

# **The Application of Semi-supervised Geobody Detection Technique using Multiple Seismic Attributes in Petroleum Exploration\***

**Lei Li<sup>1</sup>, Xianhua Ran<sup>1</sup>, Chunfeng Tao<sup>1</sup>, Zhonghong Wan<sup>1</sup>, and Shifan Zhan<sup>1</sup>**

Search and Discovery Article #41165 (2013)

Posted July 31, 2013

\*Adapted from extended abstract prepared in conjunction with poster presentation at AAPG Annual Convention and Exhibition, Pittsburg, Pennsylvania, May 19-22, 2013, AAPG©2013

<sup>1</sup>BGP, CNPC, China ([leolilei.2007@gmail.com](mailto:leolilei.2007@gmail.com))

## **Abstract**

Small fault blocks, fracture zones and organic reefs are very important in the analysis and prediction of reservoir characteristics, because they have a significant effect on the space distribution and reserves of a hydrocarbon reservoir. In this article, we present a semi-supervised methodology using seismic multi-attributes for the classification and identification of reservoir facies from seismic data. The approach, as a nonlinear and completely data-driven method, requires no accurate initial model and no priori operator linking the reservoir properties with seismic attributes and inversion data. For the real data application, the proposed method has been applied to the field data and the results show that the methodology is effective to delineate the distribution of special reservoirs.

## **Introduction**

In last decades, with the progress of exploration and research, the distribution of structural reservoirs has been clearly known on the whole, while complex stratigraphic lithologic reservoirs, represented by small fault blocks, fracture zones, and carbonate caves and reefs, are not precisely understood. In addition, conventional exploration techniques are facing more and more difficulty in finding new reservoirs, since the targeted areas have moved from middle-shallow strata to middle-deep strata, and the types of reservoir from simple structural to stratigraphic/lithologic. Consequently, the recent research and development of work is focused on the accurate and efficient recognition of special geobodies such as small fault blocks, fractures, carbonate caves and reefs, which aimed at not only heightening production efficiency, but also reducing exploration risks.

In order to use the seismic multi-attributes data to delineate the heterogeneity of the reservoir characteristics, geological bodies can be identified by the characteristics of different seismic response in these regions. This process of classification is implemented either by the manual interpretation or by some automatic techniques. The operation of manual interpretation of seismic is mainly relevant to the subjective

experience of geologists and geophysicists, which is very easily realized. However, this process is a very time-consuming process and sensitive to the subjective experience of the seismic interpreter. The problem becomes more difficult as the complexity of the seismic data increases or the number of obtained seismic attributes is increased.

Automatic methods which rely on a priori initial parameter model can be used to cluster the seismic characteristics of the reflection correlated with the reservoir interval without the need of manual intervention. Since the 1980's, many different supervised and unsupervised methods are proposed. For example, Davis (1986) developed an unsupervised K-means clustering to segment the seismic trace shapes where the Signal Noise Ratio(SNR) of target is low, and Hagan (1982) employed principal component analysis to study the lateral differences in porosity, Mathieu and Rice (1969) used discriminant factor analysis to determine the sand/shale ratio of various zones in the reservoir, Dumay and Fournier (1988) applied both principal component analysis and discriminant factor analysis to identify the seismic facies, and Saggaf et al. (2003) use back-propagation networks to estimate the porosity of the reservoir.

However, automatic methods are also facing many complex problems in practice. Firstly, supervised methods, especially neural networks, not only need initial parameter models, but often have a worse rate of convergence, and even cannot achieve training models. And they cannot overcome the problem of over-learning and local minima when solving a small sample problem with high dimension and nonlinear. Moreover, the choice of the network structure is not yet a unified and complete theoretical guidance, and generally can only be selected by the experience. On the other hand, unsupervised methods are often sensitive to incomplete data information and often unreliable with noisy data. The complex iterative process of clustering costs a lot of computing time and memory. In addition, without priori information of the number of facies and the input seismic attributes, the clustering analysis usually leads to erroneous results for huge and highly redundant seismic data sets.

## Theory

The methodology consists of the choice of the input seismic attributes, supervised sampling process based on seismic attributes, unsupervised competitive learning to optimize different parameter models of geobodies adaptively, and Bayesian classification decision of special geobodies. Moreover, the method introduces three main improvements compared to traditional supervised and unsupervised clustering methods of seismic multi-attributes. The supervised sampling process is not only decided by the information close to the well locations, but implemented on the input seismic multi-attributes sets by the interpreters. Unsupervised competitive learning (such as k-means algorithm) will be implemented repeatedly to achieve different parameter models of geobodies; and Bayesian classification decision based on the parameter models saves a lot of computing time and memory.

First of all, we begin by assuming that the input variables, i.e. the seismic attributes, can be represented by a unique vector in the space  $R^n$ ,  $x = [x_1, x_2, \dots, x_n]$ , where  $n$  is the dimension of the input vector, i.e. the number of input seismic attributes analyzed. In this proposed method, all of the seismic attributes used are chosen by seismic interpreters. And the seismic attributes are closely related to the geometric structure and geological properties of geobodies, rather than complex dimension reduction techniques. The reason is that these dimension reduction methods often inflict an unnecessary distortion on the input data representation by projecting the input space to a lesser dimension space.

In the supervised sampling process, multivariable training data are usually achieved from seismic attributes near the well locations in traditional methods. However, these data are sparse and locally measured, which may not reflect the reservoir characteristics as a whole. In addition, well data are not available at the initial phases of exploration. In contrast to sparse well data, seismic attributes can provide more information for the construction, lithology and fluid changes of the reservoir. In the proposed method, if one of the given seismic attributes can be used to distinguish a certain geobody from others in certain areas of the survey, seismic interpreters pick up the locations of sampling points on this seismic attribute by manual operation and assign their corresponding attribute patterns to a vector for this geobody. In this way, we can obtain more labeled sampling points than sparse well data. And then, different geobodies are always corresponding to different multi-variables data set,  $S_i, i=1,2,\dots,k$ . Because the process of sampling is implemented under the guidance of the seismic interpreters' experience, multi-variable data sets,  $\{S_i\}_{i=1}^k$ , where  $k$  is the number of seismic facies, represent clearer geological significance and improve the accuracy of identification.

Generally, traditional unsupervised methods based on multivariate statistics require the actual number of geological bodies, and lots of iterative computation are time-consuming. And different methods have specific assumptions and limitations. In this study, an unsupervised competitive learning method, the frequency competitive algorithm (FSCL, Ahalt, 1990), is used to adaptively optimize nonlinear parameter models of geobodies. Compared with k-means and SOM, this algorithm reduces the learning rate of the frequent winners, so that all centers have chance to win the competition. The parameter models built using one kind of labeled data are data-adaptive to characterize special geobodies, that is, for a data set  $S_i$  available from the wells or the seismic attributes, the parameter models are first trained and optimized adaptively by FSCL. In order to decide the number of clusters, we over-define the initial number of centers. Algorithm initializes the centers of clusters randomly and compute distance between multi-attribute sampling points and centers, then samples are assigned to the nearest center. Unless specific clusters are fixed, such as at well locations, the clustering operation is not biased by any input. By iteration, the locations of centers are accumulated until all points assign to the nearest center and all centers stabilized. At this moment, the sample percentage,  $\gamma_j^i$ , the center  $c_j^i$  contains the number of samples, is computed. If the percentage  $\gamma_j^i$  is less than a pre-defined threshold,  $\delta$ , the center,  $c_j^i$ , is removed from the parameter model. Thus, each kind of geobody is delineated by its parameter model,  $\{c_j^i\}_{j=1}^{k_i}$ , related to reservoir properties than to mathematical expressions. The value  $k_i$  is the number of centers of data set  $S_i$ , which is decided by FSCL automatically. In the unsupervised clustering process, the proposed methods only using sampling points are faster than traditional methods requiring all of the multi-attribute data in 3D seismic survey.

Finally, once this training is completed, the parameter models are completely specified and used to the seismic multi-attribute data to predict and estimate special geobodies. A Bayesian posteriori classification is applied to assign multi-attribute sampling points with similar seismic nature to each cluster shown by one defined color. The spatial distribution of color change is due to the change in the seismic response of different geobodies. In this case, our parameter models have been fixed, so this process of classification is to evaluate an algebraic expression for each input sample without complex iterative computations. It is the prime advantage for 3D seismic survey. The proposed workflow is very rapidly to provides an understanding of space distribution of geobodies and help qualitatively to understand the depositional environment and reservoir heterogeneity.

## Workflow

In the geological body identification workflow, multiple 3D seismic post-stack seismic volumes as input data are loaded in the system and preprocessed to reduce the influence of noise. According to the practical requirements, many seismic attributes are extracted from seismic data to define an n-dimensional vector. Usually the output of this workflow is a single 3D classification volume in which the class number of geological bodies corresponding to a pre-defined color is assigned for each sample. After supervised sampling, FSCL algorithm continuously estimates and updates parameter model for each kind of geobody. Using this method, each type of geobody can be related to one or more different parameters  $\{c_j^i\}_{j=1}^{k_i}$ . Subsequently, the detailed description of reservoir characteristics can be obtained from the 3-D seismic volume.

[Figure 1](#) shows the semi-supervised workflow proposed in the study.

## Examples

The reservoir space is mainly composed of numerous major/minor faults from East China. These fractures can enhance reservoir permeability, as well as provide conduits for diagenetic fluids. In this case, each sampling point is defined by its seismic attributes, which will be used to understand the heterogeneity of reservoir space. [Figure 2](#) shows the seismic section and the target horizon extracted from the seismic data volume. Along the seismic horizon, a 20 ms window (10 ms above and 10 ms below) is used to extract seismic attribute volumes. Next, we generated horizon slices through volumetric estimates of amplitude, sobel similarity, coherence and most positive curvature ([Figures 3a, 3b, 3c, 3d](#)). We can find that many of the minor faults are not clearly observed and distinguished either in seismic amplitude slices or in simple seismic attributes. The supervised sampling points of each kind of fault are plotted on the sobel similarity attribute shown on [Figure 4](#). For fault mapping, each (major/minor) fault has been recognized more clearly than that in the seismic attribute volume. [Figure 5](#) depicts the clustering results of the semi-supervised clustering algorithm applied to the 3D real field. The results indicate that various scales of fault blocks and fractures are identified clearly, though the thickness and volume of geobodies may be small and change fast laterally. The correct prediction of fault distribution is important for planning of oil and gas exploration and production.

In the second example, we pay attention to depict the organic reef reservoirs from Western China. The target formation of geological bodies in this study is characterized by deeply buried, compact lithological character and strong anisotropy. This kind of reef reservoirs are special carbonate rock geobodies which have unique seismic reflecting characteristics. The reservoirs are highly heterogeneous, especially the interior of the reef which is highly varied in lithology and other properties. The complex reflection characteristics make it difficult to predict the space distribution of reef laterally and vertically. Properly mapping such heterogeneities will aid in understanding the structure of the organic reef and making engineering decisions.

In [Figure 6](#), the seismic section and the target horizon extracted from the seismic data volume are shown. Along the seismic horizon, a 20 ms window (10 ms above and 10 ms below) is used to extract the multi-attribute data volumes. In order to identify different geobodies, four conventional seismic attributes (amplitude, sobel similarity, coherence and most positive curvature) are shown in [Figures 7a, 7b, 7c, 7d](#) and are used to construct the multi-attribute sample dataset.

We find that reservoir characteristics of organic reefs are not clearly observed and distinguished either in seismic amplitude slices or in simple seismic attributes. The supervised sampling points of each kind of reservoir characteristic are plotted on the sobel similarity attribute, as shown in [Figure 8](#). The classification results show three patterns highlighted with different colors in 3D views. [Figure 9](#) shows the results projected to the horizon slice. As shown in the results, the margin forms, sketch structures, and inner cavities of the reefs are clearly delineated. These areas may be significant for seismic exploration. The experiment indicates that our method is efficient at identifying the characteristics of reef reservoirs.

## **Conclusions**

The proposed workflow using the semi-supervised clustering method achieves the classification and identification process for the 3D seismic data volume. Compared to the traditional supervised and unsupervised clustering algorithms, the proposed method is simple and reduces the influence of the initial values and the number of sampling points in typical clustering algorithms. With the help of seismic attribution selected by geophysicists and supervised sampling operations, the proposed method obtains more accurate and reasonable classification results. In addition, the proposed algorithm adaptively optimizes the parameter models for each kind of geologic bodies in the target zone, which helps the interpreter investigate the geologic structure of the target area, as well as compensates for the uncertainty and subjectivity of the seismic interpreter. Finally, two real seismic data sets are used to illustrate its feasibility.

## **References Cited**

- Ahalt, S.C., A.K. Krishnamurty, P. Chen, and D.E. Melton, 1990, Competitive Learning Algorithms for Vector Quantization: Neural Networks 3, p. 277-291.
- Davis, J., 1986, Statistics and data analysis in geology, 2<sup>nd</sup> ed., Wiley, 646 p.
- Dumay, J., and F. Fournier, 1988, Multivariate statistical analyses applied to seismic facies recognition: Geophysics, v. 53, p. 1151-1159.
- Hagan, D.C., 1982, The applications of principal component analysis to seismic data sets: Geoexploration, v. 20, p. 93-111.
- Mathieu, P.G., and G.W. Rice, 1969, Multivariate analysis used in the detection of stratigraphic anomalies from seismic data: Geophysics, v. 34, p. 507-515.
- Saggaf, M.M., M.N. ToksÖzz, and M.I. Marhoon, 2003, Seismic facies classification and identification by competitive neural networks: Geophysics, v. 68/6, p. 1984-1999.

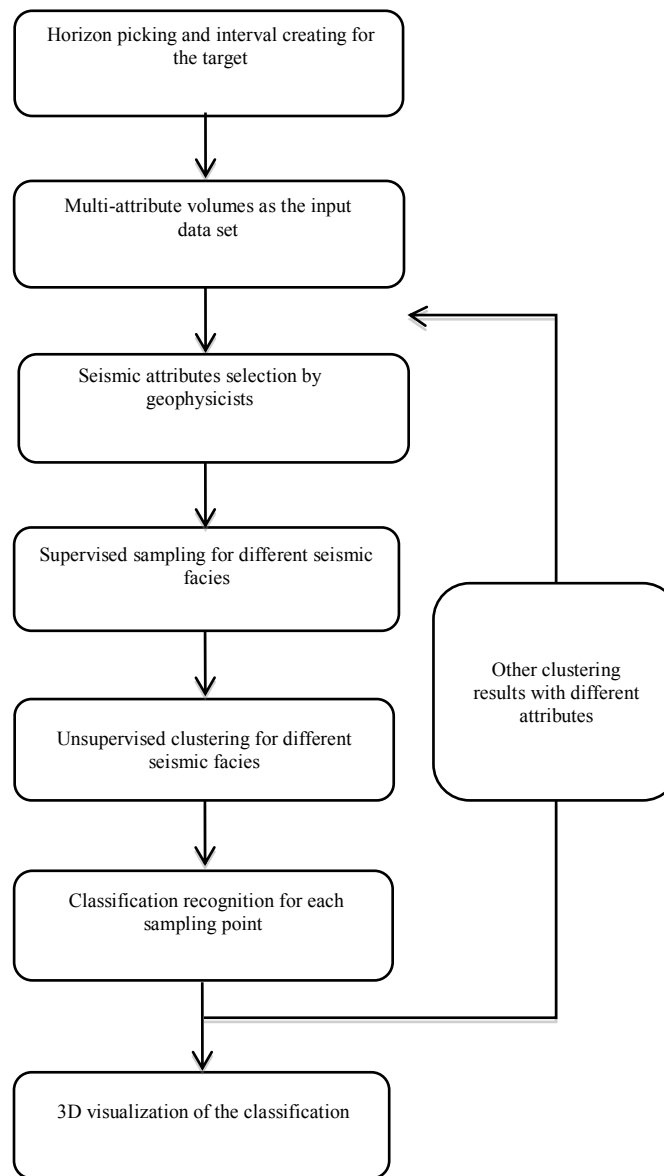


Figure 1. General semi-supervised classification workflow chart of seismic facies volume.

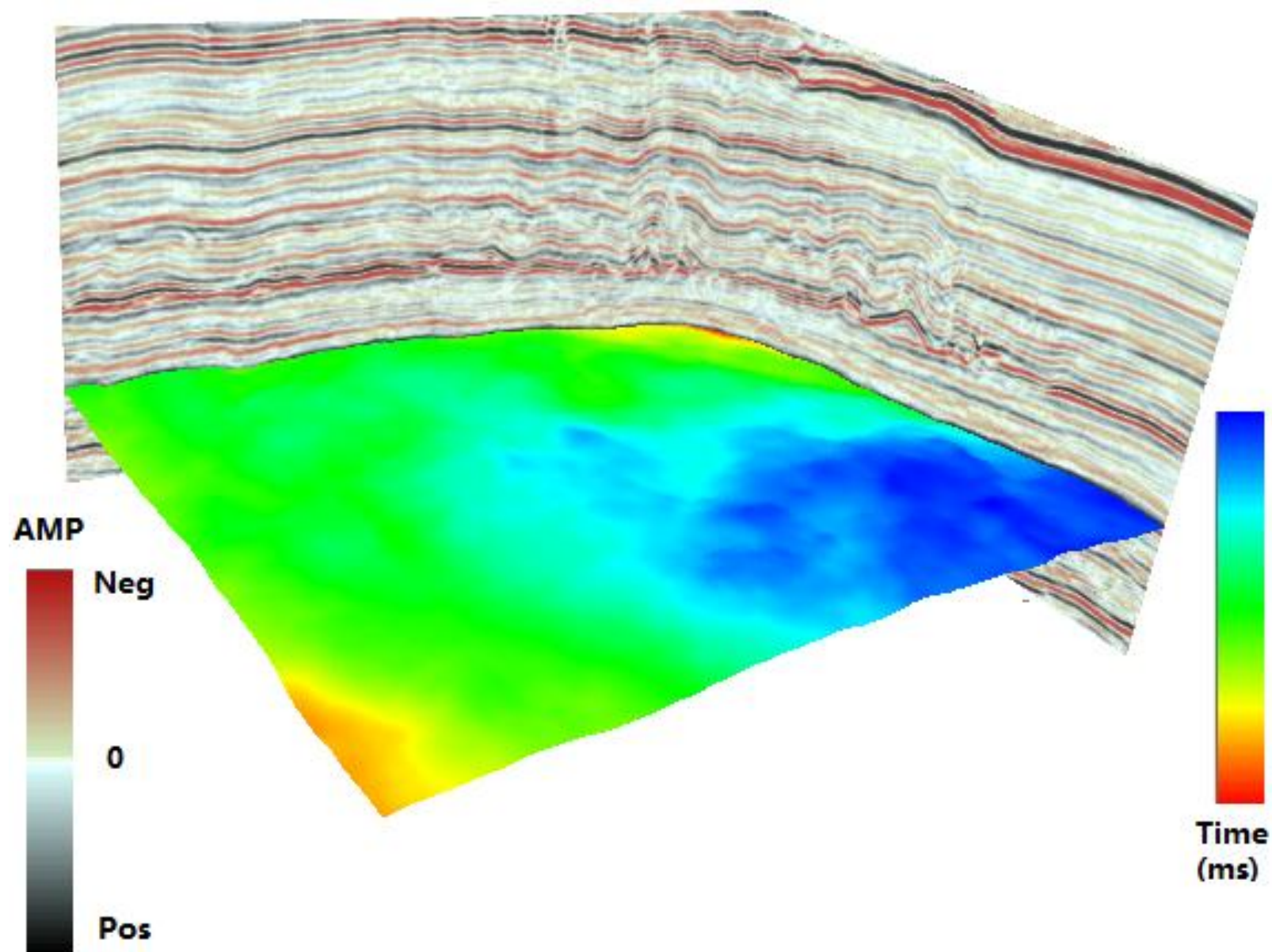
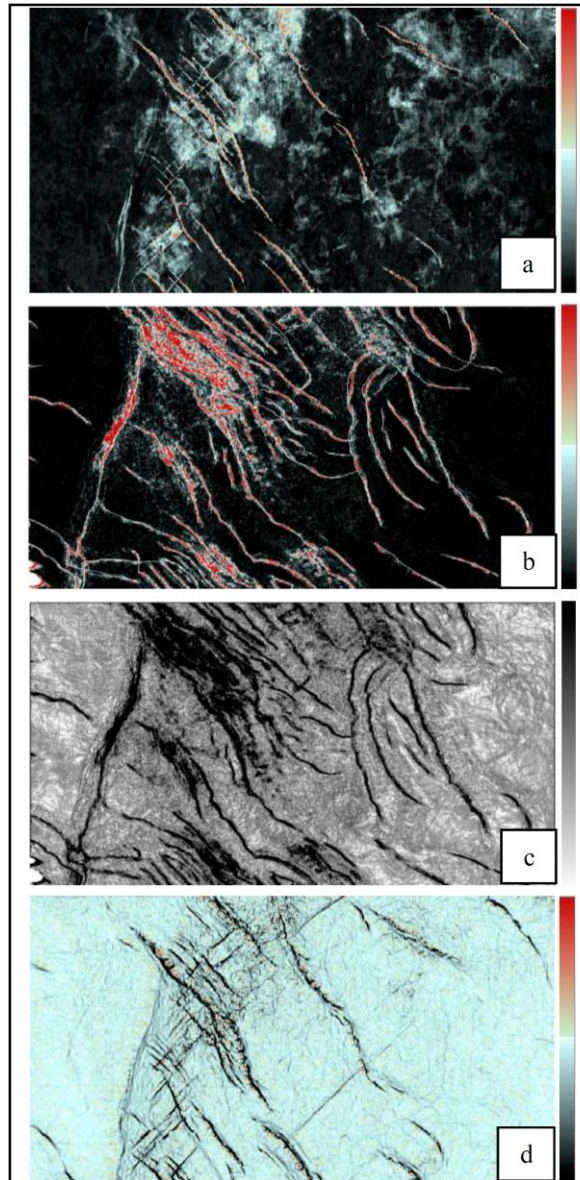


Figure 2. Horizon slice of the target reservoir shown with the seismic data. The colorbar for the horizon slice shows two-way travel time (ms). Reds and yellows are lows and blues indicate highs. The seismic amplitude colorbar shows positive peaks in black and negative troughs in red.



Figures 3a, 3b, 3c, and 3d. These figures illustrate amplitude, coherence, sobel similarity and most- positive curvature respectively. These images are shown on the target horizon in the seismic data. The amplitude is shown by the colorbar indicating that peaks are blue and troughs are red. The curvature has a similar colorbar and the positive curvatures are red and the negative curvatures are blue. In the sobel similarity seismic attribute, the dark colors represent big similarity.

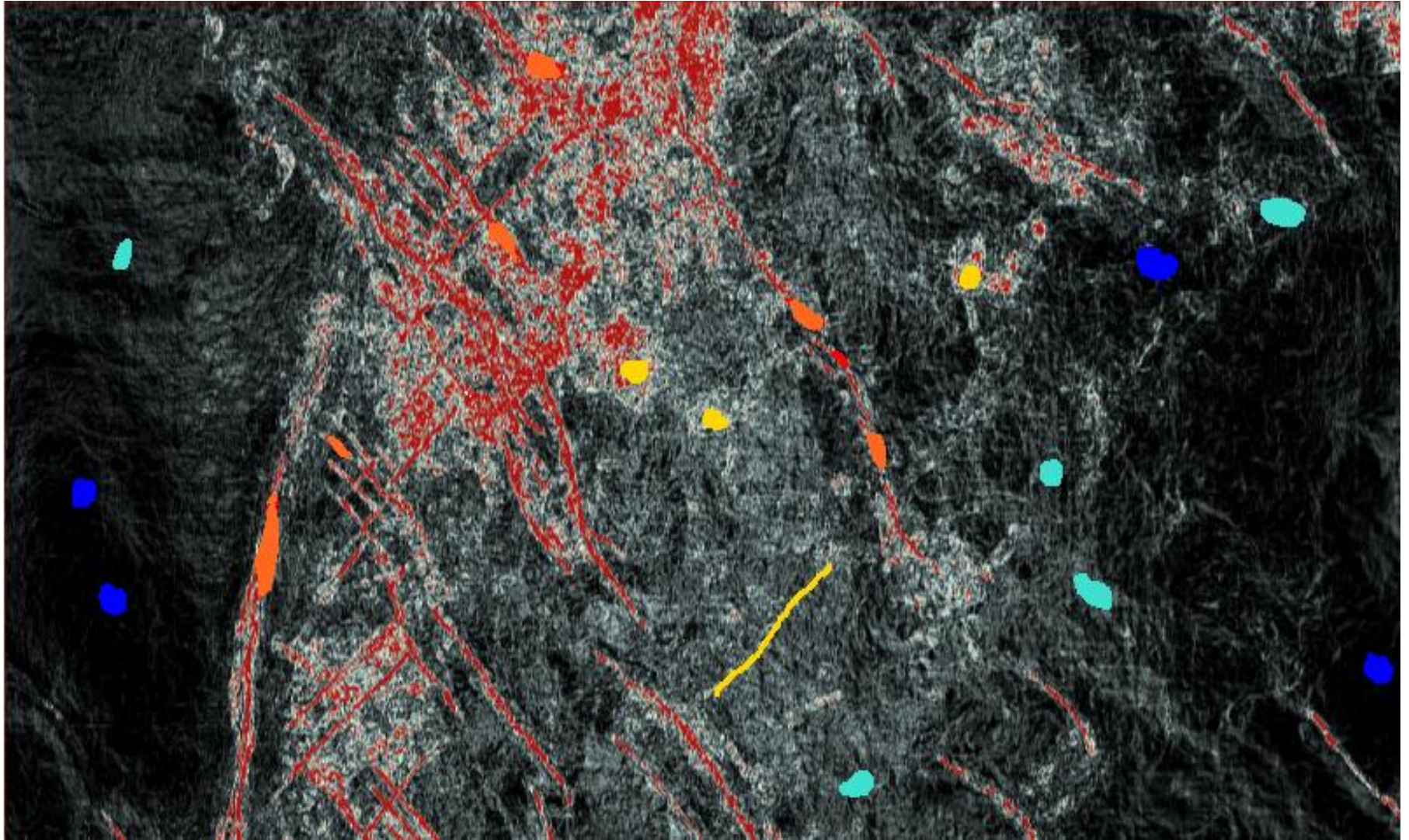


Figure 4. Supervised sampling points on the horizon slice of the coherence seismic attribute. The clustering colorbar shows various of faults shown by different colors. Orange points, yellow points, pale blue points and blue points represent the major faults, medium faults, subtle faults and background, respectively.

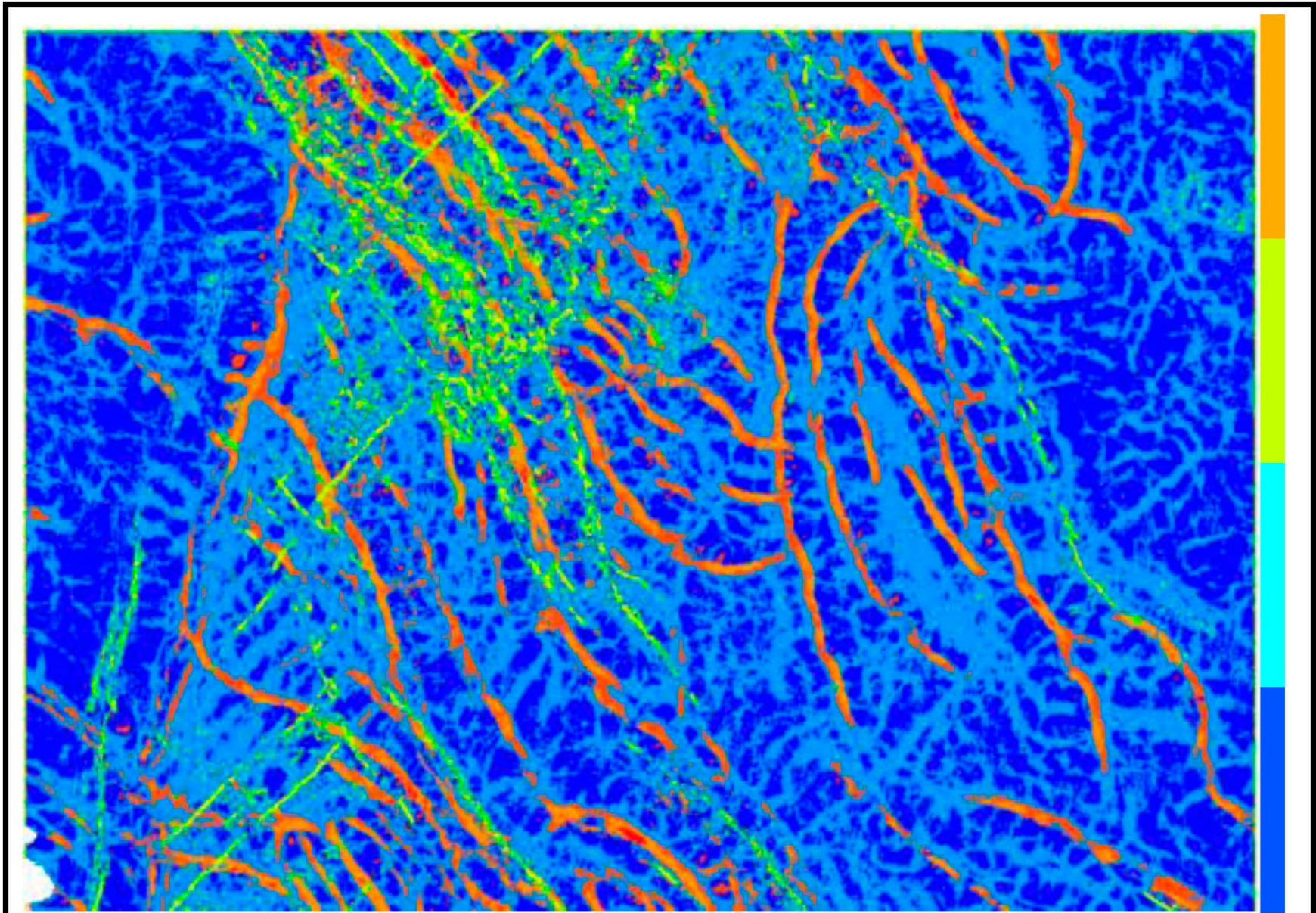


Figure 5. Classification volume projected on horizon slice of the target reservoir. The clustering colorbar shows various faults. In the original slice the major faults (orange), the medium faults (yellow) and the subtle faults (pale blue) are not clearly visible. The accuracy of this classification shows the faults significantly better.

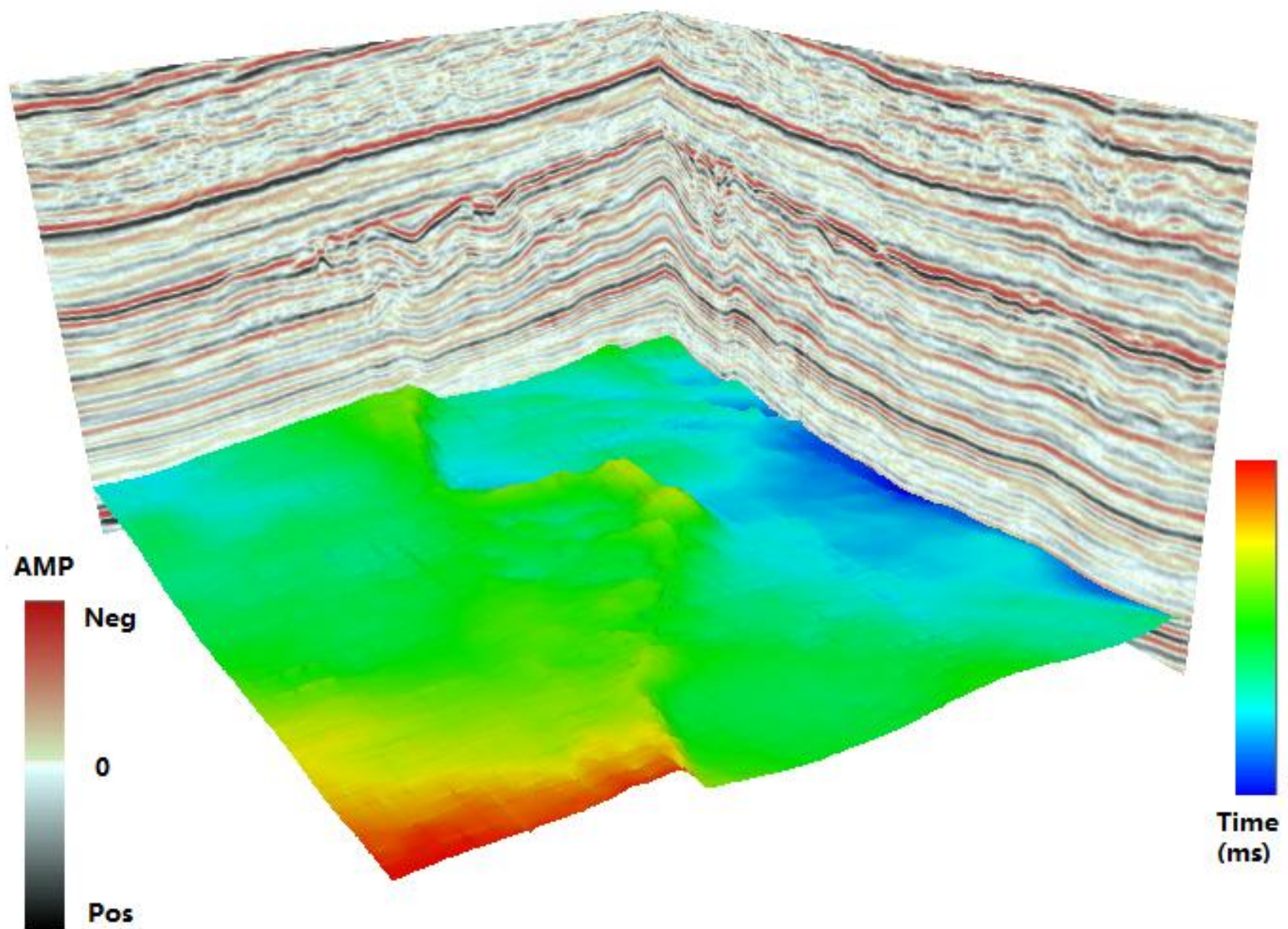


Figure 6. Horizon slice of the target reservoir shown with the seismic data. The colorbar for the horizon slice shows two-way travel time (ms). Reds and yellows are highs and blues indicating lows. The seismic amplitude colorbar shows positive peaks in black and negative troughs in red.

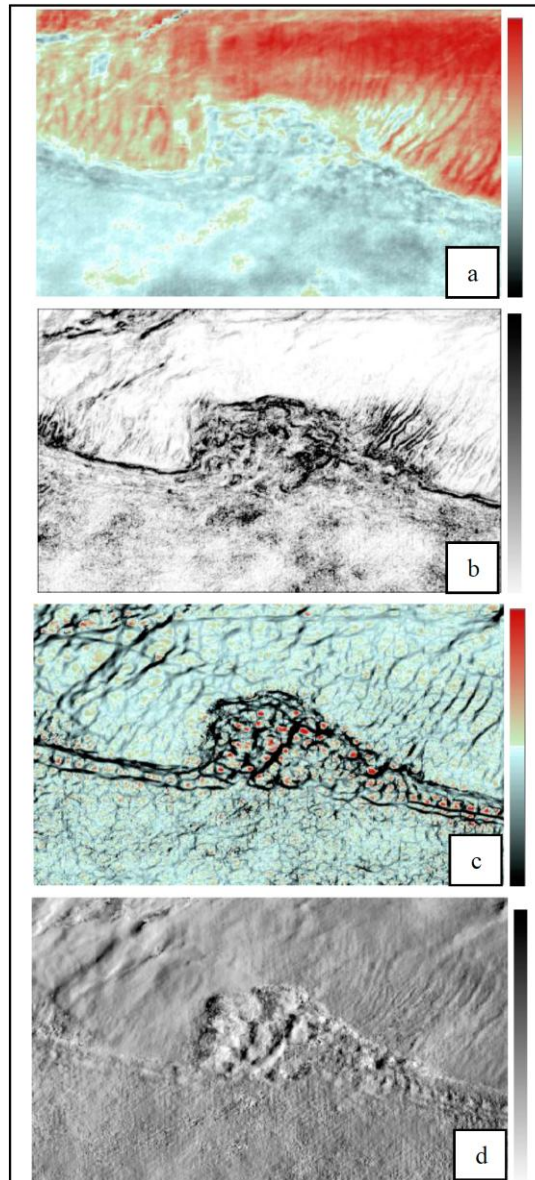


Figure 7a, 7b, 7c, and 7d. These figures illustrate amplitude, sobel similarity, most-positive curvature and coherence respectively. These images are shown on the target horizon in the seismic data. The amplitude is shown by the colorbar indicating that peaks are blue and troughs are red. The curvature has a similar colorbar and the positive curvatures are red and the negative curvatures are blue. In the sobel similarity seismic attribute, the dark colors represent big similarity.

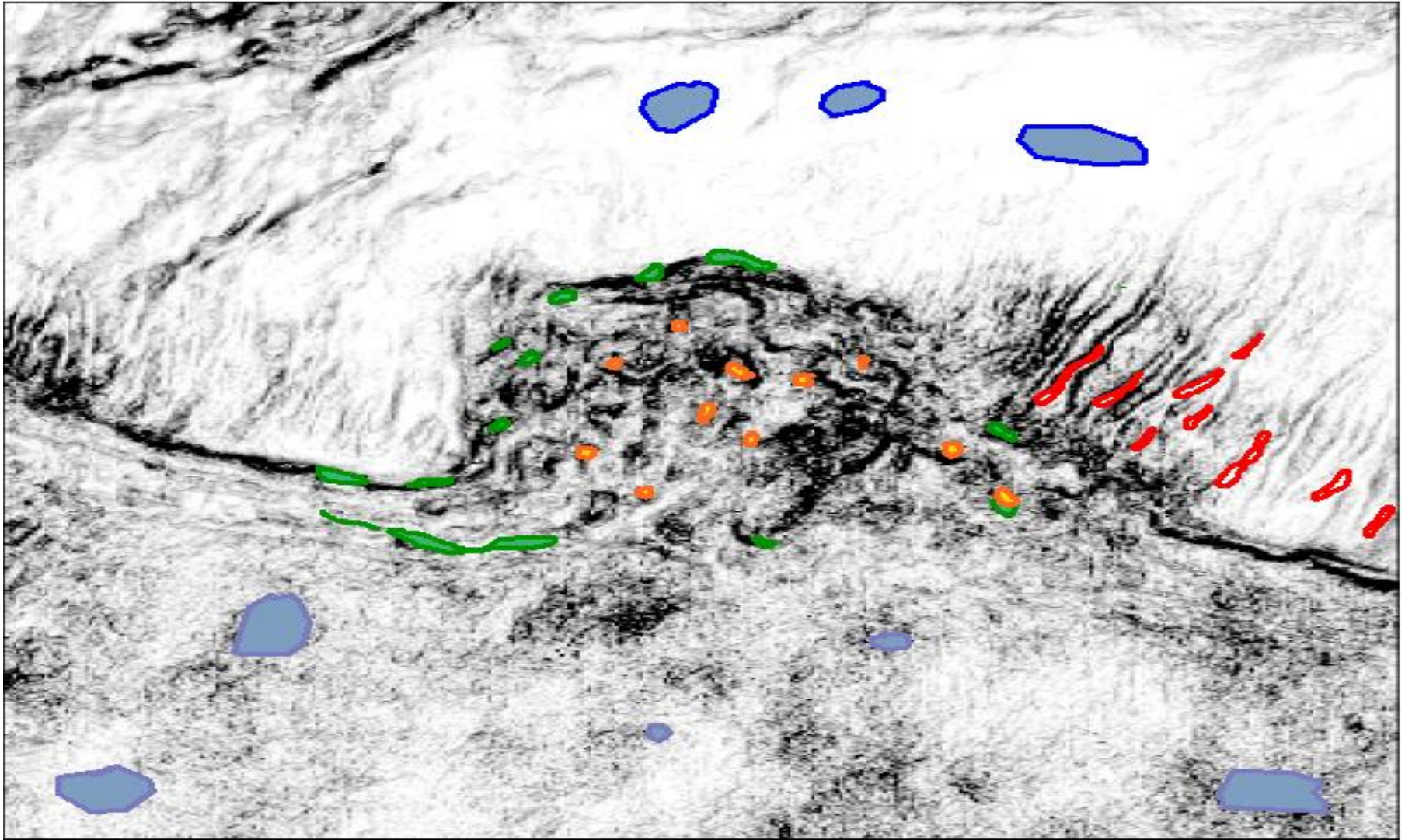


Figure 8. Supervised sampling points on the horizon slice of the soble similarity attribute. The clustering colorbar shows various of faults shown by different colors. Orange points, yellow points, pale blue points and blue points represent the maor faults, medium faults, subtle faults and background, respectively.

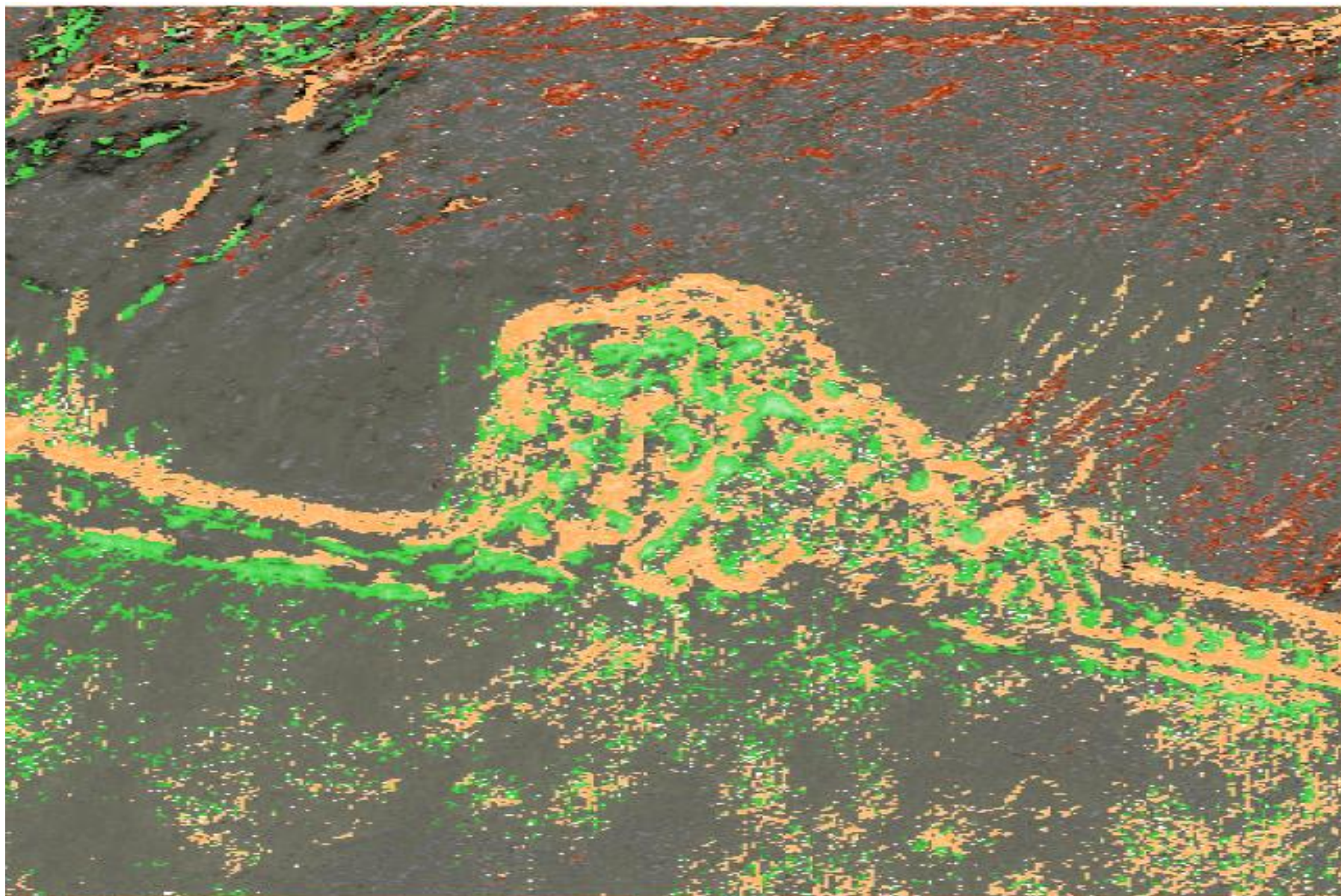


Figure 9. Classification volume projected on the horizon slice of the target reservoir.