Reliability of Seismic Data for Hydrocarbon Reservoir Characterization

Geetartha Dutta (gdutta@stanford.edu)

December 10, 2015

Abstract

Seismic data helps in better characterization of hydrocarbon reservoirs. However, seismic data is not perfect, and it is important to measure the reliability of the data. In this project, the reliability of seismic data is quantified by generating many stochastic realizations of reservoir properties from two possible geological scenarios of a synthetic hydrocarbon reservoir, forward modeling the seismic data on each of those realizations using rock physics modeling, extracting features from the seismic data, and then using classification algorithms to classify a test set of realizations into either geological scenario and computing the classification success rate.

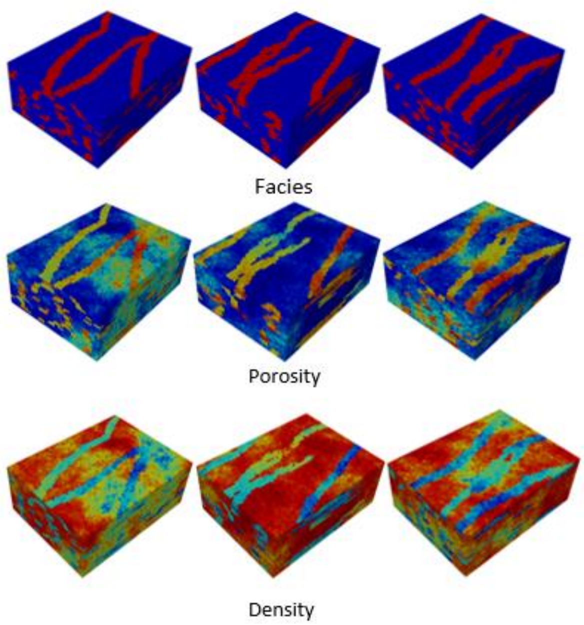
1 Introduction

To aid in making better reservoir development decisions like choosing a well location or well controls, many stochastic models of the reservoir are usually generated conditioned to seismic data. However, seismic data is not perfect as it has lower resolution than the models used to characterize the reservoir, and hence many different reservoir models could satisfy a certain seismic dataset [1]. Thus, it is important to estimate the reliability of seismic data and quantify the uncertainty in the decision making process.

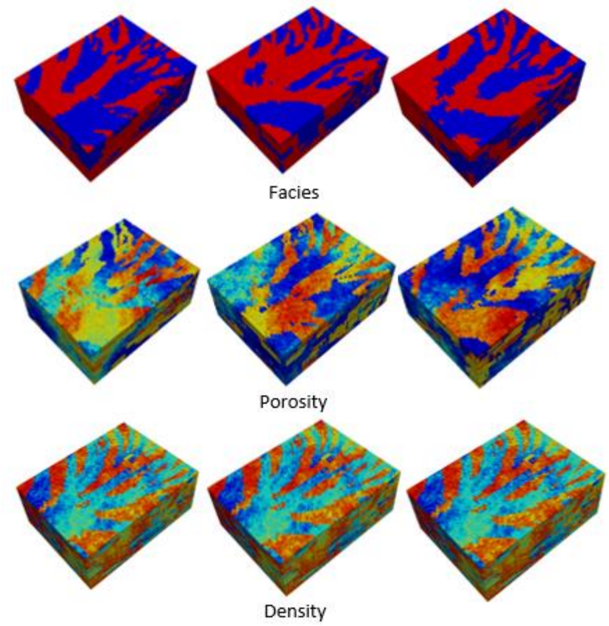
Previous work in this field has tried to model the geological scenario from seismic data using pattern similarity techniques such as Multiple Points Histogram (MPH) and Discrete Wavelet Transform (DWT) [2]. However, the reliability of such geological scenario prediction from seismic data has not been assessed quantitatively. This project aims to quantify the reliability of seismic data in geological scenario prediction using classification algorithms on forward modeled seismic data.

2 Modeling the Data

Single Normal Equation Simulation (SNESIM) [3] was used to generate 100 realizations of the reservoir property – facies – for each of two possible geological scenarios – a channel depositional system and a deltaic depositional system. Two other reservoir properties – porosity and density – were then simulated using Sequential Gaussian Simulation (SGSIM) conditioned on each facies realization. These realizations approximately define the uncertainty in the reservoir properties that seismic data might help to reduce. Figure (1) shows a few of the generated realizations for each scenario.



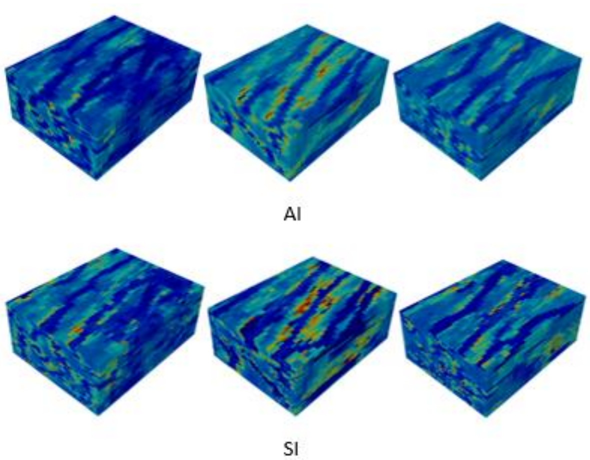
1. (a)



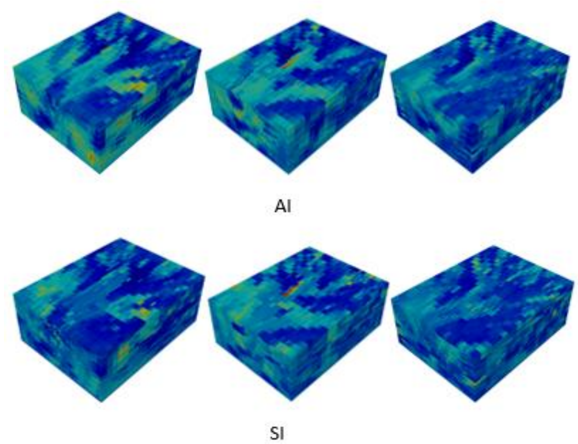
1. (b)

Figure 1. (a) Three facies, porosity and density realizations of the channel depositional system scenario, **(b)** Three facies, porosity and density realizations of the deltaic depositional system scenario.

Then, rock physics modeling was used to generate the elastic properties – acoustic impedance (AI) and S-wave impedance (SI) – at the same scale as the reservoir properties, i.e., the geostatistical scale. More specifically, the contact cement model was used with a coordination number of 7, uniform fluid mixing, cement faction = 5% and clay content = 70%. Finally, the Born filter was applied to these elastic properties at the geostatistical scale to generate the elastic properties at the seismic scale, which approximate the data that could be acquired by inverting real field seismic data. Figure (2) shows the elastic properties at the seismic scale for each geological scenario.



2. (a)

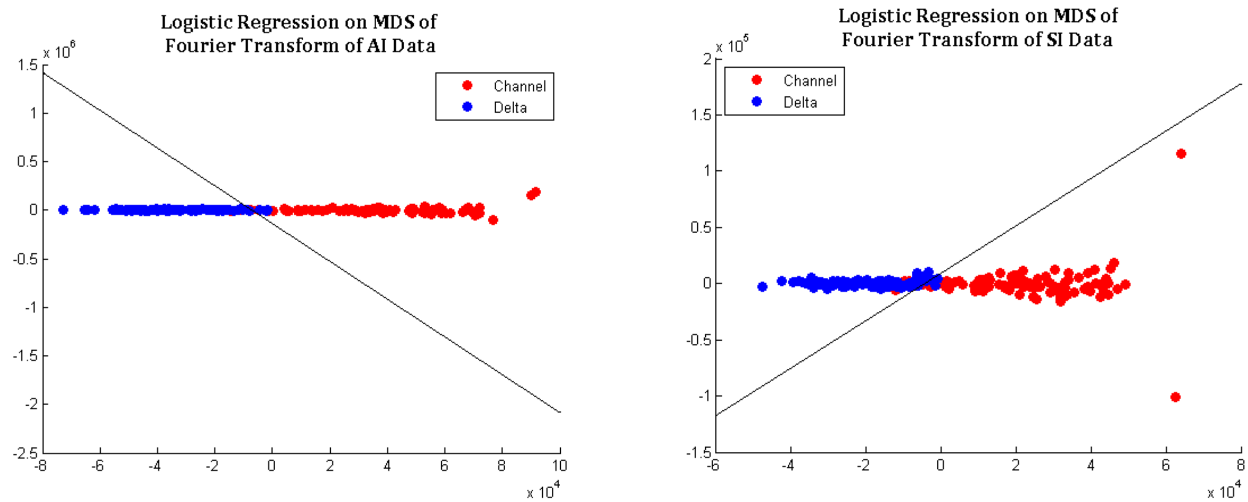


2. (b)

Figure 2. (a) Three AI and SI datasets of the channel depositional system scenario, **(b)** Three AI and SI datasets of the deltaic depositional system scenario.

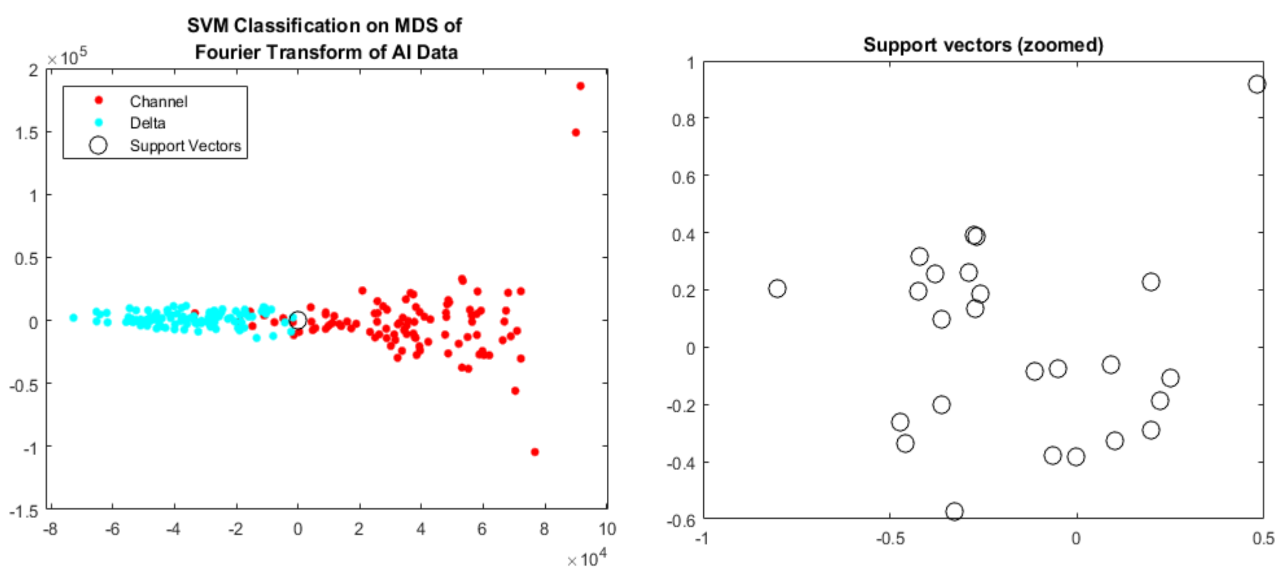
3 Extracting Features and Classification

To train classifiers to predict the geological scenario given the data, some characteristic features for each scenario need to be extracted from the data. Since the geological scenario influences the global characteristics of the models, the selected features should capture the global heterogeneity in the data, rather than the local heterogeneity such as the point-to-point Euclidean distances. The 3D Fourier transform was selected as the feature reflecting the global spatial heterogeneity. But since the 3D Fourier transform of the very high dimensional data is also very high dimensional, a Multi Dimensional Scaling (MDS) was performed on the 3D Fourier transform, and only the first two principal components were selected as features for classification.

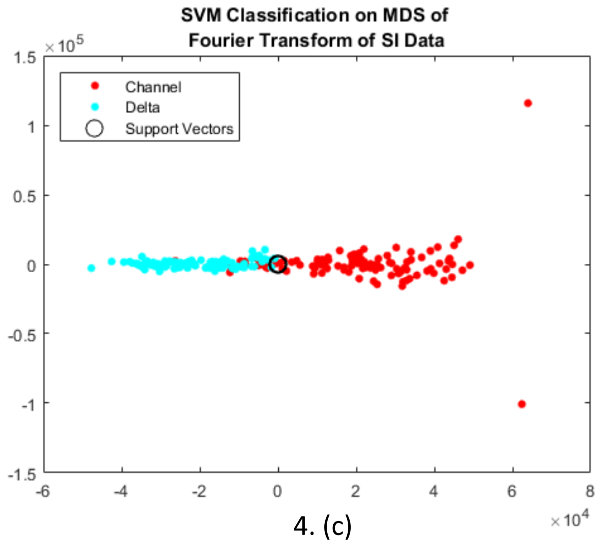


3. (a) 3. (b)

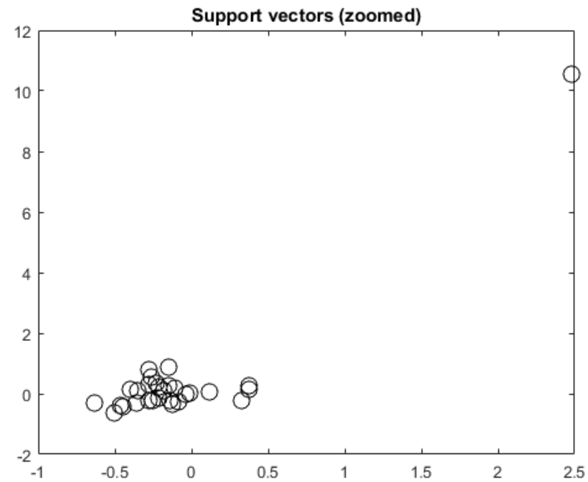
Figure 3. (a) Logistic regression applied to the MDS of the Fourier transform of the AI data, **(b)** Logistic regression applied to the MDS of the Fourier transform of the SI data.



4. (a) 4. (b)



4. (c)



4. (d)

Figure 4. (a) SVM classification on MDS of Fourier transform of AI data, **(b)** Support vectors of the SVM classifier based on AI data, **(c)** SVM classification on MDS of Fourier transform of SI data, **(d)** Support vectors of the SVM classifier based on SI data.

The features were used to train three different classifiers – Logistic Regression, Bayesian Classification and Support Vector Machine Classification – on a training set comprising 70% of the data. Each of these classifiers were then tested on the remaining 30% of the data. The process was repeated many times, each time taking a different set of data as the training / test data. Figure (3) and (4) show some of the plots obtained by applying Logistic Regression and Support Vector Machine classification to the data respectively.

4 Results and Conclusions

The average correct and incorrect classification rates of each classifier for each type of data are summarized in the form of confusion matrices shown in Figure (5).

Logistic Regression on AI

		Predicted	
Actual		Channel	Delta
	Channel	94.0%	6.0%
	Delta	4.0%	96.0%

Bayesian Classification on AI

		Predicted	
Actual		Channel	Delta
	Channel	89.1%	10.9%
	Delta	0.0%	100.0%

SVM Classification on AI

		Predicted	
Actual		Channel	Delta
	Channel	92.6%	7.4%
	Delta	3.3%	96.7%

Logistic Regression on SI

		Predicted	
Actual		Channel	Delta
	Channel	90.5%	9.5%
	Delta	3.6%	96.4%

Bayesian Classification on SI

		Predicted	
Actual		Channel	Delta
	Channel	83.6%	16.4%
	Delta	0.0%	100.0%

SVM Classification on SI

		Predicted	
Actual		Channel	Delta
	Channel	88.5%	11.5%
	Delta	1.3%	98.7%

Figure 5. Confusion matrices summarizing the results of classification using different classifiers on each type of data.

From the confusion matrices, it is seen that logistic regression and SVM classification seem to perform slightly better than Bayesian classification. This may be due to the stronger assumptions made in Bayesian classification. Also, between the two types of data, the AI data provide slightly higher classification success rates than the SI data. Overall, the seismic data are very reliable as evident from the high classification success rates.

5 References

[1] Mukerji, T., Mavko, G., & Rio, P., 1997, Scales of Reservoir Heterogeneities and Impact of Seismic Resolution on Geostatistical Integration, *Mathematical Geology*, v. 29, no. 7, p. 933-950.

[2] Jeong, C., Scheidt, C., Caers, J., Mukerji, T., 2014, Modeling Geological Scenario Uncertainty from Seismic Data using Pattern Similarity, *SEG Technical Program Expanded Abstracts 2014*: pp. 2444-2448.

[3] Strebelle, S., 2002, Conditional simulation of complex geological structures using multiple-point statistics, *Mathematical Geology*, v. 34, no. 1, p. 1-21.