TÉCNICO LISBOA

POLITECNICO DI TORINO

Department of Environment, Land, and Infrastructure Engineering
Master of Science in Petroleum Engineering

Supervisor:
Prof. Godio Alberto

Candidate:
VICTOR ENEMCHUKWU EJENAM
(S264680)

External Supervisor:
Prof. Leonardo Azevedo.

MARCH 2021

Abstract

In reservoir modeling and characterization, different seismic inversion techniques are conditioned by the available data provided by seismic surveys and the subsurface Petro-elastic properties obtained by wells. The inversion solution tries to provide a subsurface model that fits all the current observed information equally. A geostatistical framework is a natural solution to integrate both data within the same framework while assessing the spatial uncertainty. This thesis aims to develop and implement a machine learning (Functional Data Analysis) method to reduce subsurface models' computational time (acoustic impedance). The proposed method uses Principal component analysis to reduce the dimension of the data (Dimensionality Reduction). The time taken to obtain the result by the proposed methodology is compared with the time taken by a conventional approach.

Keywords: Geostatistical Seismic Inversion, Acoustic Impedance, Functional Data Analysis, Dimensionality Reduction.

DEDICATION

I dedicate this work to God Almighty and to my parents, who gave me all support.

ACKNOWLEDGEMENT

Firstly, I want to thank God Almighty for his grace and favor that allowed me to start and finish this thesis work in good health and vitality.
I would like to thank Politecnico di Torino and Instituto Superior Tecnico (Lisbon) for giving me this opportunity to execute this project. I am grateful to my thesis supervisor in the person of Professor Godio Alberto, my tutor, for his guidance and contribution throughout this work.

My profound gratitude goes to Professor Leonardo Azevedo at the Center for Natural Resources and Environment (CERENA), Instituto Superior Tecnico, Lisbon, Portugal, for providing me with the opportunity to conduct the experimental work at the laboratory of the Institute. The precious gift of their time to answer my many concerns regarding the results was admirable and highly appreciated. They kept their office doors open at all time with so much interest and enthusiasm in my work. Thank you so much. I am very much appreciating this sacrifice.

A special thanks also go to Mr. Roberto Miele a Ph.D. student in Geostatistics who also coached me extensively on geostatistical inversion.

Furthermore, I thank my parents, Mr. and Mrs. Andrew Chiedu Ejenam, who believed in me and always encouraged me throughout this master's program and to my friends for their encouragement to keep thriving, to my friends, Jonah Coleman, Saad Belmashkan, Rita Eguavoen, for their support, love for their encouragement to keep thriving. Their presence in my life always kept me going and pushing on to finish this work.

ABBREVIATIONS

GSI – Geostatistical Seismic Inversion
FDA – Functional Data Analysis
Ip – Acoustic Impedance
AVA – Amplitude Variation vs. Offset
PCA – Principal Component Analysis
FPC – First Principal Component
FPCA – Functional Principal Component Analysis
Vp – Primary Wave Velocity
Vs. – Secondary Wave Velocity
DSS – Direct Sequential Simulation
Co-DSS – Co Direct Sequential Simulation
3D – Three Dimension
2D – Two-dimension
RC – Reflection Coefficient
PDF – Probability Density Function
RC – Reflection coefficient
SGS- Subsurface Geophysical Solutions
SIS – Seismic Interpretation System

TABLE OF CONTENTS

List of Figures

CHAPTER 1

1.0 INTRODUCTION

The thesis was carried out at Institute Superior Técnico, CERENA - the Center for Natural Resources and Environment, with a real dataset provided by Petrobras Oil & Gas.

1.1 SCOPE OF STUDY

In the multimillion-dollar Oil and Gas industry, reservoir modeling and characterization are crucial during the Exploration and Production phases. An inaccurate description of the study area can result in a loss of large amounts of money by the company and its associates.

Before making any choices, it is essential to be aware of the risks. A lot of uncertainties are associated with different stages of an exploration project: from instrumentation or methodological errors, scarcity of experimental data to the vague (but sometimes helpful) assumptions. These uncertainties cannot be removed but they can be assessed, enabling them to make better decisions.

To achieve more reliable models, the companies use geophysical exploration and seismic reflection data to evaluate the underground geology properties and fluid distribution, instead of the conventional seismic interpretation approach where seismic reflection data is solely used as reservoir geometry (Bosch et al. 2010).

Transforming this seismic reflection data into quantitative rock properties that are descriptive to the reservoir is called seismic inversion.

Inversion methods estimate the subsurface petro-elastic properties, such as density, P-, and S-wave velocities. The inverted models for these properties are used to characterize the location, spatial extent, quality of hydrocarbon-bearing lithofacies and access uncertainty in production prediction therefore improving resource management and decision making.

In petroleum applications, random modeling reservoir's internal properties, such as lithofacies and sand bodies, are usually done using core and log data that locally provide comprehensive reservoir information but lack spatial data. As a result, these models have a high level of uncertainty for a well's locations. Due to this, the combination of seismic reflection data considers the properties directly measured at the wells, allowing deducing more reliable subsurface models with less uncertainty, i.e., better spatially constrained.
Geostatistical seismic inversion is the introduction of statistical techniques in the simulation of multiple prior reservoir models. It assesses uncertainties and retrieves the best fitting model

based on probability density and is sub-divided into Bayesian (linearized) and iterative geostatistical seismic inversion.

Iterative geostatistical seismic inversion is becoming a critical instrument in seismic reservoir characterization and allows the simultaneous integration information directly and indirectly about the subsurface properties of interest obtained from seismic reflection and well data, respectively. It explores the model parameter space more comprehensively by random sequential simulation as the model perturbation technique. Still, it is expensive due to computational requirements and the time taken to get results compared with the Bayesian approach.

The spatial continuity pattern and a sole probability distribution function as deduced from the existing experimental data discretely. This problem can be stopped using multi-local distribution functions, taking into account various regionalization by zones (Azevedo, 2013).

The work presented in this thesis's chapters uses FUNCTIONAL DATA ANALYSIS (Unsupervised statistical technique that uses the PRINCIPAL COMPONENT ANALYSIS) to examine the relationships between features in the dataset, thereby identifying the critical elements and compressing the variations to reduce the computation time.

1.2 MOTIVATION

Given the relevance of time in the oil and gas industry, there have always been various research on reducing time on operations and saving cost, especially during oil exploration, which is an expensive process. This work attempts to use the FDA method to reduce the computational time compared to the GSI Ip method.

1.3 OBJECTIVES

This thesis's objective is to apply the FDA method on Ip well log data. The results gotten is compared against those retrieved from Geostatistical Seismic (Acoustic Impedance) method. The proposed workflows allow a time saving reservoir model. The objectives of this work are summarized by the following:

1) Running an example of geostatistical seismic inversion (acoustic inversion) on a synthetic dataset, Analyzing and discussion of the results.
2) Running geostatistical seismic inversion (acoustic inversion) with FDA.
3) Comparing results from 1) and 2).

These algorithms' development was performed recurring to geostatistical toolboxes from CERENA/CMRP research group and MATLAB and SGEMS was used for visualization of the results.

1.4 OVERVIEW OF METHODOLOGIES

Recently, inverted acoustic model gotten from existing seismic reflection data and integrated information from existing wells have been used to obtain reservoir models that are as realistic as possible. However, to obtain petro-elastic properties from seismic reflection data, we need to solve an inverse problem. Seismic inversion problem is identified for being nonlinear, with non-unique solution and it is ill-conditioned. Thus, the inverted acoustic models are one of several models that uniformly make certain the observed seismic reflection data. Notwithstanding the method chosen to solve the seismic inverse problem, there is always the uncertainty associated with the acoustic model. It is essential to ensure that the uncertainty is continuously evaluated and spread throughout the investment process.

Seismic inversion solutions are divided into two main methods: deterministic and probabilistic (Bosch, Mukerji, and González 2010).

- Deterministic methods are based on the reduction of differences between synthetic and observed seismic reflection data based on the convolutional model and it only permit inferring a single best-fit inverse model (Francis, 2006).

- In the probabilistic setting, there are two methods of solving the seismic inversion problem: the Bayesian linearized framework and stochastic sequential simulation as the model perturbation technique.

(Grana et al. 2012) stated that the Bayesian approaches determines the propagation of uncertainties from previous distributions and experimental data (e.g., well-log data) with the probability distributions of the model parameters space. This framework assumes a linearized forward model and Gaussian, or multi-Gaussian, for the property's prior distribution function to be inverted. In this setting, the posterior distribution can be analytically expressed as multi-Gaussian.

On the other hand, stochastic inversion techniques permit a more detailed investigation of the model parameter space since the suspicion around any earlier parametric likelihood conveyance may be loose depending on the stochastic sequential algorithm utilized inside the inversion method. However, it is expensive due to computational requirements and the time taken to get results compared with the Bayesian approach due to the simultaneous random generation.

In this thesis, we applied the functional data analysis method on Ip well log data to reduce the time taken by the GSI Ip method, resulting from the random generation of the 3D Ip models process.

1.5 THESIS ORGANIZATION

Chapter one (1) investigates the background of Geostatistical seismic inversion and the background of the study. In this chapter, the reason for carrying out seismic inversion is discussed and the importance of carring out geostatistical seismic inversion in the assessment of uncertainties to mitigate risk during exploration and also on the need to reduce computational time.

Chapter two (2) deals with a Literature review on past works on some methodology involved in Geostatistical seismic inversion and FDA.

- Section 2.1 discussed spatial continuity analysis,
- section 2.2 covered the various types of variograms,
- section 2.3 discussed extensively the types of simulation used and their limitations.
- Section 2.4 seismic inversion: where seismic inversion is reviewed along with various approaches used to carry out seismic inversion.
- Section 2.5 Rock Physics and its importance in seismic inversion is discussed.
- Section 2.6 Discussed Geostatistical Seismic Inversion; this is where we analyzed the various approaches for integrating seismic reflection and well-log data reservoir modelling.
- Section 2.7 Discusses the FPCA and its goal

Chapter three (3) discusses the Methodology used to carry out the iterative GSI and FDA experiments.

Chapter four (4) outlines and analyzes and discussed the results from the GSI and FDA experiments.

Chapter five (5) compares results from the FDA with the GSI

Chapter six (6) lays conclusion on the work with some possible recommendations for future works and constraints.

CHAPTER 2

2.0 LITERATURE REVIEW

GEOSTATISTICAL EARTH MODELS

2.1 SPATIAL CONTINUITY ANALYSIS

Spatial-continuity analysis can be described as the computation and modeling of Patterns of spatial dependence which characterize a regionalized variable and can be described as a study of the experimental variogram. It is the key step before the regionalized variable is modelled using kriging or conditional simulation techniques, which requires covariance information gotten from spatial continuity analysis (Anon, n.d.).

In geostatistical methodologies, modelling the spatial behavior of a given attribute has a crucial role with two primary objectives:

- Characterizing and quantifying a reservoir property's spatial pattern, commonly designated as spatial continuity analysis in geostatistics, i.e., quantifying the spatial continuity of the important attribute and how it differs in various spatial directions.
- Provides the ground for the spatial inference/estimation, simulation, and geostatistical inversion methodologies (Azevedo & Soares, 2017).

The two primary steps to performing spatial continuity analysis are:

- Computing the experimental measures of continuity (variogram), that’s accounts for anisotropy and azimuth and
- With a continuous function, model the experimental variograms (Anon, n.d.).

2.2 VARIOGRAMS

In geostatistics, the variogram (or semi-variogram) is used to expresses the spatial relationship between points of a given attribute Z (x). It is a spatial continuity instrument used to measure the regionalized phenomenon in space, i.e., presents the variable of interest quantitatively (Soares, 2006). It is an intrinsic function that shows the circumstance understudy, its spatial structure and measures the statistical relationships such as the covariance that exist between samples of successive values consecutive distance values (h) between each experimental pair of data. The variogram is an increasing function in which the abscissa's values can vary up to a given value of h, which is known as the variogram range (Soares, 2006). Variograms with a sill are associated with static regionalized variables; otherwise, they are considered non-stationary.

The estimator of the variogram (Equation 1) gives information about the spatial continuity of the property at various distances ($Z(x + h)$), for various h values, for the calculation of this estimator. The centered covariance estimator (Equation 2) gives information about the spatial average of the product between different attributes at different distances ($Z(x)\ Z(x + h)$). The geostatistical

measures are often used and represent the correlation between samples (Zα and Zα + h) are the estimators, and can be calculated by the following expressions (Soares, 2006):

$$\gamma(h) = \frac{1}{2N(h)} \sum_{\alpha=1}^{N(h)} [Z(x_\alpha) - Z(x_\alpha + h)]^2 \quad (1)$$

By assuming h, the mean of the least-squares and the mean of the products are estimates of the variograms equation two and the centered covariance equation 3:

$$\gamma(h) = \frac{1}{2} E\{[Z(x) - Z(x + h)]^2\} \quad (2)$$

$$c(h) = E\{Z(x)Z(x + h)\} - E\{Z(x)\}E\{Z(x + h)\} \quad (3)$$

The connection between a variogram and covariance is expressed with equation 4:

$$\gamma(h) = C(0) - C(h) \quad (4)$$

Variograms are isotropic or anisotropic, representing various spatial correlations in various directions or various distances h and the difference in behavior is shown (in most cases) when variograms of horizontal directions is connected with variograms with vertical directions. The anisotropy depends on the variable or natural system under study. As an instance, the spatial continuity of a geological layer should be higher in the horizontal direction when compared with the vertical direction (Soares, 2006). The reverse situation happens when computing experimental variogram from existing well data which in this case, the vertical direction is more spatially sampled than in the horizontal direction. Therefore, as the difference between samples (|h|) increases, the average variation between pairs of samples tends to increase.

According to (Soares, 2006), If the variogram is an increasing function, for example, as the module |h| increases, the average variation between pairs of samples is prone to increase until it reaches the range and distance from which it no longer increases then it begins to stabilize. When the variogram is used in the main directions, the average dimensions of the body along respective paths can be calculated.

(Soares, 2006) stated the need to modify the data to a theoretical model when the experimental variogram from the set of samples in the region of interest is created by modeling the data through a mathematical function representing the continuity of the spatial variable and this summarizes the spatial phenomenon's structural characteristics under study, including anisotropy and degree of continuity of the variable under review.

The use of geostatistical modeling is limited to a confined set of functions that cover most natural phenomena' spatial continuity pattern. It is required to select the best suited to each case but only those that meet certain conditions by using positive-definite models that can give stable solutions. The option of possible functions for interpolating the experimental variogram samples is restricted from the start by the positive condition.

The most used variograms for theoretical models are the Spherical model, Exponential, and Gaussian. They have two parameters:

- Sill, which is the only limit that the variogram values tend to have as the distance between sampling pares grows.
- The amplitude (h = a) of the variogram is the distance from which the importance of the variogram (γ(h)) stops growing and equal to the sill and it usually coincides with the variance of the attribute in the study. It measures the distance from which the values of that attribute start to show no sign of the correlation between them.

The spherical model (Equation 5) is one of the frequently used models in modeling experimental variograms and is characterized by rapid growth at the origin and is expressed by the mathematical equation:

$$\gamma(h) = \left[1.5\frac{h}{a} - 0.5\left(\frac{h}{a}\right)^3\right] \text{ for } h \leq a \qquad (5)$$

$$c \quad \text{for } h > a$$

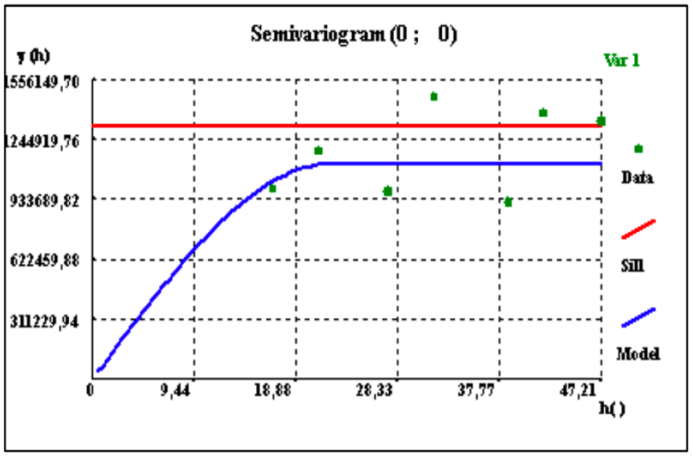


FIGURE 1: Graphical Representation of a Spherical Model (martins, 2014)

Comparing the exponential model (Equation 6) with the previous model, there is a growth near the origin and a higher spatial continuity for larger distances of h in the exponential model. This model asymptotically gets to the sill, and the distance over which the model value reaches the 95% sill is defined as the amplitude (Soares, 2006), namely (a) = 0.95C. The exponential model is expressed by the following equation:

$$\gamma(h) = C\left[1 - e^{-3h/a}\right] \qquad (6)$$

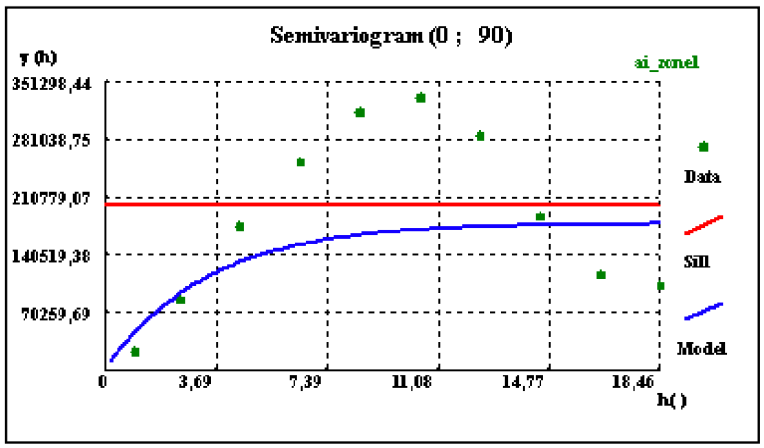


FIGURE 2: Graphical Representation of an Exponential Model (martins, 2014)

The Gaussian model (Equation 7) has a parabolic behavior close to the origin, with a slow growth favorable to the adjustment of regular phenomena, in variation to the rapid growth of the previous two models who have a more irregular behavior of the phenomena. The Gaussian model shows a great continuity of the variable under study, with the general expression:

$$\gamma(h) = C\left(1 - e^{\left(\frac{-3h^2}{a^2}\right)}\right) \tag{7}$$

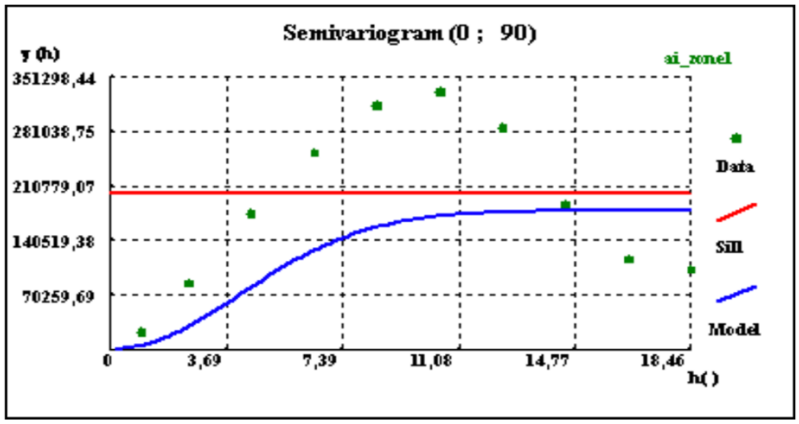


FIGURE 3: Graphical Representation of a Gaussian Model (martins, 2014)

2.3 DIRECT SEQUENTIAL SIMULATION AND CO-SIMULATION

2.3.1 DIRECT SEQUENTIAL SIMULATION

When trying to replicate very compound systems, like those with strongly distorted or multi-modal distributions, Sequential Gaussian simulation has some limitations. After the inverse Gaussian transformation, getting back the original probability distribution function will result in a variogram that is very unstructured and similar variogram reproduction problems is encountered when the joint statistics of a multivariate data set need to be replicated. (Azevedo & Soares, 2017).

According to (Journel, 1994), The direct simulation of a continuous variable was shown to replicate the covariance model if simulated values are derived from local distributions and based on simple kriging estimates with a variance corresponding to the simple kriging variance estimate.

One of the limitations regarding the practical application of the direct sequential simulation approach is that it does not guarantee the production of the histogram, which is a requirement from any simulation algorithm. This limitation is corrected by resampling with the global distribution conditioned by the local conditional mean and variance which is able reproduce the variogram and histogram of a continuous variable without calling for any transformation of the original variables.

Practically, intervals z is chosen from global $F_z(z)$, obtaining a new $F^l_z(z)$. the intervals are centered by simple kriging estimate $z(x_u)$ with a range dependent on kriging variance $\sigma^2_{sk}(x_u)$. Sampling the global distribution $F_z(z)$ by intervals is defined by the local mean and variance of the Gaussian CDF $z(x_u)$, and the value of $y(x_u)^*$ is equal to the local estimate $z(x_u)^*$. The simulated value $z^s(x_u)^*$ gotten from the interval of $F_z(z)$ is defined by $G(y(x_u)^*, \sigma^2_{sk}(x_u))$ (Soares, 2001).

$$\frac{1}{n}\sum_{i=1}^{n} z(x_i) = [z(x_u)]^* \tag{8}$$

$$\frac{1}{n}\sum_{i=1}^{n} [z(x_i) - [z(x_u)^*]]^2 = \sigma^2_{sk}(x_u) \tag{9}$$

$$y(x) = \varphi\big(z(x)\big), \text{with } G\big(y(x)\big) = F_Z\big(z(x)\big) \tag{10}$$

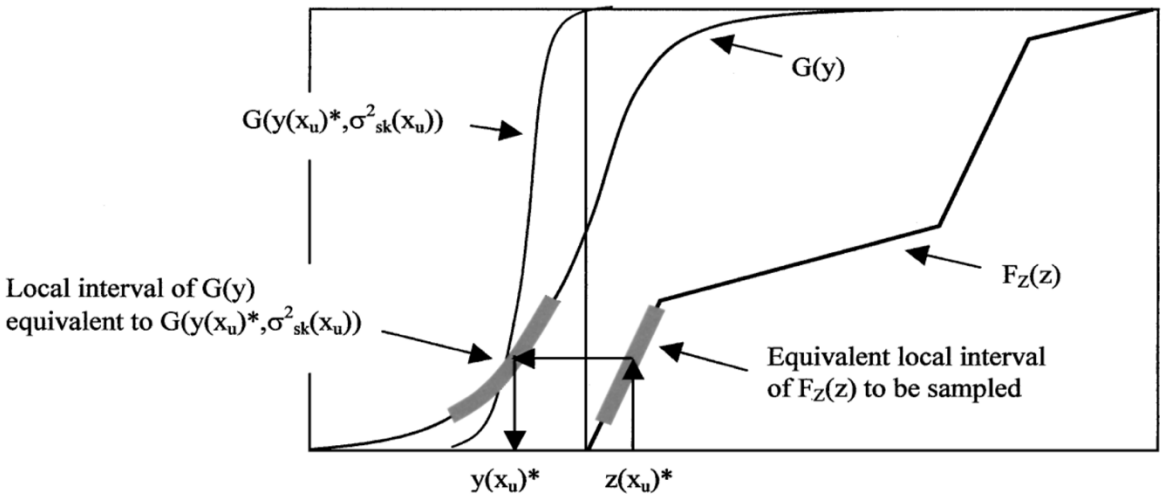


FIGURE 4: Sampling of Global Distribution Fz (z) by intervals defined by the local mean and variance of $z(x_u)$: the value $y(x_u)^*$ corresponds to the local estimate $z(x_u)^*$. the simulated value $z(x_u)^*$ drawn from the interval of fz (z) is defined by $g(y(x_u)^*, \sigma^2_{sk}(x_u))$ (Soares 2001).

The DSS can be described in the following steps (Soares, 2001; Azevedo et al., 2015):

1. A random path over the entire grid of nodes x_u, u = 1, Ns, to be simulated by generating a random seed is Defined. Ns is the total number of nodes that comprise the simulation grid.
2. Estimate the local mean $z(x_u)$ and variance $\sigma^2_{sk}(x_u)$ identified, respectively, with the simple kriging estimate conditioned to the experimental data $z(x_i)$ and the simulated values $z^s(x_i)$.
3. Define Fz(z) interval to be sampled by using the Gaussian CDF as explained in the figure above.
4. Generation of a value $z^s(x_u)$ from the CDF of Fz (z).
    a. Generate a value between 0 and 1 from the uniform distribution
    b. From $G(y(x_u)^*, \sigma^2_{sk}(x_u))$ Generate a value y^s.
    c. Return the simulated value $z^s(x_u) = \varphi^{-1}(y^s)$.
5. Loop till all N nodes have been simulated within the grid.

2.3.2 DIRECT SEQUENTIAL CO-SIMULATION

Often, by reproducing the underlying association that may inevitably occur between two or more properties, there is a need to produce the spatial realizations (e.g., porosity, acoustic impedance, permeability). It is done by joint simulations or co-simulation models (Soares, 2001; Horta & Soares, 2010), and an advantage of the proposed algorithm over conventional SIS and SGS is that it makes a joint simulation by dealing directly with the original variables, i.e., instead of simulating Nv variables simultaneously, each variable is simulated, in turn, and conditioned to the previous variables.

Two variables are co-simulated when they express dependencies and correlation through simulation between the individual distributions and variograms. According to (Gomez-Hernandez & Journel, 1993, Goovaerts, 1997a), there are basically two different approaches for

co-simulation, the one used in this thesis applies Bayes rule, where the simulation of each variable it is handled in a sequence of simulation of conditional distributions

The Bayes principle is applied in the co-simulation procedure to calculate the local conditional probability distribution at a given location. Assuming that two dependent variables: $Z_1(x)$, as the first to be simulated individually by DSS is more significant due to its more excellent spatial continuity, and the second variable, $Z_2(x)$, that is simulated conditioned by the previously simulated values of $Z_1(x)$. In location x_0 the values of $Z_2(x_0)$ are generated from the law of conditional distribution:

$$F_Z(Z_2(x_0)|Z_2(x_a) = z_2(x_a), Z_1(x_i) = z_1^s(x_i), i = 1, N) \quad (11)$$

Where $z_1{}^s(x_1)$ are the values previously simulated of $Z_1(x)$ and $z_2(x_a)$ are the experimental and the values already simulated of $Z_2(x)$ within a likelihood adjacent to the location x_0. After the simulation of $Z_1(x)$, the same DSS algorithm is now applied to $Z_1(x)$, using the images previously simulated of $Z_1(x)$ as secondary and with a spatial correlation between both variables. The values of $Z_2(x)$ are generated in any spatial location of x_0, from the laws of conditional distribution to the values previously simulated.

In the Direct sequential co-simulation, it is important to estimate the local mean and variance of $Z_2(x)$ sample and the datum at X_0 of $Z_1(X)$ to sample from the prior global probability distribution $F_Z(Z)$. This is done using the collocated sample co-Kriging estimate (Soares, 2001).

$$[z_2(x_0)^*]_{CSK} = \sum_{a=1}^{N} \lambda_\alpha [Z_2(x_\alpha) - m_2] + \lambda_\beta [z_1^s(x_0) - m_1] + m_2 \quad (12)$$

$$\sigma^2_{CSK}(x_0) = Var\{Z_2(x_u)^* - Z_2(x_u)\} \quad (13)$$

λ_α and λ_β are the weights and m_1, and m_2 is the mean for each variable $z_1(x)$ and $z_2(x)$, respectively.

The following sequence of steps can be defined as Co-DSS (Soares, 2001; Azevedo 2013):

1. Simulating first variable $z_1(x)$ for the grid with DSS.
2. Generating a path over the entire simulation grid x_0, u=1,…, N, randomly. Where N is the total number of nodes that compose the simulation grid.
3. Estimation of the local mean and variance at x_0 with a collocated simple co-Kriging estimate ($[z_2(x_0)^*]_{csk}$) and the corresponding co-Kriging variance ($\sigma^2_{csk}(x_0)$) constrained to data($z_2(x_\alpha)$) within the neighborhood, composed by the experimental, Data simulated previously, and the collocated datum $z_1(x_0)$.
4. Definition of the interval $F_{z2}(z)$ to be sampled as previously explained in the figure above.
5. Calculation of a value $z_2(x_0)^s$ from the CDF of $F_{z2}(z)$:
   a. Generation of a value p from the uniform distribution between [0, 1].
   b. Generation of a value y_s from $G(y(xo)^*, \sigma^2_{CSK}(x_o))$.

c. The simulated value $z^s_2(x_0) = \varphi^{-1}(y^s)$ is returned.

6. Add the simulated value to the set of conditioning data.
7. Loop until all the N nodes of the simulated grid have been simulated.

2.3.3 DIRECT SEQUENTIAL CO-SIMULATION WITH JOINT PROBABILITY DISTRIBUTIONS

The limitation of the traditional direct sequential co-simulation methodology, i.e., the ability to produce only linear correlations between primary and secondary variables, was corrected by direct sequential co-simulation with joint probability distributions(Horta & Soares, 2010). Like other stochastic simulations, the co-DSS procedure with joint probability distribution applies Bayes principle in a sequence of steps:

- A previously simulated model is assumed by the simulation technique and the secondary variable ($Z_1(x)$), is first determined for the whole DSS simulation grid
- The primary variable ($Z_2(x)$) is co-simulated.

The following conditional cumulative distribution function is estimated from the global bi distribution for a given $Z_1^S(x0)$ simulated at a node x_0 of a random path (Azevedo & Soares, 2017).

$$F(Z_2(x_0) \mid Z_1(x_0) = z^s_1(x_0)) \quad (14)$$

The value of z2(x0) instead of using the traditional co-DSS method is calculated using the equation below (Horta & Soares, 2010).

$$F[Z_2(x_0) \mid Z_1(x_0) = z_1(x_0)] = \text{prob}\{Z_2(x_0) < z \mid Z_1(x_0) = z_1(x_0)\}\ K \quad (15)$$

The co-DSS and joint probability distribution replicate the marginal probability distribution functions of the main variable, and the value of the experimental data is revered at its spatial locations and the spatial continuity model is replicated. The main objective of this simulation algorithm is to replicate, as calculated from the experiments, the joint probability distributions.(Horta & Soares, 2010):

Calculate the conditional cumulative distribution function $F(Z_2(x_0) \mid Z_1(x_0) = z^s_1(x_0))$ from the global joint-probability distribution $F(Z_2(x), Z_1(x))$ on the basis of the simulated values for the secondary variable, $z^s_1(x_0)$.

1. From the experimental results, estimate the global bi-distribution $F(Z_2(x), Z_1(x))$.
2. Simulation with DSS of the secondary variable $Z_1(x)$ for the entire simulation grid.
3. Estimate the local mean and variance at x_0 following a random path with a collocated simple co-Kriging estimate ($[Z_2(x_0)^*]_{CSK}$) and the corresponding co-Kriging variance ($\sigma^2_{CSK}(x_o)$) that is conditioned on the initial experimental data, the previously simulated data ($z_2(x_\alpha)$) and the second variable collocated date ($z^s_1(x_0)$).
4. Calculate the conditional cumulative distribution function $F(Z_2(x_0) \mid Z_1(x_0) = z^s_1(x_0))$ from the global joint-probability distribution $F(Z_2(x), Z_1(x))$ on the basis of the simulated values for the secondary variable, $z^s_1(x_0)$.
5. To simulate a value for $z^s_2(x_0)$ from the conditional CDF $F(Z_2(x_0) \mid Z_1(x_0) = z^s_1(x_0))$, adopt the conventional DSS method.

2.4 SEISMIC INVERSION

Over the last few years, seismic inversion techniques have played major roles in the subsurface geology's characterization, aiding professionals to make better and more efficient decisions on the oil industry. It is very crucial in reservoir modeling and description due to its potential to assess the sub-surface petro-elastic properties (Azevedo et al., 2018). Seismic-reflection data are used in reservoir characterization to get a geometrical description of the main structures of the subsurface and estimate characteristics such as lithologies and fluids. Even when it comes to data that are noise free as a result of the small frequency of registered seismic waves, the transformation of seismic data to reservoir properties is an inverse problem with a non-unique solution.
In action, the reversal of seismic data for reservoir properties complicated as a result of noise which is always present in the data, the simplification of forward modelling needed to obtain solutions within an appropriate period, and the uncertainties in well-to-seismic relations (depth-to-time conversion), the estimation of the representative wavelet, and the connection between the reservoir and the elastic properties (Bosch, Mukerji & Gonzalez, 2010).

The issue of seismic inversion includes understanding the response of the Earth to a limited collection of measurements indirectly and attempting to infer data from the modelling parameters that give rise to the solution (Azevedo & Soares, 201,7), so, it is stated as a Bayesian inference problem (Bosch, Mukerji & Gonzalez, 2010).

The main objective of the analysis related to seismic inversion methods is to derive seismic elastic properties from assessable seismic reflection data polluted by mistakes from various sources. For a specific physical system. For a specific physical system, the process of predicting response is called forward modelling, and a forward model represented in equation 16 expresses the relationship between the data observed and subsurface properties. (Tarantola, 2005; Azevedo & Soares, 2017).

$$d_{obs} = F(m) + e. \tag{16}$$

d_{obs}: represent the recorded seismic reflection data with the well log data present.
F: is the convolutional model.
m: is the model parameter space for the properties to invert, which depends on the goal of the inversion and they are acoustic and elastic impedances or density, P-wave and S-wave velocity model.

Equation 17 can describes the forward model (F) of the previous equation:

$$A = r * w \tag{17}$$

A is the recorded seismic amplitude obtained by the convolution of r, the subsurface reflection coefficients, which depend on the elastic properties (P-wave and S-wave velocities and density) of the subsurface geology, with an estimated wavelet w (Azevedo & Soares, 2017).

2.4.1 LIMITATIONS OF INVERSION METHODS

There are limitations with seismic methods due to the limited band nature, noise, forward modeling, uncertainties in-depth to time conversions simplifications, which make inversion a nonlinear problem with multiple solutions (Tompkins et al., 2011; Bosch, Mukerji & Gonzalez, 2010; Tarantola, 2005). Therefore, the solution to an inverse problem will be a set of earth-modeled configurations that match the real data within some tolerance when forward modeled into the synthetic data (Bosch, Mukerji & Gonzalez, 2010).

The different seismic inverse methodologies need to be understood because reservoir characterization includes the elastic property calculated from seismic inversion and rock physics used in estimating the reservoir properties.

The solution to the problem of seismic inversion can be split into two parts:

- Optimization or error minimization is also called deterministic. It is gradient-based and searches for the maximum of the posterior probability and
- The stochastic also called Monte Carlo or sampling approach (Bosch, Mukerji & Gonzalez, 2010).

According to (Bosch, Mukerji & Gonzalez, 2010), Various workflows combine seismic inversion and rock mechanics to model the properties of reservoirs. These workflows can be divided into two groups: sequential, joint or simultaneous workflow.

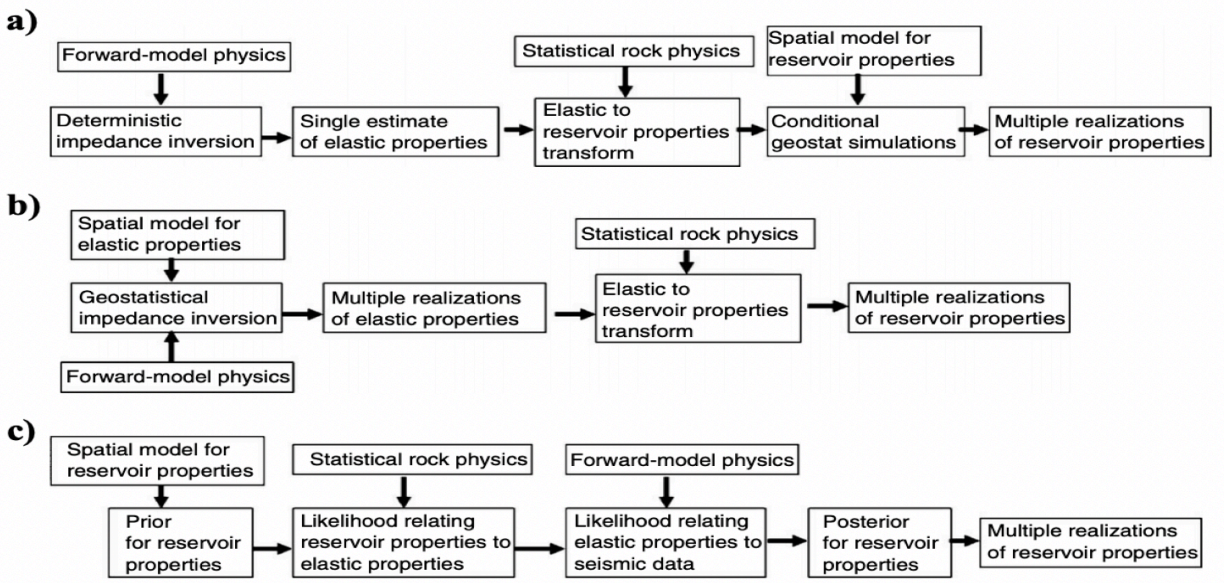


FIGURE 5: Seismic Inversion Diagram of Alternative Workflows for Reservoir Properties giving rise to Several Realizations of Seismic Data Conditioned Reservoir Properties, The Rock-Physics Model, And the Geology Imposed Spatial Continuity Model. (A, B) Two Sequential (Or Cascaded) Workflow combinations. (c) Bayesian Inversion Workflow Simultaneous. Data comes in different phases, in addition to seismic reversal data, well log data, cores, thin sections if present, and geological data gotten from outcrops (Tarantola, 2005; Russell).

2.4.2 Deterministic Approach

Deterministic methods, through an imperfect assessment and uncertainty, gives a single, local smooth estimation of elastic properties at subsurface and therefore lead to a skewed estimate of volumes and connectivity (Bosch, Mukerji & Gonzalez, 2010).

Regression models are based on deterministic methods and optimization algorithms are used to achieve a single best-fit solution, but an accurate evaluation of uncertainties resulting from the recovered subsurface model is missing.

Considering this, the uncertainty is strictly expressed by a local multivariate Gaussian and can be evaluated as a linearization around the best-fit inverse solution, usually recovered by least squares (Tarantola, 2005; Russell).

There are two common methods of deterministic or optimization-based methods of seismic inversion for elastic properties, according to (Bosch, Mukerji & Gonzalez, 2010), and they are the sparse-spike and model based (Russell & Hampson, 1991).

- Under sparse assumptions of the reflectivity array, Sparse-spike deconvolves the seismic trace where the obtained reflectivity is determined by their impedances, including low frequencies that were missing typically from well data, seismic-velocity analysis or the low-frequency trend kriging estimate (Bosch, Mukerji & Gonzalez, 2010).
- The inversion algorithm disrupts the initial model in the model-based until certain minimization conditions are met with the objective function or function to be reduced as a distinction between the data observed and the data modelled (Bosch, Mukerji & Gonzalez, 2010).

Deterministic inversion methods are cheaper and faster in terms of computational work. They provide a smooth representation of the subsurface of Earth with far less spatial variability than the geology of the actual subsurface. If the aim is to understand more global and less detailed information, they can be used.

2.4.3 Stochastic Approach

According to (Russell, n.d.), Stochastic inversion is an extension of deterministic inversion. It provides additional information such as lithology, probability, facies distribution, volumetric and petrophysical parameters, therefore quantifying uncertainties by casting a problem from the posterior distribution in terms of variables selected through processes and statistical numerical moments of the resulting group solutions (Tompkins et al., 2011) and also eliminating the need to explicitly find solution to the wide-ranging inverse problem, accounting for the nonlinearity problem, and creating uncertainty estimates. However, they do not preclude the correspondingly vast multivariate posterior space from being sampled (Haario, Saksman & Tamminen, 2001).

They avoid any parametric prior distribution of the model and errors in the data and linearization of the forward model from being assumed (Nunes, Azevedo & Soares, 2019). The probabilistic methods evaluate the spread of uncertainty from previous probability distributions, calculated from experimental data (e.g. well-log data) to model parameter space distributions (Azevedo et al., 2013).

Within the stochastic methods, Bayesian approaches makes sure that the uncertainty is propagated from the previous estimated probability distributions from the experimental data (e.g., well-log data) to the probability distributions of the space model parameters (Azevedo & Soares, 2017),. Inversion problems can therefore be described as a problem of Bayesian inference (Bosch, Mukerji & Gonzalez, 2010).

$$\text{Posterior} = \text{constant} * \text{prior} * \text{likelihood} \tag{18}$$

$$\sigma_{post}(m) = c\rho_{prior}(m)\rho_{data}(d_{obs} - g(m)) \tag{19}$$

Where the posterior probability density is $\sigma_{post}(m)$, and $\rho_{prior}(m)$ is the priori probability.

The value $\rho_{data}(d_{obs} - g(m))$ is the data probability function and relies on the dobs and their uncertainties of observation, the forward-modeling operator g that depicts the model space into the data space, and the uncertainties of modelling.

In equation 19, m defines the parameter configuration of the earth model, and c is a constant for normalizing the density of the posterior likelihood. The forward operator can be a simple function without a simple analytical expression, a matrix operator, or, more generally, an operator or a computational algorithm.

2.5 Rock Physics

Rock physics is one of the measures that succeeds seismic inversion or within a joint seismic/petrophysical formulation for quantitative seismic analysis, connecting elastic parameters and reservoir properties (Bosch, Mukerji & Gonzalez, 2010). It makes a link between parameters of the geological reservoir (porosity, clay content, sorting, lithology, saturation) and seismic properties (acoustic impedance, Vp/Vs (P-wave/S-wave velocity) ratio, bulk density, and elastic modules). In order to analyze such supposed possibilities, such as possible fluid or lithology changes, Rock-physics models are used to view observed sound and seismic velocities in terms of reservoir parameters or to forecast outside the available data range. They are also used to predict seismic reaction to the assumed reservoir and properties and conditions of overload (Avseth et al., 2010).

Several models of rock physics are available, each having its advantages and limitations, such as empirical and theoretical models, which describes elastic moduli's behavior as a function of mineral contents and porosity, type of pores and fluids present, clay content, sorting, cementation, and stress.

The application of rock-physics models gotten on the log or core scale to band-limited seismic inversion outcomes can be difficult (Doyen, 2007) because the inversion outcomes reflect aggregate seismic-scale lithologies. With suitable scale transformations for the reservoir and elastic parameters, this problem can be viewed. Rock physics validates this conversion to reservoir properties and makes it possible to improve geological process-based well-log or training data (Avseth, Mukerji & Mavko, n.d.). In order to infer reservoir properties, geostatistical algorithms can be combined with rock physics models (Doyen, 1992).

Statistical rock physics, as described by (Doyen, 1992), integrates theoretical and empirical rock physics models with statistical pattern recognition to interpret elastic properties derived from seismic inversion and use the four steps to measure interpretation uncertainty.:

- To obtain the concept of facies, study of well log data.
- Simulation using the Monte Carlo system of seismic rock properties (Vp, Vs, and density).
- Computation of statistical probability density (PDF) facies-dependent functions for the seismic attributes of interest.
- Classify the voxels inside the seismic attributes or elastic properties obtained from seismic inversion using a statistical classification technique that uses the Bayes theorem to obtain the subsequent probability of facies given the attributes and impose spatial correlation using geostatistical stochastic simulation that updates the probabilities of seismic derivation taking geological probabilities into account with reasonable spatial correlation by conditioning to the facies and fluids observed at the well locations (Bosch, Mukerji & Gonzalez, 2010).

2.6 GEOSTATISTICAL SEISMIC INVERSION

Exploration is becoming complicated in the oil industry, and targets are located very far offshore. The unavailability of well data and the improvement in the quality of geophysical data, checked in recent decades, render the use of such data inevitable for the modelling and characterization of oil reservoir activities. In the characterization of subsurface petrophysical variables the incorporation of geophysical data has become a priority aim for geoscientists, making geostatistics a big way to provide the theoretical context and necessary practical tools to combine as many different forms of reservoir modelling and characterization data as possible (in particular the integration of well-log and seismic reflection data). Geostatistical seismic inversion techniques are valuable and powerful methods used to simultaneously combine seismic reflection and well-log data to predict and classify the petro-elastic properties of subsurface lithofacies in hydrocarbon reservoirs (Daya, Cheng & Agterberg, 2018). Geostatistical seismic inversion is a technique used to simulate potential models of rock property and has many benefits that are helpful in reservoir modelling and analysis of uncertainty because it reduces tuning effects, models uncertainty and can be measured on a fine scale, but it is costly and needs large amounts of data (Cristea, 2018). It combines data from many sources and produces models that, by matching known geological trends, have better resolution than the original seismic and can be used for risk assessment and reduction. Seismic, well logs and other

input data are individually described as a function of probability density (PDF), which provides a geostatistical definition in histograms and variograms that describe the chances of a specific value at a specific location and the expected geological scale and composition throughout the modelled region (Cristea, 2018).

According to (Pereira et al., 2020), stochastic sequential simulation is used as the model generation and perturbation technique in geostatistical seismic inversion. Which uses a global variogram model to express the expected spatial continuity pattern of the subsurface elastic properties of interest. It is unsuitable for complex and nonstationary geological environments to be conditioned to a single variogram model because it results in poor inverted models which are unable to reproduce non-stationary features such as channels, folds, and faults

(Deutsch & Journel, 1996) claimed that a geostatistical seismic inversion system consists of iterative processes whereby a set of parameter realizations are generated using methods of stochastic sequential simulation and produced until a given user-specified value or the given number of fixed iterations is reached by the match of the objective function. Geostatistical inversion techniques are therefore based on the use of stochastic sequential simulation as the technique of model disruption, which ensures that the primary spatial continuity patterns are generated and that the acoustic and elastic properties of interest are jointly distributed as obtained from the current well-log data in all the models obtained during the iterative process. While enabling entry to the uncertainty related to the retrieved inverse models at the same time. There are two traditional approaches to seismic reflection integration and well-log data for modelling hydrocarbon reservoirs (Daya, Cheng & Agterberg, 2018) which are trace by trace and Global geostatistical seismic inversion.

2.6.1 Trace by Trace Geostatistical Seismic Inversion

Seismic trace within the inversion grid is individually visited according (Bortoli et al., 1993) after a predefined random path within the seismic volume. In some reported seismic reflection data, there is a low signal to noise ratio which is a major downside (Daya, Cheng & Agterberg, 2018).

2.6.2 Global Geostatistical Seismic Inversion

The Global Geostatistical Seismic Inversion (Soares, A., Diet, J. D., Guerreiro, 2007) was implemented to resolve the trace by trace method limitation by using a global approach during the stochastic sequential simulation process of the inversion procedure (Daya, Cheng & Agterberg, 2018).

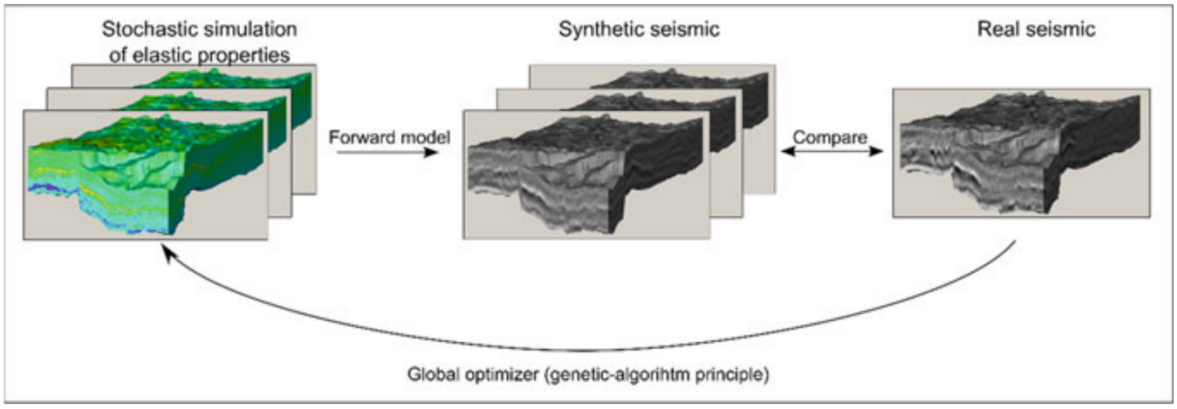


FIGURE 6: General Outline for Global Iterative Geostatistical Seismic Inversion

Low signal-to-noise ratio areas persists to be poorly balanced in the throughout the iterative inversion method in global iterative geostatistical seismic inversion processes (a group of inverted models that are best-fitted will definitely give high variability, or high uncertainty, for noisy areas that have low signal-to-noise ratio). This method was generalized for the inversion of acoustic and elastic impedance seismic reflection data, direct inversion of petrophysical properties, and inversion of seismic AVA (Amplitude Variation vs. Offset) (Daya, Cheng & Agterberg, 2018). This thesis research focuses on Geostatistical Acoustic Inversion.

2.7 FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS

Functional Data Analysis is a branch of statistics that analyses data by supplying information about surfaces varying over a continuum, i.e., a series of p-direction vectors where the i^{th} vector is the direction of a line that best matches the data while being orthogonal to the first i-1 vectors. Here, the best-fit line is called the line that minimizes the average square distance from the points to the line. These directions form an orthonormal basis where different individual data dimensions are not linearly associated with each other.

Principal Component Analysis (PCA) is the method of computing the main components by modifying the data, sometimes using only the first few main components and disregarding the remaining components.

Functional Principal Component Analysis (FPCA) is a well-known mathematical model and order reduction technique (Ramsay, J. and Silverman, n.d.). In FPCA, time series are summarized into a relatively small number of unknowns by combining principal component analysis (PCA) with functional data analysis (FDA). This makes FPCA is a broad dimension reduction method even when the presence of non-periodic data such as well-log data or petro-elastic traces are retrieved by seismic inversion.

Recently there has been interest in this method in geosciences. (Menafoglio, Alessandra.,Grujic, O., & Caers, 2016) used FPCA to predict the oil production curves of shale reservoirs. (Bottazzi

& Della Rossa, 2017) use FDA in geomechanical modeling of hydrocarbon reservoirs. However, its use as part of seismic inversion has been limited. (Nunes, Azevedo & Soares, 2019) proposed a method combining machine learning and Fourier decomposition to speed-up iterative methods of geostatistical seismic inversion. A particular case of functional data analysis can be considered to be their tool.

This study builds upon previous works on FPCA and model reduction in geosciences (Menafoglio, Della Rossa).

2.7.1 Goal of FPCA

The purpose of evaluating main components (Ramsay, 2009) is:
• When we want to find the dominant modes of variation in the results, PCA is generally used, typically after subtracting the mean from each observation.
• We want to understand how many of these types of variation are required to achieve a sufficient approximation of the original data.
• Holding only dominant modes can be believed to maximize the signal-to-noise ratio of what we hold.
• In terms that we can explain to non-statisticians, we generally want to know what these modes represent. At this stage, rotation of the principal components will help.

CHAPTER 3

3. METHODOLOGY

3.1 GLOBAL GEOSTATISTICAL ACOUSTIC INVERSION

Geostatistical seismic inversion methods have also been used to integrate seismic knowledge into stochastic fine grid models, according to (Caetano, 2012), and geostatistical inversion methods does the sequential approach in two stages:

- First, in each trace (a column of a 3D grid), acoustic impedance values are simulated on the basis of well data and spatial continuity pattern, as shown by the variograms.
- Then, the acoustic impedance values are converted into amplitudes by a convolution with a known approximate wavelet, resulting in a simulated seismogram that can be contrasted by the correlation coefficient with the actual seismic.

Based on an objective function's match, the "best" simulated trace is preserved, and another trace is visited, simulated and transformed, based on the same objective function (a function of the similarity between actual seismic trace and seismogram). The sequential approach continues until all acoustic impedance traces are simulated.

According to (Caetano, 2012), the traces of simulated acoustic impedances are imputed as "real" data for the next sequential simulation stage as long as the "best" transformed trace is adopted in each step.

The benefits of working with Ip instead of the seismic data that has been recorded are:

- This makes the layer property more geological rather than an interface property and therefore has a more physical sense.
- Secondly, all the well data is connected to the seismic data during the inversion process, providing a better understanding of the quality of both datasets.

One of the existing methods for inverting full-stack seismic reflection data for acoustic impedance (Ip) models is global stochastic inversion (Caetano, 2012; Soares, Diet, Guerreiro, 2007). The following sequence of measures, outlined in the figure below, will define the general outline of this iterative geostatistical methodology process:

1. Using direct sequential simulation (Soares, 2001), a set of Ns acoustic impedance models are simulated for the entire seismic grid, conditioned on the available acoustic impedance well-log data assuming a spatial continuity pattern as shown by a variogram model.
2. Derive a set of synthetic seismic volumes of Ns from the impedance models simulated in the previous step by calculating the corresponding normal coefficients of incidence reflection (RC).
$$RC = \frac{Ip2 - Ip1}{Ip2 + Ip1}$$
3. For that specific seismic dataset, an approximate wavelet convolves resulting RC in order to compute synthetic seismic volumes using

$A = r * w$.

4. With respect to the correlation coefficient in opposition to the actual seismic trace from the corresponding location, each seismic trace from the Ns synthetic seismic volumes is compared. The synthetic seismic with the highest correlation coefficient generated from the traces of acoustic impedance are stored in a second volume along with the value of the correlation coefficient from the ensemble simulated Ip models.
   These secondary volumes are used as secondary variables and local regionalized models to produce new sets of acoustic impedance models for the next iteration, one having the best acoustic impedance traces and the other that has the corresponding local correlation coefficients. Using direct sequential co-simulation (Soares, 2001) the new set of Ns acoustic impedance models was constructed and conditioned to the available acoustic impedance well-log data using secondary variable and local correlation coefficients of the best Ip volumes.
5. When the global correlation coefficient between the fully synthetic and actual stacked seismic volumes is above a certain threshold, end the iterative process.

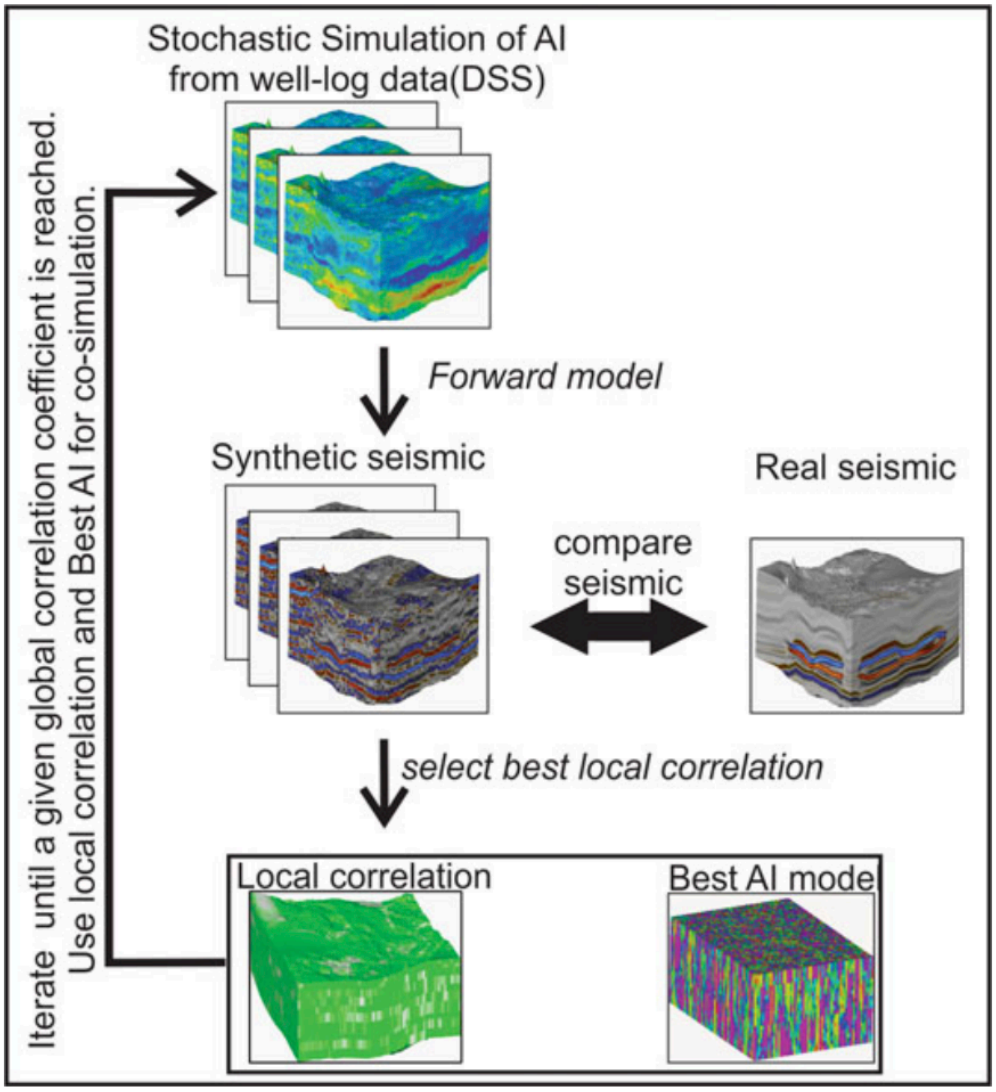


FIGURE 7: Schematic Representation of Geostatistical Acoustic Inversion (Azevedo & Soares, 2017)

The GSI methodology makes it possible to retrieve high-resolution Ip models that honor the distribution feature as estimated from the well-log data available and the spatial continuity model as obtained from a variogram model and has been successfully tested on seismic datasets from very different geological contexts with varying quality (Soares, A., Diet, J. D., Guerreiro, 2007).

3.2 Functional Principal Component Analysis

Iterative geostatistical acoustic inversion (Azevedo and Soares 2017) relies on two main ideas. First, model perturbation and update are performed using stochastic sequential simulation and co-simulation (Deustch and Journel 1994). The existing well-log data of Ip are used as conditioning data for the stochastic sequential simulation and co-simulation of the entire

inversion grid. Then, the trace-by-trace mismatch (M) between observed and synthetic seismic data (Eq. 20) drives the inversion procedure from iteration-to-iteration:

$$M^{j,r} = \frac{2\sum_{k=-n}^{n}(d_{t+k}\, d_{t+k}^{j,r})}{\sum_{k=-n}^{n}(d_{t+k})^2+\sum_{k=-n}^{n}(d_{t+k}^{j,r})^2} \quad (20)$$

where j is the iteration number, r the realization number, n is the size of the moving window used to compute the local trace-by-trace similarity, d_t and $d_t^{j,r}$ are the measured and predicted seismic traces at sample t, by construction, $-1 \leq M_t^{j,r} \leq 1$. Negative values of M are truncated at zero.

In the proposed method, we use stochastic sequential simulation and co-simulation to generate 2-D fields of FPCs as a proxy of model generation for the entire inversion grid. The generated fields are then reconstructed into the original 3-D domain. This step allows retrieving 3-D models of Ip. The proposed inversion method may be summarized in the following sequence of steps, summarized in the figure below:

1. Apply FPCA to the existing Ip well logs. Each log is summarized in a set of l FPCs, with $l < n$, n is the sample number of the original logs.
2. The l FPCs, resulting from i), are used as conditioning data to generate 2-D fields with the same spatial extent of the inversion grid. Ns independent realization is generated for each FPC.
3. Reconstruct Ns 3-D Ip fields by applying the inverse FPCA.
4. Forward model each Ns Ip model and compute Ns synthetic seismic volumes.
5. Compare, on a trace-by-trace basis, the Ns synthetic, and real seismic volumes following Equation 2.
6. For each location within the inversion grid, select the simulated FPC that ensures the maximum M. Store both the FPC and M values in two auxiliary volumes.
7. Use the stored auxiliary volumes as secondary variables to co-simulate a new set of Ns FPC models.
8. Return to step 3) and iterate until the global M between real and synthetic volumes is above a pre-defined threshold.

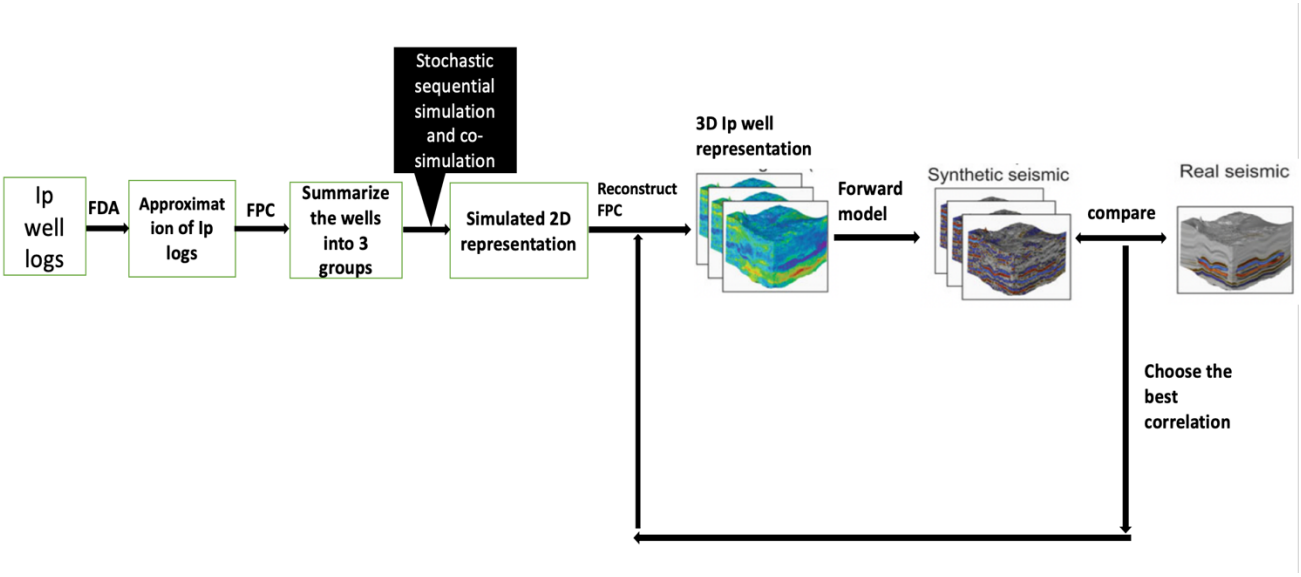


FIGURE 8: Schematic Representation of Functional Data Analysis

CHAPTER 4

4.0 RESULTS AND DISCUSSION

4.1 DATA DESCRIPTION

The field of research is a turbidite region located deep-offshore and the reservoirs have sand-prone bank deposits. As a result of their amplitude anomalies with the offset, the known reservoirs on a partial angle stack are easily identified. The seismic volumes of 321 in lines of 338 crosslines and 756ms of length with a sampling rate of 4ms were included in the available data collection, while we established an inversion grid of 101 cells in i- and j- directions and 90 vertical direction samples (k). The inversion region was delimited by a half-wavelet plus or minus the top and base reservoir ensuring better convolution within the area of the reservoir. There was a set of 13 wells accessible with Vp and bulk density logs (fig. 9). The wells are positioned preferentially drilling the geological pay formations along the study field, which imposes a bias on the recognized elastic properties (fig. 10).

4.1.1 DATA PROCESSING

The data processing was done by Professor Leonardo Azevedo. The steps he used in the processing of the well log data are as follows:
Given the available well log data,

- The high-resolution well-log data was upscaled into the reservoir grid (fig. 11), to make sure that the extreme mean values were retained, and the variance was recovered from the initial well-log after the upscaled.
- Then Acoustic impedance calculated by multiplying the upscaled density logs by the upscaled p-wave velocity logs (fig. 11).

The spatial continuity pattern of each property was inferred through the modelling of an experimental variogram computed from the upscaled well-log data in the vertical direction and from the actual partial angle stacks in the horizontal direction. Due to the great distances between wells, the horizontal spatial continuity pattern was calculated from the seismic reflection data.

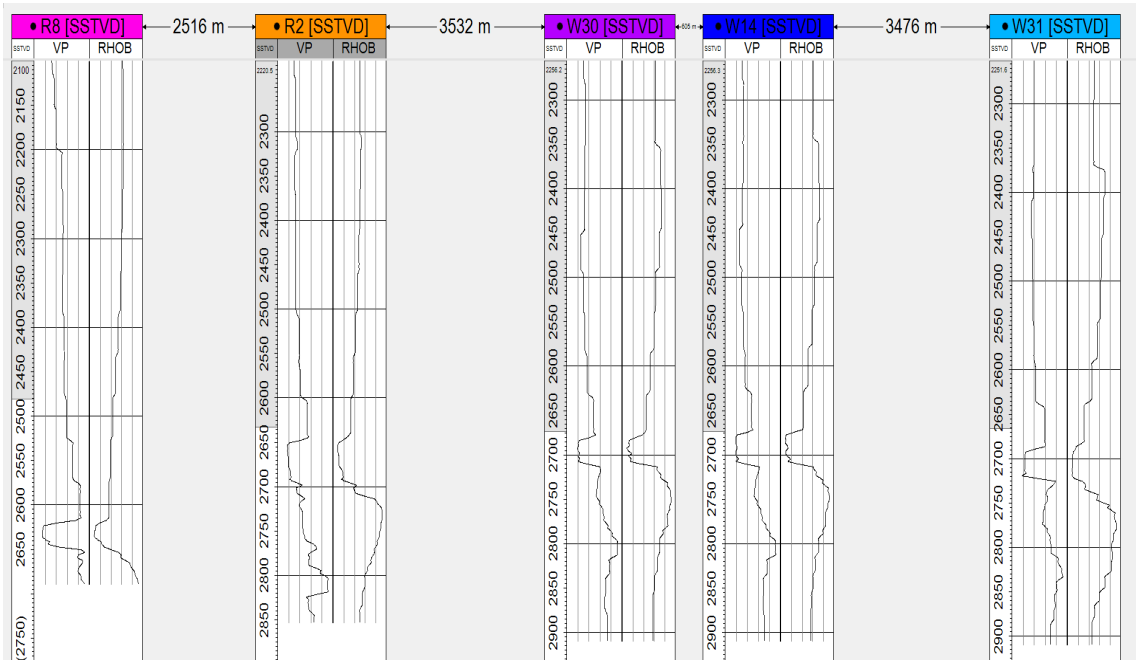


FIGURE 9: 5 of 13 Well Logs Located Along the Study Area.

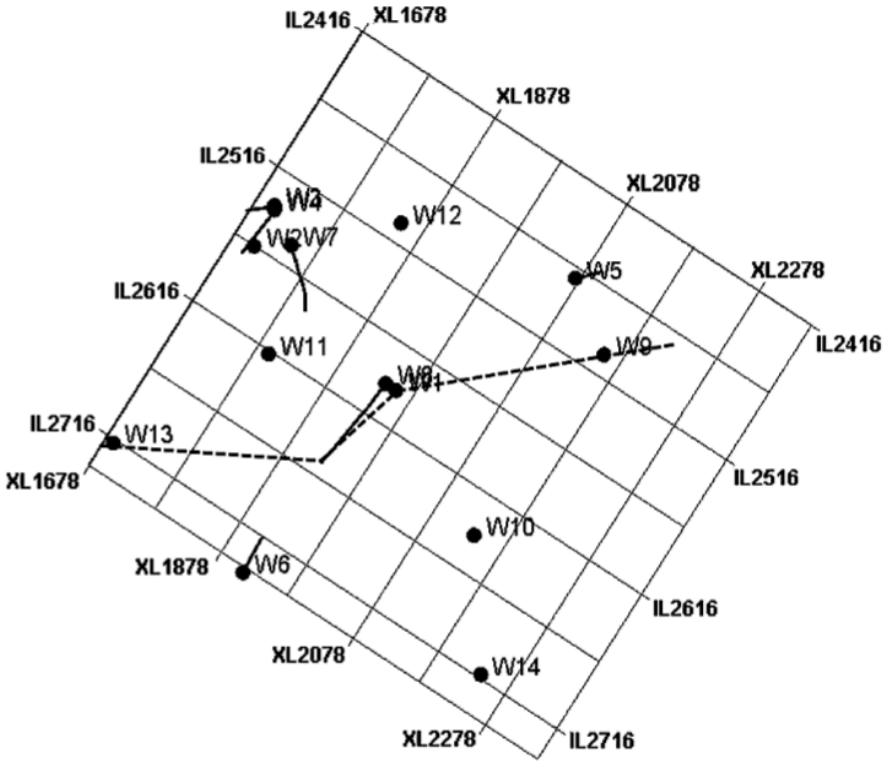


FIGURE 10: The Available Set of Wells and Its Location Within the Seismic Grid

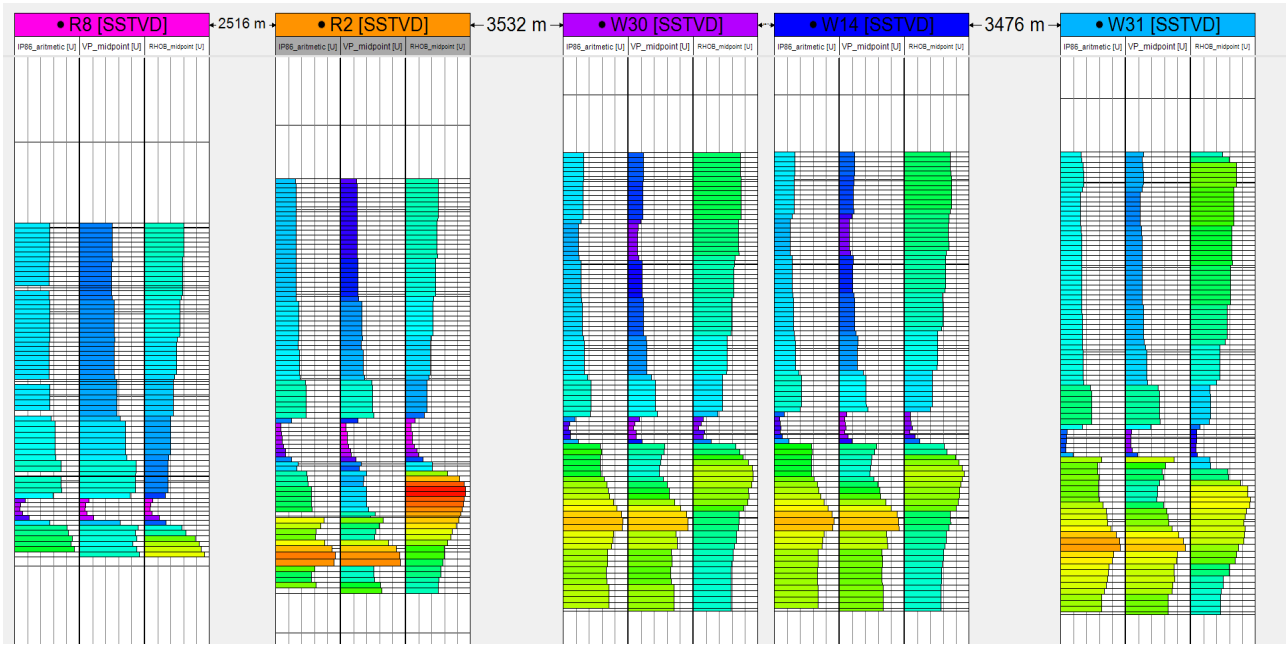


FIGURE 11: 5 of 13 Upscaled Well Logs and Ip Logs

The upscaled Ip well log in fig. 11 will be used to generate the synthetic seismic through direct sequential simulation and co-simulation through MATLAB which will be compared with the real seismic through the correlation coefficient.

4.1.2 THESIS TASKS

1. Developed MATLAB codes for GSI and FDA model.
2. Run the GSI Ip on the input data (seismic data and Ip log data processed by Professor Leonardo Azevedo using petrel) and interpret results (Best Global Correlation Coefficient, Best Ip and DSS Ip).
3. Run the FDA on the input data and interpret results.
4. Compare final image from the FDA with GSI image.

Given the unavailability of petrel software the images of the real and synthetic seismic can't be viewed, hence, the analysis of results from this work will be solely on the correlation coefficient between traces from the real and synthetic seismic.

4.2 GEOSTATISTICAL SEISMIC INVERSION USING ACOUSTIC IMPEDANCE

4.2.1 DEFINING INVERSION GRID AND IP MODELING

The seismic grid defined was 101*101*90 in i, j, and k direction in MATLAB.

The variogram experiment was done by

- Trying different variogram parameters in MATLAB Appendix 1
- Comparing the global correlation coefficient output
- Choosing the experiment with the highest Global correlation coefficient (Appendix 1, fig 3).

These variogram experiment was done in MATLAB by changing the values of the variogram range and then selecting the experiment with the highest Global Correlation Coefficient at the last iteration for analysis.

4.2.2 RESULTS FROM THE ITERATIVE GSI EXPERIMENT

The results gotten as output which are used for the analysis and interpretation are:

- Best Correlation Coefficient – shows the maximum values of traces between synthetic and real seismic
- Best Acoustic Impedance – which sums all traces with the highest correlation coefficient from all simulations in each iteration.
- DSS Ip- which gives us the image of the model.
- Global Correlation Coefficient – shows the trend of the correlation coefficient with respect to number of simulations.

4.2.3 CONVERGENCE OF THE MODELS MATLAB

The following steps describes the steps in convergence of the models:

- Using direct sequential simulation and co-simulation, a set of experiments was performed (appendix 1, fig. 3) using 40 simulations and 10 iterations on a set of Ip models on the seismic grid while conditioned to the available Ip well-log data (fig. 12).
- Through direct sequential simulation on the Ip well logs, a set of synthetic seismic volume is generated.
- This set of synthetic seismic volumes generated from the Ip model was Calculated by convolution with an estimated wavelet and compared with the real seismic automatically by the forward model.
- The Ip traces that generates the synthetic seismic with the highest correlation coefficient is stored as best Ip (fig. 17) along with the corresponding correlation coefficient (best cc) (fig. 14) in a secondary volume.

- These secondary volumes are used as secondary variables and local regionalized models to produce a new set of Ip model (fig. 20) for the next iteration.
- The steps are repeated till the 10th iteration.

Per iteration, 40 sets of Acoustic Impedance were simulated and co-simulated. After 5 iterations, the global correlation coefficient between the real and synthetic seismic trace is approximately 0.80 (fig. 13). From the graph, it is evident that the correlation between the real and synthetic seismic trace increases as the iteration proceeds. The iterative procedure stopped due to the small improvement in terms of correlation coefficient from iteration 5th to 10th.

The best simulated trace is retained for each simulation for every iteration and are interpreted from the image of best correlation coefficients between the traces from each iteration.

The image of the Best correlation coefficient (fig. 14, 15, 16) shows the maximum values of similarity of traces between the synthetic and real seismic. (fig.16 has more values with a higher correlation to the real seismic than fig. 15 and 14) for every iteration.

The image (fig. 17, 18) of the best Ip model is a combination of all best global correlation coefficient from each simulation of each iteration and shows the iteration with the highest number of correlations in each simulation considering the first and last iteration; the last iteration has a greater number of simulations with the highest number of correlation coefficients between the synthetic and real seismic trace when compared to results gotten from other iterations.

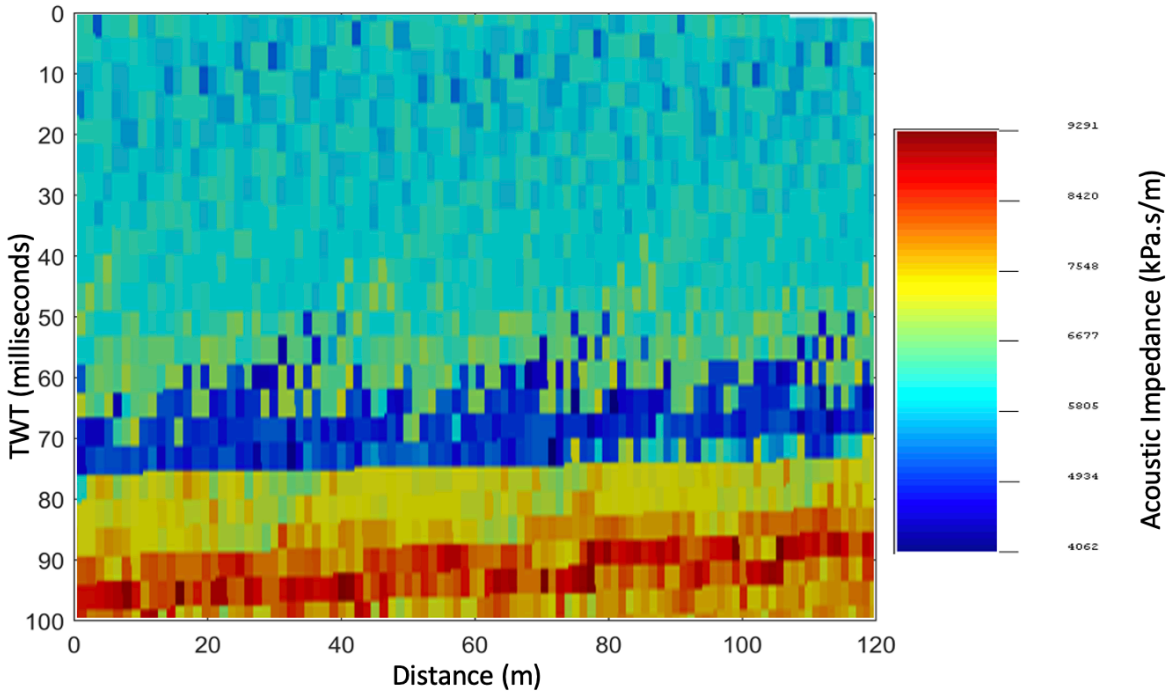


FIGURE 12: Image of Ip Well Log Data

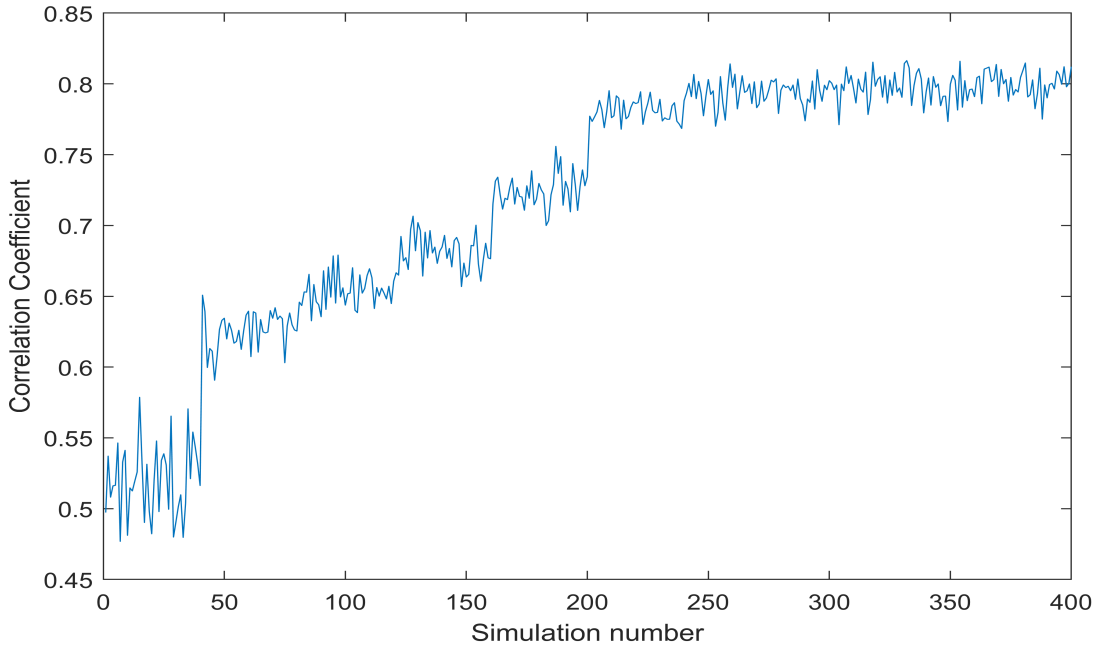


FIGURE 13: Global Correlation Coefficient between synthetic and real seismic traces

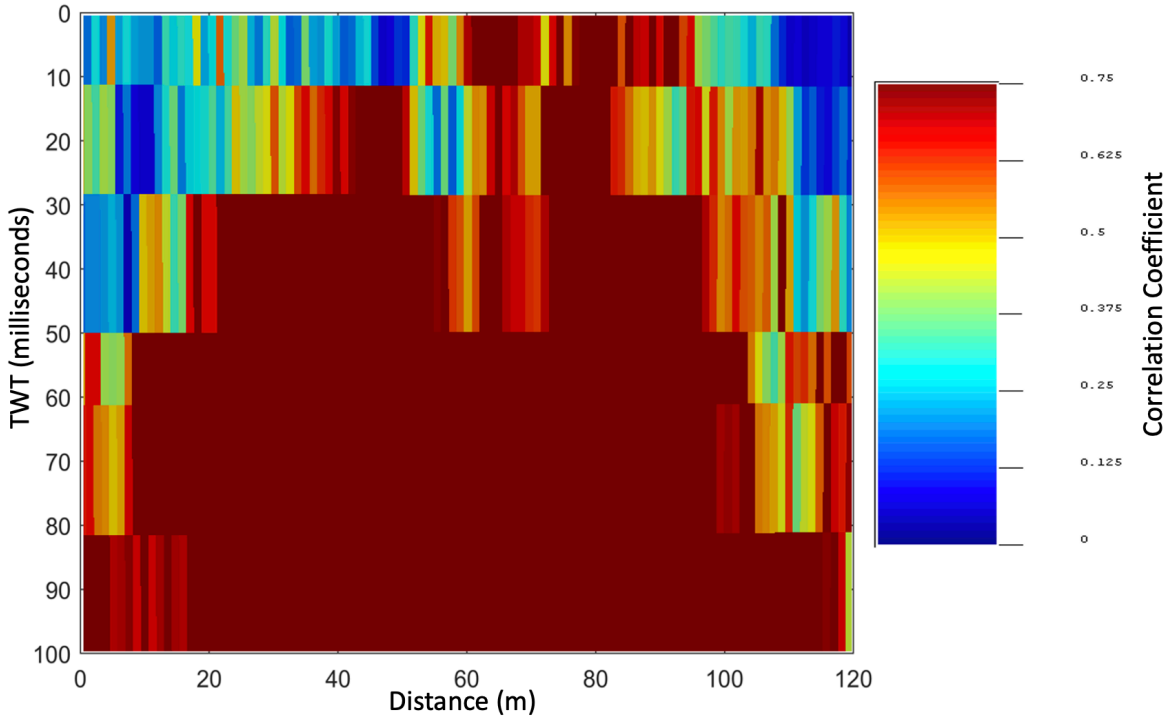


FIGURE 14: Image of Best cc from First Iteration

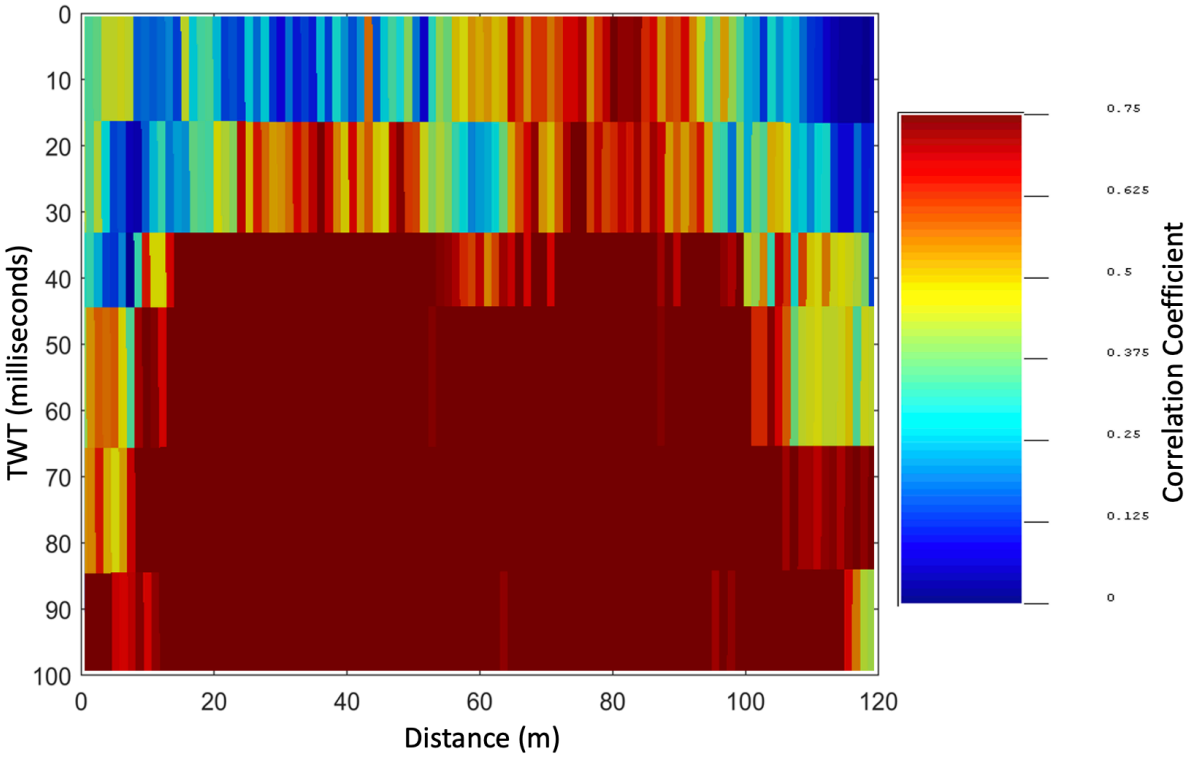


FIGURE 15: Image of Best cc from The Third Iteration

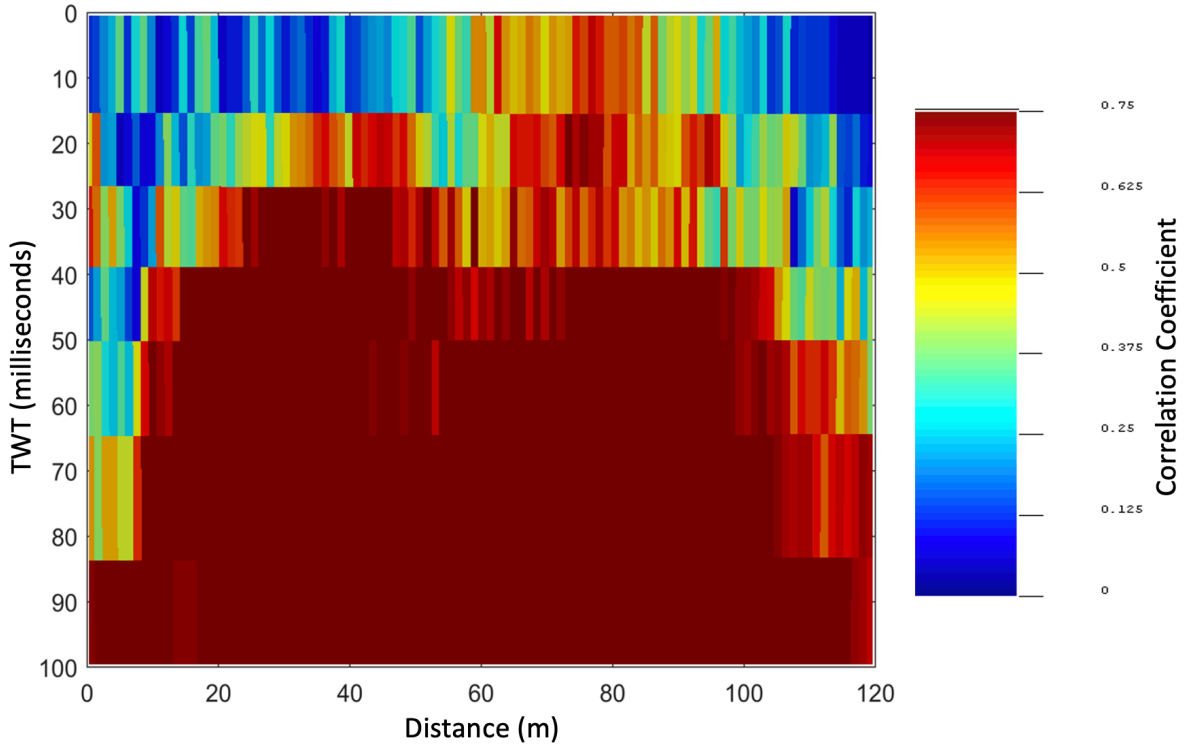


FIGURE 16: Image of Best cc of the Last Iteration

(Fig 14, 15, 16) shows the Best cc from the first, third, and tenth iteration, respectively.

The Best cc image shows the trace-by-trace matrix correlation between the real and synthetic traces. It is evident that the correlation coefficient becomes better as the iteration proceeds as the image of best cc from the last iteration has more traces with higher correlation coefficient when compared to images from the first and third iteration.

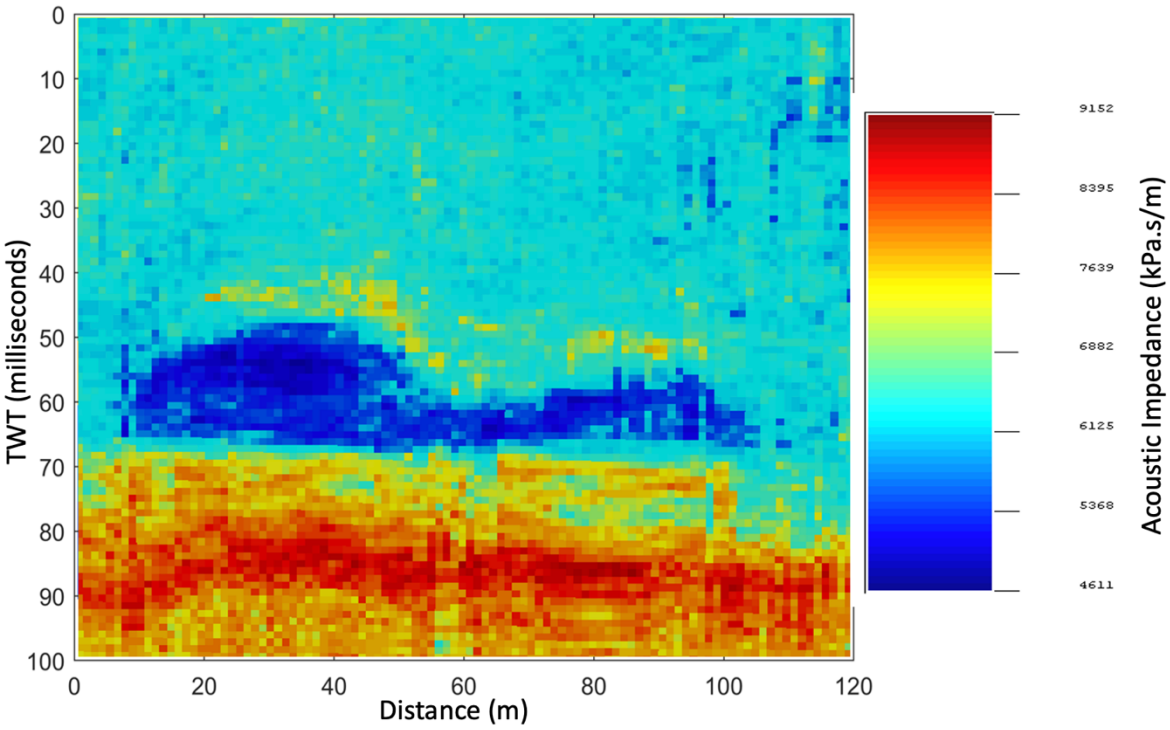


FIGURE 17: Image of Best Ip Model from First Iteration

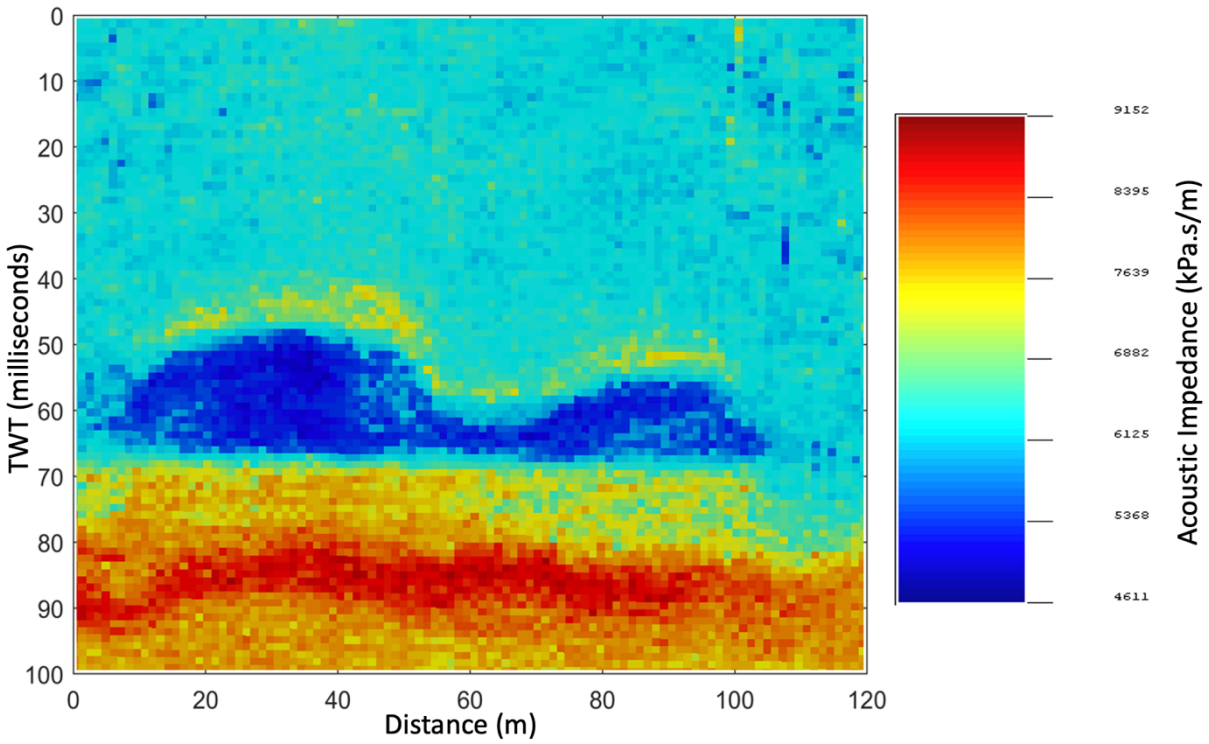


FIGURE 18: Image of Best Ip Model from the Last Iteration

Fig (17, 18) shows the best Ip model, which is obtained by the summation of all traces with the highest correlation coefficient from each simulation of each iteration.
It is obvious that the best Ip model gotten from the last iteration has more values with higher correlation hence the best Ip model gives a better image when compared with the image gotten from the first iteration this is due to the fact that the last iteration has more values with higher global correlation in each simulation. Results from the last iteration gives better image because the results get better as the iteration proceeds because best traces from previous iterations are used to generate the next synthetic images which is used as input data for the next iteration.

4.2.4 GSI Result Interpretation from DSS

Comparing the real seismic image (fig. 19) with the image direct sequential simulation results from the first, third, and last iteration (figure 20, 21, 22), it is evident that the DSS Ip image struggles to reproduce the features clearly. Still, we can see the Acoustic impedance of (6200 - 6500) occurring more frequently in the 10^{th}, which signifies the solution is close to the real data. This can be visualized using a histogram to show the distribution (fig. 25, 26, 27).

The inversion results can be assessed by interpreting the image of the mean Ip model (fig. 23, 24) from the ensemble of elastic models generated during the last iteration since it has the highest global correlation coefficient between the real and synthetic seismic.

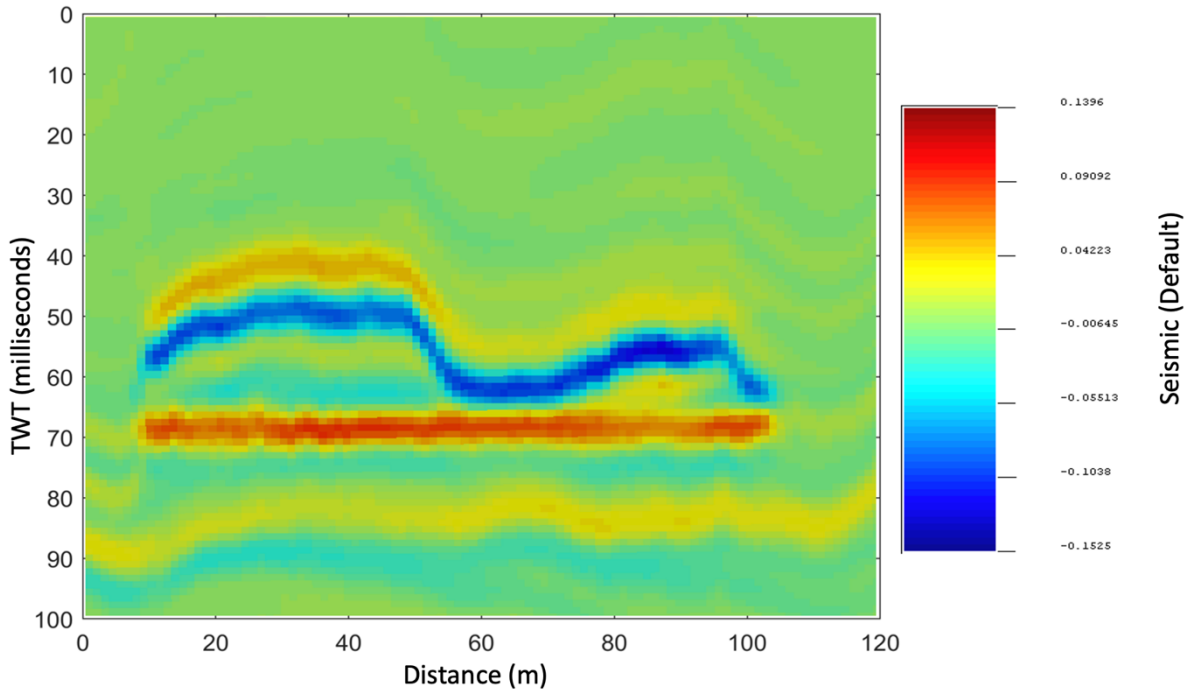


FIGURE 19: Image of True Seismic from SGEMS

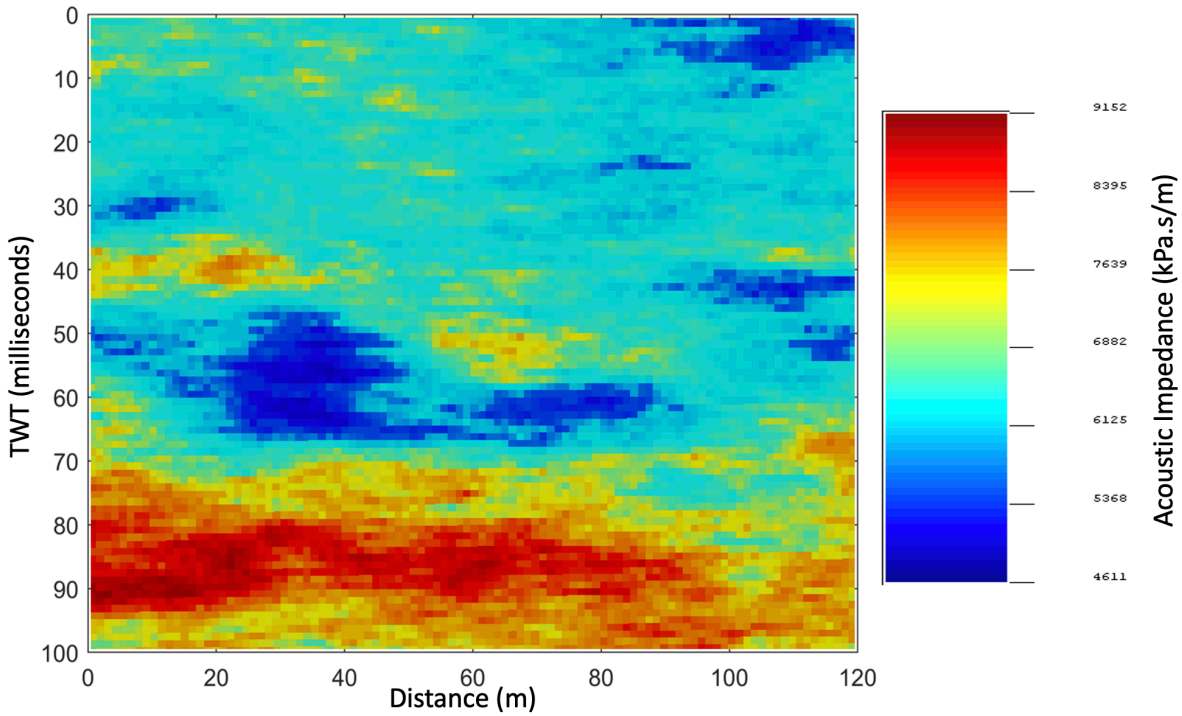


FIGURE 20: Image of Last DSS Ip model of the First Iteration

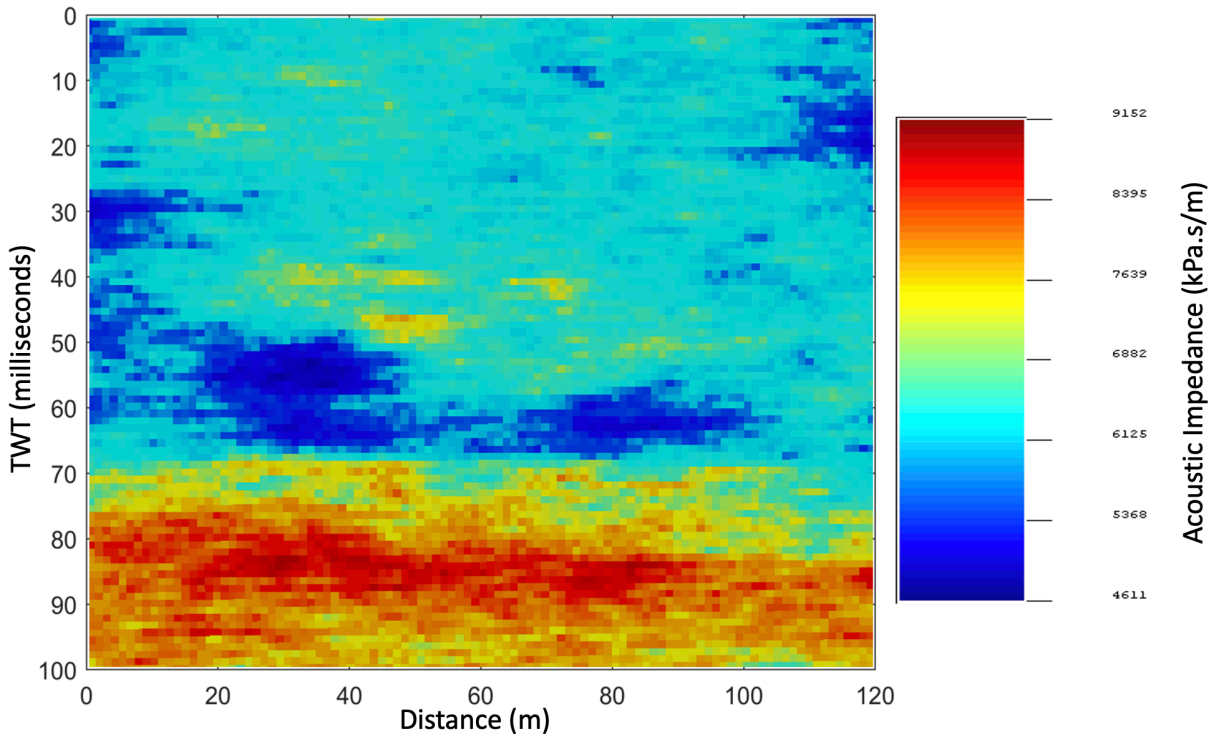


FIGURE 21: Image of Last DSS Ip model from the third Iteration

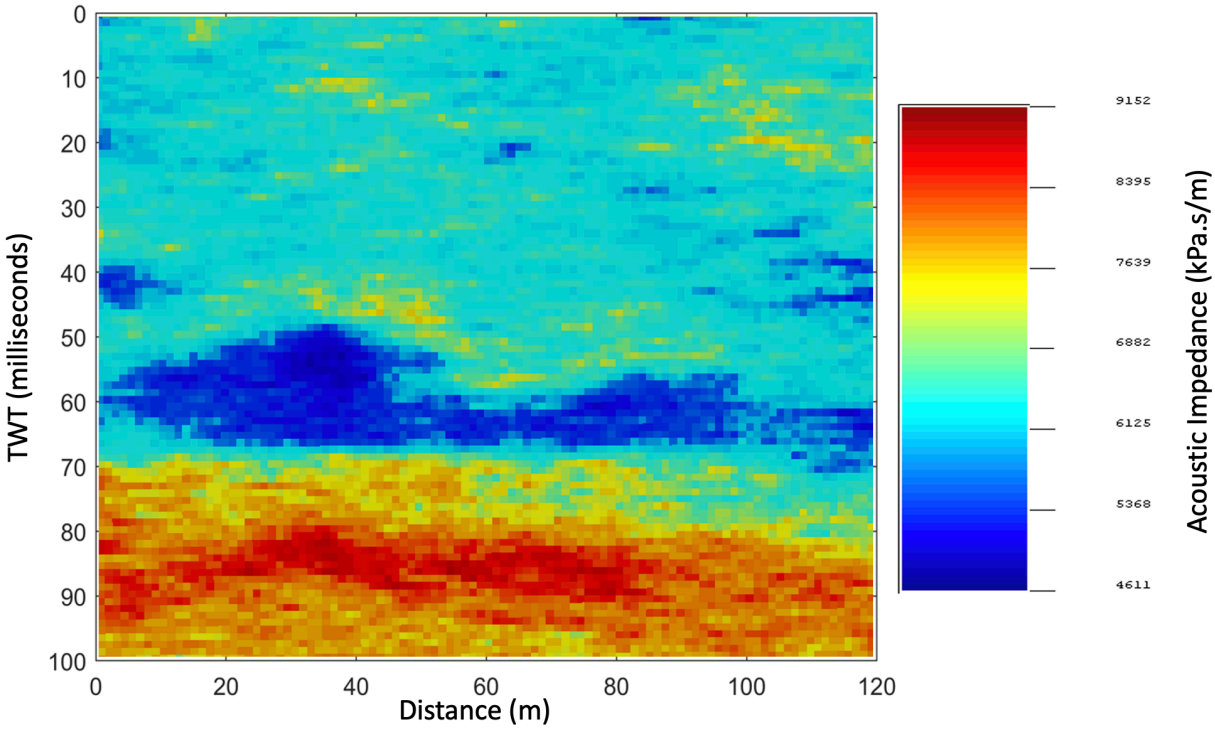


FIGURE 22: Image of Last DSS Ip model from The Last Iteration

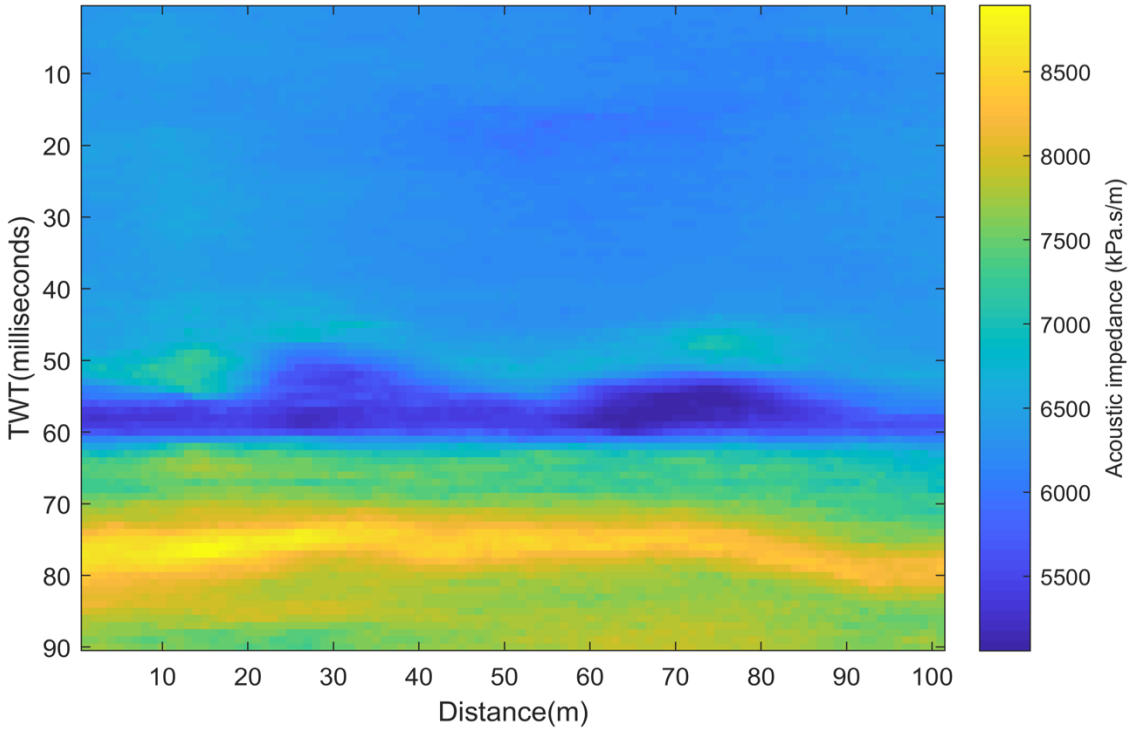


FIGURE 23: Image of Mean Acoustic Impedance Model from the Last Iteration from MATLAB

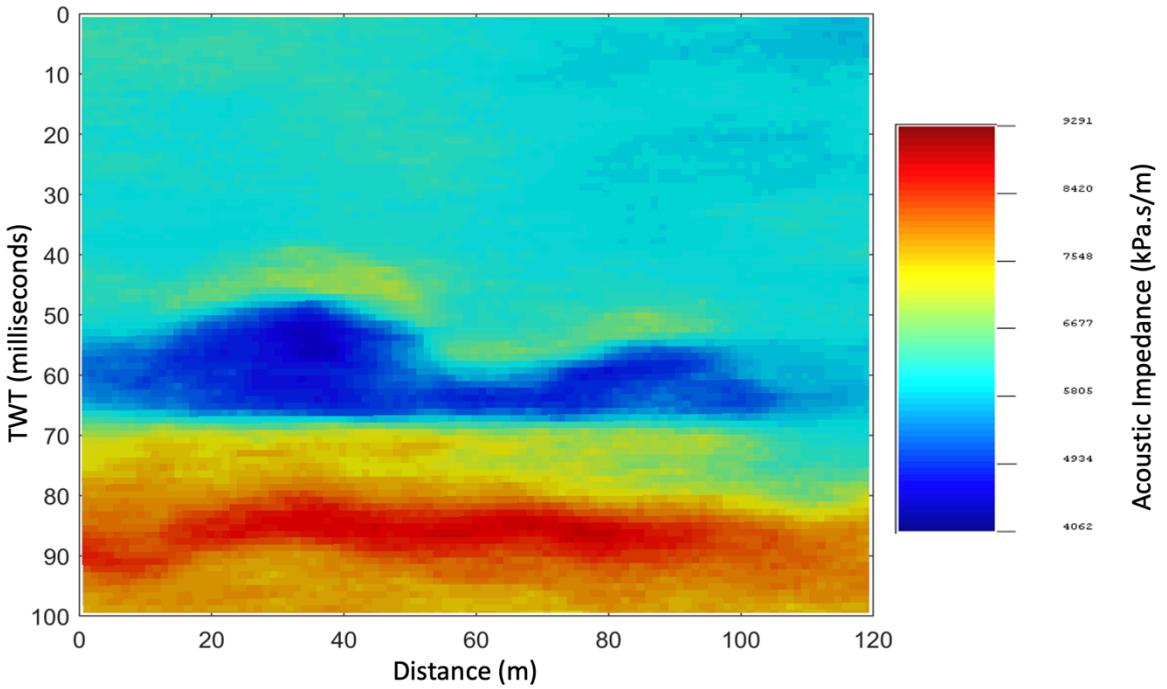


FIGURE 24: Image of Mean Acoustic Impedance Model from the Last Iteration from SGEMS

Looking at the image gotten from the mean Ip model (fig. 23, 24), it is evident that it gives a clear representation of the features from the seismic image (fig. 19). We can clearly see areas of low acoustic-impedance which are potentially hydrocarbon bearing regions.

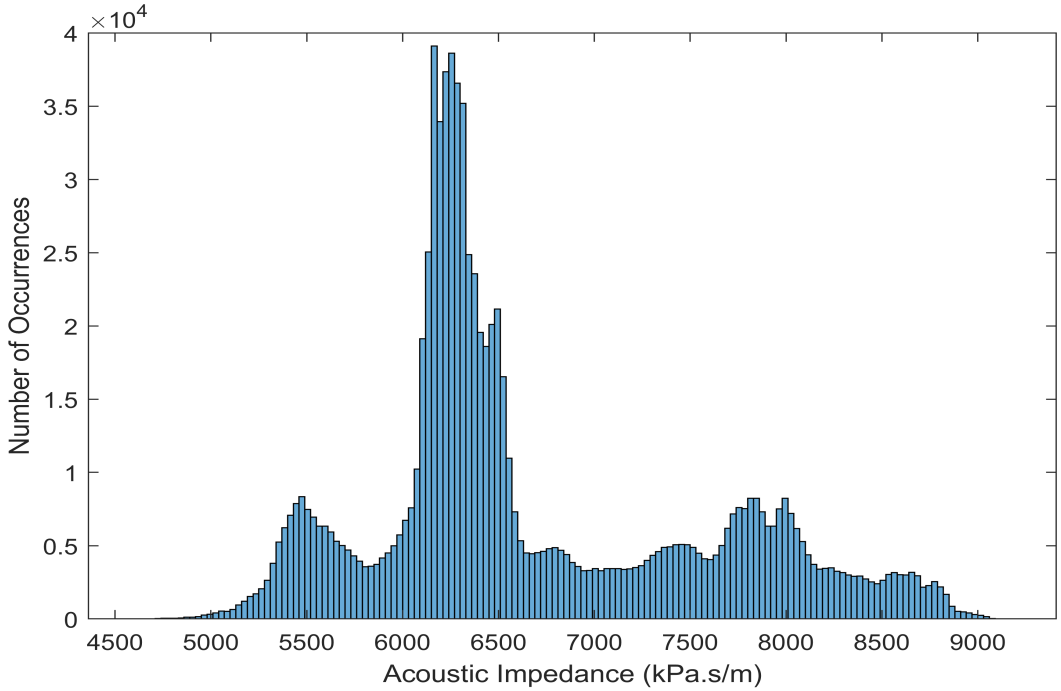


FIGURE 25: Ip Distribution Last Simulation of the First Iteration

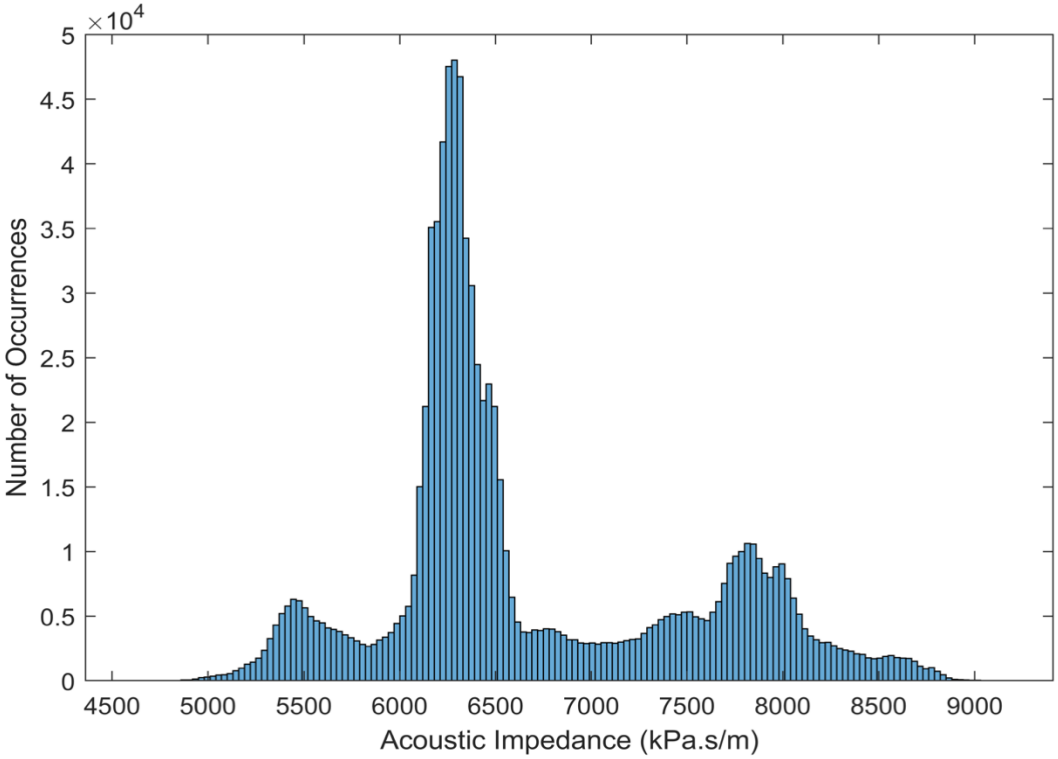


FIGURE 26: Ip Distribution Last Simulation of the third Iteration

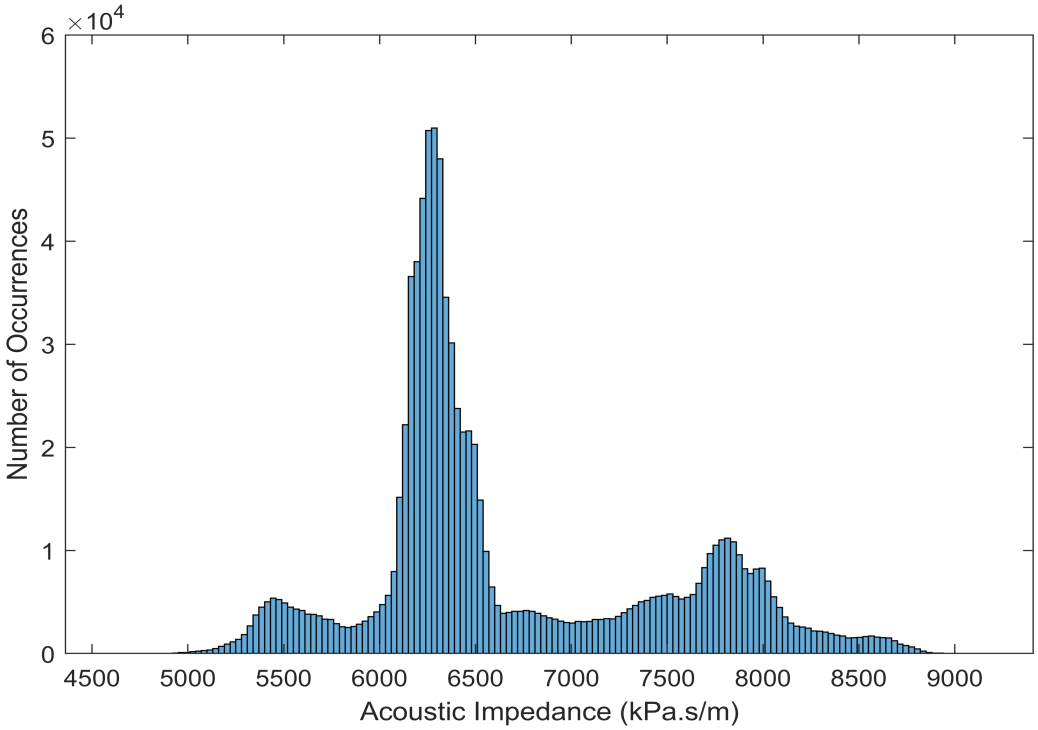


FIGURE 27: Ip Distribution Last Simulation of the Last Iteration

The rise in the peak (fig. 25, 26, 27) signifies an increase in the trace-by-trace correlation coefficient as the iteration proceeds.

4.3 FUNCTIONAL DATA ANALYSIS USING FPCS

The aim of this proposed method is to reduce the computational time of the traditional GSI Ip model. The same analysis was carried out as we did with the DSS, so only the final results and comparison as with the GSI will be carried out.

Choosing the experiment with the highest Global correlation coefficient (Appendix 1, fig 2).

4.3.1 FDA RESULT INTERPRETATION

The inversion grid was defined using 101, 101, 90 in the i, j, k direction.
Using the same data set, a set of variogram experiments was run, changing the variogram range (appendix2). Using 50 simulations and 30 iterations with a simulation time of 0.31 secs per simulation (appendix 4, fig. 32) to get a suitable variogram model, which will describe the expected spatial continuity pattern of the Functional Principal Component model.

From (fig. 28), Per iteration, 40 sets of Acoustic Impedance were simulated and co-simulated. After the 3rd iterations, the global correlation coefficient between the real seismic trace and synthetic trace is approximately 0.75.

The iterative procedure stopped due to the small improvement in terms of correlation coefficient from iteration 3rd to 10th iteration.

The result (fig. 29, 30) is obtained by plotting the 3 simulated outputs of the last iteration and the mean Principal Components; the realized model was able to produce the marked similarities with the GSI Ip model (fig. 23, 24), which shows the ability to use functional data analysis to explore the posterior distribution of the inverse solution without compromising the exploration of model parameter space.

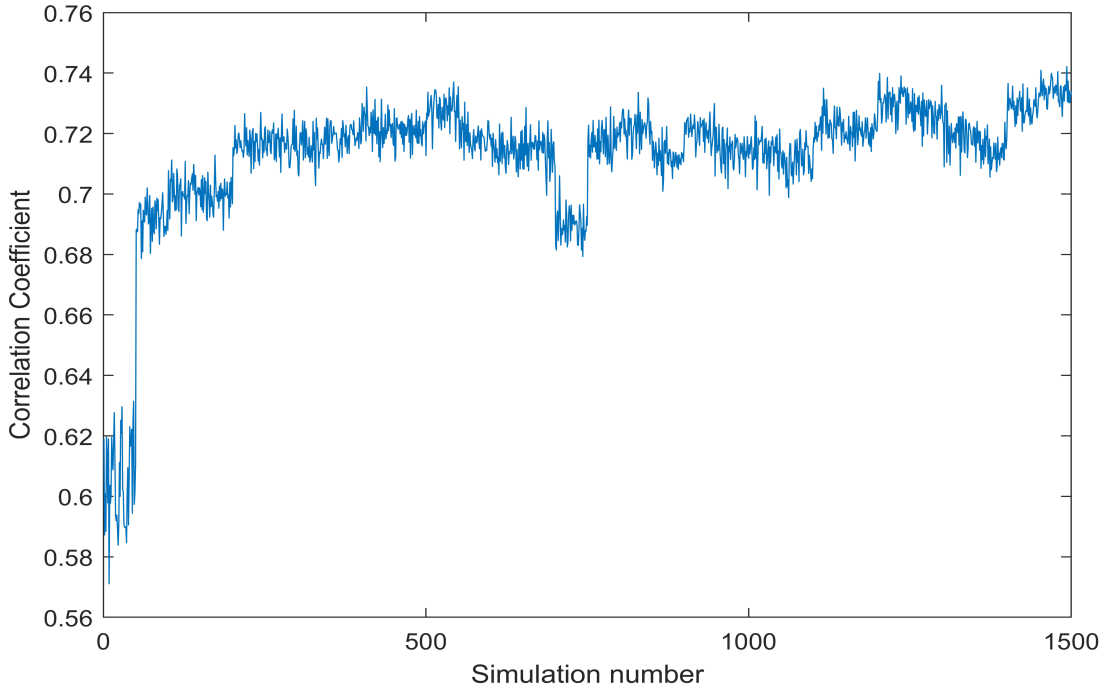


FIGURE 28: Global Correlation Coefficient between synthetic and real seismic traces from the FDA Experiment.

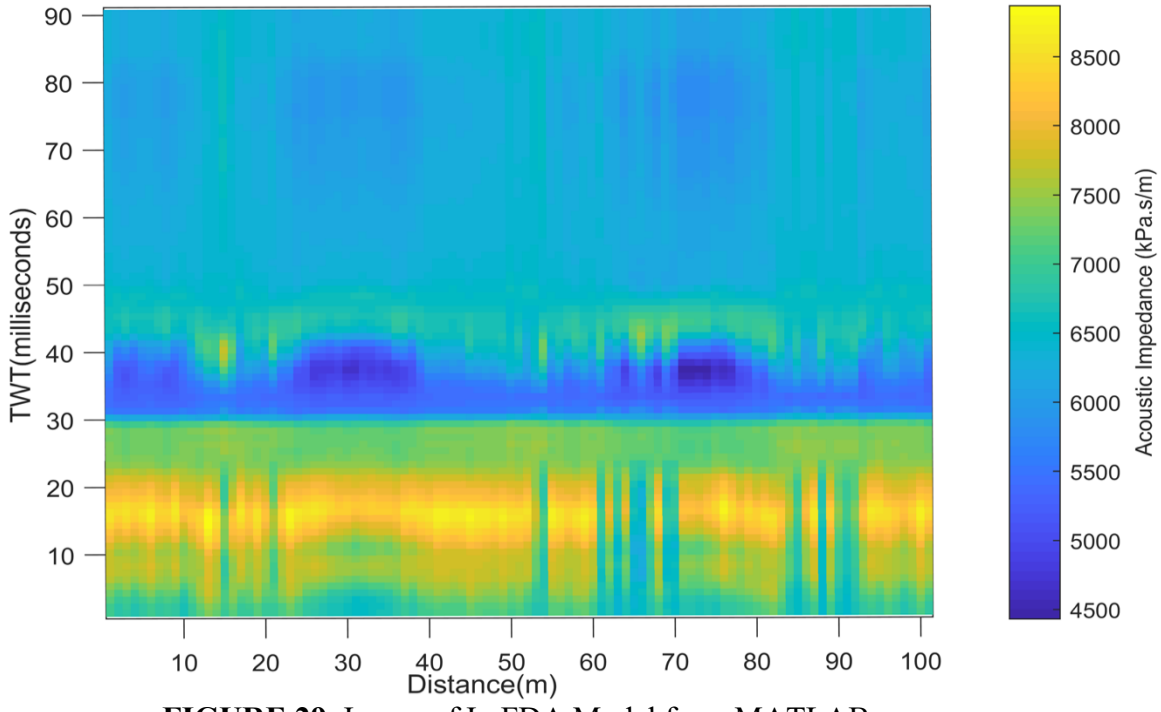


FIGURE 29: Image of Ip FDA Model from MATLAB

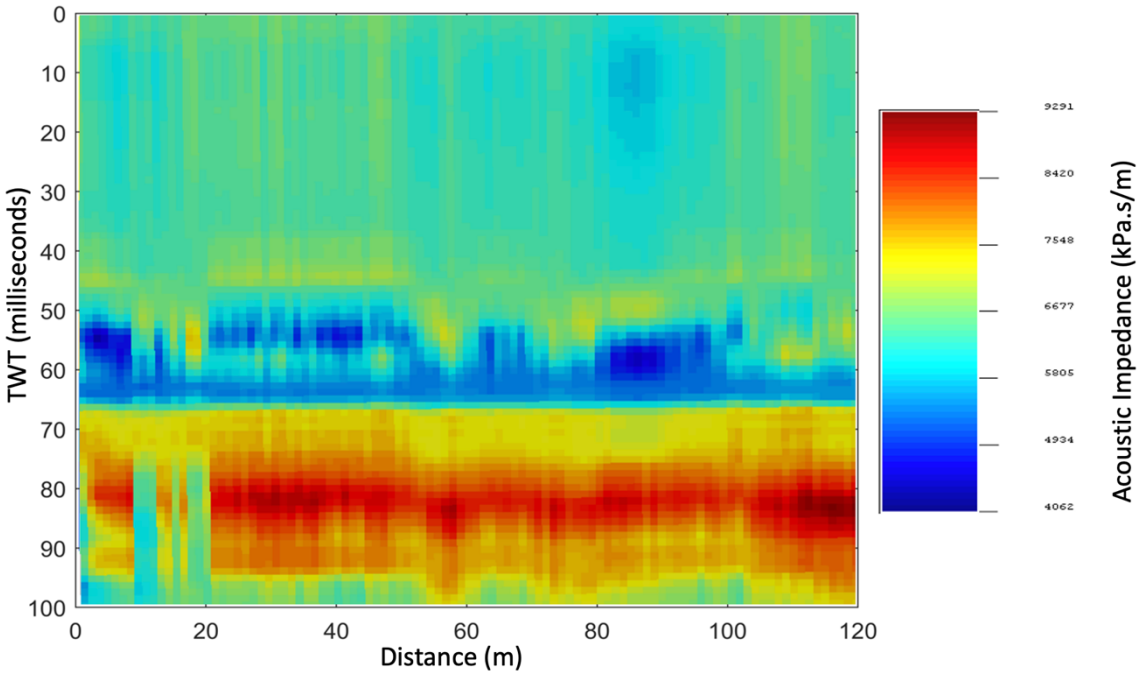


FIGURE 30: Image of Ip FDA Model from SGEMS

CHAPTER 5

5.0 COMPARISON

The goal of the proposed approach is to create a subsurface elastic properties model in order to quickly test alternative scenarios (e.g., testing various spatial continuity patterns and distributions of conditioning).

Compared to the GSI Ip process (Appendix 4, fig. 31), simulation time consumption improved significantly using the proposed FDA approach (Appendix 4, fig. 32). In addition, this method simplifies the search phase of the simulation, increasing the speed of simulation per node and reducing memory usage.

The proposed method ran in 0.47 seconds per simulation with 30 iterations and 50 simulations on a workstation with an Intel i5-8257U CPU with five cores at 1.40 GHz and 8 Gb of RAM. The conventional geostatistical acoustic inversion took 1.38 seconds to complete each simulation on the same workstation and used 10 iterations and 40 simulations.

The proposed method of inversion translates into a reduction of 66 percent of the original computational time by reducing the size of the dimension of the simulation grid.

The realized model was able to produce the marked similarities with the GSI Ip model (fig. 23, 24), which shows the ability to use functional data analysis to explore the posterior distribution of the inverse solution without compromising the exploration of model parameter space.

Difficulties with the proposed FDA method (fig. 29, 30), compared with the GSI Ip method (fig. 23, 24), include the inability to detect horizontal and directional well logs. Hence, it omits them because it looks for common patterns in well logs so horizontal wells appeared as vertical traces.

CHAPTER 6

6.0 CONCLUSION

Functional data analysis is a well-established statistical method recently extended to problems related to geosciences. FDA method allows summarizing series belonging to both space and time in a set of analytical functions. However, the application of this technique in geophysics is limited. We use functional data analysis as a proxy for model perturbation technique in geostatistical seismic inversion and to sample from the posterior distribution achieved with Bayesian linearized inversion.

The results retrieved from the FDA are reliable and shows the ability to use functional data analysis to explore the posterior distribution of the inverse solution without compromising the exploration of model parameter space at a reduced computational time.

The FDA gave a good approximation of the GSI method. However, the FDA showed vertical traces as a result of the skipping of the directional wells because it looks for common patterns in well logs as results demonstrated a considerable speedup over the traditional GSI Ip method while achieving similar performance and a good representation of the GSI impedance model.

The proposed approach was successfully used to reduce the computational time of the Ip model of Geostatistical seismic inversion. The result showed that the FDA model is capable of reproducing the same GSI Ip model results; they are therefore reliable.

Besides, the proposed FDA technique may also be used within other different seismic inversion algorithms such as AVO in order to speed up iterative geostatistical seismic inversion.

6.1 FURTHER WORKS

The FDA model can be improved by applying predictive analysis to predict the values directional well logs vertically before applying the FDA to prevent the vertical traces due to skipping of the directional well logs by the FDA. The retrieved inverse models should be free from vertical traces since they will incorporate more knowledge of the subsurface geology.

CONSTRAINTS

Due to the pandemic, I was unable to get access to Petrel software to visualize of the synthetic seismic and real seismic data, hence the work was done exclusively with MATLAB and SGEMS and the comparisons were done considering the correlations between the real and synthetic seismic traces and not with the real and synthetic seismic image.

REFERENCE

Anon (n.d.) Spatial statistics. [Online]. Available from: https://petrowiki.org/Spatial_statistics.

Avseth, P., Mukerji, T. & Mavko, G. (n.d.) Quantitative Seismic Interpretation. [Online] Available from: doi:10.1017/CBO9780511600074.

Avseth, P., Mukerji, T., Mavko, G. & Dvorkin, J. (2010) Rock-physics diagnostics of depositional texture, diagenetic alterations, and reservoir heterogeneity in high-porosity siliciclastic sediments and rocks - A review of selected models and suggested work flows. Geophysics. [Online] 75 (5). Available from: doi:10.1190/1.3483770.

Azevedo, L., Nunes, R., Correia, P., Soares, A., et al. (2013) Multidimensional scaling for the evaluation of a geostatistical seismic elastic inversion methodology. Geophysics. [Online] 79 (1), M1–M10. Available from: doi:10.1190/GEO2013-0037.1.

Azevedo, L., Nunes, R., Soares, A., Mundin, E.C., et al. (2015) Integration of well data into geostatistical seismic amplitude variation with angle inversion for facies estimation. Geophysics. [Online] 80 (6), M113–M128. Available from: doi:10.1190/GEO2015-0104.1.

Azevedo, L., Nunes, R., Soares, A., Neto, G.S., et al. (2018) Geostatistical seismic Amplitude-versus-angle inversion. Geophysical Prospecting. [Online] 66, 116–131. Available from: doi:10.1111/1365-2478.12589.

Azevedo, L. & Soares, A. (2017) Geostatistical methods for reservoir geophysics.

Bortoli, L.-J., Alabert, F., Haas, A. & Journel, A. (1993) Constraining Stochastic Images to Seismic Data. [Online] 325–337. Available from: doi:10.1007/978-94-011-1739-5_27.

Bosch, M., Mukerji, T. & Gonzalez, E.F. (2010) Seismic inversion for reservoir properties combining statistical rock physics and geostatistics: A review. Geophysics. [Online]. 75 (5). Available from: doi:10.1190/1.3478209.

Bottazzi, F. & Della Rossa, E. (2017) A Functional Data Analysis Approach to Surrogate Modeling in Reservoir and Geomechanics Uncertainty Quantification. Mathematical Geosciences. [Online] 49 (4), 517–540. Available from: doi:10.1007/s11004-017-9685-y.

Caetano, H. (2012) Integration of Seismic Information in Reservoir Models: Global Stochastic Inversion. New Technologies in the Oil and Gas Industry. [Online] Available from: doi:10.5772/52308.

Cristea, D. (2018) Combining Seismic Inversion and Geostatistics to Predict Reservoir Properties *. [Online] 42294, 1–3. Available from: doi:10.1306/42294Cristea2018.

Daya, S.B.S., Cheng, Q. & Agterberg, F. (2018) Handbook of mathematical geosciences: Fifty years of IAMG. [Online]. Springer International Publishing. Available from: doi:10.1007/978-3-319-78999-6.

Deutsch, C. V & Journel, A.G. (1996) GSLIB : Geostatistical Software LibraUser' User' s Guide Second Edition Preface to the Second Edition. 44–47.

Doyen, P.. (2007) Seismic Reservoir Characterization: An Earth Modelling Perspective. In: [Online]. p. 255 pp. Available from: doi:10.3997/9789073781771.

Gomez-Hernandez, J. & Journel, A.G. (1993) Joint sequential simulation of multigaussian fields, in Soares, A., ed., GeostatTROIA'92ROIA'92: Kluwer Academic Pub., Dordrecht, The Netherlands. 85–94.

Goovaerts, P. (1997) Geostatistics for Natural Resource Evaluation. In: Technometrics. p.

Haario, H., Saksman, E. & Tamminen, J. (2001) An adaptive Metropolis algorithm. Bernoulli. [Online] 7 (2), 223–242. Available from: doi:10.2307/3318737.

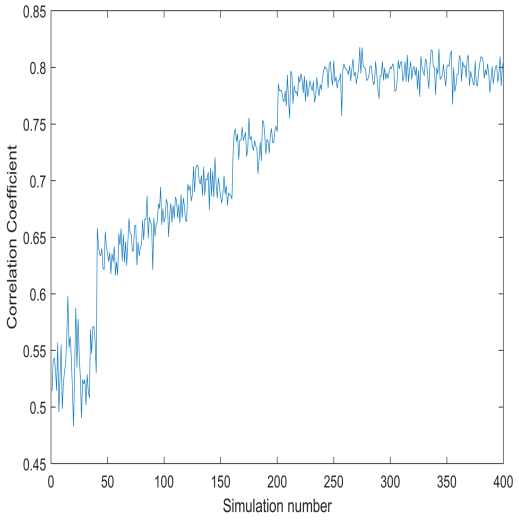
Horta, A. & Soares, A. (2010) Direct sequential Co-simulation with joint probability

distributions. Mathematical Geosciences. [Online] 42 (3), 269–292. Available from: doi:10.1007/s11004-010-9265-x.
Martins, P.M. (2014) Inversão Geoestatística de Dados Sísmicos 3D com Multi-distribuições Locais Pedro Miguel Martins Inversão Geoestatística de Dados Sísmicos 3D com Multi-distribuições Locais.
Menafoglio, Alessandra.,Grujic, O., & Caers, J. (2016) No Title. Universal Kriging of functional data: Trace-variography vs cross-variography? Application to gas forecasting in unconventional shales. [Online] Available from: doi:https://doi.org/10.1016/j.spasta.2015.12.003.
Nunes, R., Azevedo, L. & Soares, A. (2019) Fast geostatistical seismic inversion coupling machine learning and Fourier decomposition. Computational Geosciences. [Online] 23 (5), 1161–1172. Available from: doi:10.1007/s10596-019-09877-w.
P.M. Doyen, T.. G. (1992) seismic discrimination of lithology. [Online] 243–250. Available from: doi:10.1190/1.1889750.
Pereira, P., Calçôa, I., Azevedo, L., Nunes, R., et al. (2020) Iterative geostatistical seismic inversion incorporating local anisotropies.
Ramsay, J. and Silverman, B. (2005) (n.d.) No Title. Functional Data Analysis (Second ed.). Springer, New York.
Ramsay, J. (2009) Functional principal components analysis Outline.
Russell, B. (n.d.) Stochastic vs Deterministic Pre-stack Inversion Methods. [Online] Available from: https://www.cgg.com/data/1/rec_docs/3414_3412_Stochastic_vs_Deterministic_Inversion_Russell.pdf.
Russell, B. & Hampson, D. (1991) Comparison of poststack seismic inversion methods. 1991 SEG Annual Meeting. [Online] (April), 876–878. Available from: doi:10.1190/1.1888870.
Soares, A., Diet, J. D., Guerreiro, L. (2007) Stochastic inversion with a global perturbation method in Petroleum Geostatistics. [Online] Available from: doi:10.3997/2214-4609.201403048.
Soares, A. (2001) Direct sequential simulation and cosimulation. Mathematical Geology. [Online] 33 (8), 911–926. Available from: doi:10.1023/A:1012246006212.
Soares, A. (2006) Geostatistics for Earth and Environmental Sciences. [Online] Available from: https://www.wook.pt/livro/geoestatistica-para-as-ciencias-da-terra-e-do-ambiente-amilcar-soares/195467.
Tarantola, A. (2005) Inverse problem theory. [Online]. Available from: http://www.ipgp.fr/~tarantola/Files/Professional/SIAM/InverseProblemTheory.pdf%0Apapers2://publication/uuid/62A7A164-F0E3-4298-AA2B-222FF673BF4E.
Tompkins, M.J., Fernández Martínez, J.L., Alumbaugh, D.L. & Mukerji, T. (2011) Scalable uncertainty estimation for nonlinear inverse problems using parameter reduction, constraint mapping, and geometric sampling: Marine controlled-source electromagnetic examples. Geophysics. [Online] 76 (4). Available from: doi:10.1190/1.3581355.

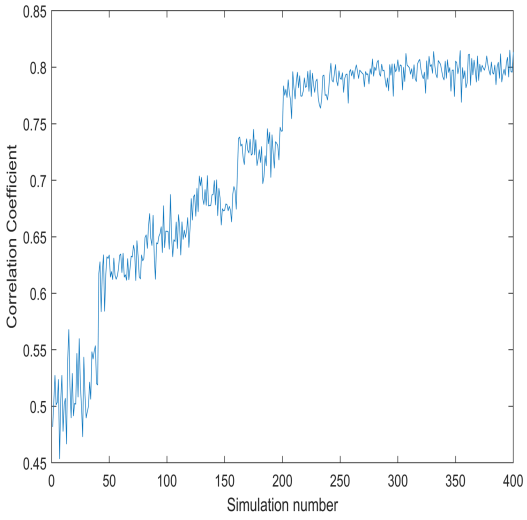
APPENDIX 1

Describes the different variogram experiments carried out for the GSI

1.

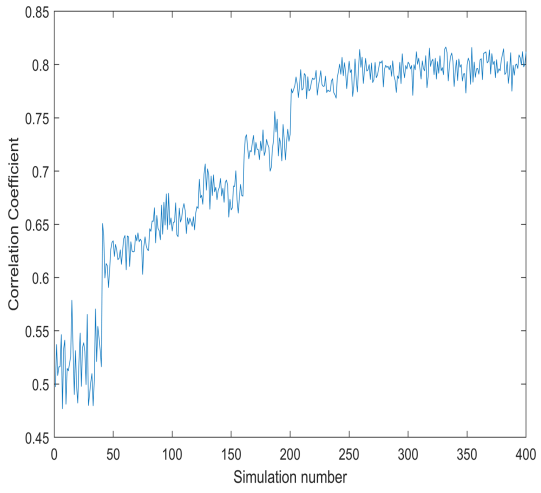


2.

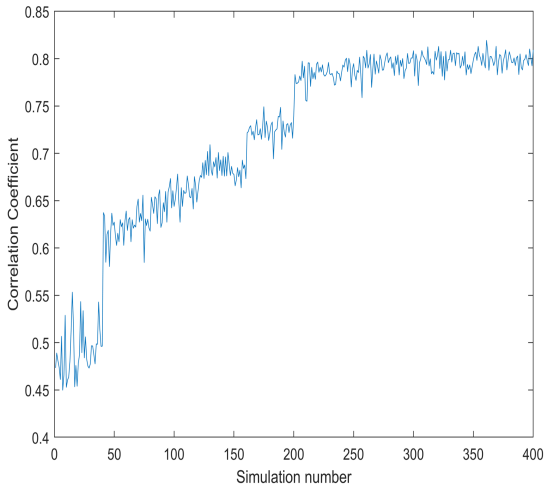


Vertical = 80	Vertical = 70
Horizontal major = 90	Horizontal major = 90
Horizontal minor = 20	Horizontal minor = 23

3.



4.

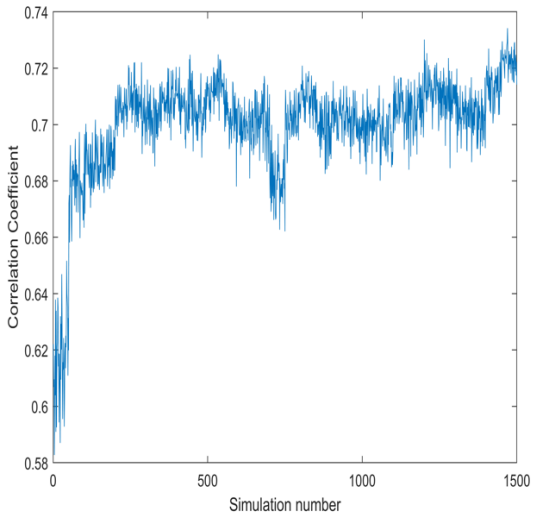


Vertical = 25	Vertical = 20
Horizontal major = 85	Horizontal major = 70
Horizontal minor = 90	Horizontal minor = 65

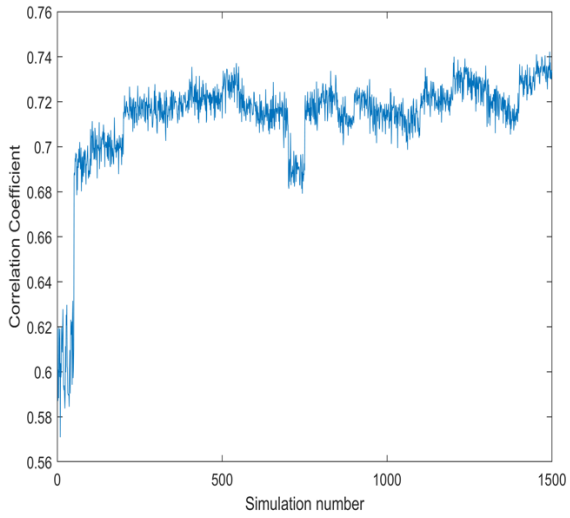
APPENDIX 2

Describes the different variogram experiments carried out for the FDA

1.

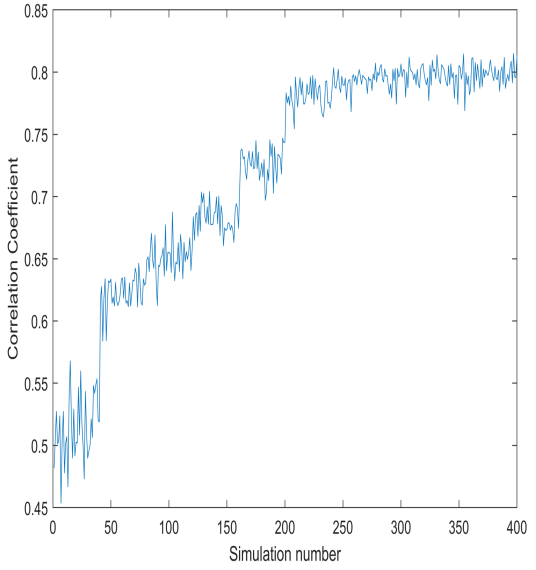


2.

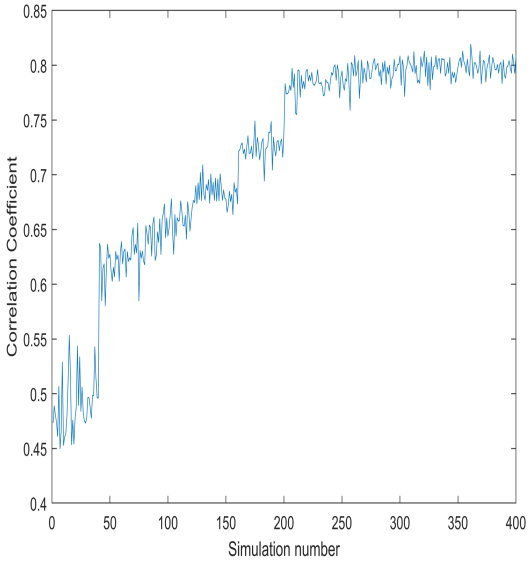


Vertical = 10	Vertical = 10
Horizontal major = 50	Horizontal major = 15
Horizontal minor = 15	Horizontal minor = 15

3.



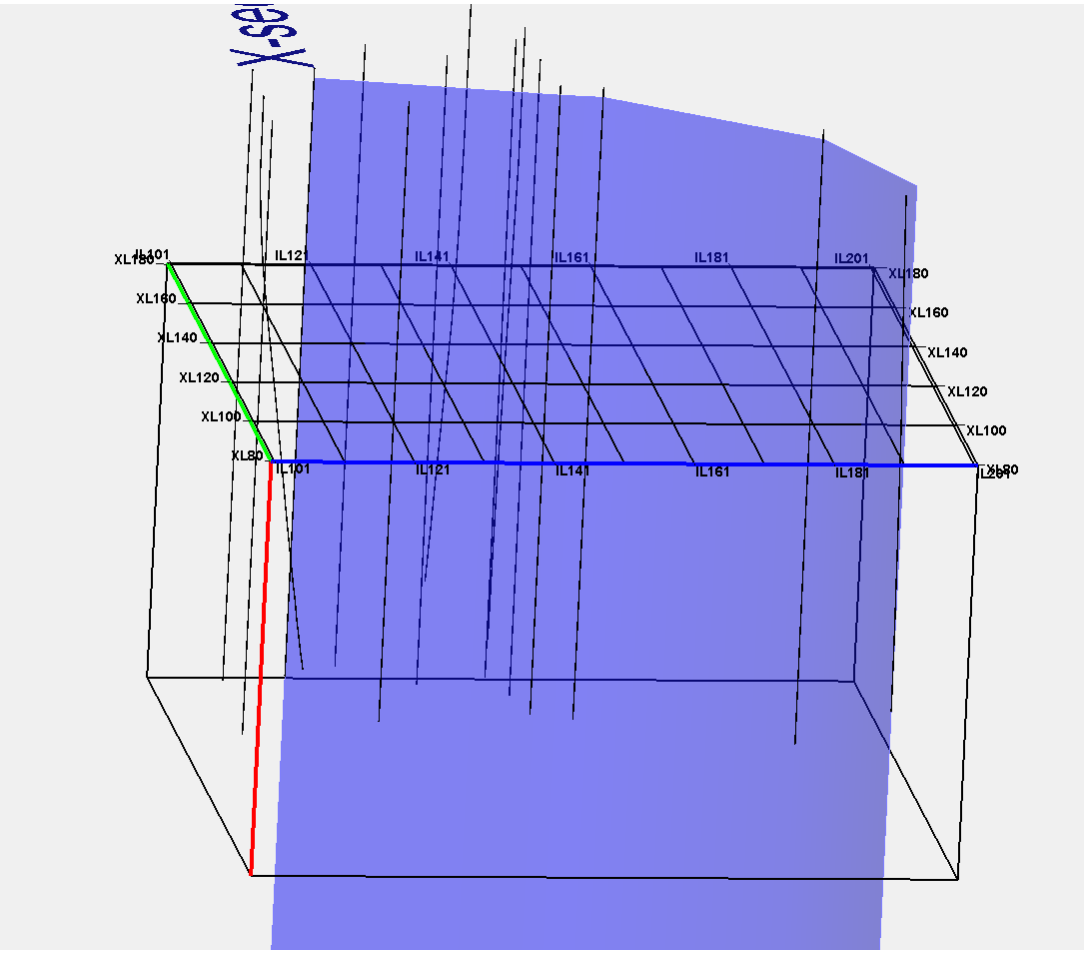
4.



Vertical = 15	Vertical = 15
Horizontal major = 50	Horizontal major = 40
Horizontal minor = 35	Horizontal minor = 30

APPENDIX 3

Shows the nature of all wells



APPENDIX 4

```
GLOBAL STOCHASTIC INVERSION - GSI
**********************************
Number of simulations per iteration 40
Number of iterations 10
**********************************
**********************************
Iteration number 1
**********************************
Simulating AI - Simulation number:  1
AI simulation - elapsed time (sec) 1.381835e+01
Forward model - elapsed time (sec) 6.987050e-02
Correlations - elapsed time (sec) 5.854860e-02
```

FIGURE 31: System Performance for GSI.

```
GLOBAL STOCHASTIC INVERSION - GSI
**********************************
Number of simulations per iteration 50
Number of iterations 30
**********************************
**********************************
Iteration number 1
**********************************
Simulating AI - Simulation number:  1
AI simulation - elapsed time (sec) 4.745528e-01
Simulating AI - Simulation number:  1
AI simulation - elapsed time (sec) 3.102675e-01
Simulating AI - Simulation number:  1
AI simulation - elapsed time (sec) 3.142332e-01
Forward model - elapsed time (sec) 8.571257e-01
Correlations - elapsed time (sec) 6.956630e-02
```

FIGURE 32: System Performance for FDA.