-learn: Machine learning in Python. Journal of Machine Learning Research 12, 2825–2830.

SAYAGO J, LUCIA M, MUTTI M, COTTI A, SITTA A, BROBERG K, PRZYBYLO A, BUONAGURO R & ZIMINA O. 2012. Characterization of a deeply buried paleokarst terrain in the Loppa High using core data and multiattribute seismic fades classification. AAPG Bulletin 96, 1843–1866, doi: 10.1306/02271211137.

TEIXEIRA L, LUPINACCI W & MAUL A. 2020. Quantitative and stratigraphic seismic interpretation of the evaporite sequence in the Santos Basin. Marine and Petroleum Geology, 122, 1–17, doi: 10.1016/j.marpetgeo.2020.104690

VAN DER MAATEN L & HINTON G. 2008. Visualizing data using t-SNE. Journal of Machine Learning Research 9, 2579–2605. URL: http://www.jmlr.org/papers/v9/vandermaaten08a.html.

WALDELAND A, JENSEN A, GELIUS L & SOLBERG A. 2018. Convolutional neural networks for automated seismic interpretation. The Leading Edge. 7, 529–537.

WIDESS MB. 1973. How thin is a thin bed? Geophysics, 38: 1176–1180. doi: 10.1190/1.1440403.

F.M: conceptualization (lead), formal analysis (lead), investigation (lead), methodology (lead), project administration (lead), software (lead), validation (lead), writing – original draft (lead); **M.C.:** conceptualization (supporting), formal analysis (supporting), investigation (supporting), supervision (supporting), project administration (equal), validation (supporting); **A.M.:** conceptualization (supporting), formal analysis (supporting), investigation (supporting), methodology (supporting), supervision (supporting), validation (equal), writing – original draft (equal); **A.B.:** conceptualization (supporting), formal analysis (supporting), investigation (supporting), supervision (supporting), methodology (equal), software (equal), writing – original draft (supporting).