






## STRUCTURE-ORIENTED FILTERING IN CROSSPLOTING AND $k$ -MEANS

André Steklain <sup>1\*</sup>, Francisco Ganacim <sup>1</sup>, Marcio R. Adames <sup>1</sup>,  
João Gonçalves <sup>1</sup>, and Danian Oliveira <sup>2</sup>

<sup>1</sup>Universidade Tecnológica Federal do Paraná - UTFPR, Computational Geophysics Laboratory, Curitiba, PR, Brazil

<sup>2</sup>Petrobras, Rio de Janeiro, RJ, Brazil

\*Corresponding author email: [asdf@asdf.org](mailto:asdf@asdf.org)

**ABSTRACT.** Several authors have proposed new techniques using multi-attribute analysis and machine learning. Studying the influence of different data treatments on such techniques is essential. We analyze the results by applying two clustering techniques, Crossplotting, and  $k$ -means, in filtered data. In particular, we use structure-oriented filtered seismic data before calculating seismic attributes. We use a migrated section of the Buzios field from the Brazilian pre-salt in the Santos Basin. We find that combining filtering and clustering techniques can improve salt identification.

**Keywords:** Seismic attributes; machine learning; mass transport complex; salt dome detection.

### INTRODUCTION

Seismic interpretation goes beyond the identification of oil or gas accumulations. The extrapolation of such data to find where the accumulations are formed and can be trapped, or where the oil flow may be interrupted by a fault are complex tasks and require the experience of human interpreters. According to [Zhao et al. \(2015\)](#), skilled professionals can classify some patterns emerging from original data without computer help. Nevertheless, as mentioned by [Chopra and Marfurt \(2018\)](#), the increasing size of the seismic surveys poses a challenge that requires not only human knowledge and skill but also technology.

Seismic attributes are used to emphasize different aspects of seismic data that may not be visible in the original data. Attributes like amplitude, phase, frequency and continuity, among several others, are widely used in the industry, as pointed out by [Subrahmanyam and Rao \(2008\)](#). In addition, recently automatized methods based on machine learning have been proposed in order to classify seismic facies. Such methods can process a large amount of data, providing information necessary so that the human interpreter can propose the regional geological model. Most of the machine learning techniques rely on seismic attributes, but, according to [Chopra and Marfurt](#)

(2018), some recent deep learning techniques propose to eliminate this intermediate step by the introduction of hidden layers in the algorithm. Teaching a computer to identify facies like an interpreter requires an adequate treatment of the data, as explained by [Qi et al. \(2016\)](#).

A key task in subsurface interpretation where machine learning has found widespread application is the identification of salt diapirs. Among the methods proposed it is frequent the use of texture attributes, as [Berthelot et al. \(2011\)](#), image segmentation, as [Halpert et al. \(2014\)](#) and edge detection, as [Asjad and Mohamed \(2015\)](#). Some other methods use combinations of attribute analysis and techniques such as map delineation, proposed by [Farrokhnia et al. \(2018\)](#) and edge detection, as [Amin and Deriche \(2015\)](#). In some cases this classification process can be challenging. For instance, differentiating salt domes from mass transport complexes (MTCs) may be easy for a trained human interpreter, but a machine learning algorithm requires a specialized treatment. The use of filters prior to the classification process can improve the effectiveness, as already shown by [Qi et al. \(2016\)](#). In their work, the proposed workflow involves the application of the Kuwahara filter on each seismic attribute, which may be a time-consuming procedure. Choosing a filter that can be used on the original data

before the generation of attributes without destroying their properties can lead to a competitive advantage as it can reduce computation time.

This work aims to study the effect of the structure-oriented filtering in two simple clustering techniques, namely Crossplotting and  $k$ -means. In this work, we use a 3D depth migrated seismic crossline from Buzios field in Brazil. We select an area with a visible salt diapir, as shown in Figure 1.

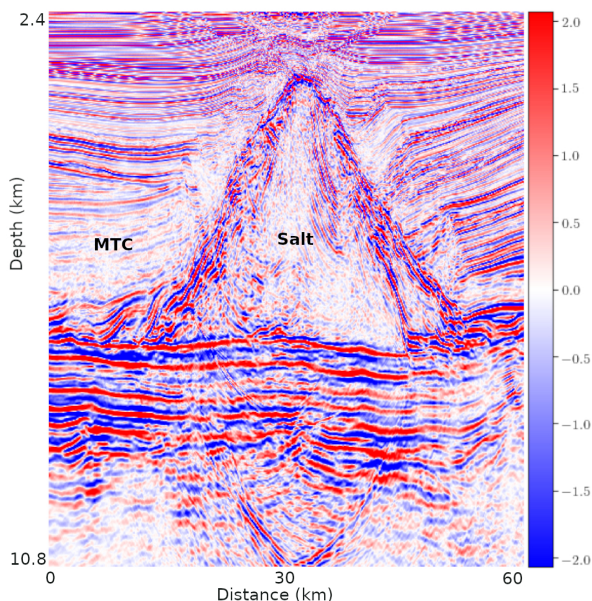


Figure 1: Seismic amplitude for a vertical slice of the original seismic data from Buzios field. In the center, there is a well-pronounced salt diapir. On the left, there is an MTC facies.

## UNSUPERVISED MACHINE LEARNING

According to Kubat (2017), supervised learning uses pre-classified (tagged) data as training data. This training data can be obtained from several sources, such as well control, human experience, and others. After this training, the model builds a mathematical relation capable to predict the classification of new data based on their attributes, as described by Chopra and Marfurt (2018). The influence of human interpreters is key to selecting and classifying the training data, and different training methods can lead to different results.

On the other hand, unsupervised learning methods do not use tagged data in order to perform a clustering scheme, in which the data is only separated into different groups (clusters). These methods can use the attributes themselves as training data. An advantage of these methods pointed by Coléou et al. (2012), is the fact that these methods are not biased by the desired output. The efficiency of these methods will depend on the capacity to identify clusters of data. Below we introduce two clustering schemes used in geophysics, namely Crossplotting and  $k$ -means.

Although these clustering schemes can be extended to use more attributes, they can be better visualized with a reduced number. In this work, we use only Semblance and Grey Level Co-occurrence Matrix (GLCM) Dissimilarity. The choice of the attributes is based on the work by Qi et al. (2016), where the authors use five attributes in order to distinguish MTC from salt. GLCM Dissimilarity presents a high response to salt and MTC and helps to distinguish different kinds of chaotic textures. Semblance can distinguish salt and MTCs from sediment. The attributes obtained from our seismic data are shown in Figure 2.

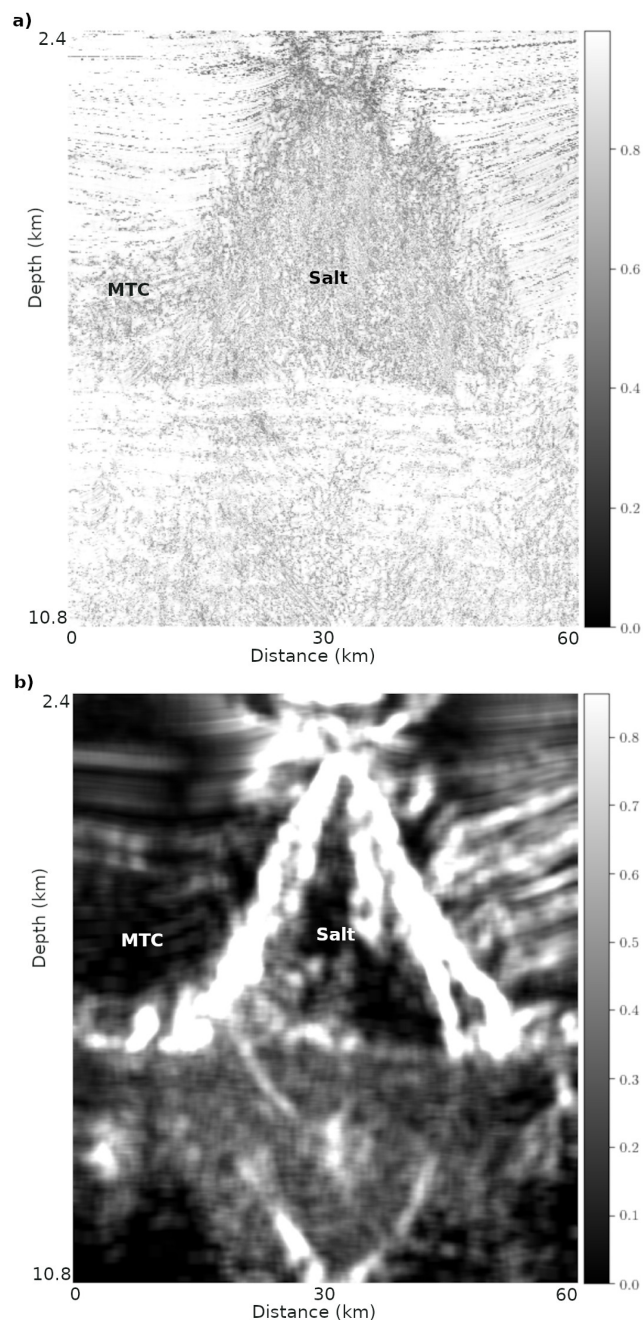


Figure 2: Attributes obtained from original data. a) Semblance. b) GLCM Dissimilarity.

## Crossplotting

Crossplotting is perhaps the most common clustering technique, as stated by Zhao et al. (2015). It is an interactive technique used by interpreters in the industry for a long time, as reported by Chopra and Marfurt (2018). In its most simple form, the method consists in selecting two (or more attributes) and producing a histogram of them. This histogram might present different regions indicating a higher correlation of certain values for the selected attributes. If the different attributes are sensitive to the same feature, the selected regions from the histogram can be used to highlight the structures in the seismic data.

Figure 3a shows a histogram generated from Coherence and GLCM Dissimilarity attributes for the seismic section. The region of maximum correlation is obtained visually from the data and is marked with a blue rectangle, including values up to 2000. In Figure 3b, the region of maximum correlation in the histogram is used to highlight the seismic section. Note that the maximum correlation region highlights the salt, but also some other structures, such as MTC.

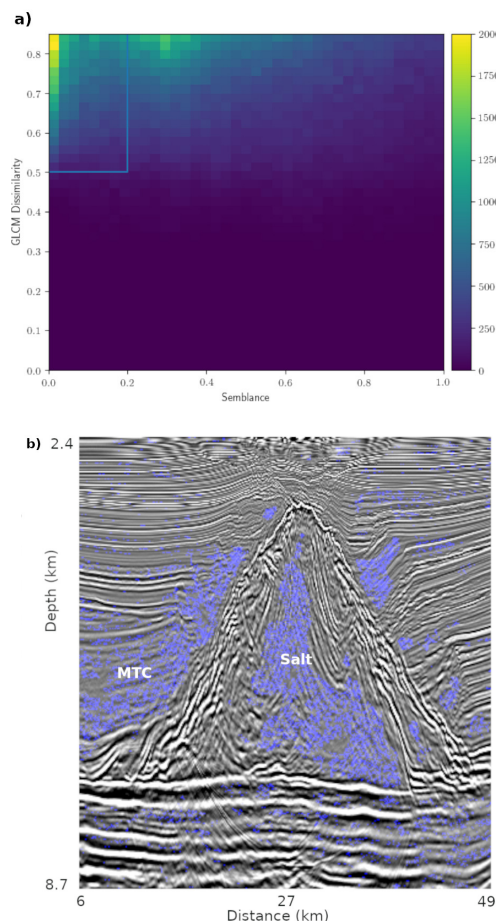


Figure 3: a) Histogram of Coherence and GLCM Dissimilarity attributes from the original seismic section. b) Original seismic data and highlighted regions according to the region of maximum correlation in the histogram of Semblance and GLCM Dissimilarity attributes.

One of the main drawbacks of Crossplotting lies in the number of attributes that can be used to compose the histograms, as it is unpractical to work with more than two attributes. Interesting geological features may be only identifiable with a larger number of attributes. In this case, Crossplotting should be substituted or complemented by other techniques.

## *k*-means

The *k*-means concept was proposed by Steinhaus (1957), and the first standard algorithm was proposed by Lloyd (1982). In essence, the goal of the method is to perform a partition of the data into *k* clusters. Each point is classified according to the cluster to which it is closest in average. In Figure 4a, we show a Voronoi diagram generated from Semblance and GLCM attributes. The partition is obtained from *k*-means with five clusters. The color code in the diagram is used to identify the same partition on the seismic image in Figure 4b.

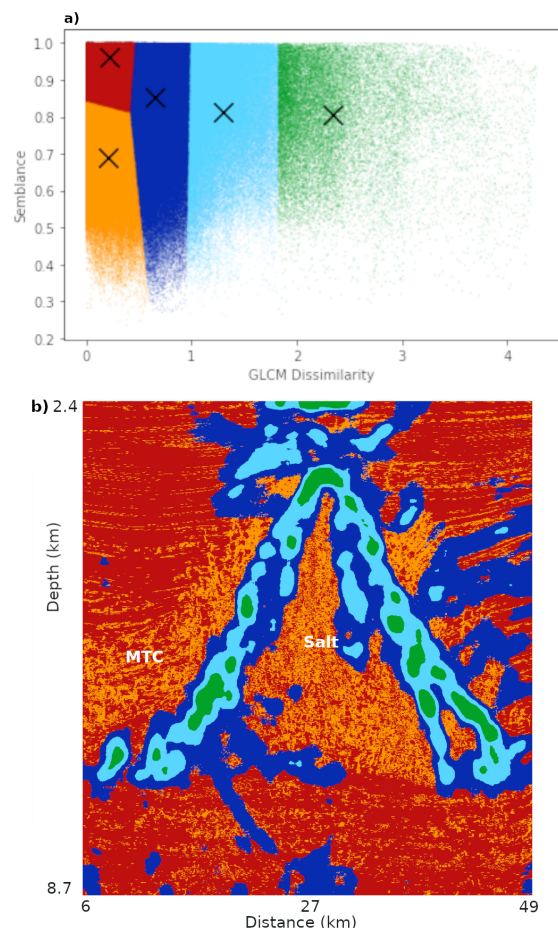


Figure 4: a) Voronoi diagram obtained from *k*-means applied on semblance and GLCM dissimilarity attributes from original data. The multiplication signs represent the center of each cluster. b) *k*-means mapped in the original data. Note that similar facies, such as the salt diapir, appear with different colors that have no relationship with each other.

The  $k$ -means method represents an improvement in comparison to Crossplotting, as it provides an automatized clustering method. The user must only select the two attributes to be used and provide the number of clusters to be used in clusterization. The optimal number of clusters is commonly determined using the elbow method, as proposed by Thorndike (1953). This is an heuristic method, and such an “elbow” may never appear, as pointed by Ketchen Jr. and Shook (1996). Nevertheless, we use the elbow method to determine the optimal number of clusters, as can be seen in Figure 5. It is possible to see that the defining feature for determining the number of clusters, the “elbow” is very subtle. We chose 5 clusters, as this configuration is the best choice in order to identify salt and MTC.

Fast and easy to implement, it is possible to use  $k$ -means in multi-dimensional data. An important downside, as pointed by Zhao et al. (2015), is the lack of structure leading to a relationship between cluster identification and proximity between clusters, causing similar facies to appear in totally different colors. In this work, we use Python implementation of  $k$ -means from the package *scikit-learn* implemented by Pedregosa et al. (2011).

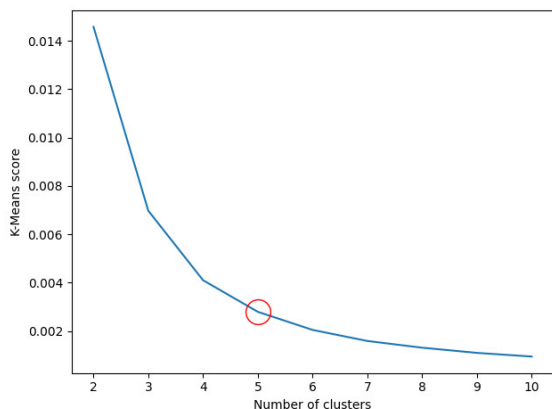


Figure 5: Elbow method used to determine the number of clusters in  $k$ -means. The “elbow” is very subtle, as the curve does not have any great variations as the number of clusters is increased. We chose five clusters, with the point marked with a red circle.

## STRUCTURE-ORIENTED FILTERING

As pointed by Hale (2009), the structure-oriented filtering is an adaptation of the coherency-enhancing anisotropic diffusion filters proposed by Fehmers and Höcker (2003). The filter works as a simulated anisotropic diffusion process (low-pass filter) that diffuses the seismic amplitude parallel to the reflections. The basic formula is given by

$$\frac{\partial u}{\partial \tau} = \nabla (\mathbf{D} \nabla u), \quad (1)$$

where  $u = u(x, y, t)$  is the seismic data,  $\tau$  is a parameter called diffusion time (not to be confused with the

time  $t$ ) and  $\mathbf{D}$  is the diffusion tensor. The diffusion tensor is very important in structure-oriented filtering, as it introduces the anisotropy to the diffusion by allowing the flow in some directions and inhibiting in others. This tensor must be constructed in such a manner that its eigenvectors are aligned with the structure of the seismic data. A proper way to achieve this is to use the structural tensor  $\mathbf{S}$  given by

$$\mathbf{S} = \nabla u (\nabla u)^T, \quad (2)$$

where the superscript  $T$  means the transpose. By choosing a scale parameter  $\sigma$  corresponding to the radius of the smoothing filter we obtain an average of this tensor, denoted by  $\mathbf{S}_\sigma$ . The spectral decomposition of this average tensor is given by

$$\mathbf{S}_\sigma = \sum_{i=1}^d \lambda_i \mathbf{v}_i \mathbf{v}_i^T, \quad (3)$$

where  $d$  is the dimension of the data,  $\lambda_i$  are the positive eigenvalues given in descending order and  $\mathbf{v}_i$  are the associated eigenvectors, which form an orthonormal basis. The diffusion tensor is defined by

$$\mathbf{D} = \sum_{i=2}^d \mathbf{v}_i \mathbf{v}_i^T. \quad (4)$$

Note that the eigenvector  $\mathbf{v}_1$  corresponding to the largest eigenvector is excluded from the sum. It means that the flow is obstructed in the direction corresponding to largest data variation, normally orthogonal to the seismic horizons. By construction the diffusion flow occurs only on the other directions, parallel to the horizons and other prominent structures.

The use of the average of the structure tensor can be problematic as a more severe filtering can eliminate not only noise, but also some important features in the data, like faults. The remedy to this situation is to introduce a continuity factor  $\varepsilon$  in Equation 1, that now reads

$$\frac{\partial u}{\partial \tau} = \nabla (\varepsilon \mathbf{D} \nabla u). \quad (5)$$

The continuity factor  $\varepsilon$  can be obtained by some measure of the continuity, as the coherence attribute.

The structure-oriented filter has three important ingredients: orientation analysis, edge detection, and edge-preserving smoothing. It is advantageous to highlight structures such as salt diapirs, where we are mainly interested in the edges. The impact on a machine learning process must consider how severe is the filtering process. The structure-oriented filter depends on the total diffusion time  $\tau$  and the scale factor  $\sigma$ . The influence of  $\tau$  is illustrated in Figure 6 and of  $\sigma$  are given in and Figure 7. The following sections compare how these different parameters can impact the clustering process.

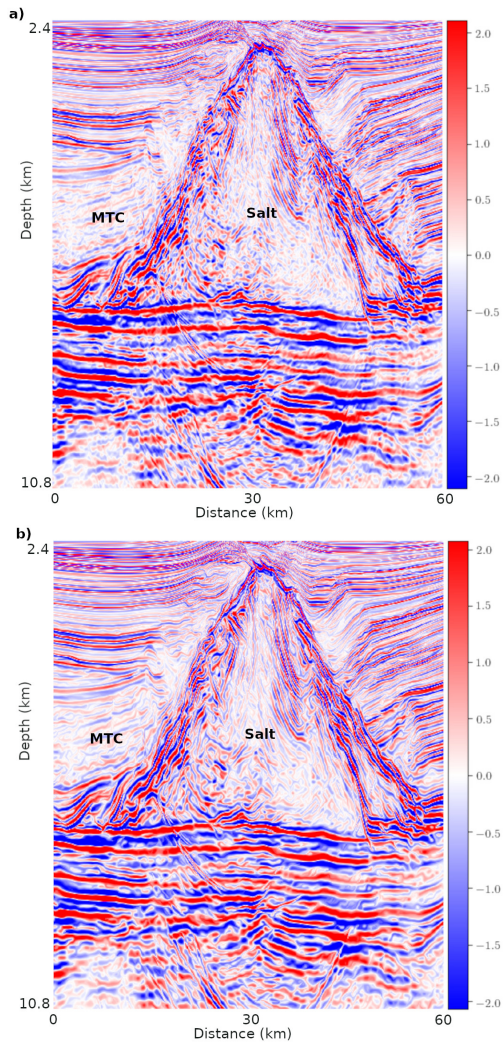


Figure 6: Structure-oriented filtering of the data with  $\sigma = 1$ . a) Data filtered until  $\tau = 100$ . b) Data filtered until  $\tau = 1000$ . As  $\tau$  increases the reflections become more continuous and incoherent noise is removed, but some artificial structures are generated in chaotic regions, such as salt and MTC.

## CLUSTERING OF FILTERED DATA

Structure-oriented filtering has visible effects on noise reduction and smoothing of seismic data, as can be seen in Figures 6 and 7. In Figure 6, as the diffusion time  $\tau$  increases, the seismic reflections become more smooth and continuous, and incoherent noise is removed. Chaotic regions, such as salt or MTC, are also affected by the smoothing process, and some artificial structures are generated in the process. In Figure 7, on the other hand, as  $\sigma$  increases, the chaotic regions are smoothed without creating new structures, but the elimination of noise is less effective. Hale (2009) combines this filter with semblance attributes to highlight discontinuities. Nevertheless, the combination of the structure-oriented filter and other attributes is also possible. The resulting combinations can be used in clustering techniques to verify if there is an improvement on clustering.

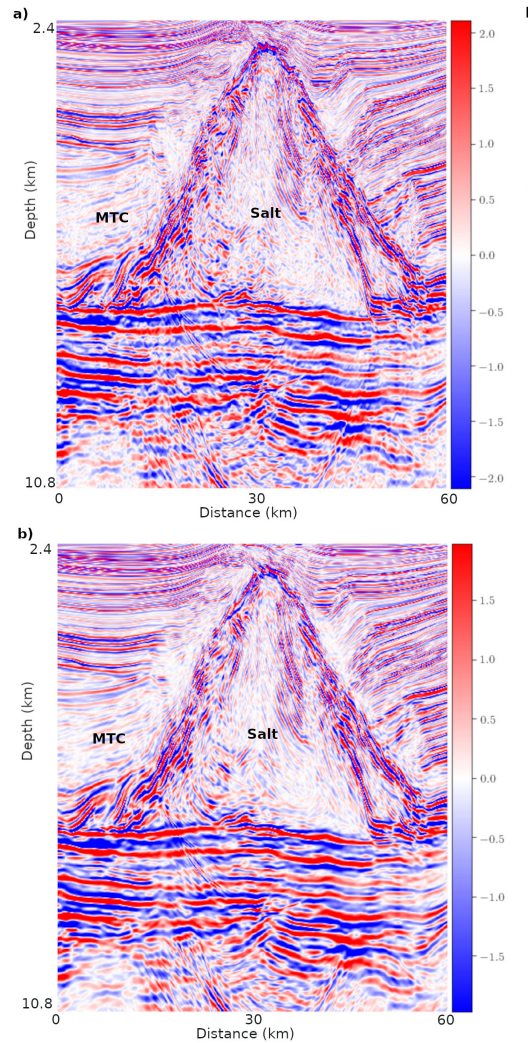


Figure 7: Structure-oriented filtering of the data with  $\tau = 100$ . a)  $\sigma = 1$ . b)  $\sigma = 20$ . As  $\sigma$  increases, there is no generation of artificial structures, but less noise is removed.

It is important to stress that we apply the filter to the original data before attribute extraction in this work. One could argue that the filter may wash away important features of the data before attribute extraction. In Figures 8 and 9, we show a comparison between attributes when the filtering is performed on the original data and directly on the attributes. It is possible to see some minor differences between these cases. In Figure 8a, there is a much-reduced noise when compared to Figure 8b. Figure 9a has a brighter central core when compared to Figure 9b, where the salt diapir has a more greyish and noise color. In the bottom left of Figure 9a filtering of the attribute leads to horizontal noise streaks. There is a very similar range of bright white values in the central diapir and outside, where is the MTC, whereas, on the right image, the center of the diapir and the outer MTC have different value ranges. Despite those differences,

filtering before or after the attribute calculation have little impact on the clustering process. We shall explore this subject in more detail in future work. Nevertheless, filtering before may provide a competitive advantage concerning computational time, as the filtering must be applied once in the original data. Nevertheless, this subject demands a more detailed investigation comprising a broader class of attributes.

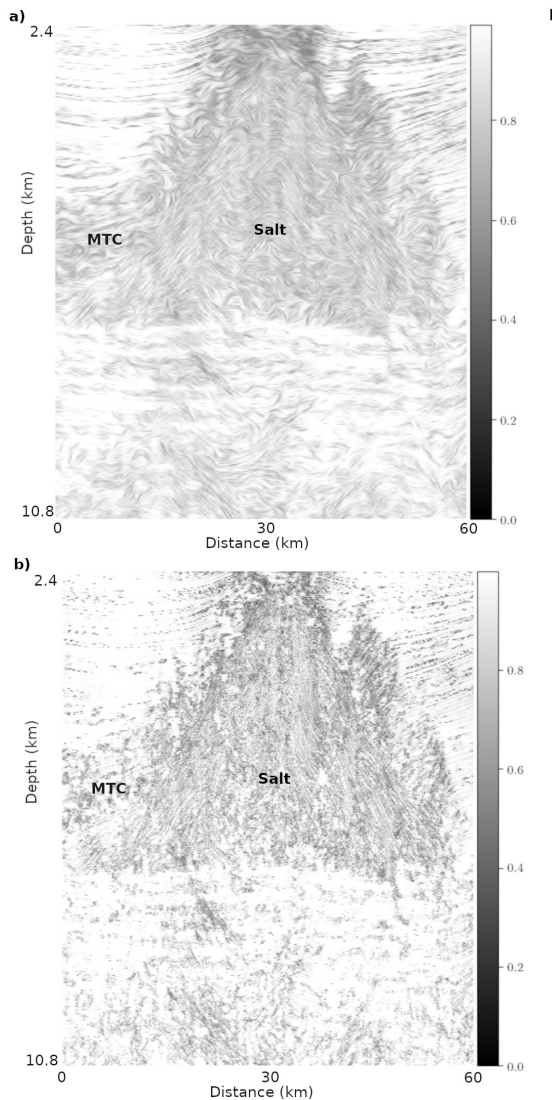


Figure 8: Semblance before and after filtering with  $\sigma = 1$  and  $\tau = 1000$ . a) Filtered semblance. b) Semblance on filtered data.

The results obtained from application of Crossplotting in the attributes from filtered results are shown in Figure 10. It is possible to see that the application of structure-oriented filtering improves the separation between MTC and salt, but some areas are still confounded. The separation between the two facies is particularly improved when we use larger values of  $\tau$ . On the other hand, with higher values of  $\tau$ , other regions near the salt diapir are also highlighted, so a severe filtering can also be a source of numerical artifacts in the data.

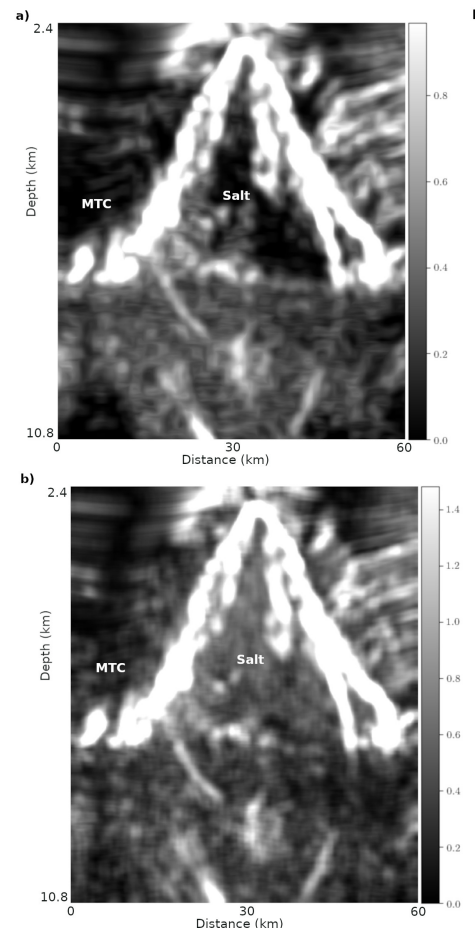


Figure 9: GLCM Dissimilarity before and after filtering with  $\sigma = 1$  and  $\tau = 1000$ . a) Filtered GLCM Dissimilarity. b) GLCM Dissimilarity on filtered data.

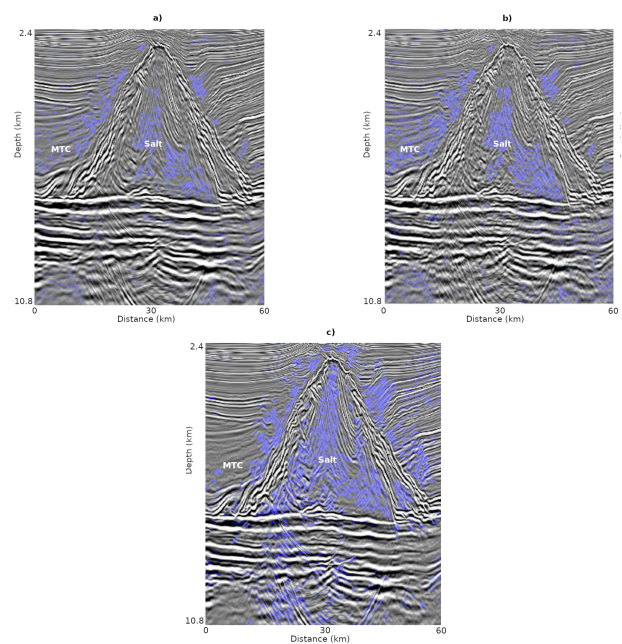


Figure 10: Crossplotting on filtered data. a)  $\sigma = 20$  and  $\tau = 100$ . b)  $\sigma = 1$  and  $\tau = 100$ . c)  $\sigma = 1$  and  $\tau = 1000$ .

The results from the application of  $k$ -means in filtered data are shown in Figure 11, with the respective clusters shown in Figure 12. For small  $\tau$ , the structure-oriented filtering has little effect considering Semblance and GLCM Dissimilarity. For a larger value of  $\tau$ , on the other hand, the separation between salt and MTC is improved just like in the Crossplotting case. Nevertheless for such larger values of  $\tau$ , we also see some regions that are misclassified near the salt diapir.

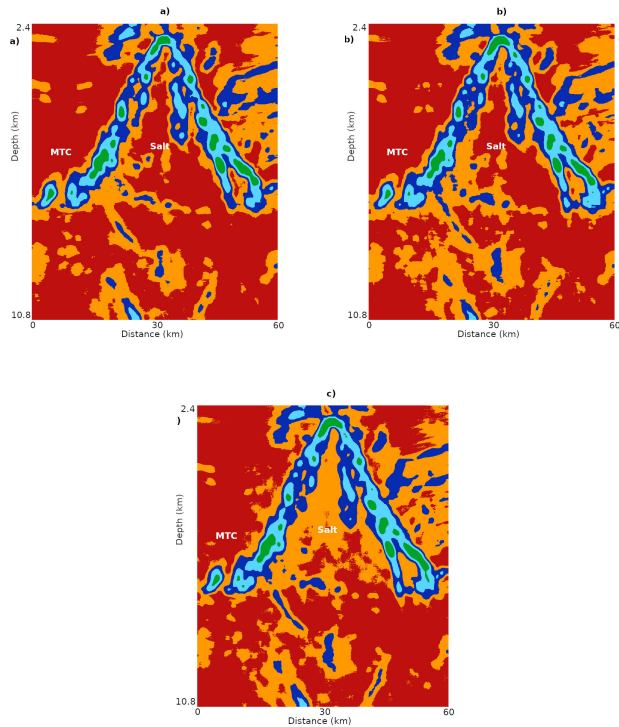


Figure 11:  $k$ -means on filtered data. a)  $\sigma = 20$  and  $\tau = 100$ . b)  $\sigma = 1$  and  $\tau = 100$ . c)  $\sigma = 1$  and  $\tau = 1000$ .

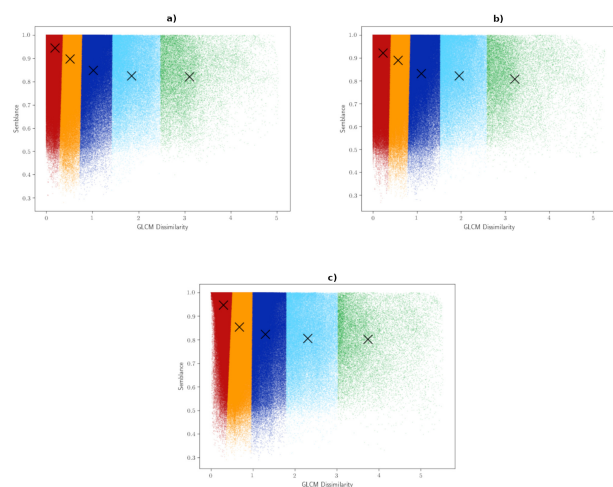


Figure 12:  $k$ -means clusters on filtered data. a)  $\sigma = 20$  and  $\tau = 100$ . b)  $\sigma = 1$  and  $\tau = 100$ . c)  $\sigma = 1$  and  $\tau = 1000$ .

## CONCLUSION

In this work, we compare the effect of structure-oriented filtering two clustering methods. The filter is applied to the original data before extracting the attribute. The evaluation of filtering on the identification of facies by machine learning techniques becomes relevant because structure-oriented filtering is often used in the interpretation workflow and can improve unsupervised and semi-supervised classification process.

Clustering methods that use only two attributes, as Crossplotting and  $k$ -means, may have a good improvement from filtering when the parameters  $\tau$  and  $\sigma$  are well tuned. It is important to stress that the tuning of such parameters is performed based on the knowledge of the geology of the area. The parameters are obtained for small data and can be extended accordingly to a larger area. A quantitative method to find the best values must be established. Severe filtering can improve the differentiation, but can also cause distortions that can be visually observed on the data. The parameters used in filtering must be calibrated to avoid these distortions. The manual calibration of such parameters can be a time-consuming process that have an impact on the workflow, but once established it can be applied to a large amount of data without further calibration.

This result may be a consequence from the choice of attributes showing that these methods have a strong dependency on the attributes used. On the other hand, different clustering techniques like SOM and GTM are more likely to improve facies differentiation between MTC and salt. These techniques will be explored in a future work.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge ANP (Brazil's National Oil, Natural Gas, and Biofuels Agency) for making data available for research and education. We also thank the reviewers for all their valuable contributions to work.

## REFERENCES

- Amin, A., and M. Deriche, 2015, A hybrid approach for salt dome detection in 2D and 3D seismic data: 2015 IEEE International Conference on Image Processing (ICIP), IEEE, 2537–2541. doi: [10.1109/ICIP.2015.7351260](https://doi.org/10.1109/ICIP.2015.7351260).
- Asjad, A., and D. Mohamed, 2015, A new approach for salt dome detection using a 3D multidirectional edge detector: Applied Geophysics, **12**, 334–342, doi: [10.1007/s11770-015-0512-2](https://doi.org/10.1007/s11770-015-0512-2).
- Berthelot, A., A. H. S. Solberg, E. Morisbak, and L. J. Gelius, 2011, Salt diapirs without well defined boundaries - a feasibility study of semi-automatic detection: Geophysical Prospecting, **59**, 682–696,

- doi: [10.1111/j.1365-2478.2011.00950.x](https://doi.org/10.1111/j.1365-2478.2011.00950.x).
- Chopra, S., and K. J. Marfurt, 2018, Seismic facies classification using some unsupervised machine-learning methods: SEG Tech. Progr. Expand. Abstr. 2018, Society of Exploration Geophysicists, 2056–2060. doi: [10.1190/segam2018-2997356.1](https://doi.org/10.1190/segam2018-2997356.1).
- Coléou, T., M. Poupon, and K. Azbel, 2012, Unsupervised seismic facies classification: A review and comparison of techniques and implementation: The Leading Edge, **22**, 921–1056, doi: [10.1190/1.1623635](https://doi.org/10.1190/1.1623635).
- Farrokhnia, F., A. R. Kahoo, and M. Soleimani, 2018, Automatic salt dome detection in seismic data by combination of attribute analysis on CRS images and IGU map delineation: Journal of Applied Geophysics, **159**, 395–407, doi: [10.1016/J.JAPPGEO.2018.09.018](https://doi.org/10.1016/J.JAPPGEO.2018.09.018).
- Fehmers, G. C., and C. F. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293, doi: [10.1190/1.1598121](https://doi.org/10.1190/1.1598121).
- Hale, D., 2009, Structure-oriented Smoothing and Semblance: CWP Report, **CWP-635**, 261–270.
- Halpert, A. D., R. G. Clapp, and B. Biondi, 2014, Salt delineation via interpreter-guided 3D seismic image segmentation: Interpretation, **2**, T79–T88, doi: [10.1190/INT-2013-0159.1](https://doi.org/10.1190/INT-2013-0159.1).
- Ketchen Jr., D. J., and C. L. Shook, 1996, The application of cluster analysis in strategic management research: an analysis and critique: Strategic Management Journal, **17**, 441–458, doi: [10.1002/\(SICI\)1097-0266\(199606\)17:6<441::AID-SMJ819>3.0.CO;2-G](https://doi.org/10.1002/(SICI)1097-0266(199606)17:6<441::AID-SMJ819>3.0.CO;2-G).
- Kubat, M., 2017, An introduction to machine learning: Springer International Publishing.
- Lloyd, S., 1982, Least squares quantization in pcm: IEEE Transactions on Information Theory, **28**, 129–137, doi: [10.1109/TIT.1982.1056489](https://doi.org/10.1109/TIT.1982.1056489).
- Pedregosa, F., G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 2011, Scikit-learn: Machine learning in Python: Journal of Machine Learning Research, **12**, 2825–2830, doi: [10.5555/1953048.2078195](https://doi.org/10.5555/1953048.2078195).
- Qi, J., T. Lin, T. Zhao, F. Li, and K. Marfurt, 2016, Semisupervised multiattribute seismic facies analysis: Interpretation, **4**, SB91–SB106, doi: [10.1190/INT-2015-0098.1](https://doi.org/10.1190/INT-2015-0098.1).
- Steinhaus, H., 1957, Sur la division des corps matériels en parties: Bull. Acad. Polon. Sci., 801–804.
- Subrahmanyam, D., and P. H. Rao, 2008, Seismic Attributes- A Review: 7th International Conference & Exposition on Petroleum Geophysics.
- Thorndike, R. L., 1953, Who Belongs in a Family?: Psychometrika, 267–276, doi: [10.1117/12.620399](https://doi.org/10.1117/12.620399).
- Zhao, T., V. Jayaram, A. Roy, and K. J. Marfurt, 2015, A comparison of classification techniques for seismic facies recognition: Interpretation, **3**, SAE29–SAE58, doi: [10.1190/INT-2015-0044.1](https://doi.org/10.1190/INT-2015-0044.1).

**Steklain, A.:** Conceptualization, Methodology, Data Curation, Writing - Original Draft, Supervision;  
**Ganacim, F.:** Software, Formal Analysis, Investigation;  
**Adames, M.:** Methodology, Software, Investigation;  
**Gonçalves, J.:** Methodology, Software, Investigation;  
**Oliveira, D.:** Conceptualization, Data Curation.

Received on November 10, 2021 / Accepted on August 16, 2022



- Creative Commons attribution-type BY