Geological scenario identification using seismic impedance data

Geetartha Dutta
Stanford University
California, USA
gdutta@stanford.edu

Abstract

Geological scenario characterization is very important for making reservoir development decisions because the outcome of those decisions depends on it. The goal of this project is to identify the geological scenario of petroleum reservoirs given seismic impedance data using a convolutional neural network (CNN). A CNN is ideal for this task because it takes into account the spatial correlation in the data, which is crucial for discriminating between different geological scenarios. Here, the CNN is trained on synthetic seismic data generated using rock physics forward modeling. To approximate the low resolution real seismic data that is usually collected in the field, an averaging filter is applied to the generated impedance at the geostatistical scale. Both high quality and low quality seismic data are considered. The CNN is found to work quite well on both these types of data, and, as expected, a higher classification accuracy is obtained for the high quality seismic data. This methodology is more robust than other methodologies for geological scenario characterization using geophysical data, which require different transforms or distance measures.

1. Introduction

The geological scenarios of different subsurface petroleum reservoirs are different. The geological scenario of a conventional petroleum reservoir is usually determined by the depositional environment under which the sediments composing the reservoir were deposited over geologic time. For example, the depositional environment could be fluvial or deltaic, and with or without the formation of crevasse splays. Also, the orientation and position of these channels and deltas could be altered over time by processes such as aggradation and progradation. Even after deposition and consolidation of sediments has ceased, processes like erosion can dramatically change the geomorphological features of the rocks. Therefore, correctly identifying the geological scenario of petroleum reservoirs is a very challenging problem.

It is important to identify the geological scenario of petroleum reservoirs because the outcome of reservoir development decisions, like drilling new wells or setting well control parameters, depends on it. A correctly identified geological scenario might make the difference between drilling a well at the right place and finding oil, and drilling at the wrong place and incurring huge losses. This is due to the complex spatial heterogeneity that characterizes petroleum reservoirs. A typical petroleum reservoir consists of many different lithofacies, each with different petrophysical properties such as porosity, permeability and density, distributed spatially in a complex manner. Different geological scenarios require different production strategies to optimize oil production, and hence to maximize profits. For example, a single well might be sufficient to extract oil from a reservoir with conducting, highly permeable sands, while multiple wells are required in compartmentalized reservoirs with impermeable boundaries. Therefore, it is very important to identify the geological scenario of the reservoir before making reservoir development decisions.

Different types of geophysical data can be used to identify the geological scenario of a reservoir. But this is very challenging because usually spatial geophysical data have low resolution, of the order of tens of meters, and cannot resolve small-scale variability in reservoir properties. Well log data have high resolution, of the order of a meter, but they are collected only at well locations and hence are sparse, and therefore do not inform much about the large scale spatial variability in reservoir properties characteristic of a particular geological scenario. Spatial geophysical data, such as seismic data or electromagnetic data, can inform us about the large scale spatial variability in the reservoir, even though they are not perfect information because of their lower resolution.

This project aims to classify the geological scenario of subsurface petroleum reservoirs using convolutional neural networks on seismic acoustic impedance data. A synthetic dataset has been created using geostatistical simulation to generate the reservoir properties, and then using rock physics modeling to generate the synthetic acoustic impedance data. The acoustic impedance data thus generated, and the associated geological scenario labels, comprise the dataset which will be used to train and test a convolutional neural network. Convolutional neural networks can be expected to give better classification results than other classification techniques because they can capture complex patterns in the data. To evaluate the results, the classification accuracy on the test data is a good metric to use. By testing the trained convolutional neural network on the test data, a confusion matrix can be obtained which will indicate how well the network performs on the dataset.

2. Related Work

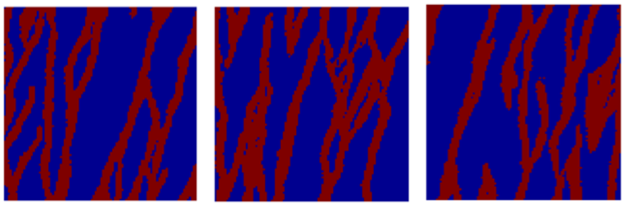
Previous work in identifying geological scenarios from geophysical data has been done by Trainor-Guitton (2010), who used directional variograms on inverted images obtained from electromagnetic data to classify the realizations into different geological scenarios. Transient or time-domain electromagnetic (TEM) data, which are magnetic field responses over time, were inverted to obtain a layered model of electrical resistivity and thickness values. The electrical resistivity maps thus obtained were used to generate lithological images, considering different electrical resistivity values for different lithologies. These lithological images were used to infer the channel direction and hence the geological scenario. However, the possible scenarios considered consisted of only three different orientations of channels, which is probably too simplistic. In addition to uncertainty in the orientation, there might be uncertainty in the depositional environment, types of lithofacies present, and their proportions and spatial distributions. Also, the resolution of land-based TEM measurements is very low because of the high in-line measurement separation, and hence the classification accuracy obtained is quite low (about 50%) for this three-class classification problem.

Yang et al. (2015) used different distance measures to discriminate geological and rock physics scenarios from time-lapse seismic data. Seismic data has higher resolution than electromagnetic data, but still its resolution is quite low compared to the geostatistical scale at which the reservoir properties are simulated. Besides, the time-lapse signature might be caused by various factors such as fluid saturation change, pressure change, etc. So, it is not trivial to retrieve meaningful information from time-lapse seismic data which could help in making better reservoir development decisions. Therefore, a sensitivity analysis using Distance-based Generalized Sensitivity Analysis (DGSA) was first performed to determine the rock physics and seismic parameters that influence the time-lapse signature the most, to better model the uncertainty in the forward time-lapse seismic response. They considered slightly more complex scenarios; in addition to different orientations, they considered different shapes of geological bodies (channels and ellipsoids) and different lithofacies (channel sand, splay sand and background shale). To discriminate between these different scenarios, different distance measures were used to capture the complex and non-stationary spatial patterns in the realizations, including Principal Component Analysis (PCA), Fractal Dimension and Clustered Histogram of Patterns (CHP). Then Bayesian classification was used to classify the realizations based on these distance measures and the test accuracy was evaluated for each distance measure. It was found that CHP gave the highest overall test accuracy of about 70%. The performance of this methodology depends on the problem at hand because one particular distance measure might work well for discriminating some particular types of geological scenarios, but might not effectively distinguish between other types of scenarios. In short, there is no distance measure that works well universally for all types of geological scenarios.

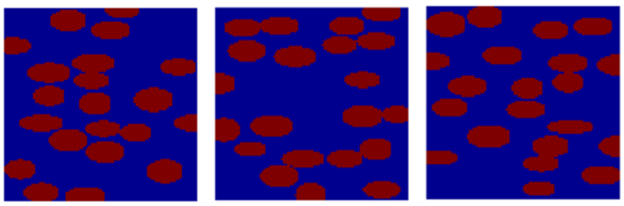
It is clear from the above discussion that a more robust classification technique that performs well over many different geological scenarios incorporating the complex spatial and lithological variability in the reservoir is required. The technique should be flexible enough to distinguish between non-stationary spatial patterns, yet powerful enough to detect large-scale regional variations in the data. It should be able to learn from features at multiple scales to successfully discriminate lithological and structural differences between different geological scenarios. A Convolutional Neural Network (CNN) seems ideal for this purpose, as it takes into account the spatial correlation in the data, and has numerous parameters that can be learned by proper training to build a classifier which can distinguish between both large-scale and small-scale features in the data.

3. Dataset Generation

To model the dataset, we need to first model the uncertain reservoir properties corresponding to each geological scenario. In this project, two different geological scenarios, having different depositional environments, are considered: channels and mounds. 1000 facies realizations (samples) are simulated corresponding to each of these scenarios on a 100 x 100 grid. The facies realizations for the channel scenario are simulated using the multiple-point geostatistical algorithm Single Normal Equation Simulation or SNESIM (Strebelle, 2002), with 65% floodplain facies and 35% channel facies (SGeMS, 2011). On the other hand, the realizations for the mound scenario are simulated using object based Boolean simulation with ellipsoidal objects (Deutsch and Journel, 1998). The major and minor axes of the ellipsoids are sampled from Gaussian distributions with means 10 and 5 respectively, and variances 1 and 0.5 respectively. The mounds represented by the ellipsoids occupy 30% of the volume of the reservoir, and there is no overlap between them. Figure 1 shows three facies realizations of each geological scenario thus generated.

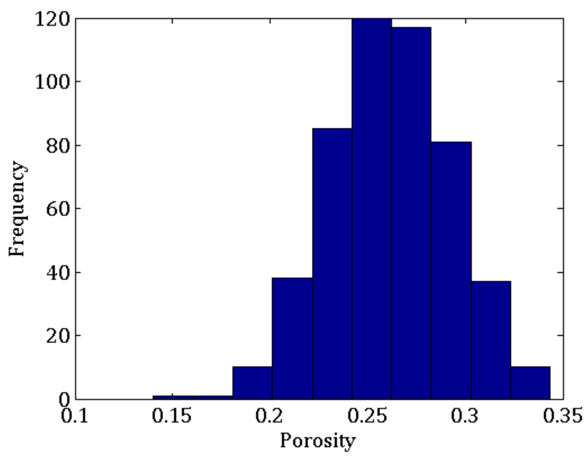


(a)

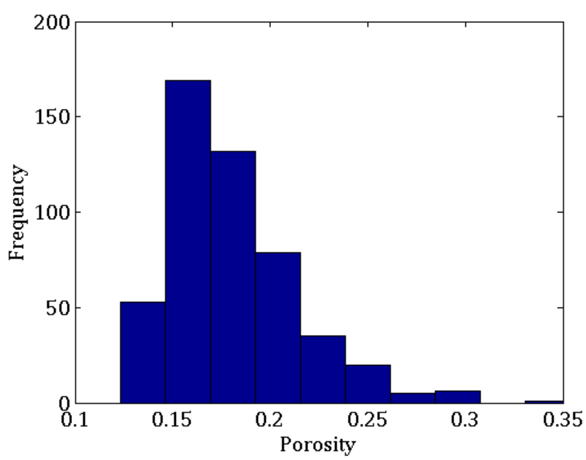


(b)

Figure 1: Three facies realizations each of (a) the channel geological scenario, and (b) the mound geological scenario.



(a)



(b)

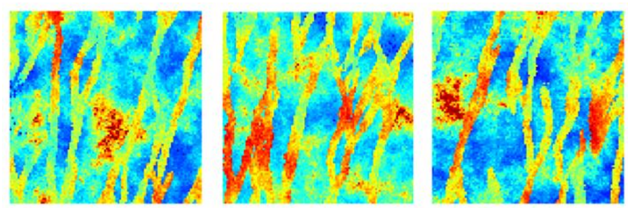
Figure 2: Histogram of porosity in (a) the sand facies, and (b) the background facies.

The porosity is then simulated conditioned to the facies, using the two-point simulation algorithm Sequential Gaussian Simulation (SGSIM). To do so, we basically require the following information:

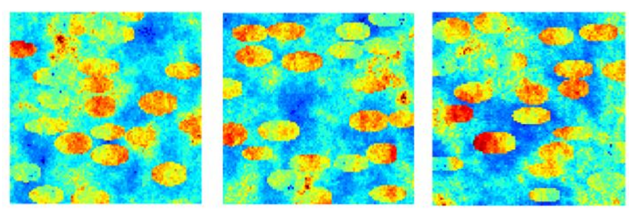
a) Histograms of the sand and background facies which represent the distributions of porosity in these two facies.
b) Variograms representing the spatial correlation of the porosity distribution in both the geological scenarios.

The histograms of the porosity distributions in each facies are obtained from prior knowledge of the reservoir. Figure 2 shows the histograms of porosity in each facies that are used for this purpose.

The generated porosity realizations for the two facies are then combined using the Cookie-Cutter method to obtain the porosity realizations for the reservoir. It is necessary to model the porosity of the reservoir because it affects the density of the reservoir and the velocity of seismic waves through the reservoir, both of which affect the acoustic impedance in turn. Figure 3 shows the porosity realizations corresponding to the facies realizations shown in Figure 1.



(a)



(b)

Figure 3: Porosity realizations corresponding to the facies realizations in Figure 1 for (a) the channel geological scenario, and (b) the mound geological scenario.

Given the porosity and the constituent minerals for each facies, the density of the reservoir can be computed as the weighted average of the densities of its constituents, weighted by their fractions. Table 1 shows the constituents of each facies along with their fractions. In addition, the sand facies is assumed to contain 1% cement. Also, the water saturation and the oil saturation in the sand facies are assumed to be 15% and 85% respectively, and those in the background facies are assumed to be 30% and 70% respectively.

Facies	Constituent	Fraction
Sand	Quartz	0.65
	Feldspar	0.20
	Rock fragments	0.15
Background	Clay	0.85
	Quartz	0.15

Table 1: The constituents of the two facies and their fractions.

Using the constituent densities shown above, the density realizations of the reservoir are modeled using the porosity and the facies realizations. Figure 4 shows the density realizations corresponding to the facies realizations of Figure 1. The density is one of the two reservoir properties that directly affects the acoustic impedance, the other being seismic velocity (the acoustic impedance is defined as the product of the density and the velocity). So, to model the seismic data, i.e. the acoustic impedance, the seismic velocity need to be forward modeled. To do so, the constant cement model (Avseth et al., 2000) is employed. This model assumes that the reservoir rocks have gone through the same diagenesis process, which is a reasonable assumption if the vertical depth range of the reservoir is not very high, like in our reservoir model. The constant cement model first computes the elastic moduli of the dry rocks and then applies the Gassmann equation (Gassmann, 1951) to compute the elastic moduli for the saturated rocks and hence their velocities (Mavko et al., 2009).

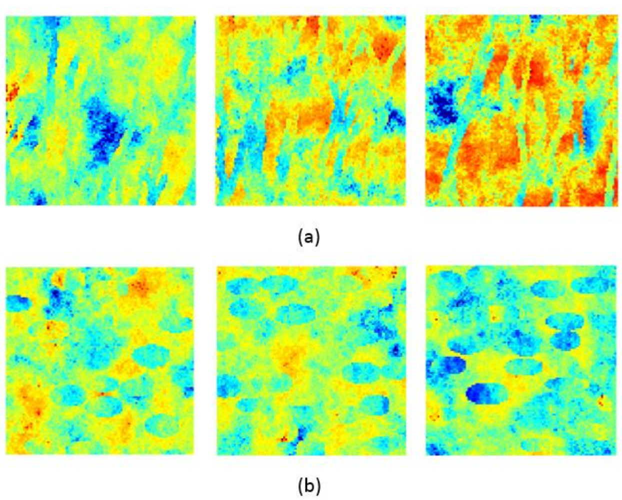


Figure 4: Density realizations corresponding to the facies realizations in Figure 1 for (a) the channel geological scenario, and (b) the mound geological scenario.

To model the seismic data, the acoustic impedance (AI) for each realization is first modeled at the geostatistical scale using the density and the velocity realizations already generated. However, the AI obtained from inverting real seismic data is usually of much lower resolution than the AI at the geostatistical scale. Therefore, a moving mean filter is applied to the AI at the geostatistical scale to approximate the AI at the seismic scale. Here, we assume two kinds of seismic data – high quality and low quality. The high quality seismic data is obtained by applying a moving mean filter with window size of 2x2, and the low quality seismic data is obtained by applying a moving mean filter with window size 4x4. Figure 5 shows the AI at the geostatistical scale, and Figure 6 shows the AI at the seismic scale for the low quality seismic data.

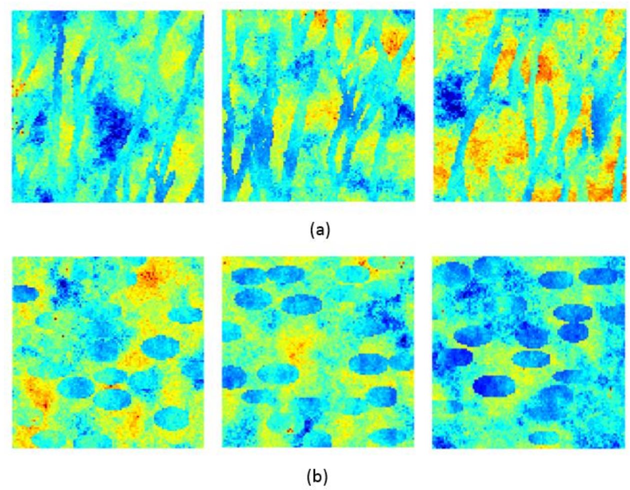


Figure 5: AI at the geostatistical scale corresponding to each facies realization shown in Figure 1.

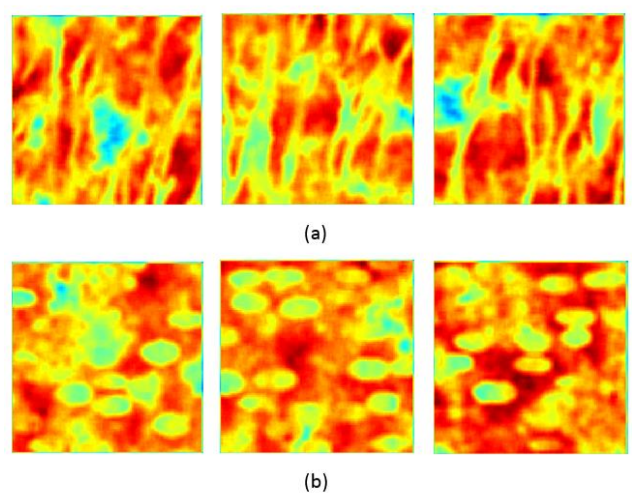


Figure 6: AI at the seismic scale corresponding to each facies realization shown in Figure 1 for the low quality seismic data.

Thus, we have 2000 samples of the AI at the seismic scale, which are the data, and their corresponding labels (the geological scenario). Each of the AI samples is on a grid of dimension 100 by 100, which can be stretched out into a vector of size 10000. Thus, we have a data matrix of dimension 2000 by 10000, and their associated labels. This dataset is split into a training set containing 1800 samples (900 samples of each scenario), and a test set containing the remaining samples. This will be used to train and test a CNN to classify each realization into one of the two geological scenarios. The accuracy of the classification will be determined using a confusion matrix, which shows the percentage of samples belonging to each scenario that are classified as each scenario.

4. Methods and Experiments

A CNN with the following architecture is trained on the dataset (MATLAB 2017a):

[32 channel 29x29 CONV, stride 1, padding 2]
[ReLU]
[3x3 Max Pooling, stride 2, padding 0]
[64 channel 29x29 CONV, stride 1, padding 2]
[ReLU]
[3x3 Max Pooling, stride 2, padding 0]
[64 channel Fully Connected]
[ReLU]
[2 channel Fully Connected]
[Softmax]

The input image is 100x100 with 1 channel, and the output is the probability of each geological scenario. A large kernel size is used for the convolutional layers because we would like to capture the large scale spatial correlation in the data which is characteristic of each geological scenario, and hence very important for discriminating them. This CNN is first trained on the low-quality training data using stochastic gradient descent with momentum, with a momentum of 0.9, learning rate of 0.001, decay rate of 0.95, L2 regularization of 0.004, and minibatch size of 128. After training for 40 epochs, a training classification accuracy of 81.2% is obtained. The first layer weights after training is shown in Figure 7.

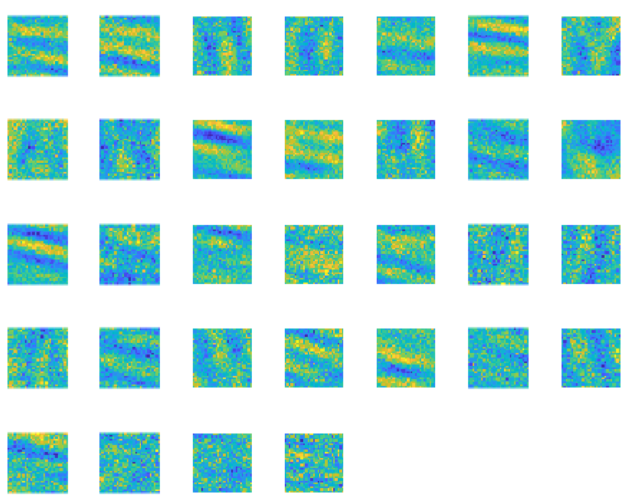


Figure 7: First CONV layer weights after training on the low-quality data.

This trained CNN is evaluated on the test set and the results are summarized in the confusion-matrix in Table 2. It is seen that the overall test accuracy is 81.5%, with perfect classification for the mound geological scenario and not very good classification for the channel geological scenario. The reason why some of the channel realizations are classified as mounds by the CNN might be because of the blobs present in some of the channel realizations which look like mounds.

	Predicted channel	Predicted mound
True channel	63	37
True mound	0	100

Table 2: Confusion matrix showing test accuracy of the CNN when trained on the low-quality dataset.

The CNN is then trained on the high-quality seismic acoustic impedance dataset. The same architecture and set of hyperparameters is used. The first layer weights for the CNN trained on the high-quality data is shown in Figure 8. It is seen that the weights learn some elongated channel-like features and some blobby mound-like features. This trained CNN is again evaluated on the test data and the results are shown in the confusion matrix in Table 3. It is seen that the overall test accuracy is 91.5%, with good classification accuracy rates for both the geological scenarios. The increase in the classification accuracy for the channel geological scenario can be attributed to the higher resolution in the dataset. On the other hand, the classification accuracy for the mound geological scenario is almost the same as that using the low-quality data. This shows that the resolution of the seismic impedance data is important for detecting channels, but not very important for detecting mounds.

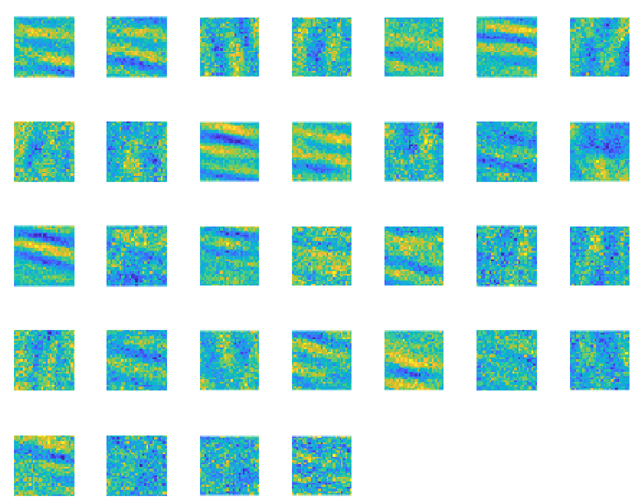


Figure 8: First CONV layer weights after training on the high-quality data.

	Predicted channel	Predicted mound
True channel	84	16
True mound	1	99

Table 3: Confusion matrix showing test accuracy of the CNN when trained on the high-quality dataset.

5. Conclusions and Future Work

A robust and flexible methodology to characterize the geological scenario of petroleum reservoirs from seismic impedance data has been demonstrated using a CNN. This methodology is robust because, unlike other methodologies discussed in the Related Work section, it does not involve computing different types of transforms or distance measures on the data and can make predictions based on the raw spatial data. This methodology is also flexible because the same network can be used to classify based on different types of spatial geophysical data.

The CNN used in this project is trained on both high-resolution and low-resolution seismic acoustic impedance data, and the effect of the change in resolution is evident in the classification accuracy rates. This type of analysis could be very useful prior to actually collecting the data to determine the minimum resolution required for the classification to give satisfactory results and hence help in proper design of seismic surveys.

In this project, only seismic acoustic impedance data is used to classify the geological scenario, but future work might combine various types of geophysical data to make predictions on the geological scenario. These include various seismic attributes such as the seismic S-wave impedance, velocities of seismic P and S waves (V_p and V_s), and the V_p/V_s ratio. So, instead of a “one to one” classification, it would be a “many to one” classification, and can be performed using Recurrent Neural Networks (RNNs) instead of CNNs. The classification accuracy rates might increase if multiple data sources are used.

Another direction for future research is to perform a value of information (VOI) analysis (Howard, 1966) on various geophysical spatial data for geological scenario characterization. To do a VOI analysis, a decision situation need to be framed, and the prospect values evaluated for each possible decision alternative. Then, the effect of the geophysical data on the outcome of our decision alternatives need to be evaluated using an information reliability measure (Trainor-Guitton, 2010). The VOI is computed before actually collecting the data, and is a monetary estimate of the additional value resulting from acquiring the data before making the reservoir development decision. Thus, a VOI analysis can tell us whether it is worthwhile to collect the data, and if so, how much and what type of data to collect. These are important questions because geophysical data, such as seismic data, are expensive to collect and process.

References

[1] P. Avseth, J. Dvorkin, G. Mavko, and J. Rykkje. Rock physics diagnostic of North Sea sands: Link between microstructures and seismic properties. Geophysical Research Letters, v. 27, no. 17, p. 2761-2764, 2000.

[2] C. V. Deutsch, and A. G. Journel. GSLIB: geostatistical software library and user’s guide: 2nd ed. Oxford University Press, 369 p., 1998.

[3] F. Gassmann. Uber die Elastisitat Poroser Medien. Vierteljahrschr. Naturforsch. Ges. Zurich, 96, 1-23, 1951.

[4] R. A. Howard. Information Value Theory. IEEE Transactions on Systems Science and Cybernetics, v. 2, no. 1, p. 22-26, 1966.

[5] MATLAB and Neural Network Toolbox Release 2017a, The MathWorks, Inc., Natick, Massachusetts, United States. 2017.[https://www.mathworks.com/downloads/web_downloads/get_release?release=R2017a].

[6] G. Mavko, T. Mukerji, and J. Dvorkin. The Rock Physics Handbook: Tools for Seismic Analysis of Porous Media, 2nd ed. Cambridge University Press, 2009.

[7] SGeMS 2.5beta, Stanford Center for Reservoir Forecasting (SCRF), Stanford University, California, USA. 2011. [http://sgems.sourceforge.net/?q=node/77]

[8] S. Strebelle. Conditional simulation of complex geological structures using multiple-point statistics. Math Geol, 34:1-21, 2002.

[9] W. J. Trainor-Guitton. On the value of information for spatial problems in the earth sciences. Stanford University, 2010.

[10] G. Yang, and J. Caers. Analysis of distance for discriminating geological and rock physics scenarios from 4D seismic data. SCRF annual affiliates meeting, 2015.