USE OF TOPOLOGICAL DATA ANALYSIS IN RESERVOIR ENGINEERING: APPLICATION TO INVERTED 4D SEISMIC DATA

A Thesis

by

ABDULHAMED ABDULAZIZ ALFALEH

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	John E. Killough
Committee Members,	Maria A. Barrufet
	Eduardo Gildin
Head of Department,	A. Daniel Hill

December 2014

Major Subject: Petroleum Engineering

ABSTRACT

Data analysis is one of the most important topics in any industry. In petroleum engineering, the complexity of reservoir data presents a challenge for engineers to study and make decisions. A new approach to analyze complex data is called topological data analysis, which aims to extract meaningful information from such data. It relies on the concept that complex data have shapes and these shapes can be translated to information.

The objective of this research was to use topological data analysis in studying reservoirs connectivity and compartmentalization. This topic is an essential component of reservoir engineering because it ensures the accuracy of forecasts and development plans, the correctness of reservoir simulation, and the success of performance diagnostics and optimization. In addition, introducing topological data analysis to reservoir engineering allows identification of reservoir engineering data behavior, detection of anomalies and events, and minimizing uncertainties.

Topological data analysis had been applied on inverted four-dimensional (4D) time-lapse seismic datasets. Two simulation models were used to generate the datasets: Brillig, and Norne. First, data were prepared for topological data analysis. Then, similarity distance function and lenses were defined and used to create topological data analysis graphs. Once completed, graph features were identified and analyzed. Lastly, the results were validated.

Topological data analysis was able to compartmentalize the reservoir models with various process configurations. It identified regions that matched the actual reservoir compartments in the simulation model. It has been proven to extract valued information from petroleum engineering data.

To my lovely family

Your love and support will always be remembered,

fondly in good times, and as encouragement in bad.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, Dr. John E. Killough for his valuable guidance and advice. This thesis would not have been possible without his tremendous contribution. It was a great honor to be a student member of his group. Additionally, I am also grateful for Dr. Maria A. Barrufet and Dr. Eduardo Gildin for serving in my committee and for providing his valuable time. I would like to thank Saudi Aramco for sponsoring my graduate studies, and Dr. Umar A. Nahdi for being a great mentor. Special thanks goes to Ayasdi Inc, especially Dr. Damir Herman for providing me with resources and support through the time of this research. The authors would like to thank Statoil (operator of the Norne field) and its license partners ENI and Petoro for the release of the Norne data. Further, the authors aknowledge the Center for Integrated Operations at NTNU for cooperation and coordination of the Norne Cases. I am also thankful to Dr. Killough group members. Finally, I would like to thank my parents, brothers and sisters for their usual support and prayers.

The view expressed in this research are the views of the author and do not necessarily reflect the views of Statoil and the Norne license partners.

GLOSSARY

2D	Two-Dimensional
3D	Three-Dimensional
4D	Four-Dimensional
GOC	Gas Oil Contacts
OWC	Oil Water Contacts
PE	Petroleum Engineering
P	Pressure
RE	Reservoir Engineering
Sw	Water Saturation
TDA	Topological Data Analysis
VNE	Variance Normalized Euclidean

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION AND LITERATURE REVIEW

1.1 Introduction

Data analysis is one of the most important topics in any industry. In petroleum engineering, the complexity of reservoir data presents a challenge for engineers to study and make decisions. In the following section, a background about analysis of such data in petroleum engineering is provided. In addition, a review of a new data analysis technique called topological data analysis (TDA) is presented, as well as a review of the compartmentalization problem that is used to explore the capabilities of TDA. This chapter concludes with the research objectives.

1.2 Background

Because of the advancement of technology in recent years, the problem has shifted from the lack of data to having significant and uncontrollable volumes of it. With the current high-frequency data collection tools, large and complex datasets (big data) have become one of the main topics in all industries including medical, financial, and energy industries. In the petroleum industry, many areas from exploration to production are subject to significant data collection and need tools to take advantage of such data. For example, real time sensors from wells provide high-frequency records of well performance data. This creates a different set of problems that cannot be solved by the traditional analysis techniques. This research explores a new technique called topological data analysis to analyze such data.

1.2.1 Data Mining

Data mining is the process used to extract meaningful patterns and information from complex data in quantity, quality, heterogeneity, and non-linearity. It has

the capability to extract profound insights from massive and diverse data streams to assist the engineers in making better decisions. It improves the value of the data by applying different processes such as clustering, classification, summarization, regression, deviation detection, and dependency modeling (Fayyad, 1997).

Data mining has been used in petroleum engineering in many areas such as seismic data analysis, reservoir surveillance, and prediction (Bailey et al., 2013; De Jonge et al., 2002; Marroquín et al., 2008). In reservoir engineering (RE), the applications include threat management, production enhancement, performance diagnostics, and optimization. Traditional data mining methods such as self organizing maps and back propagation neural network were used for data quality control, well rate estimation, well-to-well interaction, and event detection (De Jonge et al., 2002). Data mining techniques have also been used to study well-to-well communication and injector-producer relationships to generate better sweep efficiency (Bailey et al., 2013).

1.2.2 Topological Data Analysis

A new approach in data mining to analyze complex data is called topological data analysis (TDA), which aims to extract meaningful patterns from such data. It relies on the concept that complex data have shapes and these shapes can be translated to information. Shape value can be recognized as data breakage to clusters. Also, it can be seen in flairs that represent a gradual divergence from the central core values. In addition, the shape can be loops that shows a recurrent behavior of the data (Carlsson, 2009). The shape is built based on a metric that is a distance function that attempts to measure similarity between data points. Distance formulas that satisfies the triangular inequality can be used like euclidean distance or correlation (Carlsson, 2009).

Topological data analysis is based on topology, which is the mathematical

formalism of measuring and representing shape (Carlsson, 2009). Since the Swiss mathematician Leonhard Euler started working on Topology in the 18th century (Euler, 1741), its applications were focused on defined shapes. Recently, the use of topology has expanded to data analysis of complex multidimensional data (Carlsson, 2009).

The main characteristics of topology that enable topological data analysis are coordinate-free representation, deformation invariance, and compressed representation. Coordinate-free representation means the constructed shape does not depend on the selected coordinates but the distance function, which allows the comparison of data derived from different systems. Second, deformation invariance provides the ability to detect the shape of the data regardless of deformation and noise. Finally, compressed representation retains the main shape features of the infinite or complex data points with a finite network of points (Lum et al., 2013).

Complex multidimensional shape is represented by an abstract network of points (graph) that reflects the structure of the data. It provides a global summarization of a high-resolution local constraints. The structure of the data depends on a distance function to create a complex multidimensional object. (Figure 1.1) shows an example of a simple two-dimensional dataset in which data points are positioned using a distance function. The graph is a two-dimensional projection of the complex shape acquired by applying lenses on the data. Lenses are functions that generate a real value for each point in the dataset (Figure 1.2). Lenses can be statistical calculated functions like mean, variance, density, or centrality. Also, any property from the dataset can be used as a function.

Using the function value of each data point and their distances, overlapping ranges of data points are set (Figure 1.3). The size of the ranges is decided by defined value called resolution in this research. Increasing lens resolution means

smaller ranges. Also, another value is defined for ranges overlap to control their shared covering. Resolution and overlap are important in getting a compressed representation. Based on the selected similarity metric, the data are clustered into nodes in each range. Then, an edge is created between any two nodes if they share the same data point (Figure 1.4). This generates a network of points (graph) that shows a two-dimensional projection of the complex shape (Ayasdi, 2013). The TDA approach is summarized by a three-dimensional (3D) object example in (Figure 1.5).

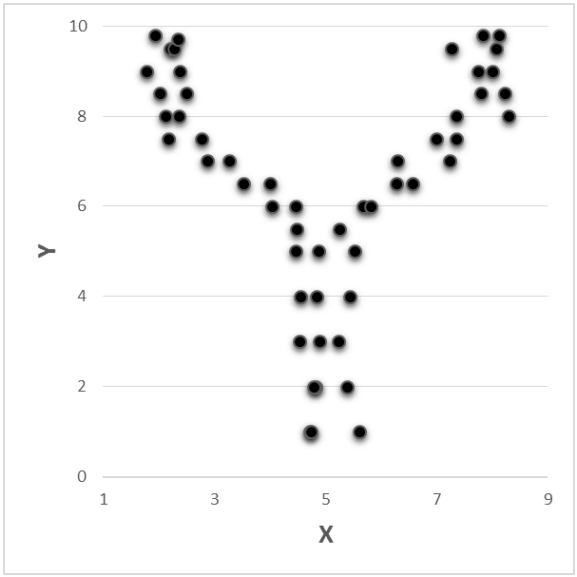


Figure 1.1: TDA First Step: Using Distance Function (Metric). The first step in TDA is to use a distance function (Metric) such as euclidean distance to position data point in space.

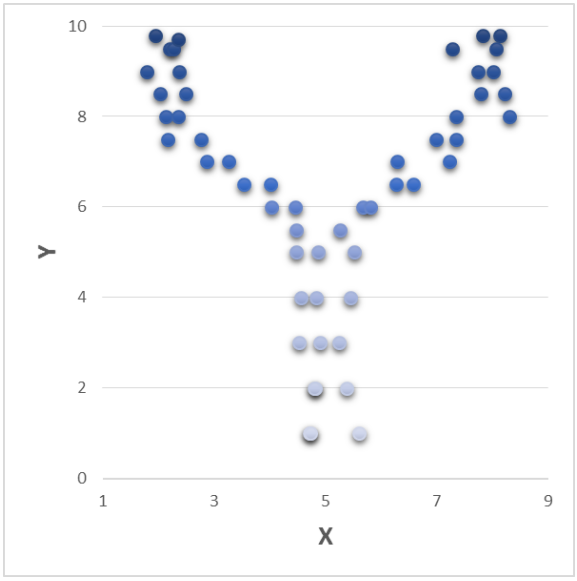


Figure 1.2: TDA Second Step: Applying Filter Function (Lenses). The second step in TDA is to calculate and color by a filter function value (Lens) such as Y value for each data point. More complex functions such as variance or L-infinity centrality can be used to color data points.

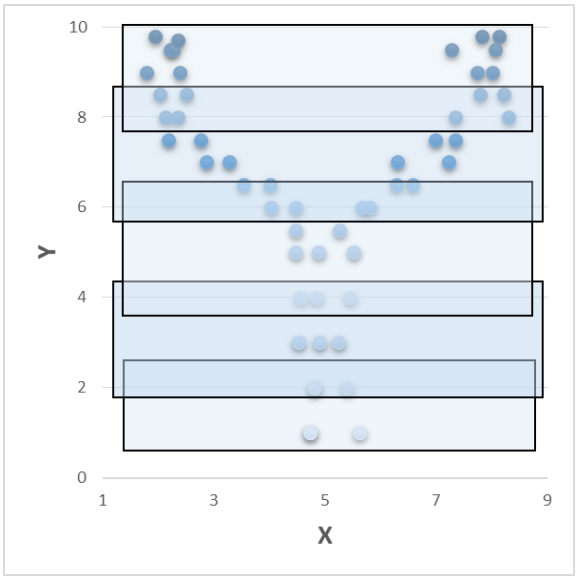


Figure 1.3: TDA Third Step: Grouping Data Points. The third step is to create overlapping groups based on the used filter function value (Lens).

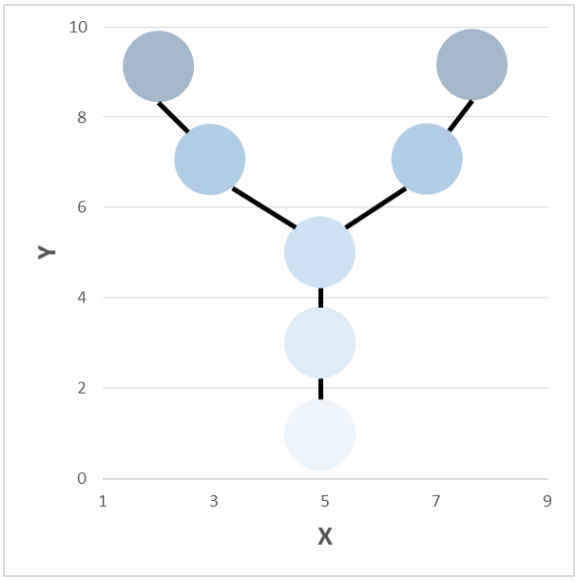


Figure 1.4: TDA Forth Step: Linking and Creating the Shape Graph. The fourth step is to cluster data points in each filtered group to create nodes in the shape graph. Finally, edges are created between nodes if they contain shared data points.

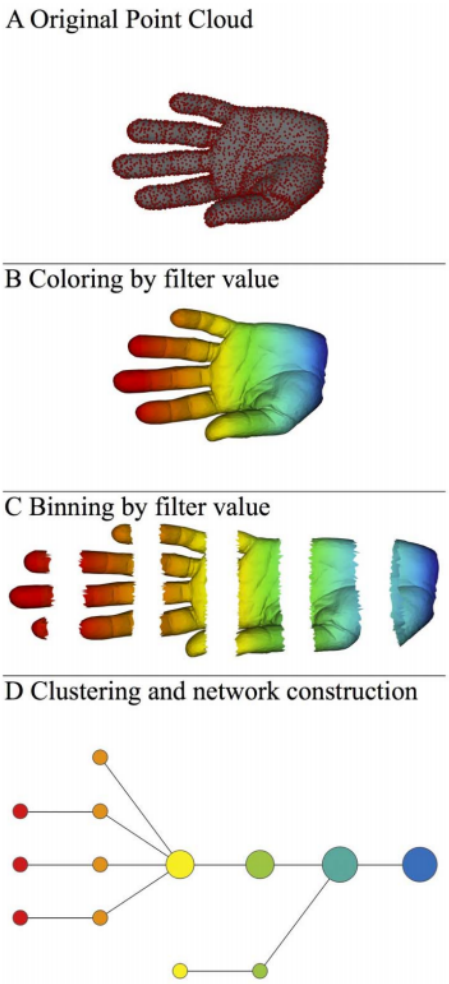


Figure 1.5: TDA Approcach Summary. TDA approach is summarized using a 3D object (hand) represented by a points cloud (Lum et al., 2013).

TDA has many successful applications in the real world. In the medical field, a study of the gene expressions of patients with breast cancer leads to the identification of patients subsets relevant to their survival rates (Figure 1.6). Another study shows the relationship between patients with leukemia and their genes, which results in better patient biomarkers (Lum et al., 2013, 2012; Nicolau et al., 2011). In the United States House of Representatives, TDA was applied to study the members'

voting data and implicit networks of their behavior was generated (Lum et al., 2013).

TDA success stories in the industry encourage its application in petroleum engineering. TDA is expected to complement other techniques used in petroleum engineering. It is a simple but powerful tool for the reservoir engineer to use while analyzing complex data. In comparison to clustering techniques, it provides an additional value by retaining the dataset geometric structure in the created graph. It shows internal connections of the nodes inside created clusters, which helps in studying continuous datasets.

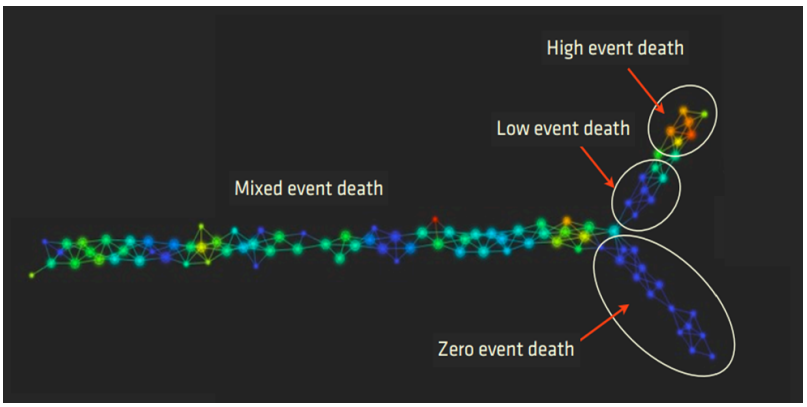


Figure 1.6: Patients with Breast Cancer Survival Graph. (Lum et al., 2012).

1.2.3 Compartmentalization

In this work, reservoir compartmentalization was selected to test topological data analysis and study its capabilities in petroleum engineering. A compartment is defined as a trap with no inner boundaries, which results in equilibrium of fluids at different depths (Snedden et al., 2007). Reservoir connectivity is an appropriate petroleum engineering and data mining problem to test TDA ability to identify connected reservoir regions (Figure 1.7).

The study of compartmentalization and its inverse reservoir connectivity is an essential area of research in reservoir description. Identifying reservoir compartments is very important for the correctness of reservoir simulation and the accuracy of forecasts and development plans.

Compartmentalization has been analyzed using different reservoir datasets. Pressure responses (drawdown and buildup) were used to model compartmentalized systems using material balance (Stewart et al., 1989). Others used 4D seismic datasets to find and evaluate the flow barriers in the reservoir (Almaskeri et al., 2005). In this work, the analysis of reservoir connectivity was done on an inverted 4D seismic data generated from reservoir simulation. Inverted 4D seismic data include pressure and saturation changes across the 4D time period. There are many techniques and correlations to estimate the pressure and saturation changes from 4D seismic data. It can be approximated by transforming 4D seismic properties in a cross-plot domain to pressure-saturation domain (Lumley et al., 2003). Rock physics forward modeling is also used in the inversion of 4D seismic data (Cole et al., 2002).

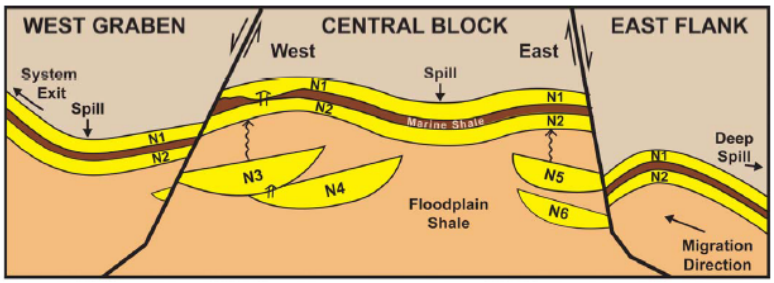


Figure 1.7: Compartmentalization. Hypothetical field compartments. (Snedden et al., 2007)

1.3 Research Objectives

The main purpose of this research was to investigate the capabilities of TDA in petroleum engineering. The objective was to study the application of TDA to an RE problem. Inverted 4D time-lapse seismic data were used to study reservoir connectivity and compartmentalization. Through this problem, the shapes of RE data was explored and the data's relationship to meaningful information is explained.

2. METHODOLOGY

2.1 Introduction

This chapter describes the methodology used in this work to apply topological data analysis in an RE problem. It explains the process used to generate and analyze inverted 4D time-lapse seismic data for studying reservoir compartmentalization. The first section covers data preparation for TDA, which includes input and output data generation. Then, the steps used in the application of TDA on this dataset are presented.

2.2 Data Generation and Processing

This section details the preparation of an inverted 4D seismic dataset for TDA. To generate this dataset, reservoir simulation was selected, which provides easier control and a validation for the results. First, a description of the compartmentalization problem input data is provided. Then, relative outputs properties that can be used to validate the results are discussed. Finally, the steps to ensure that the data are compatible with TDA are discussed.

2.2.1 Input Properties: Inverted 4D Seismic

Inverted 4D time-lapse seismic data generated by simulation were used as an input for TDA. These data include differences in pore pressures and water saturation in addition to the location coordinates of these measurements. For each simulation case, four (4) years of 4D time lapse seismic data were simulated which was enough time for wells to create variations in pore pressure and water saturation over the reservoir. A description on each input property from the simulation model is presented (Table 2.1).

Property	Description	unit
X-Coordinate	The absolute distance value in x-axis to the origin point of the simulation model	meter
Y-Coordinate	The absolute distance value in y-axis to the origin point of the simulation model	meter
Z-Coordinate	The absolute distance value in z-axis to the origin point of the simulation model	meter
Pressure Difference	Change of pore pressure in 4 years	bar
Water Saturation Difference	Change of water saturation in 4 years	%

Table 2.1: Input Properties Description. Each data point generated from the simulation model has these five input properties.

2.2.2 Output Properties

In addition to the input properties, output properties from simulation were generated for the purpose of validating the results. Output properties were used to relate the shape of the data with the actual behavior of the reservoir. Regarding the main problems in this work which is reservoir connectivity and compartmentalization, a value for each point that represent the actual compartment number to which this data point belongs. Output properties are divided into three groups: grid cells definition, reservoir properties, and well output properties. In the following tables, a description of the output properties that were generated in the dataset are presented.

First, grid cells properties are static measurements of the simulation model grid cells. They provide a way of relating the TDA shape of the input data to the simulation model (Table 2.2).

Second, output reservoir properties include static and dynamic values of reservoir

Property	Description	unit
i	Cell index in I-direction in the corner-point grid.	
j	Cell index in J-direction in the corner-point grid.	
k	Cell index in K-direction in the corner-point grid.	
Cell X-dimension	Length of the cell.	meter
Cell Y-dimension	Width of the cell.	meter
Cell Z-dimension	Depth of the cell.	meter

Table 2.2: Grid Cells Output Properties Description. Output properties are generated for each data point.

simulation model (Table 2.3). They provide the input data TDA shape relationship to the actual reservoir properties. For the analysis of the inverted 4D seismic data, the compartmentalization number was used to validate the resulted connectivity map of the reservoir.

Third, well output properties are defined for the cells that are penetrated by wells. It relates the performance of the wells to the TDA shape of the input data (Table 2.4).

2.2.3 Data Processing

Using the input and output properties dataset described in the last section, a couple of steps must be performed before applying TDA. First, all data properties must be migrated into a tabular format. In this table, each row represents a point of interest and each column represents a property of that point. For the inverted 4D seismic dataset, each row is a point in the simulation model and columns are values associated with that point like location or compartment number. Second, another version of the input data with a unified scale is created in which all input variables have the same weight in building the TDA shapes. Finally, properties are reviewed

Property	Description	unit
Porosity	Value of porosity at the cell of interest.	%
Permeability	Value of permeability at the cell of interest in horizontal and vertical directions.	mD
Pressure	Value of the pressure for a specified date.	bar
Oil Saturation	Value of the oil saturation for a specified date.	%
Gas Saturation	Value of the gas saturation for a specified date.	%
Water Saturation	Value of the water saturation for a specified date.	%
Fluid in Place	The amount of oil, gas or water in place.	cubic meter
Compartment Number	Compartment number for the cell of interest.	
Faults	Value define faults existence and its multiplier value.	

Table 2.3: Reservoir Output Properties Description. Output properties are generated for each data point.

Property	Description	unit
Well Name	The name of the well that penetrates the selected cell.	
Well Type	The type of the well that penetrates the selected cell.	
Oil Production Rate	Maximum, minimum, average and total oil production rates.	cubic meter
Gas Production Rate	Maximum, minimum, average and total gas production rates.	cubic meter
Water Production Rate	Maximum, minimum, average and total water production rates.	cubic meter
Gas Injection Rate	Maximum, minimum, average and total gas Injection rates.	cubic meter
Water Injection Rate	Maximum, minimum, average and total water Injection rates.	cubic meter

Table 2.4: Wells Output Properties Description. Wells output properties would be defined for cells that are penetrated by a well.

and edited based on the study objective. The main focus was to avoid null values and create discrete versions of output properties, which allows for a clear validation process.

2.3 Topological Data Analysis

The objective of this section is to give an overview on the application workflow of topological data analysis used in this work. The process of building shape graphs and the way to analyze them is introduced. Then, These steps are then summarized through the compartmentalization problem.

2.3.1 Metrics and Functions Analysis

The objective of this step is to evaluate different metrics and lenses performance considering compartmentalization. The focal point of this research was finding combinations of input variables, metrics, and lenses that exposes the targeted outcome.

First, the similarity metric between data points of the input data needed to be defined. Similarity criteria were defined by finding the appropriate distance function and applying it to the input data. Choosing which input features to include in defining similarity is a big open problem and studying which ones would work well with the selected distance function is important.

Then, a function or a set of functions that uses the calculated matrix of distances between data points needed to be defined. It transforms the data into two dimensional graphs. By testing a variety of statistical and data functions, the effect of each one was investigated for the outcome of the study. The resulted graphs should satisfy stability where clusters of data points travel together across different graphs.

Finally, outstanding clusters of data points in generated graphs such as flairs,

cycles, or separation represent a behavior of the data that was focused on in the study. By analyzing and comparing the values of these groups of data points, any deviation of the data can be identified and these interesting shapes can be related to the actual behavior.

2.3.2 Output Overlap and Validation

With shape graphs established for the input problem, the analysis could be confirmed if the outcome variable was one of the distinctive features of outstanding clusters. In addition, the outcome variables could be used as functions to investigate the properties that are correlated with the outcome variable. With the gained knowledge, the analysis was continued and additional shape graphs were generated to expose more of the input data behavior.

2.3.3 Compartmentalization TDA Algorithm

For the compartmentalization problem, two simulation models were used that were compartmentalized by faults and geological features. Reservoir connectivity for both models were studied by applying the TDA process (Algorithm 1).

Algorithm 1 Topological Data Analysis

Input: Inverted 4D time-lapse seismic data (X, Y, and Z Coordinates, Pressure and Saturation Difference).

1: Migrate all data into a tabular format and create a scaled version of the input data.
2: Study all metrics and data points input variables for best similarity criteria.
3: Explore data and statistical functions that expose compartmentalization.
4: Apply statistical comparisons for outstanding clusters of data points to prove reservoir compartmentalization.
5: Validate the results and overlap with the actual reservoir compartments.

Output: TDA described reservoir compartmentalization and its results are validated by simulation model.

3. APPLICATION

3.1 Introduction

The objective of this chapter is to present the application of TDA on the selected datasets and related results. Simulation models were used to generate the input datasets in this work. Model data pre- and post-processing were done using PETREL(Schlumberger, 2013). In addition, different cases studied in this research were simulated using ECLIPSE (Schlumberger, 2010). Two simulation models have been used to test the capabilities of TDA: Brillig and Norne. Brillig is one of the ECLIPSE original models used to provide cases to study compartmentalization. Norne is a real simulation model that was used to prove the findings in this work from the synthetic model. A description of each model and TDA application process are provided in the following sections.

3.2 Models and Data Description

3.2.1 Brillig Simulation Model

This section provides an overview of the Brillig model. First, the simulation model is described with respect to the compartmentalization problem. The generation and processing of the inverted 4D seismic data and problems related to the output data from the simulation model is also discussed.

3.2.1.1 General Description

Brillig is a synthetic model provided in the ECLIPSE simulator test data files. It is used to test many options, such as wells' group control and tracer tracking of the original gas cap-gas, and injected water. The model has an anticline structure that has oil trapped between a gas cap on the top and water aquifer in the bottom. This

3D model has a total of 2400 cells, which consist of 20 cells in I-direction, 15 cells in J-direction, and 8 layers in K-direction (Figure 3.1). The average porosity of the model is 25% and the average horizontal permeability of the model is 572 md.

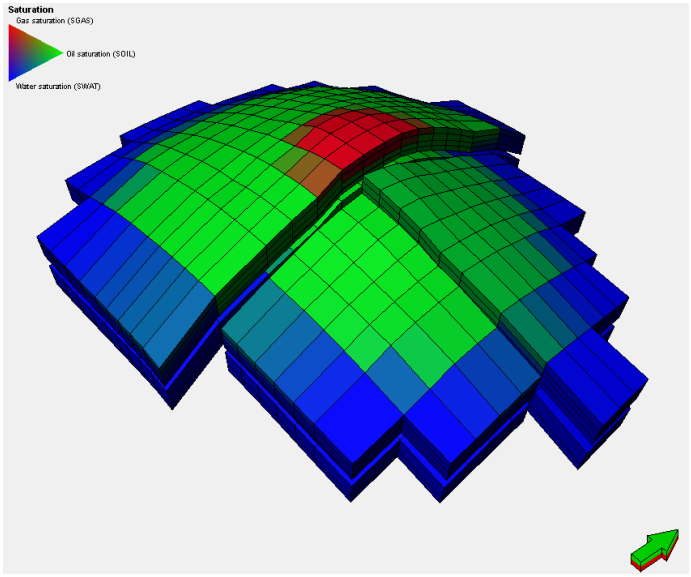


Figure 3.1: Brillig 3D Model. The model is colored by fluid saturation. (water=blue, oil=green, and gas=red)

Brillig Model is using a development strategy of 10 years between January 1990 and December 1999. It has 19 wells including 10 oil producers, 9 water injectors and 1 gas injector. Water injection wells are peripheral and they are set to a voidage replacement that is equal to the total production of reservoir. For the gas injector, it is re-injecting the produced gas volume back to the reservoir in the gas-cap (Figure 3.2). The production from this field is set to a group rate of 150000 rb/day with restrictions to water and gas production.

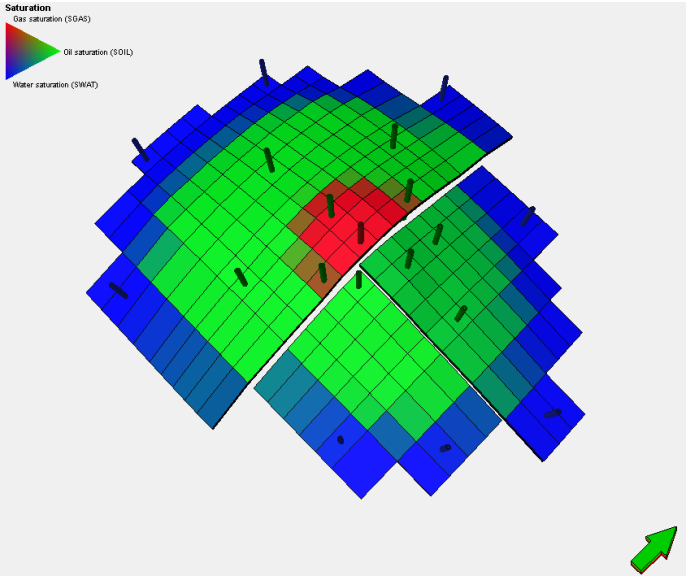


Figure 3.2: Brillig Wells. Brillig has 19 wells including 10 oil producers, 9 peripheral water injectors, and one gas injecotr in the gas-cap.

In terms of reservoir compartmentalization, the model has three regions that are created by two faults. The first fault crosses the whole model in the middle where cell index I is between twelve 12 and 13. The second fault split one of the halves generated by the first fault in the I-direction where cell index J is between 9 and 10 (Figure 3.3). The half which have not been split represent region 1. The split regions created by the second fault are 2 and 3 (Figure 3.4). The two faults have caused three regions of different depths. The top layer of region 3 is in contact with the bottom layer of region 1. The depth of region 2 depth is between the depth of the other two regions.

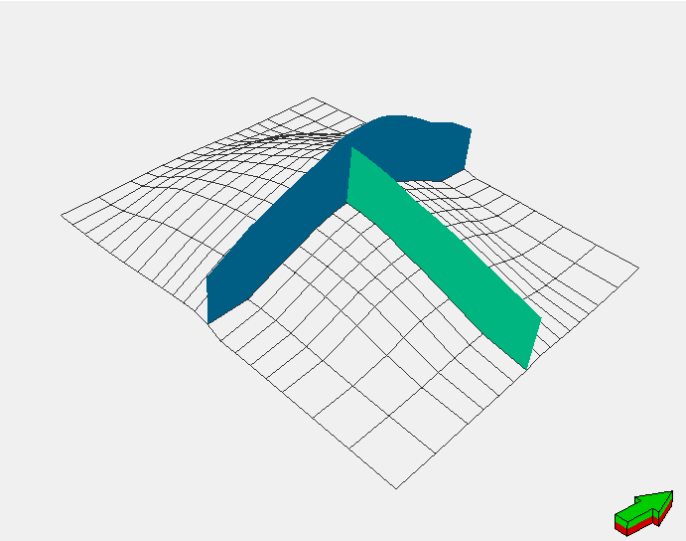


Figure 3.3: Brillig Faults. Brillig has two faults which divide the model into three regions.

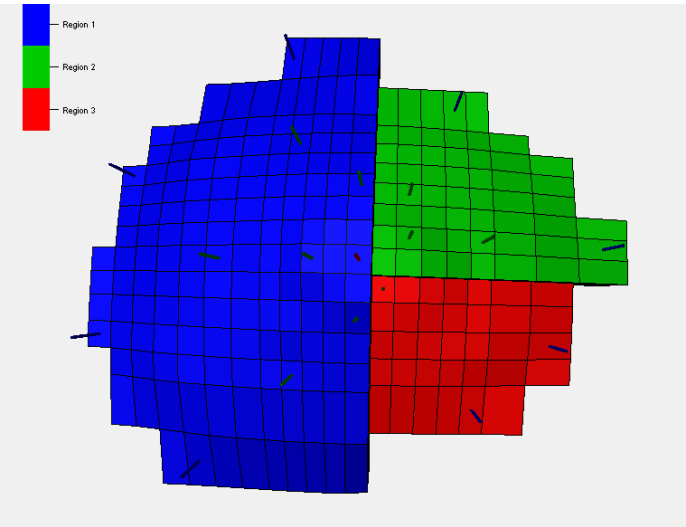


Figure 3.4: Brillig Regions. Brillig is divided by the faults into three (3) regions.

Two tests were applied using this model. The first one was to run the

simulation model with impermeable faults, which creates three compartments of different pressure behavior. The second test have transmissible faults, which allow for pressure and saturation to be similar across the faults in the model. All other properties are exactly the same in the simulation cases. The two tests show the relationship between the connectivity in the reservoir and the shape graphs that are created by TDA.

3.2.1.2 Data Generation

Using the model cases described in the last section, a dataset was generated that includes both input and output properties. The input data for the compartmentalization problem in this work were inverted 4D seismic data. It was a dataset of five dimensions: 3 location coordinates and 2 properties for pressure and saturation difference. For the first 3 components of the inverted 4D seismic, we are using cells X, Y, and Z global coordinates representing the location of each data point. The other two components are calculated based on a 4 years difference of pressure and saturation values.

The initial pressure and water saturation used to create the input data are the same in the permeable and impermeable faults cases. In 1990, the pressure distribution over the reservoir did not have any sharp changes of values between adjacent cells (Figure 3.5). Also, water saturation values did not have any sudden changes between connected cells (Figure 3.6).

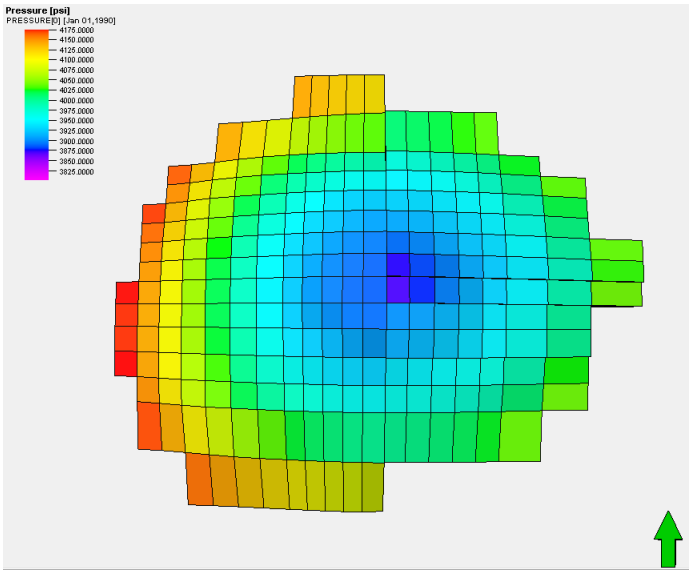


Figure 3.5: Brillig Pressure in 1990. Initial pressure did not have any sharp changes of values between adjacent cells (Regions Communication Layer).

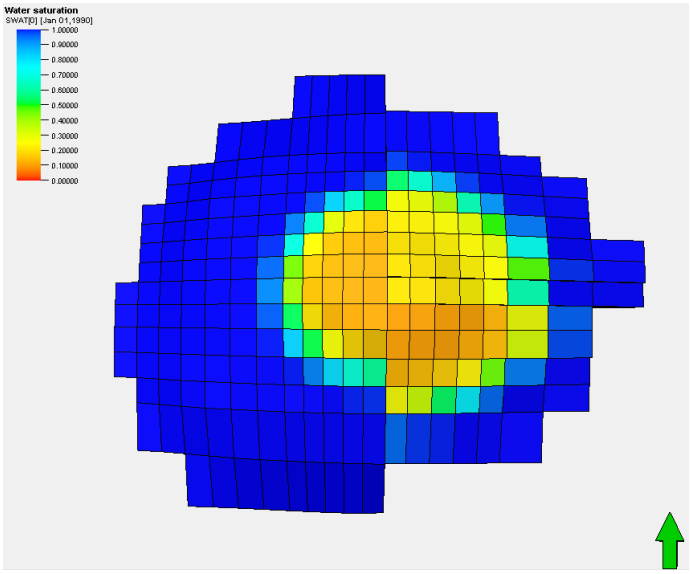


Figure 3.6: Brillig Water Saturation in 1990. Initial water saturation did not have sharp changes between adjacent cells (Regions Communication Layer).

The first simulation case with impermeable faults, simulated pressure, and saturation distribution in 1994 across the reservoir was different than the initial distribution. In 1994, the pressure value showed the effect of impermeable faults in restricting fluid movement and balancing the pressure across the reservoir (Figure 3.7). Although water saturation changes are not as important in identifying barriers, reasonable changes are visible between connected nearby cells and they are also considered in the input dataset (Figure 3.8).

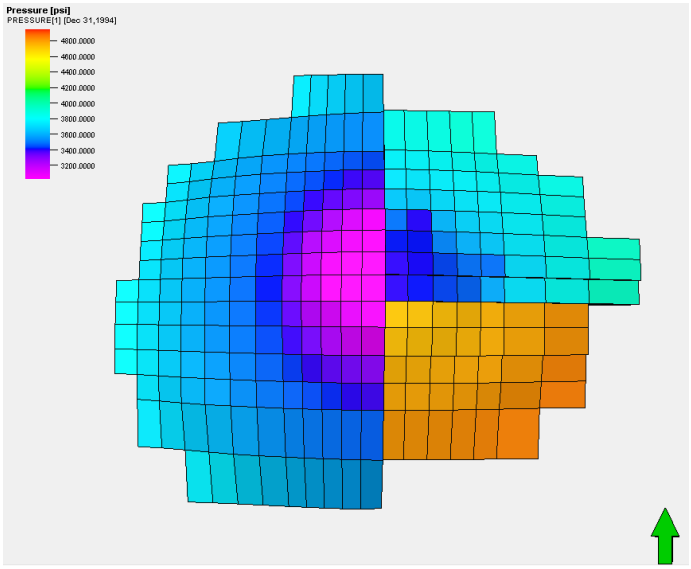


Figure 3.7: Brillig Pressure in 1994 (Impermeable Faults Case). Pressure values in 1994 shows a clear sharp change between adjacent cells along the fault line (Regions Communication Layer).

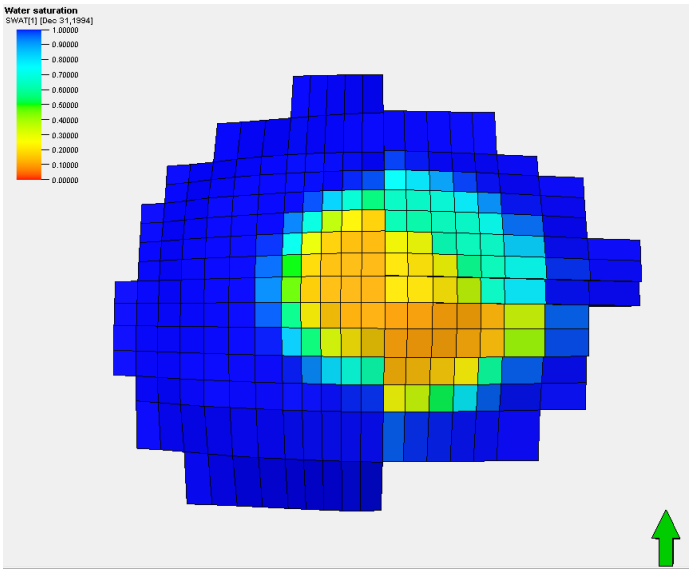


Figure 3.8: Brillig Water Saturation in 1994 (Impermeable Faults Case). Reasonable water saturation changes between adjacent cells in the fault line have developed. However, water saturation have areas in which all regions have the same value, like aquifers.

The input dataset generated from this case include the deference between 1994 and 1990 pressure and water saturation values (Figures 3.9 and 3.10). Compartments that are separated by the fault barrier are defined through the unreasonable sharp change of pressure in their boundaries. The boundary between regions 1, 2 and 3 is clearly shown in the pressure deference values in the communication layer. Therefore, it makes an excellent test case for TDA to study reservoir connectivity.

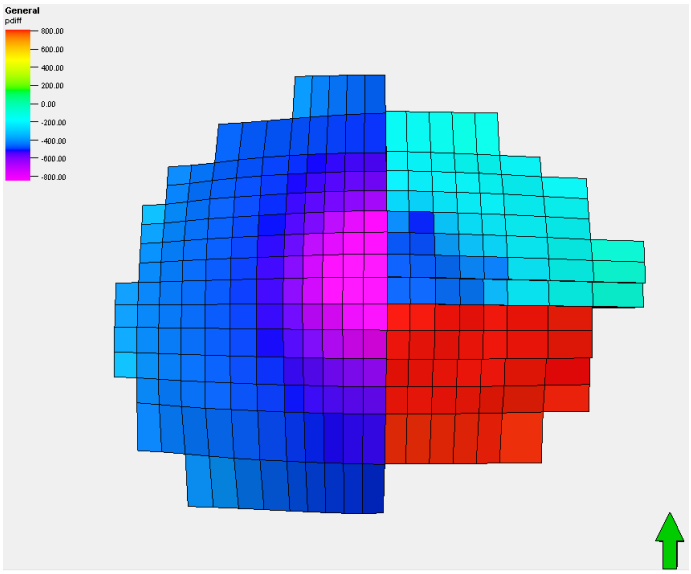


Figure 3.9: Brillig Pressure Difference between 1994 and 1990 (Impermeable Faults Case). Pressure difference defines the regions' boundaries clearly with clear sharp changes between adjacent cells along the fault line.

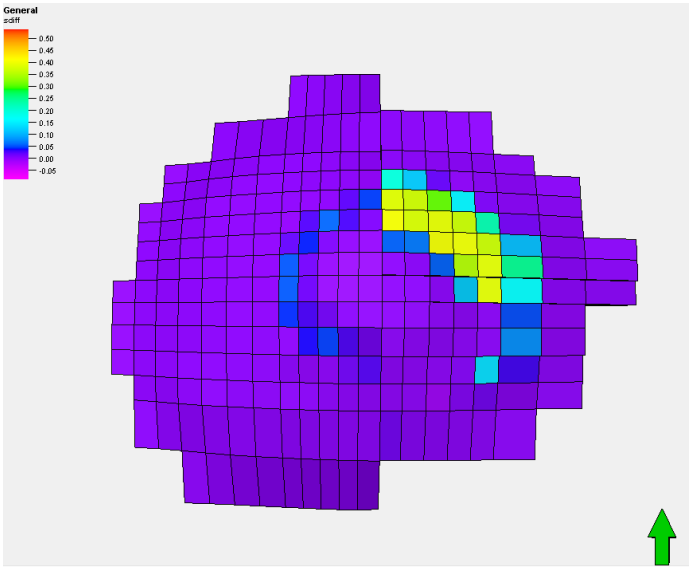


Figure 3.10: Brillig Water Saturation Difference Between 1994 and 1990 (Impermeable Faults Case). Different water saturation changes in each region have developed. However, water saturation difference is not a clear distinguisher between regions in this case. It would introduce a challenge for TDA to overcome.

The second simulation case with transmissible faults, simulated pressure, and water saturation do not show any discontinuity in their values across the reservoir in 1994 (Figures 3.11 and 3.12). The changes in pressure values are smooth across the fault area in the contact layers. It shows that there is a clear communication of fluids between these regions in this case (Figure 3.13). Water saturation changes were higher between water injectors and oil producers, which is irrelevant to our compartmentalization problem and presents a challenge for TDA (Figure 3.14).

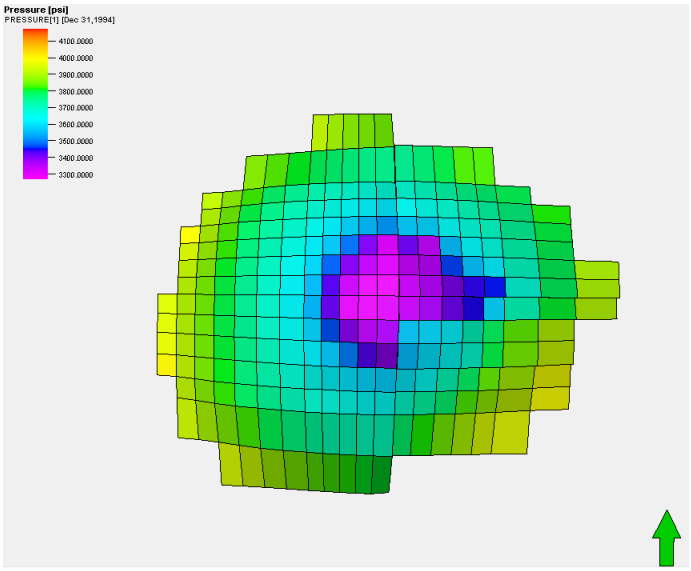


Figure 3.11: Brillig Pressure in 1994 (Permeable Faults Case). Pressure values in 1994 shows smooth changes between adjacent cells along the fault line.

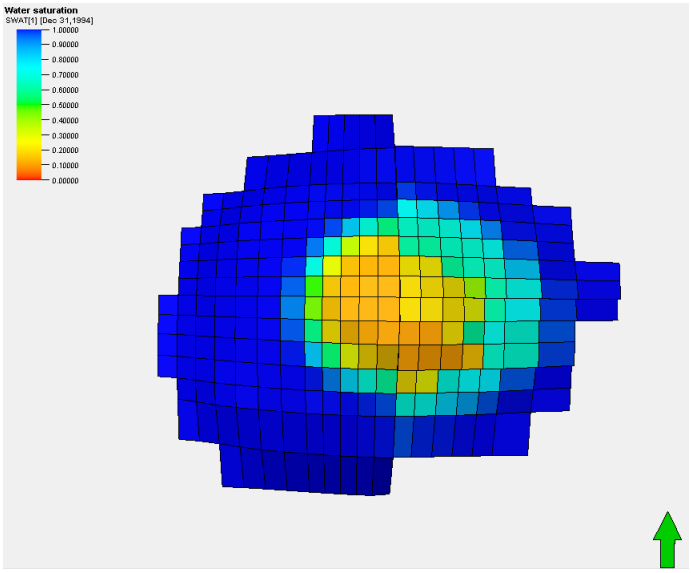


Figure 3.12: Brillig Water Saturation in 1994 (Permeable Faults Case). Water saturation changes between adjacent cells along the fault line are also smooth. However, the effect of producers and injectors on water saturation can disturb the analysis.

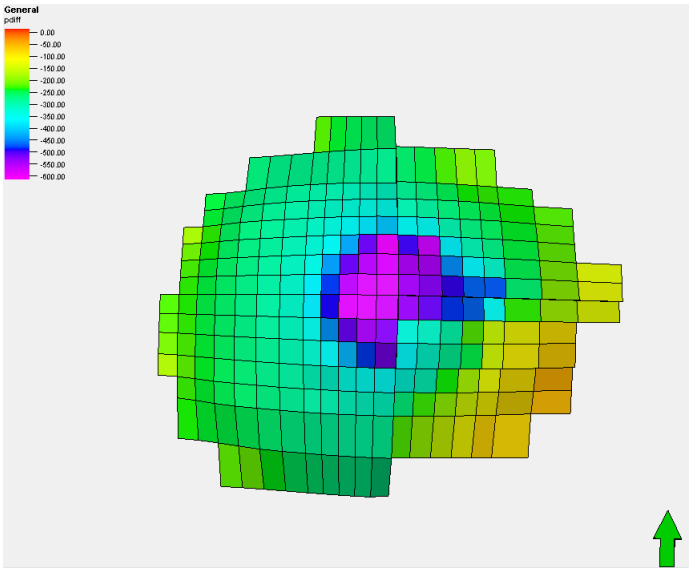


Figure 3.13: Brillig Pressure Difference between 1994 and 1990 (Permeable Faults Case). Pressure difference shows a clear communication between adjacent cells along the fault line.

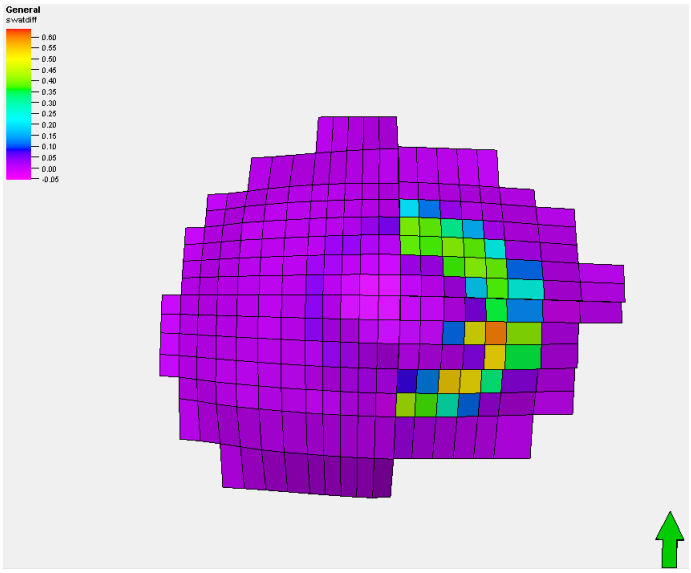


Figure 3.14: Brillig Water Saturation Difference Between 1994 and 1990 (Permeable Faults Case). Changes of water saturation in each region have developed. The effect of producers and injectors on water saturation is shown clearly on regions 2 and 3 which is irrelevant to the compartmentalization problem.

In addition to the input dataset represented by inverted 4D seismic, output data were also included in the dataset for both cases. The main output variable included in this study was the compartment number for each data point. Each cell in the simulation model was tagged with the compartment to which it belonged and added to the input dataset (Figure 3.4). Other output data such as the indices of these cells and the real simulation properties were also included.

The result of the data generation process for the two simulation cases are a table containing both input and output data. Each row is a data point representing a cell in the simulation model. The columns are the variables associated with this data point such as input and the output properties. Another version of the input properties is normalized to balance the effect of the pressure and saturation changes and location coordinates. Using this dataset, TDA was used on inverted 4D seismic

to study reservoir connectivity and its results were validated using output data.

3.2.2 Norne Simulation Model

This section provide an overview on the Norne model and data generation. The Norne simulation model was described considering reservoir connectivity and compartmentalization. The inverted 4D seismic dataset generated by this model and the expected output after TDA analysis was also discussed.

3.2.2.1 General Description

Norne is an offshore oil field located in the Norwegian Sea (Norwegian University of Science and Technology, 2012). It was discovered in 1991 and production started in 1997. The wells were drilled from two floating vessels. Time lapse seismic surveys were performed multiple times between 2001 and 2006. The Norne simulation model was used in this work to test and confirm the TDA results. It provides a natural case to generate an inverted 4D time-lapse data and use it to study the compartmentalization of the reservoir.

The simulation model has 113344 cells (I= 46, J=112, K=22). Oil exists between a gas cap and a water aquifer. The simulation model development strategy contains 35 wells including twenty 27 producers and 8 injectors (Figure 3.15).

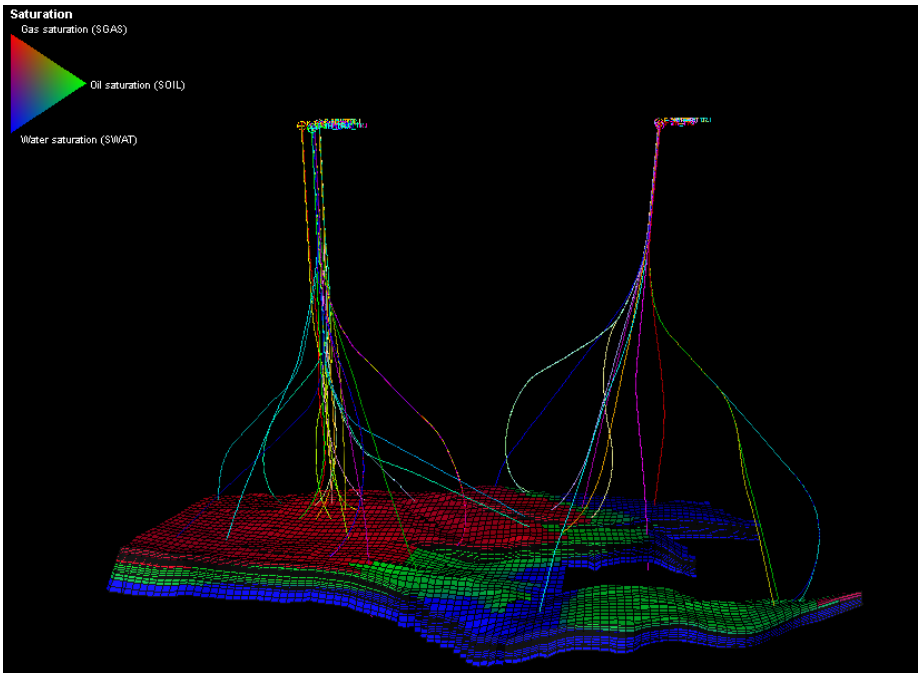


Figure 3.15: Norne 3D Model. The model is colored by fluid saturation. (water=blue, oil=green, and gas=red). Wells are also shown starting from two locations above the sea level.

The Norne simulation model has two main oil compartments. compartment 1 contains segments C, D, and E in (figure 3.16). Compartment 2 is segment G. The difference in the oil water contacts (OWC) and gas oil contacts (GOC) confirms these two compartments (Table 3.1). It is also important to note the non-reservoir clay stone layer that prevents communications between Garn Formation and Ile formation on most areas in the reservoir (Figure 3.17).

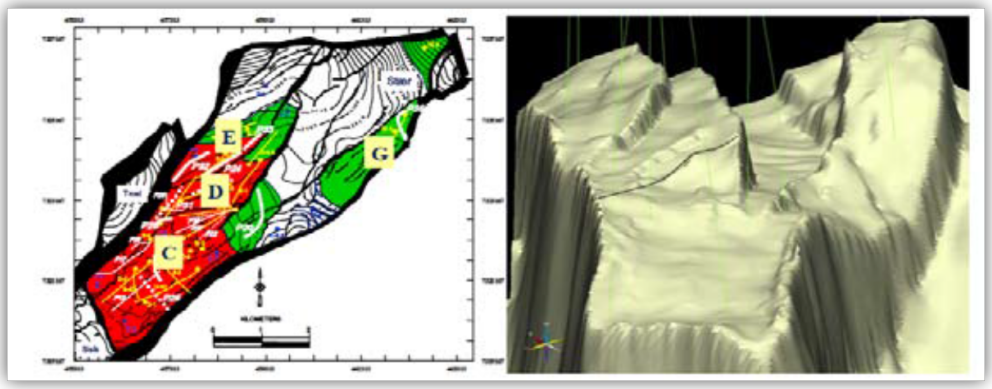


Figure 3.16: Norne Structure and Compartments. Norne has two main oil compartments. Compartment 1 contains C, D and E. Compartment 2 is segment G (Norwegian University of Science and Technology, 2010).

Formation	CSegment		DSegment		ESegment		G-Segment	
	OWC	GOC	OWC	GOC	OWC	GOC	OWC	GOC
Garn	2692	2582	2692	2582	2618	2582	2585	No gas cap
Ile	2693	2585	2693	2585	2693	2585	Water filled	Water filled
Tofte	2693	2585	2693	2585	2693	2585	Water filled	Water filled
Tilje	2693	2585	2693	2585	2693	2585	Water filled	Water filled

Table 3.1: Norne Regions and Formations OWC and GOC. Only G-segment has a different OWC than all other segments (Norwegian University of Science and Technology, 2010).

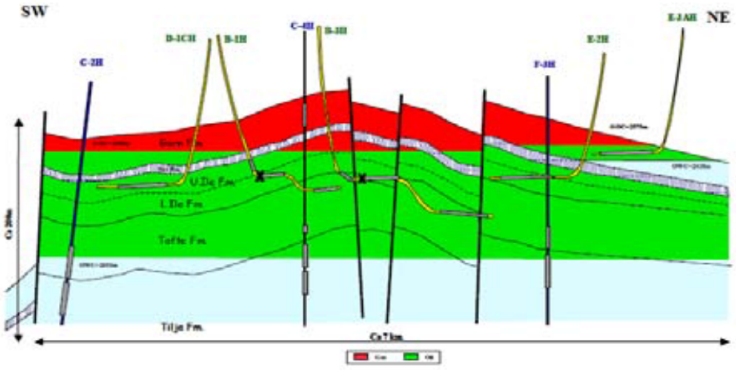
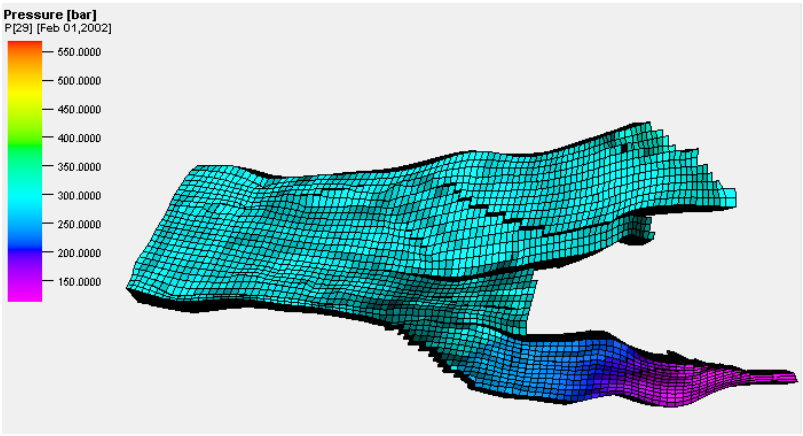


Figure 3.17: Norne Formations Cross-section (SW-NE). Norne has two main oil compartments mentioned previously. In addition, note the clay stone layer that serves as a barrier between the top layer (Garn Formation) and (Ile formation) almost all over the field (Norwegian University of Science and Technology, 2010).

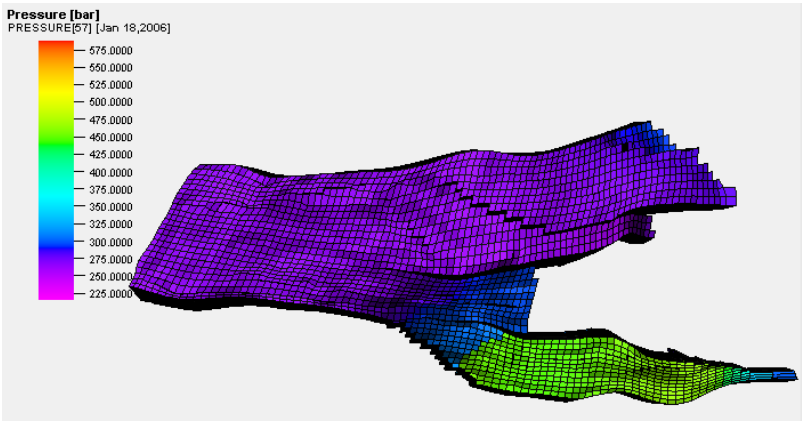
3.2.2.2 Data Generation

Using the Norne simulation model described in the last section, an inverted 4D seismic dataset was created along with model compartmentalization property. Data specifications are similar to the ones described for Brillig case. The inverted time-lapse seismic data were generated between 2002 and 2006.

First, pressure data in 2002 does not have sharp changes between cells across the reservoir. However, pressure values in 2006 have drastic changes between connected cells across different regions (Figure 3.18). Both were used to generate the pressure difference and to create the first component in the inverted 4D seismic dataset. Water saturation data were also used to generate another component of the inverted 4D seismic dataset (Figure 3.19).

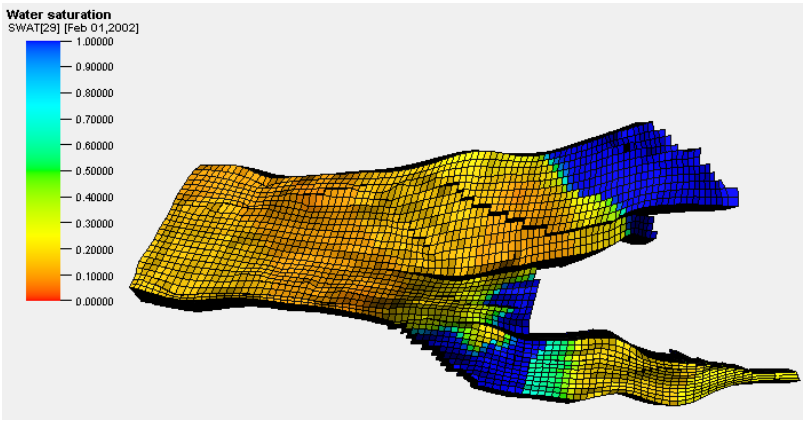


(a)

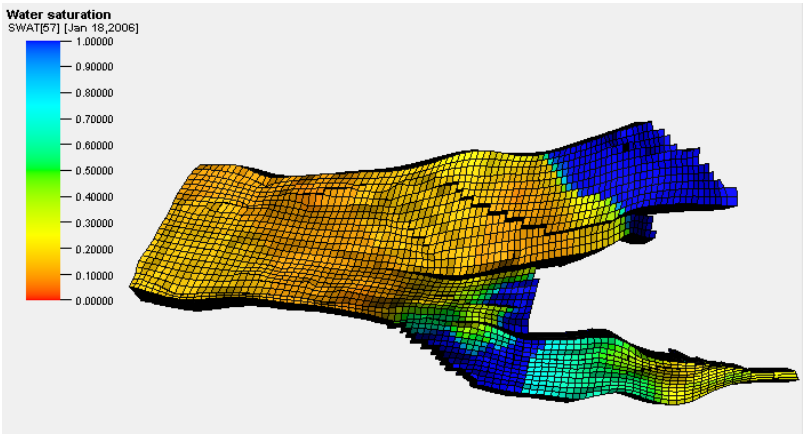


(b)

Figure 3.18: Norne Pressure in 2002 and 2006. Initial pressure (a) had smooth changes of values between adjacent cells across different regions. Pressure values in 2006 (b) shows a clear sharp change between adjacent cells between regions.



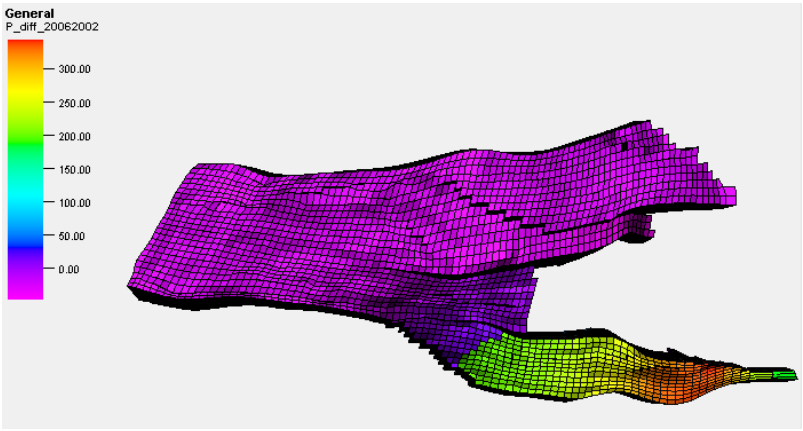
(a)



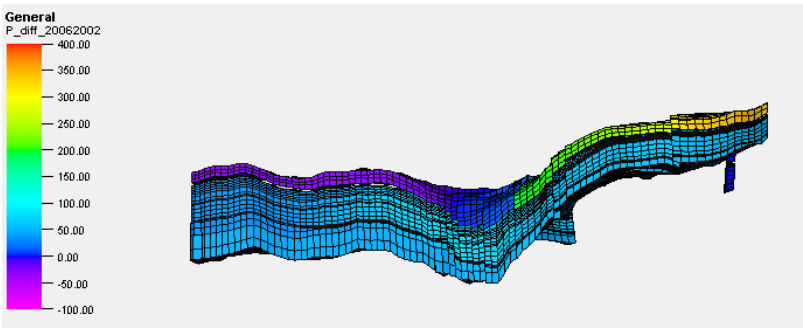
(b)

Figure 3.19: Norne Water Saturation in 2002 and 2006. Initial water saturation is shown in figure (a). Water saturation values in 2006 (b) shows a smoother transition to the aquifer cells.

Pressure difference between 2006 and 2002 shows a clear separation between regions 1 and 2 in the top view in (Figure 3.20). Also, region 3 has a distinct pressure difference that is apparent in the side view. However, small areas in the middle has communication between regions 1 and 3, which shows a gradual transition between the two regions.



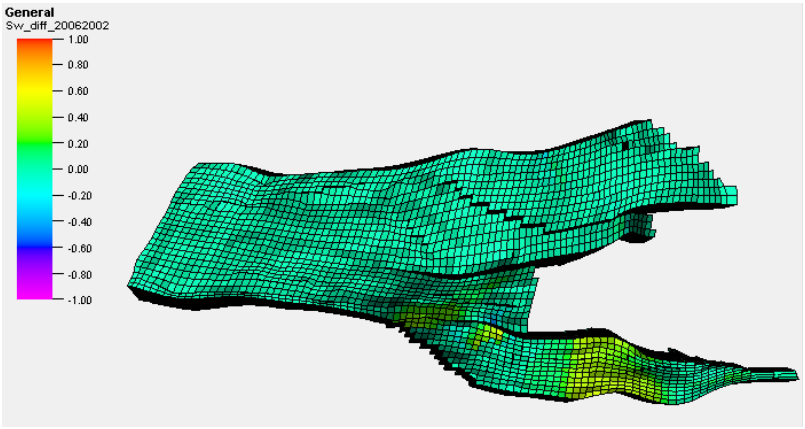
(a)



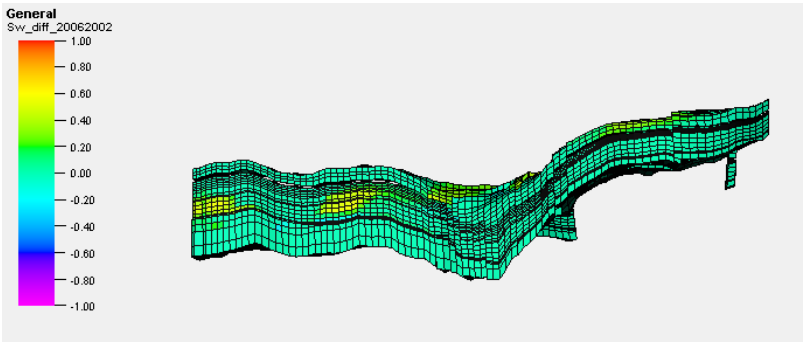
(b)

Figure 3.20: Norne Pressure Difference between 2002 and 2006. Pressure difference (a) defines the regions' boundaries clearly due to the sharp changes between adjacent cells along the lines between different regions. The side view (b) show pressure difference transitions between different regions. The sharp changes between the three defined regions can be seen. Also, a small contact area between regions 1 and 3 where the transition is smoother.

Water saturation difference is more affected by other variables (Figure 3.21). For example, well production and injection can create higher changes in water saturation than the ones created by barriers. It can disturb this compartmentalization study with data that are not relative to the objective.



(a)

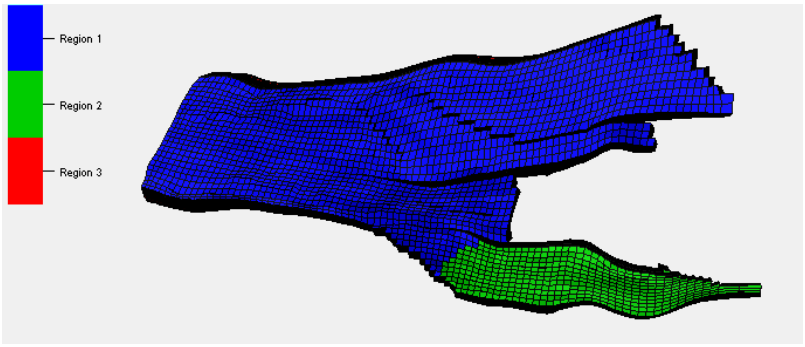


(b)

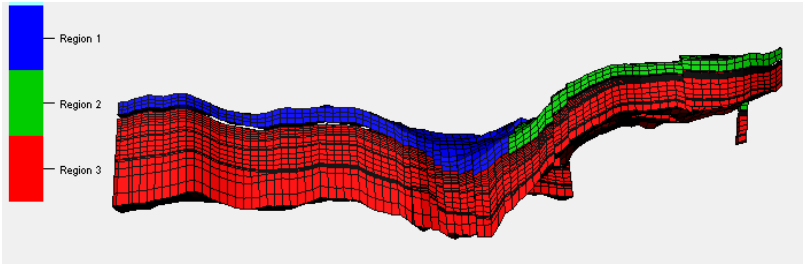
Figure 3.21: Norne Water Saturation Difference between 2002 and 2006. Water saturation difference (a) shows a minimal change in water saturation values. The side view (b) also shows minimal changes that are attributed to wells. It seems that water saturation changes in this case have lower effect on compartmentalization.

Norne has two oil compartments in addition to a sealing layer that prevents communication in most areas of the reservoir. We have defined three regions accordingly. Region 1 covers most of the top formation (Garn) sealed by the clay stone layer. Region 2 is the region that is defined by G-segment. Region 3 is the rest of the reservoir, which is the lower three formations. It can be also noted that the small communication between regions 1 and 3 which affects the solution to the

compartmentalization problem (Figure 3.22).



(a)



(b)

Figure 3.22: Norne Compartments. Norne top view (a) shows region 1 and region 2. The side view (b) shows the two regions along with region 3 which covers the lower layers of the reservoir.

3.3 Topological Data Analysis

The application process of topological data analysis can be divided into two main parts: graph creation and its analysis. Using the simulation model described in the previous section, The steps that were taken to create the graph that represent the shape graph of the data are explained. The analysis of these graphs are presented and the results are validated.

Topological data analysis figures were generated by performing TDA with the Ayasdi software platform (ayasdi.com). Nodes in the graph represent clusters of grid cells and edges connect nodes that contain samples in common.

3.3.1 Graph Creation

Building a graph representing the shape of the data requires two main choices: a similarity metric and lenses. A similarity metric defines the distance between all the data points in our dataset. Then, a lens is a projection function that is applied to transform the multidimensional dataset to a 2D graph. An embedded choice in both is feature selection. It is related to selecting input properties that can be used in either calculating the similarity metric or applying lenses. The choices and steps taken to create the graph for the compartmentalization problem is discussed in the following paragraphs. This process is explained using the simulation case with impermeable faults.

The first step in building the shape graph was to define how to measure the similarity between data points. A distance function to measure similarity between data points was used. All input variables were considered when measuring similarity. Equation 3.1 calculates the distance between data point A and B in a defined vector space where a and b are vectors of all input properties of data points A and B. Using euclidean distance for the input dataset, the similarity distance function is shown in

equation 3.2.

$$d(A, B) = \|a - b\| \tag{3.1}$$

$$d(A, B) = \sqrt{(P_A - P_B)^2 + (Sw_A - Sw_B)^2 + (X_A - X_B)^2 + (Y_A - Y_B)^2 + (Z_A - Z_B)^2} \tag{3.2}$$

Many distance functions were tested in regard to the input dataset, for example, euclidean, correlation, angle, cosine. These distance functions failed in working with the input data with the original scale. This failure occered because the data contained a huge scale difference that corrupted the calculation of the distance (Figures 3.23 and 3.25). The solution was to use a normalized version of the data with a similar scale (Figure 3.24). Another solution was to utilize distance functions with embedded normalization of the data, like variance normalized euclidean (Figure 3.26). The result of this step was a distance matrix of all data points with respect to each other.

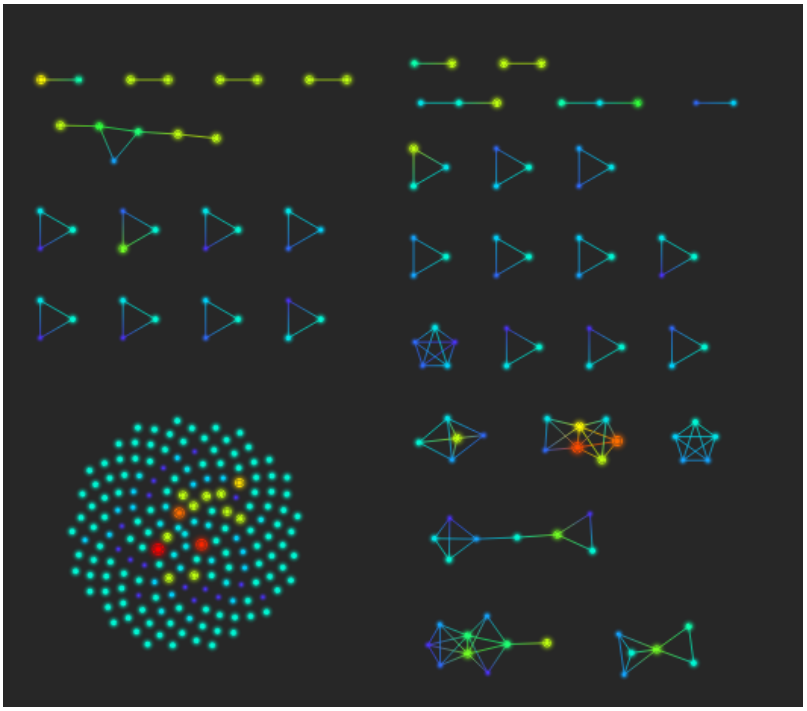


Figure 3.23: Distance Functions with Non-Scaled Dataset. Higher scale data variables are going to take over and corrupt the results of TDA.

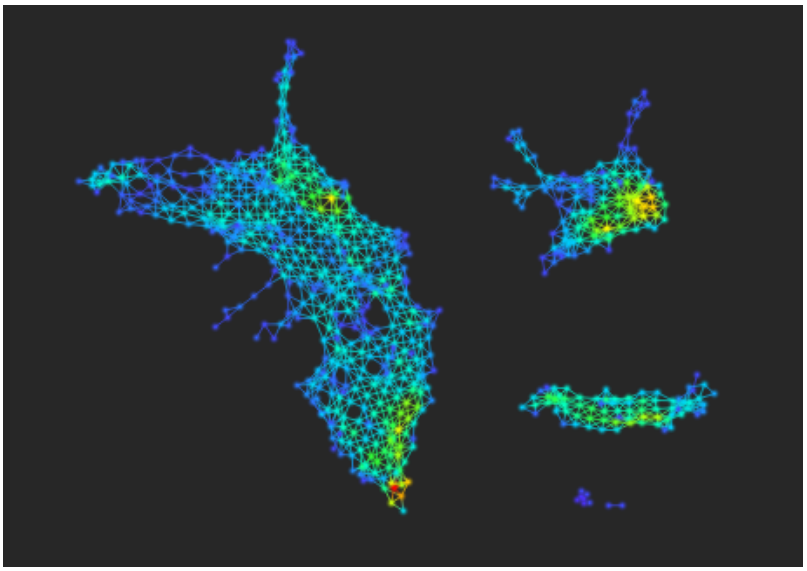


Figure 3.24: Distance Functions with Scaled Dataset. All data variables have the same scale and they have the same effect on TDA results.

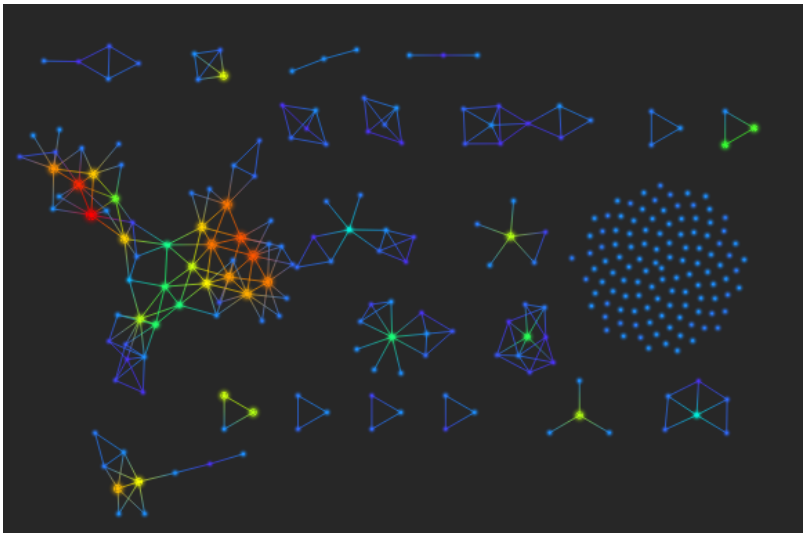


Figure 3.25: Non-Normalized Euclidean Distance Function. The use of a non-normalized distance function on data variables with different scales will generate corrupted graphs that are dictated by the highest scale variables

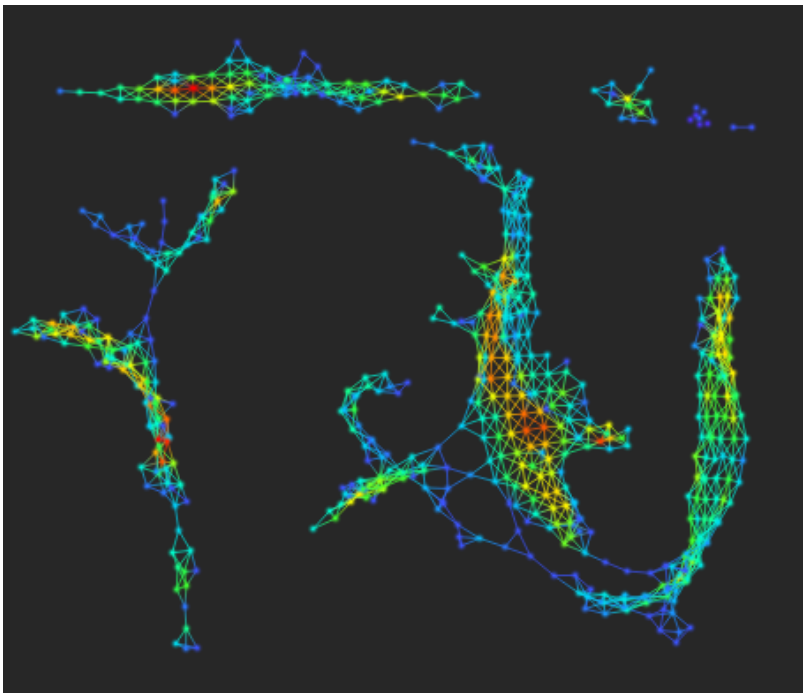
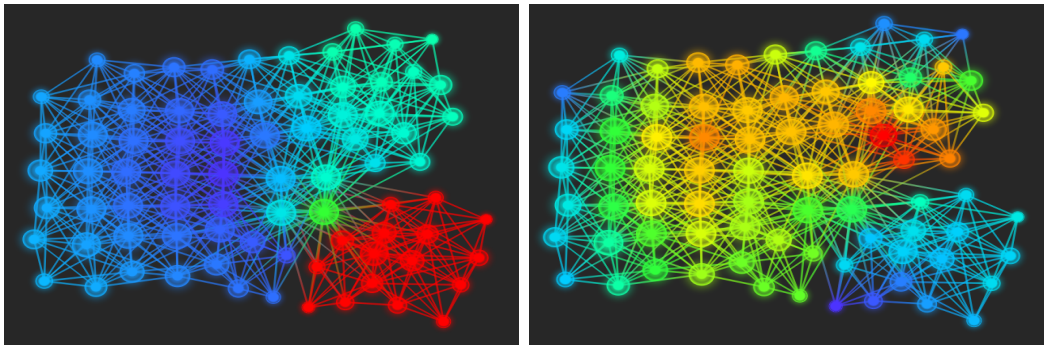


Figure 3.26: Variance Normalized Euclidean Distance Function. Using a normalized version of the distance function can eliminate the effect of variable scale difference.

Second, one or more lenses were applied to the dataset including statistical and data functions. Functions were tested with different specifications of resolution and overlap to expose data clustering with regards to compartmentalization. Successful lenses created stable clusters or graph features that traveled together among multiple graphs. Also, it recognized the effects of sharp changes in pressure and water saturation.

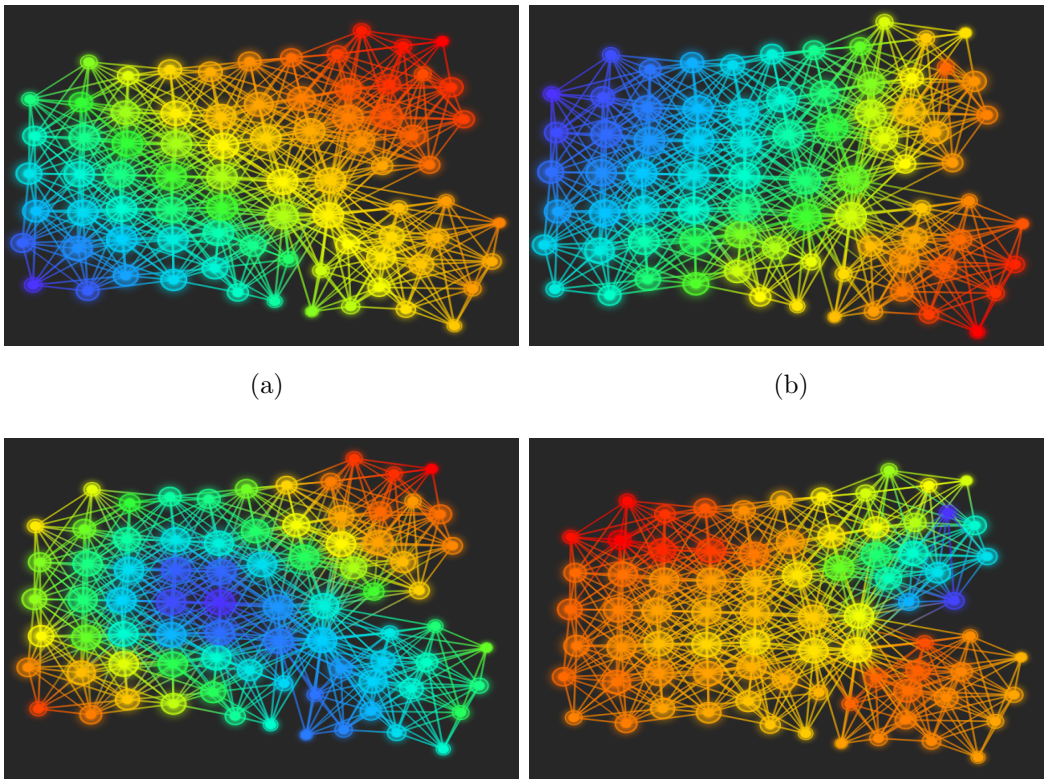
To test different statistical functions, the input dataset of Brillig model was projected on x and y coordinates. Other functions were used to color the generated graph and interesting functions were identified (Figure 3.27). Many statistical functions were eliminated because they did not expose regions of different pressure behavior using the same input dataset, for example, mean, variance, approximate

kurtosis, and entropy (Figure 3.28). Although some of the other functions looked promising at this stage, they did not succeed at the end in creating graphs, which exposed compartmentalization. For instance, median and L-infinity centrality (Figure 3.29). Projection functions that succeeded include principle component analysis (PCA) and neighborhood lenses. In addition, data functions like pressure changes were successful.



(a) (b)

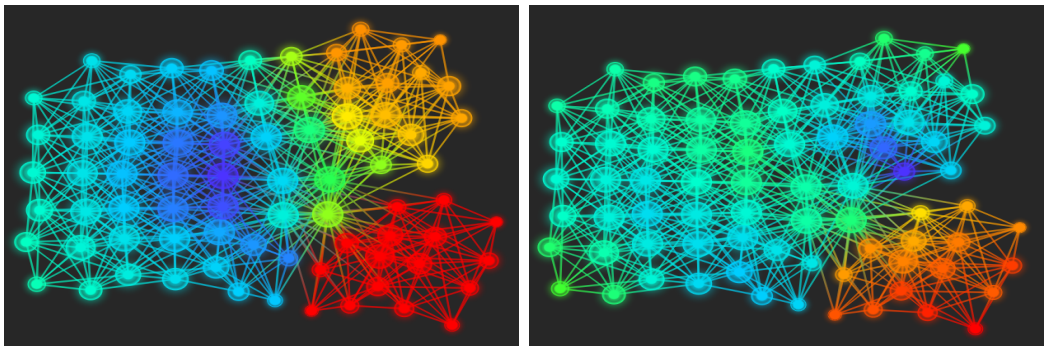
Figure 3.27: Projection of Input Data on X and Y Coordinates. The projection is colored by pressure difference in (a) and water saturation difference in (b).



(a) (b)

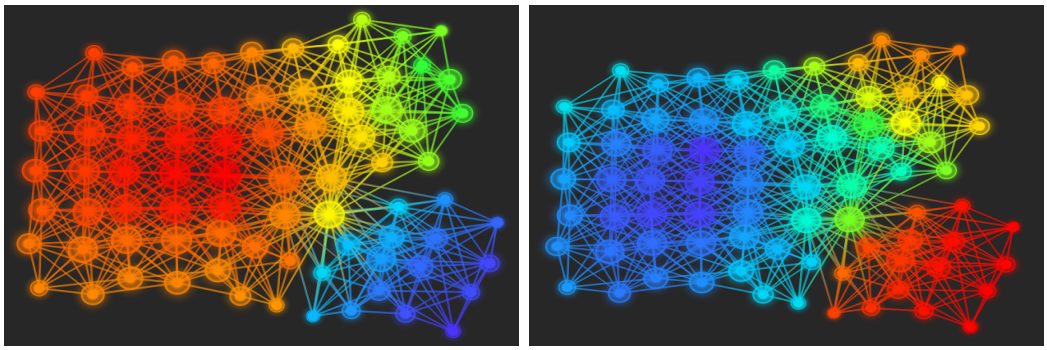
(c) (d)

Figure 3.28: Eliminated Projection Functions. Projection of input data on X and Y coordinates is colored by statistical functions, like mean (a), variance (b), approximate kurtosis (c), and entropy (d).



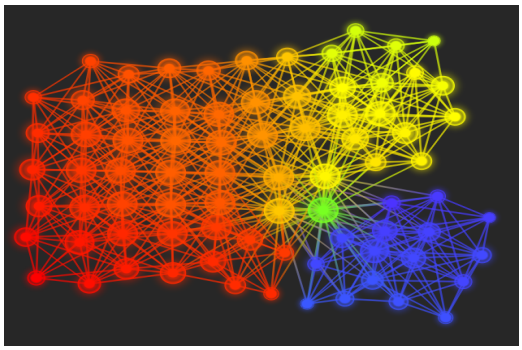
(a) (b)

Figure 3.29: Unsuccessful Projection Functions. Projection of input data on X and Y coordinates is colored by statistical functions, like median (a), and L-infinity centrality (b). Although they seem to offer good detection of compartmentalization, finding all reservoir compartments was not successful.

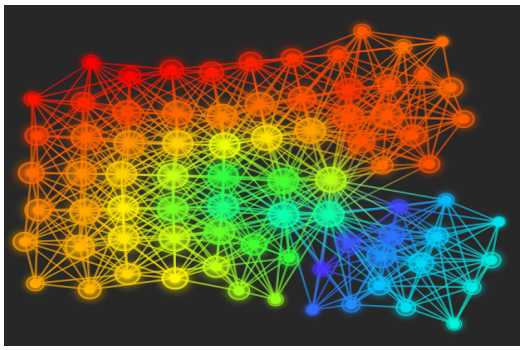


(a)

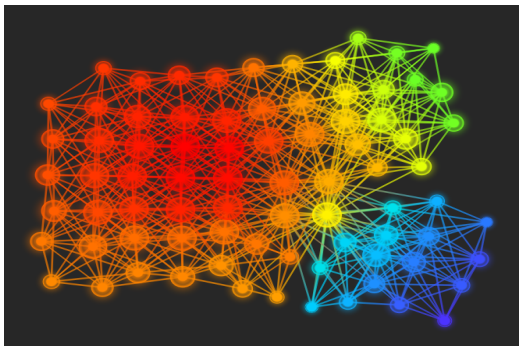
(b)



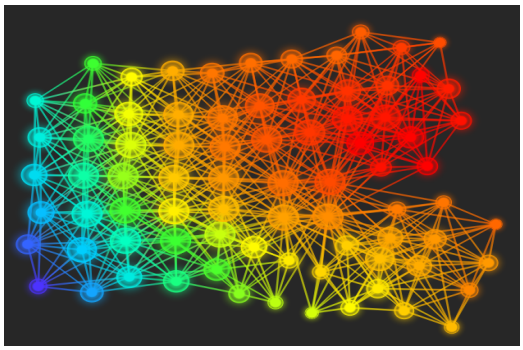
(c)



(d)



(e)



(f)

Figure 3.30: Successful Projection Functions. Projection of input data on X and Y coordinates is colored by statistical functions, like first and second principle components (a and b), topological neighborhoods lenses (c and d), and multidimensional scaling coordinates 1 (e and f).

Although similarity metric and lenses are the key choices in building the shape, selecting which input data to use in each one is also a challenge. Three combinations of input data were tested. The first combination was to use location coordinates, and pressure and saturation difference in defining the similarity between data points and projecting using statistical functions. The second combination was to use location coordinates for similarity and pressure changes for the lenses. The last one was to use pressure and saturation changes for similarity metric and project using the coordinates. Graphs were created that were helpful in studying reservoir connectivity using only the first two combinations.

3.3.2 Graph Analysis

To analyze the various graphs generated in the last section, TDA needed a few steps. First, graphs were scanned for features like clusters or flairs. Then, an explanation was required for the existence of these features using statistical comparisons. Targeted features are the ones that can be explained with respect to reservoir connectivity. After that, validation of the explanation was acquired using the output dataset. Once the results and explanations were validated, the same process was used to test other cases.

The first step was to identify graph features such as cluster and flairs (Figure 3.29). These features were generated by the behavior of the inverted 4D seismic data when applying selected metrics and lenses. For each interesting graph, these features were identified and data points were grouped based on the features (Figure 3.31). Clusters were expected in the impermeable faults case and discontinuity in the generated shape was expected to exist. This is because the reservoir pressure is totally different between the faults. In the preamble faults case, the shape was expected to have clusters that were weakly connected in the

communication layers.

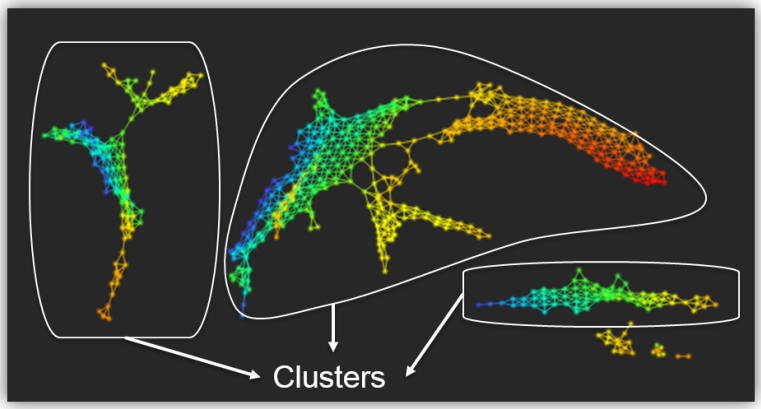


Figure 3.31: Brillig Graph Feature Identification. Three main clusters of data have been identified. By analyzing their data, the relationship to compartmentalization was easy to identify.

Different clusters identified in the previous step were examined and studied. An explanation for their existence is obtained considering input data and the configurations of metrics and lenses. By running statistical comparisons between the nodes groups that represented graph features, the properties that distinguish them and the properties creating the shapes were identified (Figure 3.32). Popular statistical comparisons tools between groups of nodes were used, like KolmogorovSmirnov test (KS test) and Student's t-test.

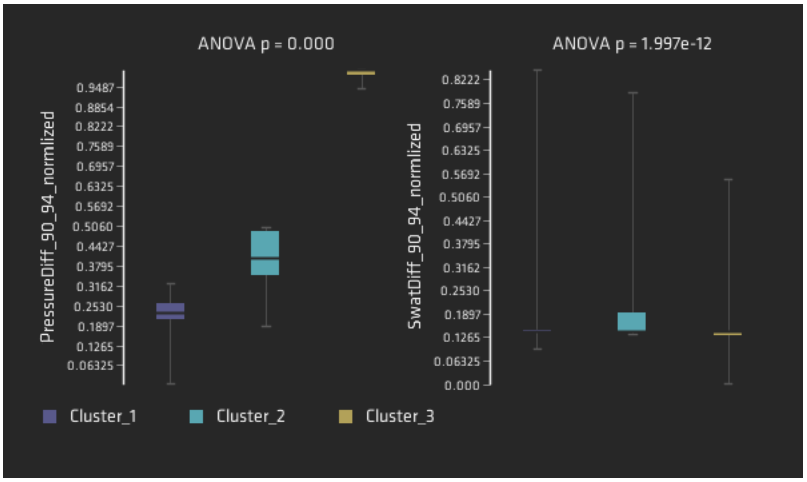


Figure 3.32: Brillig Clusters Comparison. The three cluster identified in Figure 3.31 have different pressure difference ranges.

Using the graph generated by variance normalized euclidean and neighborhood lenses as an example, data clusters were caused mainly by the pressure difference in the input data. Each cluster of data contained connected sub-clusters and flairs (Figure 3.33). This sub-clustering of nodes was found to be caused by location coordinates changes (Table 3.2).

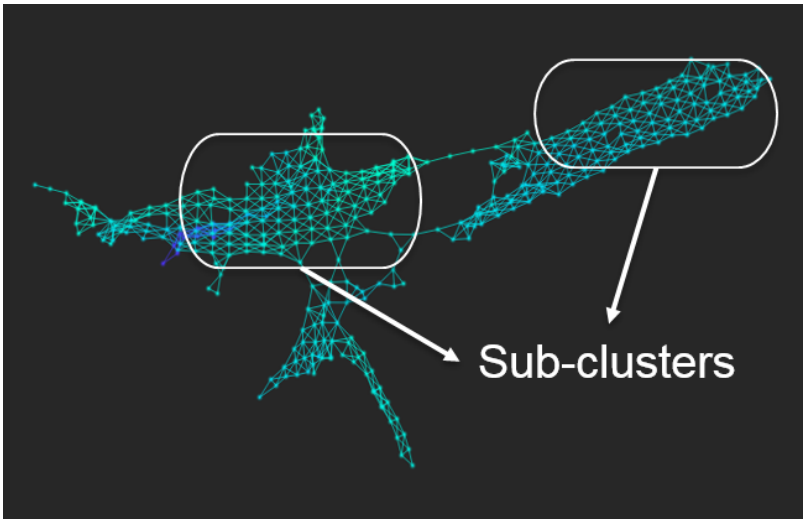


Figure 3.33: Brillig Graph Internal Features. For each clusters in Figure 3.31, flairs and sub-cluster have been identified.

Property	KS score	KS p-value
Pressure 90	0.955	5.37E-13
Z coordinate	**0.955**	**5.37E-13**
Water Saturation 90	0.955	5.37E-13
Water Saturation 94	0.947	5.37E-13
Pressure 94	0.824	5.37E-13
Water Saturation Difference 90_94	0.769	5.37E-13
Pressure Difference 90_94	0.614	5.37E-13

Table 3.2: Brillig Graph Internal Features Comparison. The main difference between sub-cluster identified in Figure 3.33 is the Z-Coordinates value. All the other values have low KS-scores or not part of the input.

The last step is in this analysis was to validate the compartmentalization results. Using the same graph, each node was colored in the graph by the compartments number to which it belongs. It was found that TDA created clusters matched the

actual reservoir regions from the simulation model. The result section covers various successful combinations of metrics and lenses validated by compartments numbers.

4. RESULTS AND DISCUSSION

4.1 Introduction

This chapter presents and discusses the results of using TDA on inverted 4D seismic data. By using different setups and configurations of TDA, compartmentalization was achieved in the three simulation model cases. First, the results for the Brillig case with impermeable faults were shared followed by the permeable faults case results. Then, Norne TDA results with different configurations were presented. Finally, a general discussion of the results and how TDA was able to generate them was offered.

4.2 Results

4.2.1 Brillig

Brillig simulation model was used to create two cases. The first case was with impermeable faults which, generated 3 distinct compartments. The second case had permeable faults, which would have resulted in one connected region containing the three compartments.

4.2.1.1 Impermeable Faults

The first successful TDA analysis was using original data and a normalized similarity metric. Three clusters are separated clearly when using variance normalized euclidean for similarity metric on original data (Figure 4.1). The lens used was topological neighborhoods with (Resolution= 51 and Overlap=2). To validate the results, the graph was colored by the compartment numbers. The generated three clusters matched the actual compartments in the simulation model.

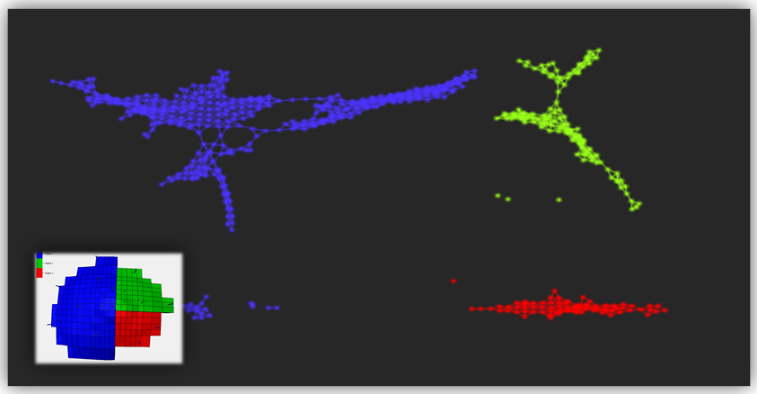


Figure 4.1: Brillig Compartmentalization Results (VNE Metric).

A similar TDA analysis was to apply a regular distance function on a normalized dataset. Compartments have separated clearly in terms of clusters when using euclidean for similarity metric on normalized data (Figure 4.2). The lens used was topological neighborhoods with (Resolution= 51 and Overlap=2). Coloring by the compartment number, it can be observed that TDA was able to create clusters that matched the original compartmentalization.

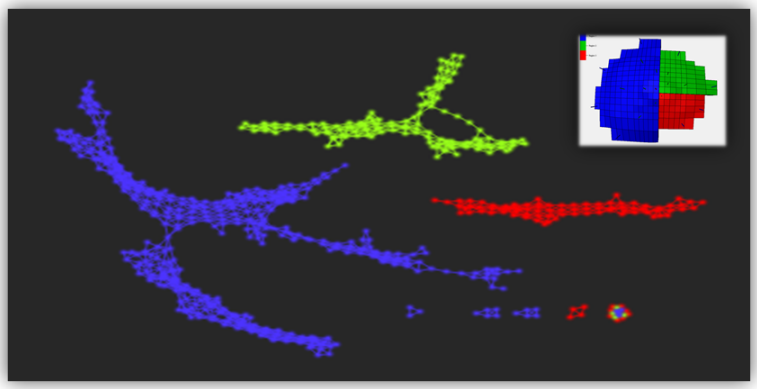


Figure 4.2: Brillig Compartmentalization Results (Euclidean Metric).

The last successful configuration to compartmentalize the reservoir was using Pearson correlation as a similarity metric between normalized data points (Figure 4.3). Three clusters of nodes in the graph were generated. It used principle component analysis (PCA) with (Resolution= 30 and Overlap=2 equalized) as lenses to create the graph. With this configuration, TDA was able to extract reservoir regions from the inverted 4D seismic data.

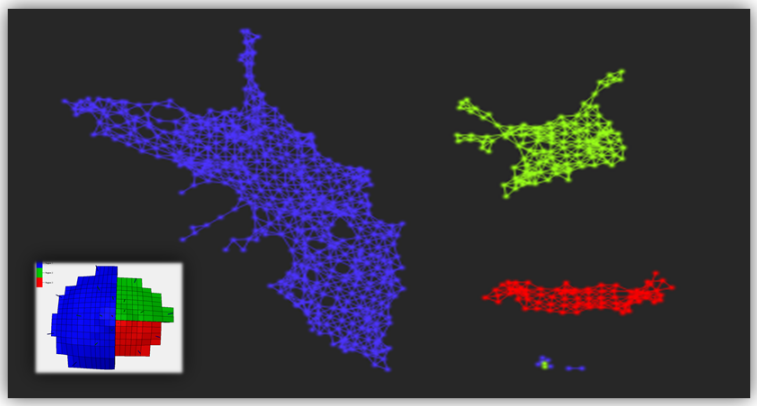


Figure 4.3: Brillig Compartmentalization Results (Correlation Metric).

4.2.1.2 Permeable Faults

For the permeable faults case, one cluster that contains the three compartments should exist. This was tested using correlation and euclidean similarity distance functions on normalized data.

The first graph was built using correlation for similarity metric on normalized data (Figure 4.4). The graph contains a one major cluster nodes that represents the full model with 3 dense areas. The three sub-clusters represent the three regions and they are in communication with each other because of having permeable faults.

The lens used was principle component analysis (PCA) with (Resolution= 30 and Overlap=2 equalized). Coloring by the compartments in the original brillig case, the three compartments did not separate as a result of fluid and pressure communication.

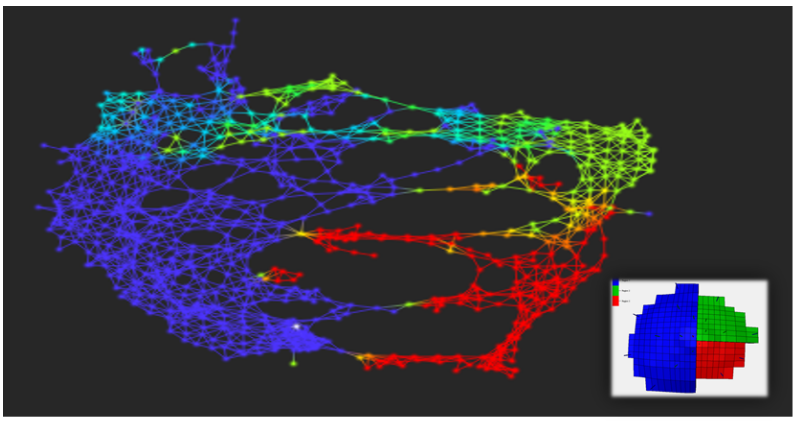


Figure 4.4: Brillig Compartmentalization Results (Correlation Metric).

Using euclidean for similarity metric on normalized data, one main cluster emerged because of having communication between regions (Figure 4.5). The lens used was topological neighborhoods with (Resolution= 51 and Overlap=3).

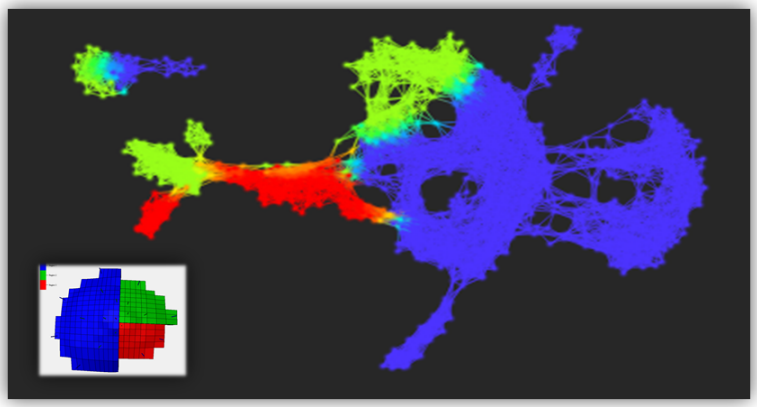


Figure 4.5: Brillig Compartmentalization Results (Euclidean Metric).

4.2.2 Norne

Norne had two main oil compartments. In addition, a non-reservoir layer existed between the top layer and lower layers, which prevents most communication between the top and bottom regions. It was decided to use three compartments in this model.

Using variance normalized euclidean to measure the similarity between the input data, the three main clusters were created and validated by coloring the graph by compartment number (Figure 4.6). The lens used was topological neighborhoods with (Resolution= 51 and Overlap=2).

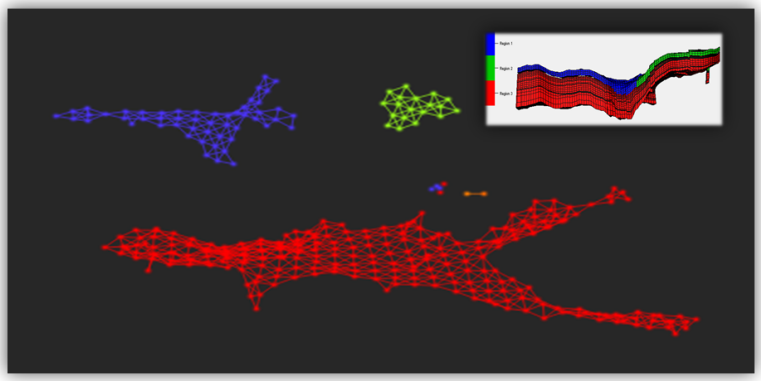


Figure 4.6: Norne Compartmentalization Results (VNE Metric).

Second, compartments separated clearly in terms of clusters when using euclidean for similarity metric on normalized data (Figure 4.7). The lens used was topological neighborhoods with (Resolution= 51 and Overlap=2).

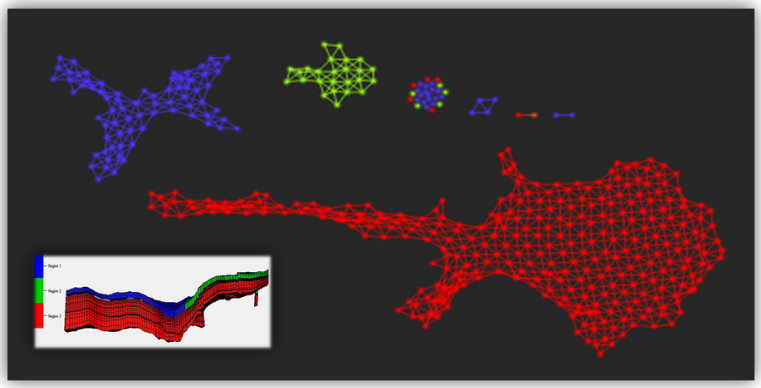


Figure 4.7: Norne Compartmentalization Results (Euclidean Metric).

Third, three compartments represented by three nodes clusters emerged when using euclidean for similarity metric on normalized coordinates data (Figure 4.8).

Pressure difference was used as a lens with (Resolution= 100 and Overlap=3).

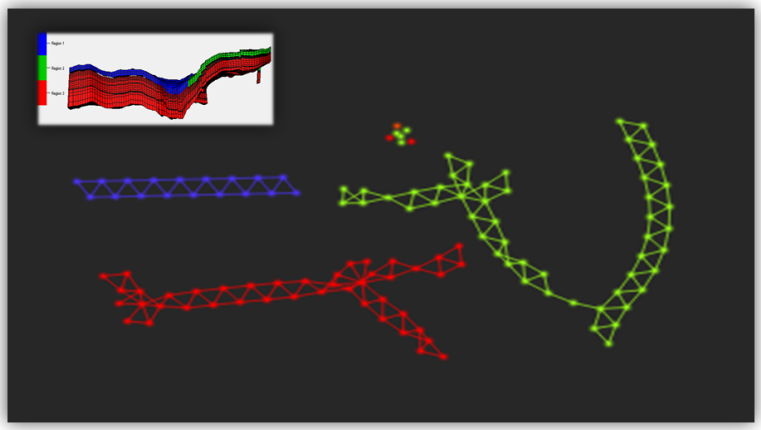


Figure 4.8: Norne Compartmentalization Results (Euclidean Metric and Data Projection Function).

Finally, two clusters of nodes were created when using correlation for similarity metric on normalized data (Figure 4.9). The lens used was principle component analysis (PCA) with (Resolution= 51 and Overlap=2 equalized). Region 2 was isolated in a cluster while regions 1 and 3. It can be explained by the low difference in pressure data in small area between regions 1 and 3. The separation layer between the two regions in that area was eliminated. It showed that some combinations of metrics and lenses are capable of identifying some information others not able to discover.

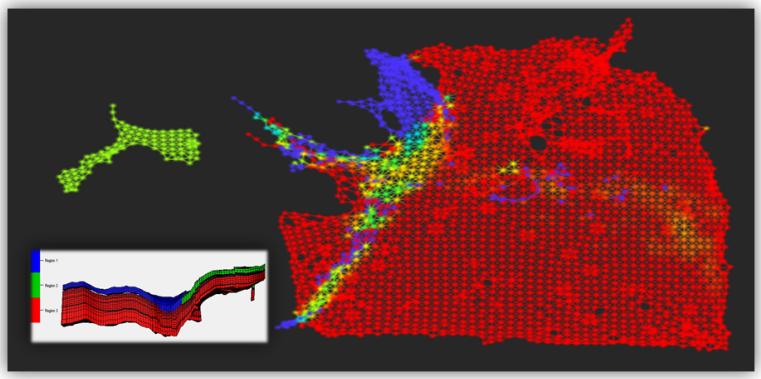


Figure 4.9: Norne Compartmentalization Results (Correlation Metric).

4.3 Discussion

From the results in the last section, TDA was able to find clusters of data points which were in communication. It is expected because of how TDA works. Using TDA to cluster the data, grouping are generated by fine tuning lenses specifications. When clear clusters are created, we run statistical comparisons to understand the reason behind them. If the reason is related to our objective, the generated clustering is considered. In the following paragraphs, TDA process is discussed from compartmentalization perspective.

First, TDA uses the user expertise in defining a concept of similarity between data points. In this study, TDA used both the location coordinates and the difference of pressure and saturation to measure the distance between data points. It had to satisfy both in order to establish a connection between any two points. Points had to be close to each other in terms of location and they had been through similar changes of pressure and saturation.

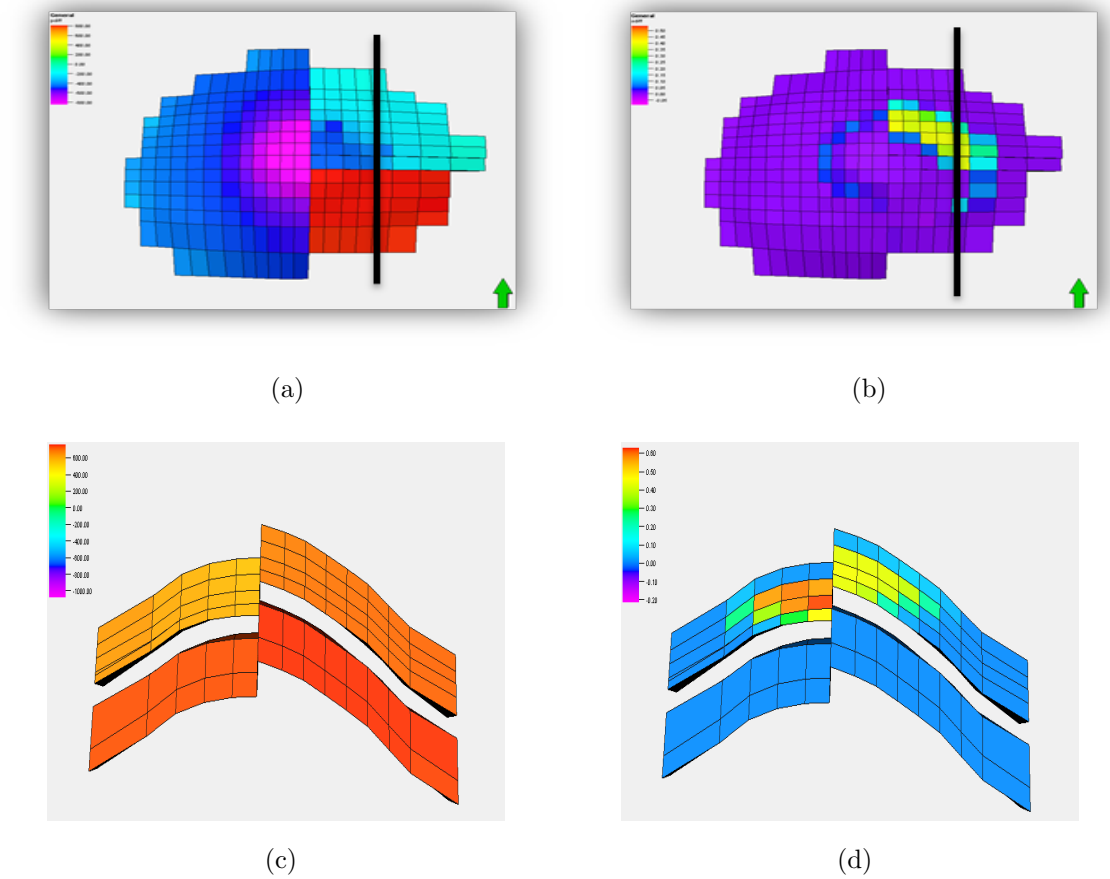
Selected lenses transforms all inputs variables into one. Using one or more function, TDA creates overlapping groups of data points in which the similarity

distance function decides if the data points belong to the same node. TDA was successful with defining both statistical functions and data functions.

Successful Statistical functions like principle components analysis (PCA) and topological neighborhoods created overlapping groups of similar changes and coordinates. Then, the similarity metric separated data points in these groups to different nodes if they are not similar. Moreover, edges created the shape of the data by linking nodes with the same data point. Thus, it ensured a compressed representation of complex data.

by looking into cross-sections of the model, it can be recognized that the points across the impermeable fault line are not similar. TDA has the capability to consider this group of data points and clusters them according to the defined metric. Pressure and saturation changes are considered using two cross sections from both permeable and impermeable faults Brillig cases in the following paragraphs.

First, a cross section over I17 is presented for the impermeable faults case (Figure 4.10). For the pressure difference, it can be observed that a clear and sharp change between the two sides of the faults. This led to a discontinuity in the generated graph. For the water saturation changes, the top layers were separated based on the fault line. However, the aquifer layers in the bottom have similar change and would be connected if water saturation were used alone.

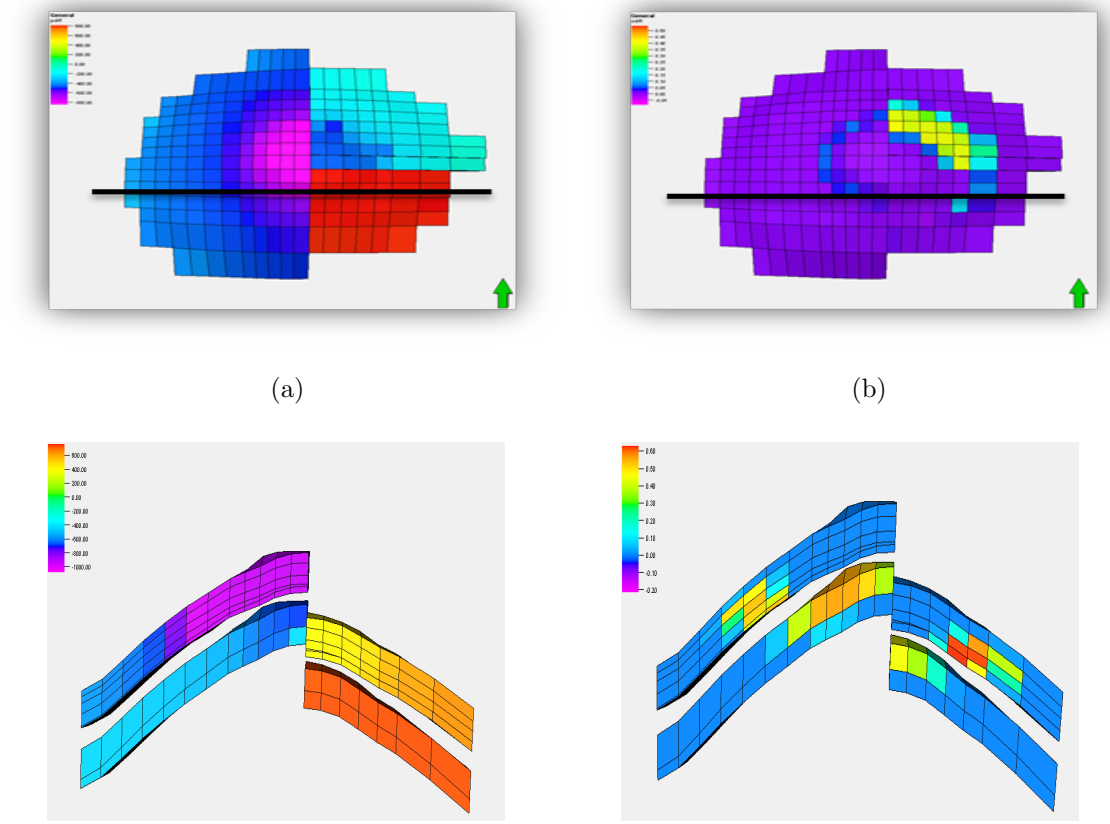


(a) (b)

(c) (d)

Figure 4.10: Cross Section I17 Location on Impermeable Faults Case (Pressure and Water Saturation Difference)

Second, a cross section over J10 is presented for the impermeable faults case (Figure 4.11). It is clear that the pressure changes split the data points between the two sides of the faults. Thus, the generated graph had disconnected nodes representing the two sides. For the water saturation changes, it was similar to the last case and it shows that water saturation changes were not effective in compartmentalizing the reservoir.

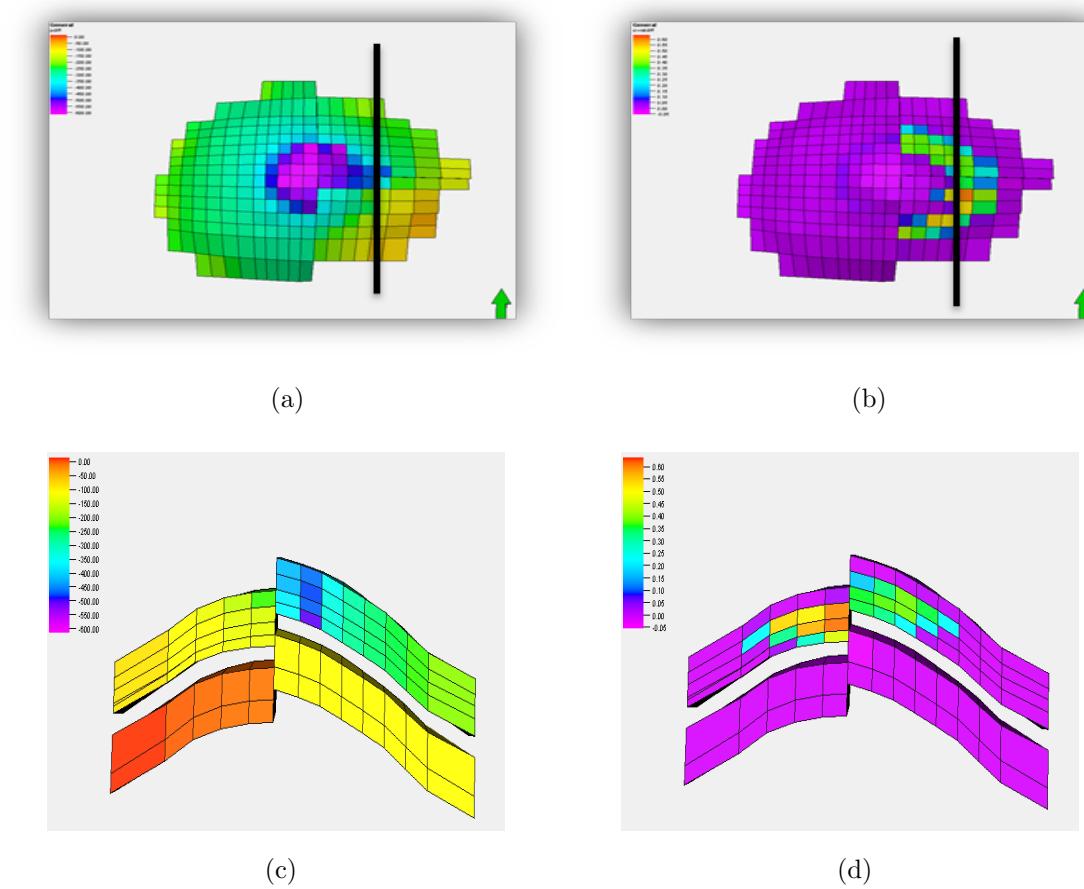


(a) (b)

(c) (d)

Figure 4.11: Cross Section J10 Location Impermeable Faults Case (Pressure and Water Saturation Difference)

For the permeable faults case, a cross section over I17 was considered (Figure 4.12). For the pressure difference, it can be observed that a clear communication between the two sides of the faults. This mean a continuous cluster of nodes in the generated graph. For the water saturation changes, the data is not changing between the two sides and it will not create a discontinuity in the graph.

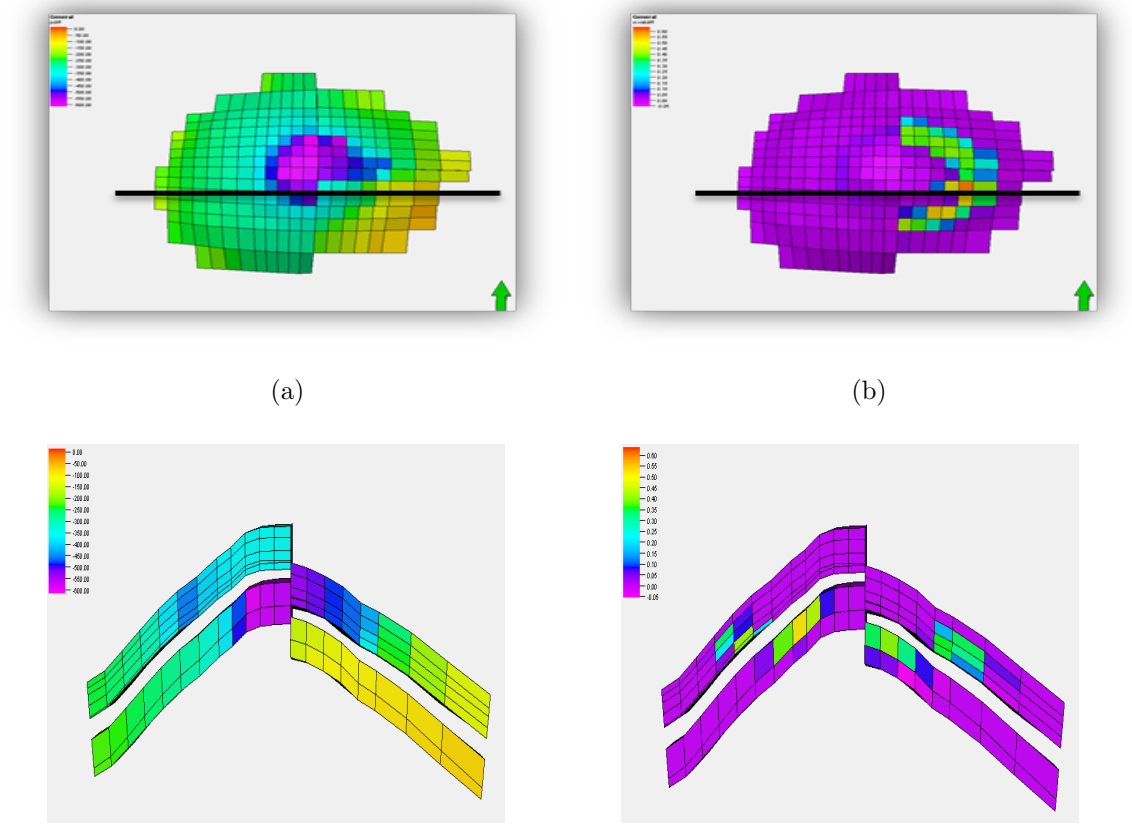


(a) (b)

(c) (d)

Figure 4.12: Cross Section I17 Location on Permeable Faults Case (Pressure and Water Saturation Difference)

Finally, a cross section over J09 is presented for the permeable faults case (Figure 4.13). It is clear that the pressure changes continue through the two sides of the fault. Thus, the generated graph had a connected nodes representing them. For the water saturation changes, it is similar to the pressure changes and the generated graph was connected.



(a) (b)

(c) (d)

Figure 4.13: Cross Section J=09 Location Permeable Faults Case (Pressure and Water Saturation Difference)

In summary, TDA was able to identify irregular transitions between data points through similarity metrics (distance function) and lenses. Lenses creates local areas where data points are comparable and a distance function defines regularity of the transitions.

5. CONCLUSION

This study have introduced topological data analysis to reservoir engineering. It was applied to inverted 4D seismic datasets in order to study reservoir connectivity. Reservoir compartmentalization is a key part of reservoir engineering because it ensures the accuracy of forecasts and development plans, the correctness of reservoir simulation, and the success of performance diagnostics and optimization. In addition, introducing topological data analysis to reservoir engineering allows identification of reservoir engineering data behavior, recognition of new opportunities, detection of anomalies and events, and lastly minimizing uncertainties.

Topological data analysis aims to extract meaningful information from complex data. It relies on the concept that data has shapes and these shapes can be translated to information. Topological data analysis has been proven to extract valued information from petroleum engineering data.

Two simulation models were used to generate the input datasets: Brillig and Norne. Topological data analysis was able to compartmentalize the reservoir models with various process configurations. Variance normalized euclidean and topological neighborhood function were used successfully to compartmentalize the reservoir model. Using normalized input dataset, correlation and principle component analysis also created similar compartments. Last, pressure changes have been used successfully as a lens.

Topological data analysis capability to discover meaningful patterns is attributed to: 1) defining a similarity distance function (metric) that represent the objective of the study, 2) finding one or more lenses that exposes the data, and 3) selecting the right combination of input data, metrics and lenses.

REFERENCES

Almaskeri, Y., MacBeth, C., et al., 2005. Location and evaluation of flow barriers using 4d seismic. In: 2005 SEG Annual Meeting. Society of Exploration Geophysicists.

URL https://www.onepetro.org/conference-paper/SEG-2005-2472

Ayasdi, 2013. Introduction to topological data analysis. Retrieved March 28, 2014.

URL http://www.ayasdi.com/social/TDAintroduction/

Bailey, R., Shirzadi, S., Ziegel, E., et al., 2013. Data mining and predictive analytics transforms data to barrels. In: SPE Digital Energy Conference. Society of Petroleum Engineers.

URL http://dx.doi.org/10.2118/163731-MS

Carlsson, G., 2009. Topology and data. Bulletin of the American Mathematical Society 46 (2), 255–308.

URL http://dx.doi.org/10.1090/S0273-0979-09-01249-X

Cole, S., Lumley, D., Meadows, M., Tura, A., et al., 2002. Pressure and saturation inversion of 4d seismic data by rock physics forward modeling. In: 2002 SEG Annual Meeting. Society of Exploration Geophysicists.

URL https://www.onepetro.org/conference-paper/SEG-2002-2475

De Jonge, G. J., Stundner, M., et al., 2002. How routine reservoir surveillance with neural networks and simplified reservoir models can convert data into information. In: European Petroleum Conference. Society of Petroleum Engineers.

URL http://dx.doi.org/10.2118/78334-MS

Euler, L., 1741. Solutio problematis ad geometriam situs pertinentis. Commentarii academiae scientiarum Petropolitanae 8, 128–140.

URL https://www.math.dartmouth.edu/~euler/pages/E053.html

Fayyad, U., 1997. Data mining and knowledge discovery in databases: implications for scientific databases. In: Scientific and Statistical Database Management, 1997. Proceedings., Ninth International Conference on. IEEE, pp. 2–11.

URL http://10.1109/SSDM.1997.621141

Lum, P., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., Carlsson, G., 2013. Extracting insights from the shape of complex data using topology. Scientific reports 3.

URL http://dx.doi.org/10.1038/srep01236

Lum, P. Y., Paquette, J., Singh, G., Carlsson, G., 2012. Patient stratification using topological data analysis and iris. Retrieved March 28, 2014.

URL http://www.ayasdi.com/_downloads/Patient_Stratification_using_Topological_Data_Analysis.pdf

Lumley, D., Meadows, M., Cole, S., Adams, D., et al., 2003. Estimation of reservoir pressure and saturations by crossplot inversion of 4d seismic attributes. In: 2003 SEG Annual Meeting. Society of Exploration Geophysicists.

URL https://www.onepetro.org/conference-paper/SEG-2003-1513

Marroquín, I. D., Brault, J.-J., Hart, B. S., 2008. A visual data-mining methodology for seismic facies analysis: Part 1 testing and comparison with other unsupervised clustering methods. Geophysics 74 (1), 1–11.

URL http://dx.doi.org/10.1190/1.3046455

Nicolau, M., Levine, A. J., Carlsson, G., 2011. Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival. Proceedings of the National Academy of Sciences 108 (17), 7265–7270.

URL http://dx.doi.org/10.1073/pnas.1102826108

Norwegian University of Science and Technology, 2010. Introduction to the norne

field. Retrieved June 15, 2014.

URL http://www.ipt.ntnu.no/norne/wiki/doku.php?id=english:fieldinformation

Norwegian University of Science and Technology, 2012. Norne simulation model. Retrieved June 15, 2014.

URL http://www.ipt.ntnu.no/norne/wiki/doku.php

Schlumberger, 2010. Eclipse black oil simulation. Retrieved December 2, 2014.

URL http://www.software.slb.com/products/foundation/pages/eclipse-black-oil.aspx

Schlumberger, 2013. Petrel e&p software platform. Retrieved December 2, 2014.

URL http://www.software.slb.com/products/platform/Pages/petrel.aspx

Snedden, J. W., Vrolijk, P., Sumpter, L., Sweet, M. L., Barnes, K. R., White, E., Farrell, M. E., et al., 2007. Reservoir connectivity: Definitions strategies and applications. In: International Petroleum Technology Conference. International Petroleum Technology Conference.

URL http://dx.doi.org/10.2523/11375-MS

Stewart, G., Whaballa, A., et al., 1989. Pressure behavior of compartmentalized reservoirs. In: SPE Annual Technical Conference and Exhibition. Society of Petroleum Engineers.

URL http://dx.doi.org/10.2118/19779-MS