Masters
Program
in **Geospatial**
Technologies

Stratigraphic Interpretation of Well-Log data of the Athabasca Oil Sands Alberta Canada through Pattern recognition and Artificial Intelligence

Onyedikachi Anthony Igbokwe

Dissertation submitted in partial fulfilment of the requirements
for the Degree of *Master of Science in Geospatial Technologies*

UNIVERSIDADE NOVA ISEGI
UNIVERSITAT JAUME·I
WESTFÄLISCHE WILHELMS-UNIVERSITÄT MÜNSTER

TITLE PAGE

Stratigraphic Interpretation of Well-Log data of the Athabasca Oil Sands of Alberta Canada through Pattern recognition and Artificial Intelligence

Final Thesis
Master of Science in Geospatial Technologies

by

Onyedikachi Anthony Igbokwe
Westfälische Wilhelms-Universität Münster (WWU)
Institute for Geoinformatics (ifgi), Münster Germany

Main Supervisor
Prof. Dr. Edzer Pebesma
Westfälische Wilhelms-Universität Münster (WWU)
Institute for Geoinformatics (ifgi), Münster Germany

Co-Supervisors
Prof. Dr. Jorge Mateu
Universitat Jaume I (UJI), Castellón,
Dept. Lenguajes y Sistemas Informaticos (LSI), Castellón, Spain

Prof. Ana Cristina Costa
Universidade Nova de Lisboa (UNL),
Instituto Superior de Estatística e Gestão de Informacão (ISEGI), Portugal

February, 2011

AUTHOR’S DECLARATION

I hereby declare that this thesis is my original work.

I also certify that, to the best of my knowledge, my thesis does not infringe upon anyone’s copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices

This thesis has not been submitted for any other degree program, and is submitted exclusively to the Universities participating in the Erasmus Mundus Master program in Geospatial Technologies.

……………………………………………. Munster, 25th February, 2011

Onyedikachi Anthony **IGBOKWE** **Place and Date.**

ACKNOWLEDGEMENTS

I wish to acknowledge first Dr. Tobias Rudolph who afforded me the opportunity to research in this area by allowing the use of the sample data and contributing an idea for the possibility of automating the stratigraphic interpretation of Athabasca Oil Sand.

I wish to thanks Juliana Exeler- A Ph.D Research student of the Institute of Geoinformatic University of Muenster Germany, who mediated for the provision of this data and her invaluable assistance in editing, proof reading and sharing ideas in many aspect during the course of this thesis.

I wish to acknowledge also my main supervisor and co-supervisors at Westfälische Wilhelms-Universität Münster (WWU), Institute for Geoinformatics (ifgi), Münster, Germany, Prof. Dr. Edzer Pebesma, Universidade Nova de Lisboa (UNL),Instituto Superior de Estatística e Gestão de Informacão (ISEGI), Portugal Prof. Ana Cristina Costa and Universitat Jaume I (UJI), Castellón, Dept. Lenguajes y Sistemas Informaticos (LSI), Castellón, Spain., Prof. Dr Jorge Mateu for their immense contribution through ideas and advice.

My thanks go also to my Erasmus Mundus colleagues Ermias, Chima, Maia, Teshome, and Neba for the friendship we shared together during this period.

I wish to acknowledge my parents for their prayers and support.

Finally, I wish to acknowledge my beloved wife Divina for her immense moral support and encouragement. Divina thanks for always being there for me. You are blessed.

ABSTRACT

Automatic Stratigraphic Interpretation of Oil Sand wells from well logs datasets typically involve recognizing the patterns of the well logs. This is done through classification of the well log response into relatively homogenous subgroups based on eletrofacies and lithofacies. The electrofacies based classification involves identifying clusters in the well log response that reflect 'similar' minerals and lithofacies within the logged interval. The identification of lithofacies relies on core data analysis which can be expensive and time consuming as against the electrofacies which are straight forward and inexpensive. To date, challenges of interpreting as well as correlating well log data has been on the increase especially when it involves numerous wellbore that manual analysis is almost impossible.

This thesis investigates the possibilities for an automatic stratigraphic interpretation of an Oil Sand through statistical pattern recognition and rule-based (Artificial Intelligence) method. The idea involves seeking high density clusters in the multivariate space log data, in order to define classes of similar log responses. A hierarchical clustering algorithm was implemented in each of the wellbores and these clusters and classifies the wells in four classes that represent the lithologic information of the wells. These classes known as electrofacies are calibrated using a developed decision rules which identify four lithology -Sand, Sand-shale, Shale-sand and Shale in the gamma ray log data. These form the basis of correlation to generate a subsurface model.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

The geophysical wireline log measurement acquired from the drilling of densely distributed wellbore in the Athabasca Oil Sands of Alberta, Canada requires expert geological knowledge for its interpretation. It requires manual analysis which is time consuming especially when the wellbores are numerous thereby making an automatic analysis more important in Oil and Gas exploration. Pattern recognition methods are used to identify the patterns of the wireline log reading from wellbore. This thesis explores the opportunity of automating the interpretation and analysis of well log data of gamma ray logs using a statistical pattern recognition approach.

1.1. Problem Statement and Motivation

A crucial problem in oil and gas exploration industry is the interpretation of the measurement of the physical properties of particular underground rocks such as density, electrical resistivity, sound transmission, radioactivity etc. These measurements are called logs and are characteristic of the various rocks penetrated by the drill. However, wellbore analyses from the massive volume of data as a result of different types of surveys and continuous well logging are constantly practices in the oil and gas exploration industry. The idea of efficient techniques to process such large volume of data have brought about the concept of an automated technique to refine the data (trace editing and filtering), select the desired event types (first-break picking) or automated interpretation (horizon tracking) for effective geological data interpretation. These have been proven to be particularly interesting to apply in stratigraphic interpretation as well as correlation. Nikravesh and Aminzadeh (2003) explain how the mining and fusion of reservoir data are utilized in the soft computing techniques for intelligent reservoir characterization. Most existing computerized well-log interpretation systems deal with maintenance of a well-log database and evaluation of formation fluid, but rarely with a stratigraphic interpretation (Wu and Nyland 1987). Over the past two decades, geosciences have experienced developments and application of numerous quantitative approaches ranging from spectral to cluster analysis. Agterberg and Bonhan-Carter, (1990) and Agterberg and Griffiths, (1991) describes how these approaches have found wide application throughout geosciences and have been used for process simulation, modeling, mapping, stratigraphic analysis and correlation, classification and prediction.

The expert geological knowledge in the analysis and interpretation of the oil sand well logs are of paramount importance. However, the manual analysis can be time consuming especially when the well log data are obtained in a large number of densely distributed wellbores. Therefore, the possibilities of automating the interpretation and correlation of these well-log datasets to build a subsurface model are very helpful in Oil and Gas exploration. It is also possible that computer-assisted correlations may suggest zonal matches of interest and originality that might not have been considered during the manual analysis. Different approaches have been explained by several authors ranging from statistical methods to Pattern recognition and Artificial Intelligence, all leading towards an automated extraction of information from signals. The term 'signal' is broadly interpreted to include well log data, waveforms, images and other survey data. This work will utilize the approaches of pattern recognition and artificial intelligence in an attempt to automatically interpret and correlate well-log datasets from Athabasca Oil Sands of Alberta, Canada. This area has about 250,000 wellbores and obviously is a real challenge in performing manual analysis.

Sequel to the above, the sample well data set of Exeler,(2009) work and the general geological knowledge of Athabasca Oil Sands were utilized for the prototypical implementation of an automated stratigraphic interpretation using well log data set within this Master thesis.

1.2. Oil Sand in Athabasca

The term Oil Sand can be defined as crude deposits which are substantially heavier (more viscous) than other crude oils consisting of sand, bitumen, mineral rich clays and water. Bitumen is a product of the oil sands that requires upgrading to synthetic crude oil or dilution with lighter hydrocarbons to make it transportable by pipelines and usable by refineries. Hence, the crude bitumen together with the reservoir rocks where it is found is known as Oil Sands (C&C Reservior 2007).

The base of the Canadian Oil Sand Reservoir is located in the north-east Alberta which comprises three major deposits, illustrated in (Figure 1): the Peace River Oil Sands, the Athabasca Oil Sands and the Cold Lake Oil Sands. According to the Alberta Department of Energy (ADE, 2006) the Oil Sands deposits were evaluated to contain approximately 1.7 trillion barrels of bitumen in-place, of which 173 billion barrels are proven reserves that can be recovered using current technology. The proven oil reserve in Alberta accounting up to 15% of the world reserve is second after Saudi Arabia (ADE, 2006).

HILLS, (1974) explain that the Oil Sand formations of Northern Alberta represent a vast source of fossil fuel whose exploitation is of major economic importance compared to the conventional crude oils.

Figure 1. The major deposit: the Peace River Oil Sands, the Athabasca Oil Sands and the Cold Lake Oil Sands Adapted from (C&C Reservior 2007)

In Athabasca, the area of interest for development contains thick bitumen-saturated sands which generally have good lateral continuity and excellent permeability. The geological formation that contains this bitumen is explained in section 3.2. Recovery of these bitumen (reservoir sands) can be done by surface mining which follows sand and clay mineral separation by warmflotation (in areas where overburden thickness is less than 75 meters) and by *in situ* thermal techniques like the steam-assisted gravity-drainage (SAGD) process especially in areas where the Oil Sands are deep suited.

Generally the recovery factor for surface mining and *in situ* thermal techniques are estimated to be between 80% and 60% respectively (C&C Reservior 2007). It is essential to note that around 250,000 wells have been drilled in this area since 1945 which have been cored and measured with petrophysical wireline logging.

1.3. Objectives

The general objective of this thesis aims at recognizing the patterns of the available well log dataset (Gamma ray data) to describe and interpret the stratigraphy by automating the classification and possible correlation process. This is done by considering the coherent grid of well logs from 21 wellbores made available for the study area. This spatial data may allow in deriving a model of the subsurface by integrating the geological knowledge of the area. The analysis will focus on determination of the electrofacies of the individual wells and delineating the sandbodies which serve as the reservoir rock for the bitumen. In the first step, the geophysical log data of each individual well will be examined, analyzed and partitioned, here gamma ray logs, using the developed clustering algorithm. The clustering aims to classify the data and subsequently depict different rock types. The difference of each gamma ray data will be re-computed and clustered for each well in order to recognize all the hidden patterns of the log data, which the original log data and the first clustering fails to recognized. In the second step, the gamma ray log data limit of each clustered well will be computed. This will be based on the similarity matrix index of each well. The optimal number obtained which is explained in section 5.2 will be utilized in determining the electrofacies of each individual wells and subsequently construct the lithology of each well. The whole process will be automated so that all the wellbores are clustered automatically and the electrofacies of each well will be generated. In the third step, a set of decision rules will be generated based on the mean gamma ray and the maximum value of gamma ray in each cluster. The idea is to automatically assign lithology facies in each electrofacies classification. The electrofacies and the lithology will identify the sand, shale and intermediate sediment. The spatial relations between the individual wells will be determined and integrated into the analysis process.

1.4. Scope of the Research

The thesis investigates a pattern recognition approach utilizing clustering algorithms for the interpretation of geological well log datasets. This approach tries to classify different rock types in the subsurface leading to the determination of a refined geological stratigraphy that will identify shale, sand and intermediate sediments from the well log dataset automatically. The implementation focuses on the depositional environment of the study area and this can be applied to different depositional environments in the future. It can be argued that the result of an automatic subsurface model from well log cannot replace human expertise due to the complexity of geological processes, but will definitely speed up the analysis process especially in large data sets from numerous wellbore.

1.5. Research Hypothesis and Question

1.5.1. Hypothesis

The hypothesis of this study is stated as: Pattern recognition approaches based on utilizing clustering algorithm can be used in classifying and interpreting stratigraphic column geologically, which promotes synergy for well correlations.

1.5.2. Research Questions

In order to meet the hypothesis and objectives of this study, the following research questions were set.

- Are the patterns from the well log data of the study area to recognizable?
- How can these patterns be recognized and automatically be utilized in the subsurface stratigraphic interpretation?
- Does this form the basis for correlation?

1.6. Thesis Structure

In this thesis the problem of pattern recognition of well log data utilizing clustering algorithm for the identification, classification and interpretation are considered. The reminder of the thesis is structured as follows:

The second section will provide an overview of related work in the field of well log cluster analysis, interpretation and correlation. It will also discuss the techniques and methodology of pattern recognition. In section three, the study area and general geological background will be discussed. These include an overview of the geological setting and environment, Petroleum system and stratigraphy. Section four will explore the dataset provided for this thesis, it will describe the well location and spatial distribution of the wells, considering the dimension and resolution of the wells. In section five the methodological framework of the stratigraphic interpretation of oil well log data will be described. It will elaborate on the steps of clustering analysis and the computation processes. Section six will illustrate the prototypical implementation of the described methods with the statistical programming language R. The results will also be evaluated in this section. Lastly, the discussion and conclusion will be discussed in section seven.

2. RELATED WORKS

The recognition of signal patterns generated during wellbore measurements is paramount in exploring for economically viable accumulations of hydrocarbon. Stratigraphic signals require analysis of its large volume of data. This enables the development of an extensive knowledge of its interpretation within reservoir distributions. Several techniques have been proposed and developed to interpret these signals (well logs) computationally, as the basis for reservoir model constructions from well log data. It comprises classification of logs, determination of electrofacies and inter-well correlation of the identified facies.

2.1. Well Log Cluster Analysis

A well log can be described as a record of the characteristics of rock formation against depth (see section 4.3 for further details). Cluster analysis starts with the partitioning of data into meaningful subgroups, when the number of subgroups and other information about their composition may be unknown. The methods range from those that are largely heuristics to more formal procedure based on statistical model (Fraley and Raftery, 1998).

Euzen, *et al*.,(2010) utilizes a method in seeking high density areas (clusters) in the multivariate space log data, in order to define classes of similar log responses thereby determining and classifying electrofacies. This was applied to develop unconventional gas prospect in the Upper Mannville incised valley fills of the Western Canadian Sedimentary Basin. Performing an electrofacies zonation based on attempt to identify clusters of log values coming from similarity level with the same characteristics has been carried out in the works of Wolff and Pelissier-Combescure, (1982); Delfiner, et al., (1987); Lim, et al., (1997); Lee, et al., (2002) and Lim, (2003). They attempt to utilize all available wireline log data set, which was against the conventional norms of using only one log, often resistivity log or gamma ray log. These logs data sets are corrected and are selectively reduced via a procedure known as principal component analysis. The first principal component computed which is a dimensionless log containing the largest common part of variances of the input logs was clustered. The clustering attempts to reduce the input log to a set of clusters which are meaningful and each cluster can be related to a specific geological facies.

Gill, et al., (1993) Partitioned a suite of well logs into geologically meaningful zones by a proposed numerical multivariate clustering. Zones and logfacies were discriminated by hierarchical clustering algorithm that defined clusters so that within-cluster dispersion is minimal. Lee, et al., (2002) proposed a hierarchical agglomerative clustering technique known as Model-based clustering to classify log data. The author noted a better performance than single-link (nearest neighbor) and k-mean clustering which often fail to identify groups that are either overlapping or of varying sizes and shapes

An automated pattern classification of well log data through different neural and non-neural techniques such as self-organize vector quantization to categorize lithological profile and determine electrofacies have been proposed by (Hassibi, et al., 2003). Vector quantization is an unsupervised clustering technique based on distance functions within Euclidean space.

2.2. Well log Correlation

After the clustering and electrofacies determination or zonation of well logs, logs can be correlated to build a geological model. These correlations are generally done manually (Schaefer, 2005). However, different methodologies have been proposed to automate this process (Gradstein et al., 1985; Tipper, 1988; Olea, 1994; Hassibi et al., 2003).

Conventional method that utilizes mathematical correlation in the space and frequency domains is reviewed in (Hoyle, 1986). Olea (1994) developed a rule based Expert Systems for automated correlation. The same approach was proposed by Lim et al., (1999) in their rule-based inference program in correlating zones between wells. Wu and Nyland (1987) applied dynamic sequence matching by coding sequence into lithofacies.

Exeler, (2009) Proposed a topological approach for the interpretation of geological well log data by integrating geological knowledge and well topology into an automatic classification and correlation process. Hassibi et al., (2003) performed similarity characterization of reservoir via pattern recognition approach which delineates dominant reservoir compartmentalization. Lateral continuity correlation of logs was performed by an Expert System.

2.3. The Theory of Pattern Recognition

Pavlidis, (1977) defines pattern recognition as involving the identification of particular structures which a given object is composed of. These objects are inspected for "*recognition*" process which turns into *classification* (Friedman and Kandel, 1999). Pattern class reflects a set of patterns that have in common some similar characteristics. To recognize a pattern, one can use a model such as self-organizing networks (Kohonen, 1997) or fuzzy c-means techniques (Bezdek, 1981; Jang and Gulley, 1995). Self-organizing networks and fuzzy c-means techniques can learn to recognize the topology, patterns, or seismic objects and their distribution in a specific set of information.

The science of pattern recognition is concerned with three major issues (Pao, 1989).

- The appropriate description of objects, physical or conceptual, in terms of representation space;
- The specification of an interpretation space; and
- The mapping from representation space into interpretation space

Furthermore, pattern recognition sciences can be exemplified as follows:

a. Classification: This tries to assign input values to one of a given set of *classes.* It implements a procedure that learns to classify new instances based on learning from a training dataset of instances with correct classes. Commonly known as supervised classification. The corresponding unsupervised procedure is known as Clustering which involves grouping of data into classes or finite set of categories according to their similarity relations.
b. Regression: This assigns a real-valued output to each input
c. Sequence labeling: This assigns a class to each member of a sequence of values (for example, part of speech tagging which assigns a part of speech to each word in an input sentence); and
d. Parsing: This assigns an input sentence a parse tree describing the syntactic structure of the sentence.

From a broad perspective, pattern recognition techniques can be classified into two major categories—the conventional approach and the artificial intelligence (AI) based approach. Conventional techniques are based upon two major methodologies known as statistical and structural pattern recognition (Nandhakumar, and Aggarwal, 1985).

2.3.1. Statistical Pattern Recognition

A statistical Pattern Recognition scheme provides for the classification of the signal into one of a finite number of classes for each of which a multivariate probability distribution function is assumed to exist, especially when the various distribution functions of the classes are known. The method tries, in most cases, to model the class distributions and also to find a discriminant function that minimizes the classification error. Modeling class distribution is based by two approaches:

i. Supposing that the class distribution comes from a known family of distributions
ii. Letting the data from the class distributions.

These approaches are categorized under parametric and non parametric method as explained in (Duda and Hart, 1996).

It is essential to note that a statistical classifier must be able to evaluate risk associated with every classification which measures the probability of misclassification.

The Bayes classifier based on Bayes formula from probability theory minimizes the total expected risk (Friedman and Kandel, 1999). The distribution function may be known *a priori* or it may be estimated from a training datasets. A classifier uses the feature values evaluated for a particular signal to assign the signal to a class. Typically, the classifier is designed with the criterion of minimizing the Bayesian error probability, or a cost measure based upon it. There are some exceptions to this which is termed template matching. Nandhakumar and Aggarwal, (1985) explain how the data being examined and the template that is being used are considered to be vectors, which utilizes a metric (e.g. the Euclidean norm) in measuring the similarity, or distance, between the two vectors. Statistical Pattern Recognition techniques are domain independent in that the algorithms can easily be transported to different domains provided that some encoded heuristics are followed.

Wu and Nyland (1987) explain that stratigraphic interpretation begins with zonations and zone correlation in which the statistical algorithms are utilized and the major interpretation is based upon maximum cross-correlation of zones in two wells. This supports the explanations of Huang and Williamson (1994) which affirm that the quantitative approaches applied in geosciences in analyzing log datasets are statistical in nature. Its application has permitted systematic, rapid and objective analysis and processing of dataset.

2.3.2. Structural Pattern Recognition

Structural Pattern Recognition schemes are based on defining primitives (substructure relationships) and identifying allowable structures. It represents an attitude rather than a specific set of procedures which involves the following processes (Nandhakumar and Aggarwal, 1985):

- identifying and extracting morphs[1] (segmentation, symbol designation);

[1] Morphs is a Greek word Morphe which means shape or form

- identifying relationships between morphs in allowable structures (defining the syntax/semantics);
- designing an algorithm for recognizing the occurrence of a structure of morphs in terms of the derived relationships (designing a parsing strategy)

It is essential to note that a signal may be considered to be made up of an arrangement of morphs (-segements of specific shapes like parabolas, straight lines etc) and the final goal of a structural pattern recognition is the detection of structures.

2.3.3. Artificial Intelligence

Artificial Intelligence (AI) is seen as a collection of advanced computing techniques developed to solve problems that humans can easily solve but are very difficult for conventional computing techniques (Baker, 1989) and also the development of computational models of intelligent behavior, including both its cognitive and perceptual aspect (Duda and Shortliffe 1983). It involves the description of abstract concepts (represented by several/ hierarchical levels of abstraction) and the recognition of instances of the signals (Nandhakumar and Aggarwal, 1985). Researchers from different disciplines have been attracted to this study. They have considered the fascinating power of the brain formed by very simple cells called neurons in controlling body action, processing signals, making decision, and information storage. A neuron can therefore be said to be a specialized cell capable of processing the incoming information and conducting it to the next neuron. Kandel, *et al.,(*1991) and Nicholls*, et al.,* (1992) explained in details the processing phase. The idea of copying the brain and neuron forms the basis of AI and is pioneered by the works of McCulloch and Pitts, (1943); Pitts and McCulloh, (1947), Rosenblatt, (1958) and Hebb, (1949).

It is essential to note that Artificial-intelligence researchers have tried also to cast the visual perception problems in the domains of symbolic representation, symbolic structures, and symbolic processes to facilitate the symbolic representation of arbitrary objects and the relationships among them. This has resulted to the broad division of Artificial Intelligence into two basic categories:

a. Rule-based (expert) systems
b. Adaptive (neural) systems

2.3.3.1. Rule-based (expert) systems

This is commonly known as *knowledge-based,* which involves reasoning. Reasoning, however, involves drawing inferences from information, provided that data are in an appropriate representation scheme. The reasoning procedures work as programs manipulating data syntactically to deduce new programs following pre-specified rules of inferences. Hence, the computer programs formulated this way exhibit what is generally considered as intelligent behavior. This type of computer software system is called a rule-based system (Startzman, *et al*., 1987). This author also noted its successful application in many areas, including computer configuration, diagnosis of infectious diseases, mineral deposits prospecting, log interpretation, and drilling-mud consultation.

2.3.3.2. Adaptive (neural) system

Artificial neural systems, or neural networks, are physical cellular systems that can acquire, store, and use experiential knowledge. The knowledge is in the form of stable states or mapping, embedded in networks that can be recalled in response to the presentation of cues. Hence, unlike a digital, sequential computer with a central processor that can address an array of memory locations, neural networks store knowledge in the overall state of the network after it has reached some equilibrium condition (stable state), thereby storing not in a particular location (Mohaghegh, *et al.,* 1996). Neural networks have pattern recognition and adaptability as its proven strong points. The essence of pattern recognition is the concurrent processing of a body of information, all of which are available at the same time (Mohaghegh, S., *et al* 1996)

The geological application of Artificial Intelligence introduced in the work of (Simann and Aminzadeh, 1989) have recorded success to the numerous challenges posed in the quantitative analysis and interpretation of geological data and had since been further developed.

Wu and Nyland (1987) state that a computerized well-log stratigraphic interpretation system based on artificial intelligence can be seen as two steps, contact recognition and interval identification which considers geologic environment for effective interpretation. The system (well-log interpretation using artificial intelligence techniques) developed by Wu and Nyland is illustrated in Figure 2 indicating how the results can perform.

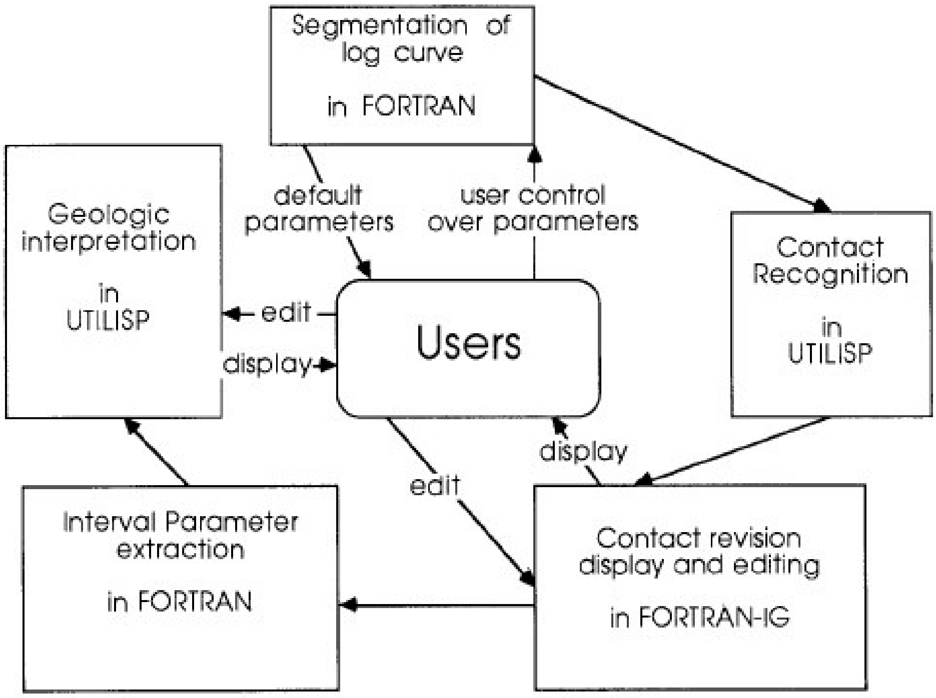


Figure 2 A schematic of the design of Artificial Intelligence in well-log interpretation Adapted from (Wu and Nyland 1987)

In this study, Statistical Pattern recognition and Rule-based system (Artificial Intelligence) methods will be used. This is because of the nature of the sample data.

Secondly, statistical pattern recognition allows a *priori* geologic knowledge to be inserted into the algorithm.

3. THE STUDY AREA

This section describes the sedimentary environment of deposition, the general geological setting, and the sequence stratigraphic framework of Athabasca Oil Sand where the exploration is done through drilling sizeable quantities of densely distributed wellbores. The output of the drilling amounts to large datasets that requires quantitative analysis.

Sequel to the objective set for this thesis which explores the concepts of an automated interpretation of well log data from Athabasca Oil Sand through pattern recognition, the detailed knowledge of the geology is significant for the interpretation approach described in section 5.

3.1. The Sedimentary Environment and General Geological Setting

A sedimentary Environment describes the combination of physical, chemical and biological processes associated with the deposition of a particular type of sediment. It also describes the rock type formation after lithification if the sediment is preserved in the rock record.

The Sedimentary environment can be divided into various classes which includes alluvial fans, rivers and flood plains, marginal-marine (deltas, alongshore sand bodies), and marine (shelf, submarine fans, turbidite sequences) (see figure 10). According to Hassibi *et al.,* (2003) deep marine environments are mostly formed by turbidity flows and its facies especially those comprising fan channels and lobes, constitute some important hydrocarbon reservoirs worldwide. It is important to note that heterogeneities are bound to occur in turbidite reservoirs which can be as a result of the lobes to the source material. The nature of sand lobes determines the vertical and horizontal variations in the sequence thereby making sand continuity and lateral correlation an important issue.

The Athabasca Oil Sands Area is located in the Western Canada Sedimentary Basin, north-east Alberta Canada (Figure 1, Section 1.2). The area is one of the several bitumen-producing areas that occur along the eastern margin of the western Canada Sedimentary basin (C&C Reservoir, 2007). The Western Canada Sedimentary Basin (WCSB) covers an area of 1, 400,000 Km^2 and stretches from the Proterozoic crystalline basement of the Canadian Shield in the north-east to the Cordilleran fold-thrust belt in the south-west. According to Alberta Geological Survey, WCSB can be divided into two distinct parts, reflecting sedimentation in two profoundly different tectonic settings.

i) The carbonate rocks dominated Paleozoic to Jurassic platformal succession

ii) The clastic rocks dominated overlying mid-Jurassic to Paleocene foreland basin succession

The former was deposited on the stable craton adjacent to the ancient (dominantly passive) margin of North America while the later was formed during active margin orogenic evolution of the Canadian Cordillera. This explains the three phases of the geological history of the Western Canada Sedimentary Basin (WCSB) according to (C&C Reservoirs, 2007) which includes the following: :

1. Cratonic platform (Precambrian-Middle Jurassic)
2. Retro-arc foreland basin (Middle Jurassic-Eocene)
3. Intracratonic basin (Eocene-present).

During the Precambrian to Middle Jurassic, the WCSB lay on the western flank of the North American continent as part of the cratonic platform. Price (1994) stated that the WCSB was affected by block faulting and volcanism associated with subduction beneath an oceanic volcanic arc lying outboard of the continental margin during the Lower Paleozoic period and in this phase clastic detritus came mainly from the east. The Antler Orogeny in the USA was linked to the basin's extension and rapid subsidence in the late Devonian-Early Mississippian. In the middle Jurassic, Columbian Orogeny occurred as a result of eastward subduction resulting in regional angular unconformity. It is essential to note that the locus of maximum subsidence migrated north-east as the accretionary prism prograded eastward onto the flank of the continental craton. An episode of crustal extension in the Eocene in the Cordillera marked the transition to the present-day intracratonic tectonic regime. From the Eocene onward, the WCSB was progressively uplifted, with as much as 1 km of uplift occurring by the Quaternary (C&C Resevior, 2007). The basin contains a wedge of Middle Proterozoic to Eocene sediments that pinches out into the Canadian Shield and thickens to >6 km adjacent to the thrust front in SW Alberta (Wright et al., 1994).

3.2. Petroleum System and Stratigraphy

The Pre-Cretaceous regional angular unconformity of Athabasca lies below the Lower Cretaceous (Aptian-Albian) Mannville Group deposit (Fig. 3A). The north-northwest trending regional valley was created as a result of the dissolution of salt in the Elk Point Group and collapse of carbonates in the overlying Beaverhill Lake Group within the subcropping Middle Devonian section which accompanied erosional downcutting of the unconformity surface

(Figures 3 and 4). This valley is commonly called the Main Valley which bordered to the west by a carbonate-cored anticline, known as the Athabasca Anticline and served as a major feature in the Athabasca area in the Early Cretaceous (figure 5). During a marine transgression in which the Boreal Sea invaded the Alberta Basin from the north-west, sediments of the Lower Mannville Group (Aptian McMurray Formation) were deposited in the erosional valley. At this moment, the Alberta Basin was bounded on the west by the rising Cordillera and continental sediments eroded from this mountain chain entered the Main Valley from the south, beyond the southern terminus of the Athabasca Anticline (figure 5). Above the unconformity surface, fluvial channel-fill sandstones of the lower McMurray Formation were directly deposited by the river flowing northward along the valley axis.

As the sea level rises, the Upper Mannville Group (Wabiskaw Member of the Clearwater Formation) was deposited above the McMurray Formation which was covered by upper Clearwater marine shales. The great majority of bitumen in the Athabasca area is entrapped in the McMurray/ Wabiskaw interval. The Clearwater shales are overlain by marine shelf/shoreface, delta-front and delta-plain deposits of the Grand Rapids Formation, which comprises the upper portion of the Mannville Group and consists of multiple prograding deltaic cycles topped by marine flooding surfaces. (C&C Reservoir, 2007).

Figure 3A depicts the bitumen generation of Athabasca from the Devonian-Mississippian Exshaw Shale, which occurs beneath the regional angular unconformity that underlies the Mannville section. The Exshaw was deposited on an anoxic marine shelf containing Type II kerogen with TOC values of 10-20% and hydrocarbon indices of 400-600. During the uplift and uncapped episode in the mid-Tertiary, hydrocarbon expulsion ceased. However, most bitumen trapped in the Athabasca migrated out of basin as oil and entered shallow stratigraphic and structural traps in the Late Cretaceous-Paleocene Laramide deformation. The oil was subsequently water-washed and biodegraded to form bitumen.

The occurrences of bitumen in three reservoir rocks of Athabasca have been identified in the Lower Cretaceous McMurray, Wabiskaw Grand Rapids deposits and the Devonian Grosmont/Nisku deposit. The heavy viscous oil is contained in a shallow stratigraphic-type trap formed by the McMurray Formation and the overlying Wabiskaw Member of the Clearwater Formation, both of Early Cretaceous age.

The Lower McMurray fluvial succession is preserved mostly parallel with lows on the sub-Cretaceous unconformity, and contains mainly bottom water. Figure 3 shows the development such as the in-situ projects going on in these areas. Flach and Hein, (2001) described the Joslyn Creek *in-situ* project which target bitumen in Lower McMurray braided river-sand reservoirs where the reservoir comprise sand-dominated, channeland- bar complexes, having high porosities and permeability, high interconnectivity and lacking internal barriers or baffles

The overlying Upper McMurray succession is a transgressive systems tract that contains some of the richest bitumen reservoirs within the Athabasca deposit, hosted mainly within amalgamated or stacked estuarine channel-and-point bar complexes. These reservoirs include thick (up to 58-m) estuarine channel sands with no laterally extensive shale breaks.

In summary, the Cretaceous McMurray Formation of the Athabasca Oil Sands was deposited on the eastern, low-accommodation side of the foreland-basin. Reservoirs, 10–90 m thick, occur in tidally influenced meandering point-bar and tidal-bar deposits. The reservoir occurs at depths of 0 to 400 m a combination of structural-stratigraphic trap by viscous immobility, depositional pinchout, and subtle anticlinal closure (C&C Resevior, 2007).

The sediments were mainly derived from exposed craton to the east and northeast with the minimum sediment burial and early oil migration resulting in 30–35% porosity and multi-Darcy permeability. The variation of the reservoir and the bitumen parameters are identical.

According to Norsk Geologisk Forening, microbial biodegraded bitumen of Athabasca varies from 6–8º API gravity with greater than 1,000,000 cP viscosity, which can also vary by an order of magnitude over 50 m vertical and 1 km laterally. However, the shale layers (mud plugs), extensive tidal flat or muds on laterally accreting point bars posses a manageable challenge to bitumen development. The oil source is likely Mississippian shale of the Exshaw Formation (see Fig. 4) from the underlying passive margin succession, which reached maturity during foreland basin compression. Work by others suggests that the Exshaw oil may have been generated and emplaced ~ 112 Ma ago, just after deposition of the McMurray Formation..

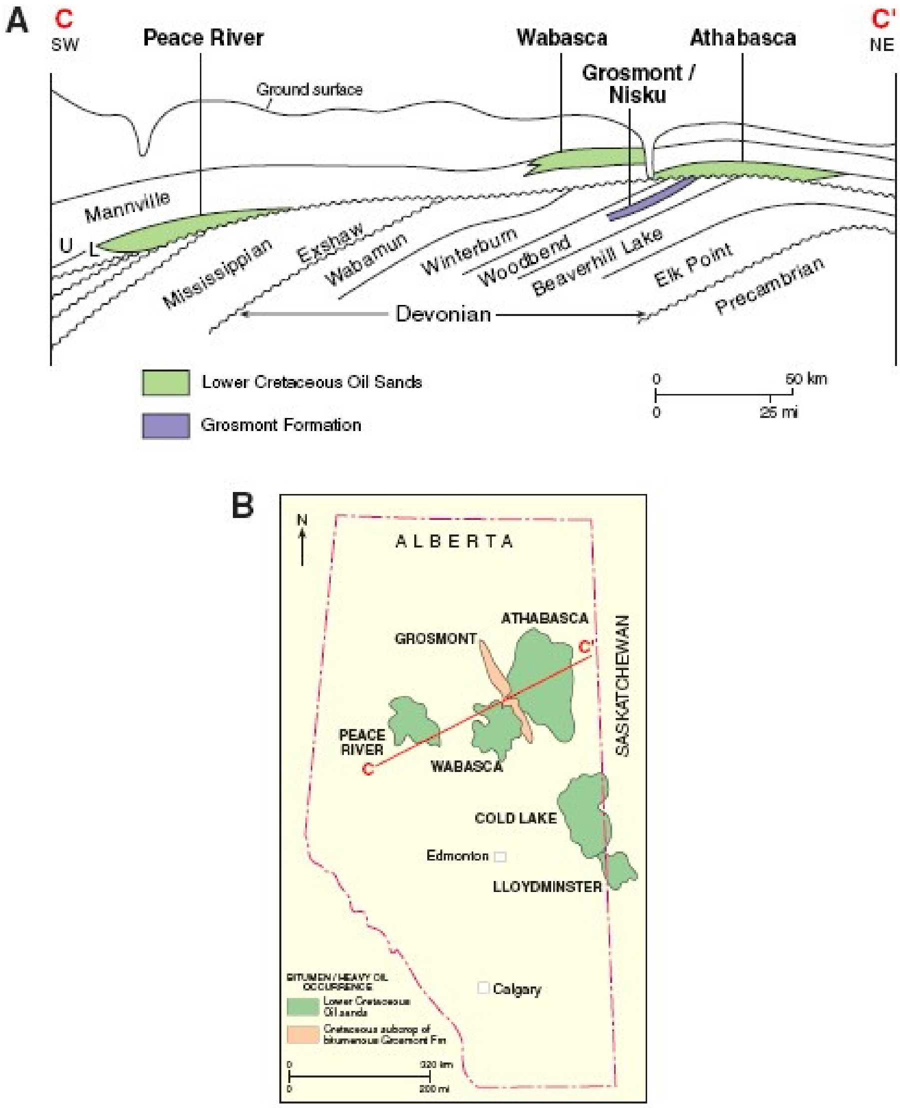


Figure 3 – (A) SW-NE schematic structural cross section showing distribution of Athabasca, Wabasca and Peace River oil sand in the Lower Cretaceous Mannville Group and their relationship to subcropping bitumen-impregnated carbonates of the Grosmont/Nisku deposit in the Devonian Woodbend Group. Section line shown below (Cutler, 1982). (B) Major bitumen and heavy oil of Alberta. The Grosmont/Nisku deposit partly subcrops beneath the western portion of the McMurray/Wabiskaw deposit (Hallam et al., 1989)

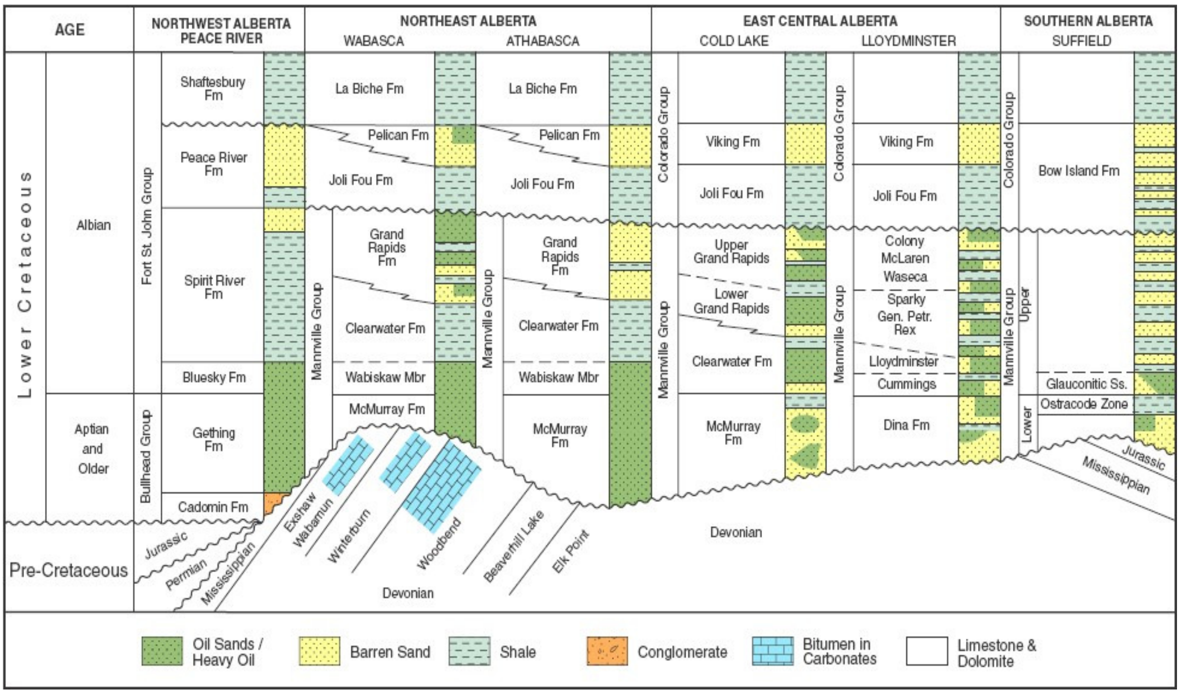


Figure 4 Lower Cretaceous stratigraphy of Alberta, showing unconformable relationship of the Cretaceous section to the underlying Paleozoic section. Adapted from (Keith et al., 1988)

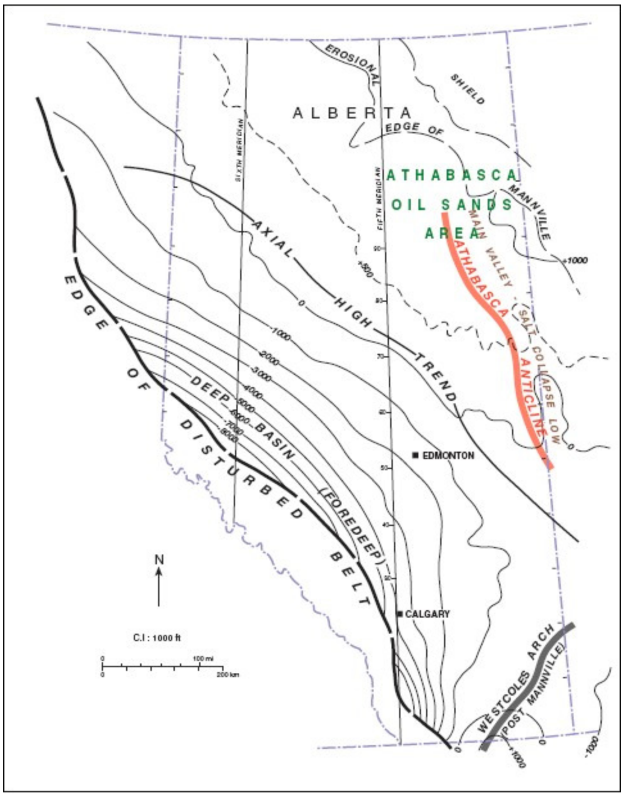


Figure 5 Depth-structure map drawn on pre-Cretaceous unconformity, Alberta, Canada.

Table 1 shows the percentages of bitumen reserve represented as 68% and 32% respectively, contained in the Lower Cretaceous sandstone of the Mannville Group and the underlying Devonian carbonate of the Grosmont and Nisku formation. The Mannville Group of Athabasca oil Area comprises, in ascending order, the Aptian McMurray Formation, the Albian Clearwater Formation and the Albian Grand Rapids Formation. The McMurray is dominated by fluvial/coastal-plain sandstones and the Clearwater by marine-shelf shales and was deposited directly upon a Cretaceous angular unconformity underlain by Paleozoic carbonate rock (Figure 3) (C&C Resevior, 2007). Wabiskaw Member directly overlies the McMurray Formation which comprises a thin interval of marginal-marine glauconitic sandstones and shales. Hence, in the oil sand of the Lower Cretaceous Manville Group, 94% of the bitumen reserves occurred in the McMurray/Wabiskaw deposit while the remaining 6% are found in the Grand Rapid reservoir.

Table 1 In-place reserve for Canada's bitumen deposits (Alberta Energy and Utilities Board, 2006)

Area	Deposit	Bitumen in Place	
		($10^9 m^3$)	(10^9 bbls)
Athabasca	Grand Rapids	9	55
	McMurray/Wabiskaw	148	932
	Grosmont/Nisku	61	383
		218	**1370**
Cold Lake	Grand Rapids	17	109
	Clearwater	9	59
	McMurray/Wabiskaw	4	27
		30	**195**
Peace River	Bluesky/Gething	10	62
	Debolt/Shunda	10	65
		20	**127**

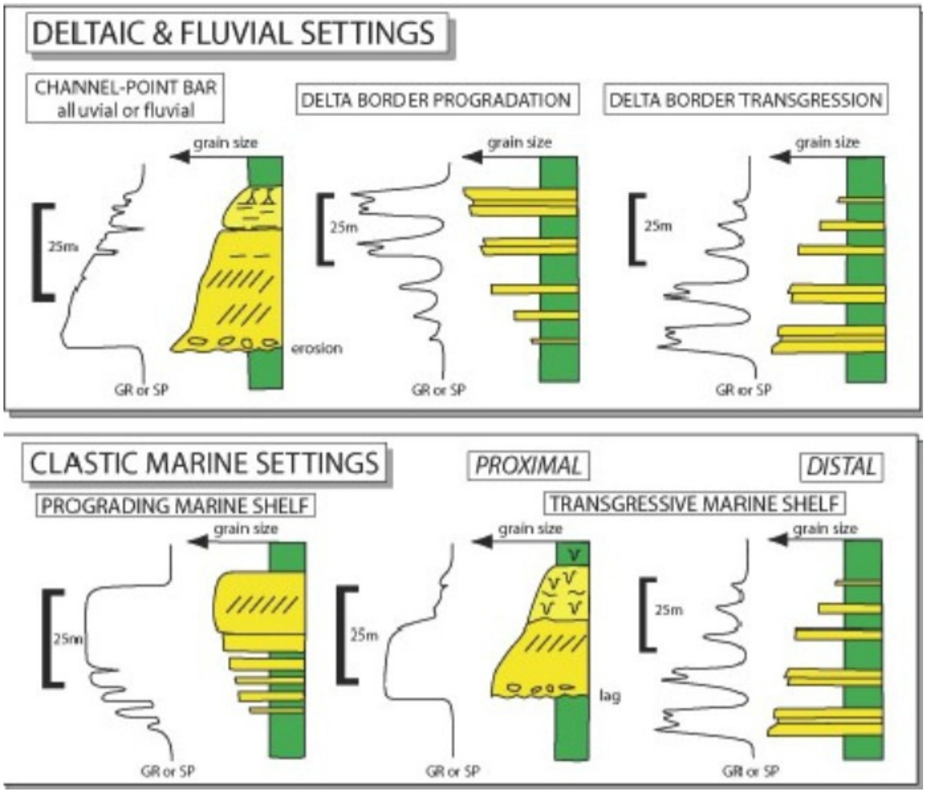


Figure 6 Response of gamma ray to different depositional environment (Modified from Tsai-Bao Kuo. 1986).

Athabasca deposit is hosted within fluvial, estuarine, and marginal marine deposits of the Lower Cretaceous Wabiskaw-McMurray succession (Hein and Cotterill, 2006). Figure 6 shows different response of gamma ray to different depositional environment. It essential to note that different depositional environment have different characteristic log response and the higher the gamma ray units the sandy the signature.

Flach and Mossop (1985) describe the stratigraphic pattern of the northern portion of the Athabasca Oil Sand Area has the McMurray Formation divided into lower, middle and upper members. The thickness of the lower member is between 0-60m with an overall fining-upward profile and consists of medium- to coarse-grained sands, arranged in 5-10 m thick, fining-upward packages, interbedded with carbonaceous mudstones and thin coals that are common near the top (Figure 7). The middle members is approximately 40m thick and consist of channel facies and off-channel facies and the upper member between 15-30m thick and is composed of bioturbated, argillaceous sand and shales that are local truncated by channel-fills. Some sand intervals near the top of the upper member were interpreted as marine offshore bars (Figure 7B)

Figure 8 describes the detailed interpretation of a gamma log. Thus, this thesis will explore the possibilities of recognizing the signature of the sand bodies (sand intervals) with different grain size and also shale intervals (which are more laterally extensive) from the given gamma ray datasets in order to investigate the connections. This will enable the complete stratigraphic interpretation and correlation.

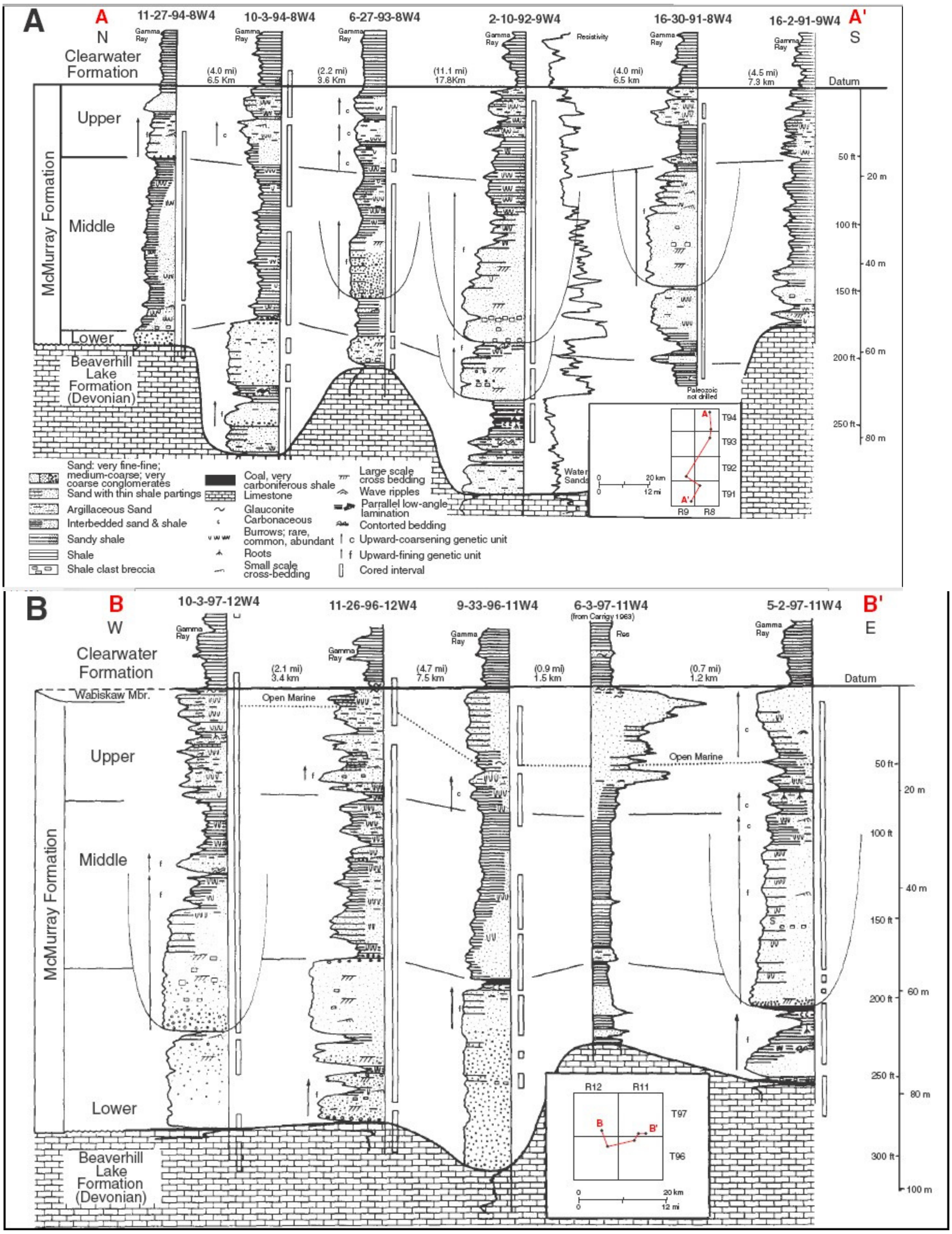


Figure 7 – N-S (A) and W-E (B) stratigraphic cross-sections of the McMurray Formation/Wabiskaw Member in the Athabasca Oil Sands surface mineable area. McMurray Formation is dominated by fluvial and estuarine deposits. Marine influence becomes more important upwards, with shallow-marine facies developed in the Wabiskaw Member. Adapted from (C&C Resevior, 2007)

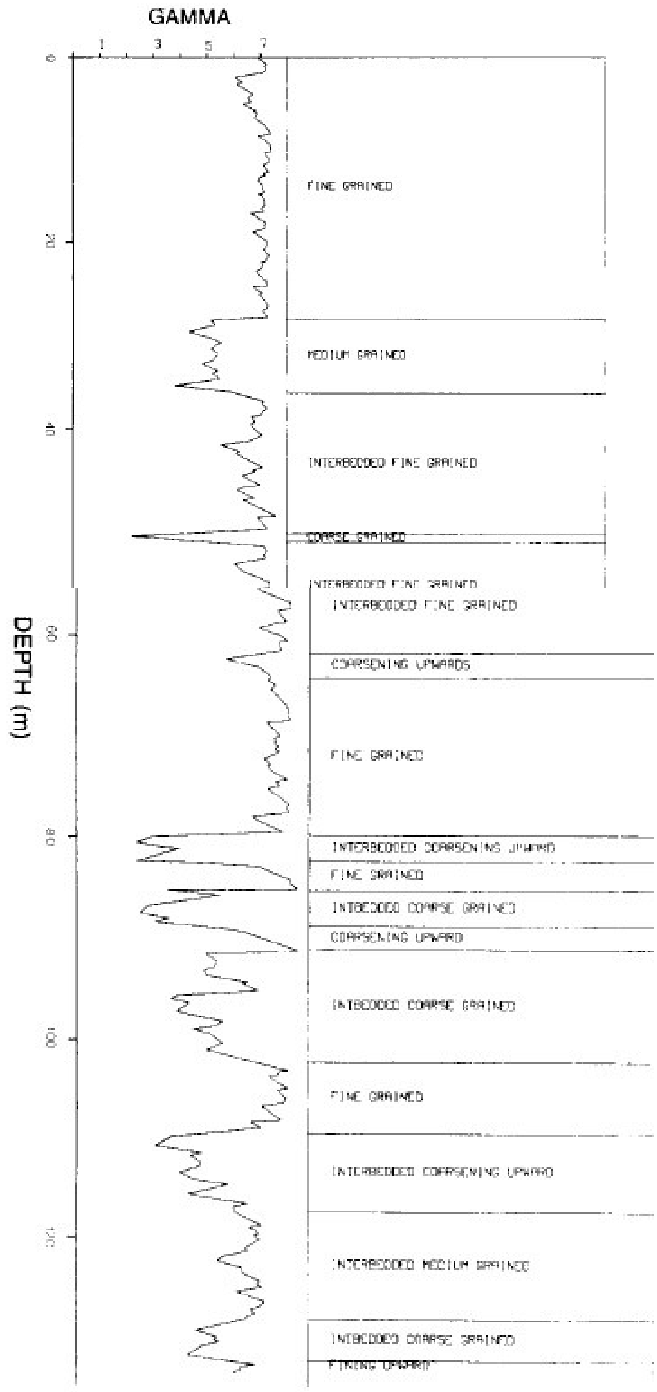


Figure 8 An automated interpretation of gamma log. Adapted from (Wu and Nyland 1987)

4. STUDY DATA

This section provides a brief description of the sample data available from the study area. These sample dataset were used to test the electrofacies determination/ interpretation approach of this thesis. The well logging techniques utilized in the wells of the study area are also described in this section.

4.1. The Sample Data Set

The sample dataset available for this study are the well logs of Gamma ray, Dipmeter and Azimuth. It is provided from 21 wellbores from the study area, with the down-hole penetration (depth) varying between 75 and 101 meters. The depth values were adjusted to a common reference level and hence remain a relative depth representation. The log of each wells are read approximately at every 10cm down-hole in which the readings of gamma radiation, the azimuth and the dip are recorded. However, for the purpose of this study, only the dataset of gamma ray were utilized after an extensive discussion with an expert geologist from Shell International Exploration and Production, The Netherlands.

4.2. Well Distribution

The sample data is unevenly distributed of the sample dataset in an area of about 15Km^2(Figure 9). The distance between two adjacent wells varies to a large extent between 300 and 2000 meters. The sample dataset were georeferenced in a UTM projection based on WGS84, given the well location in a metric Cartesian coordinate system (Exeler, 2009). It is essential to note that the coordinates represented in this thesis are not original absolute coordinates, instead it shows relative positions and the original coordinates were changed for anonymity reasons.

4.3. Well logging

A well log is simply described as a recording of characteristics of rock formation against depth. It is carried out for reservoir characterizations where measurements of the physical properties of surrounding rocks are with a sensor located in a borehole are recorded (Telford et al., 1990). The principal aims of these tasks include the following

a. Identification of geological formations
b. Identification of fluid formation in the pores
c. Evaluation of the production capabilities of a reservoir formation.

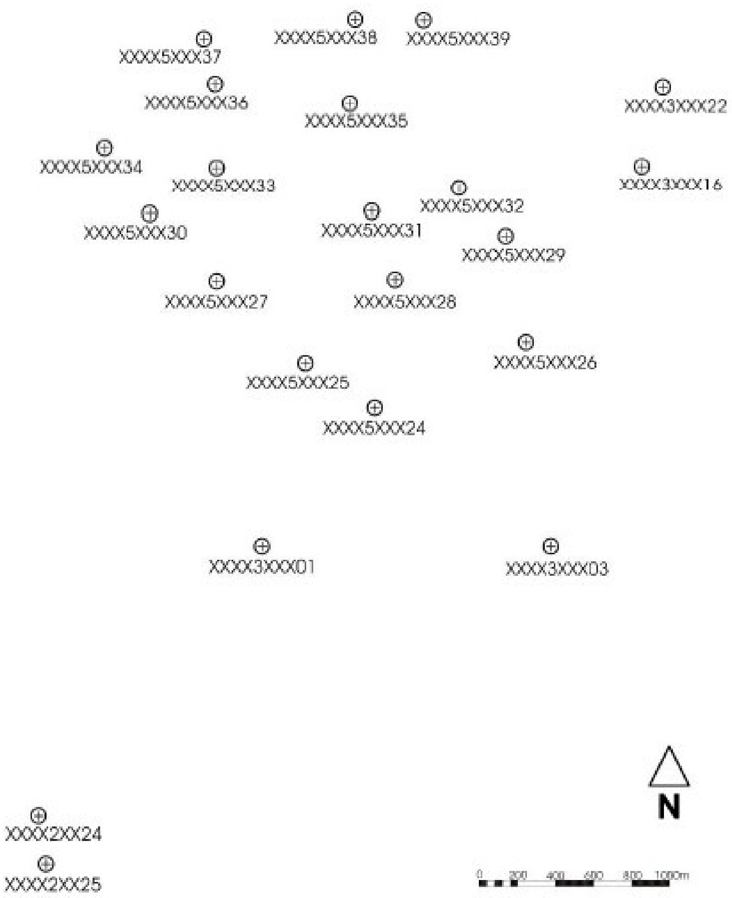


Figure 9 Relative well locations of the sample data. Adapted from Exeler, (2009)

The geological investigation of formation thickness (Lithology), porosity, permeability, saturation water and hydrocarbon usually combines about five well logging such as Electrical resisitivity logging, Radioactive logging (Gamma ray and density logging), Auxiliary logging (Includes sonic logging), Dipmater logging, Azimuth logging etc. (Telford et al., 1990). The changes which are projected on a well log imply that well log signals are function of sedimentary patterns. Figure 10 depicts a sedimentary pattern. It shows a 3D model that indicates the development of a submarine fan and the variation of sedimentation in different location. For the purpose of this work, well log data of Gamma ray, Dipmeter log and Azimuth log were provided.

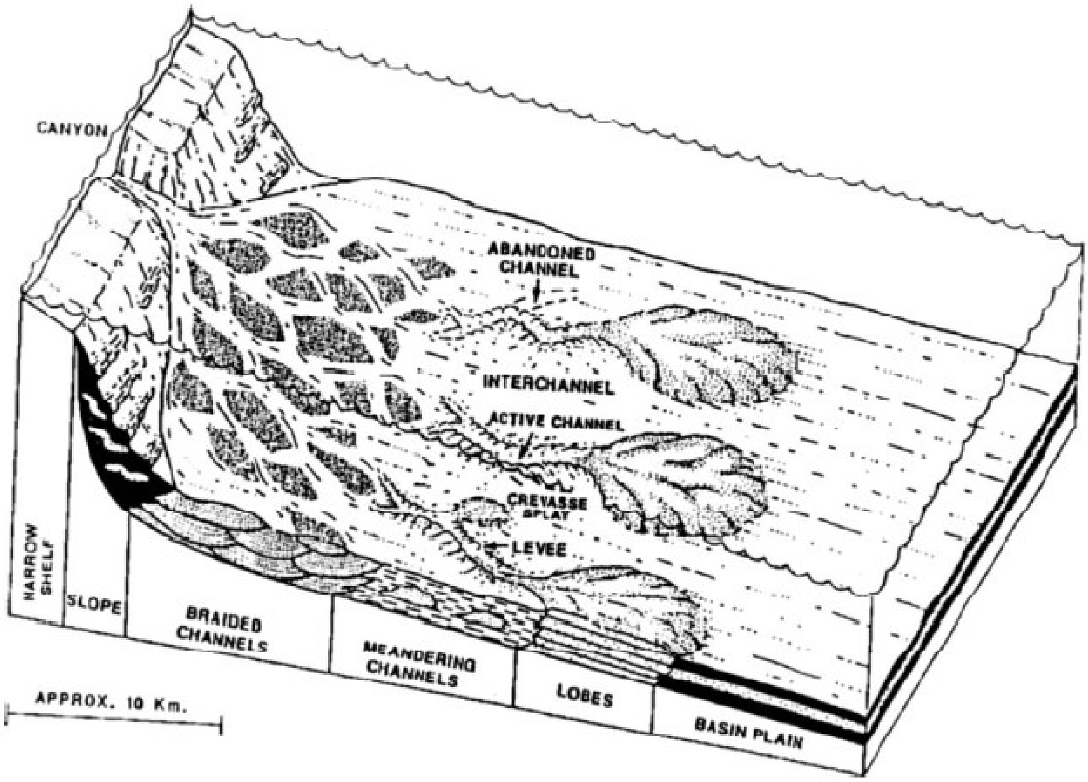


Figure 10 3D model of a sub marine fan. (Adapted from Shanmugam et al., 1988)

4.3.1. Gamma ray logging

Gamma radiation is used for qualitative evaluation of shaliness or clay content of a formation by measuring the natural radioactivity of the formation adjacent to the wellbore. Shale emits more gamma radiation than non shale sediments such as shale free- sandstones and carbonates because its gamma ray reading is high, since the concentration of radioactive material is high. Gamma ray logs are used in identifying lithologies, correlating formations and calculating volumes of shale. It is recorded in a relative scale using American Petroleum Institute (API) units. Figure 8 and 11 show an interpretation of gamma ray log.

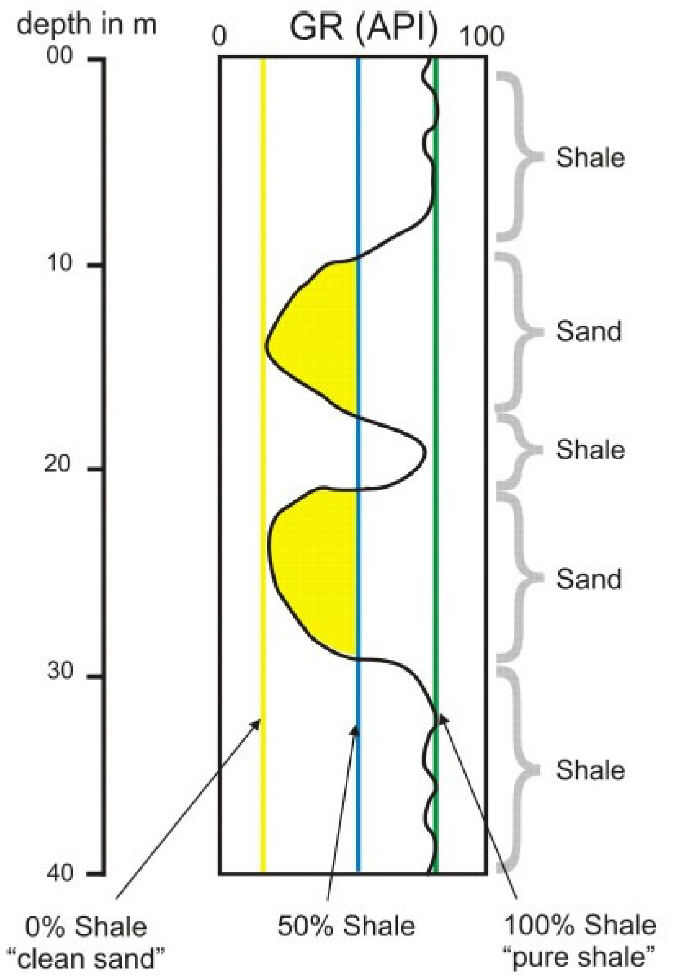


Figure 11 Interpretation of a gamma ray log. Adapted from Exeler, (2009)

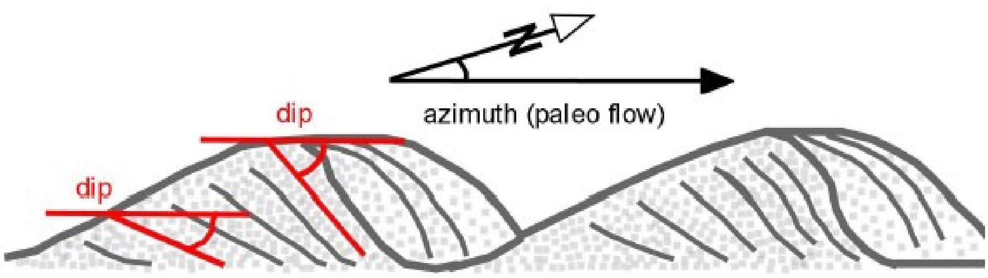


Figure 12 Azimuth and dip angles in a ripple structure deposited by channel flow. Adapted from Exeler, (2009)

4.3.2. Dip and Azimuth

Dipmeter logging tools records high-resolution conductivity curves from multiple pads pressing against the borehole wall (Baker, 1989). Dipmeter tools determine the structural dip and azimuth in wells with an accuracy of plus or minus two. Dipmeter logs have a high vertical resolution of 3 to 5 cm with the depth of investigation is approximately 1 cm. Figure 12 describes the azimuth and dip angles as it concerns the study area.

5. METHODOLOGICAL FRAMEWORK FOR THE STRATIGRAPHIC INTERPRETATION OF WELL LOG DATA

The stratigraphic interpretation procedures of the well log data applied to this thesis is described in the flow chart of figure 13. The first two steps examine the individual wellbore. The explanation of the processes and the logic behind applications is explained below.

5.1. Recognition of Lithofacies and Electrofacies

Recognition of lithofacies is a common practice in drilled wells where suitable well logs and core samples are available. Pattern recognition techniques such as hierarchical and k-means cluster analysis can be used for classifying well log data into discrete classes. The automation of well log correlation using both multivariate statistical techniques including principal component analysis (PCA) and rule based system for efficient and reliable pattern recognition for well-to-well correlation can be established. Lim, (2003) used the first principal component log, since it has the largest common part of variance of all available well log data.

The key to an automated interpretation of pattern is to explore the logic of human interpreters and follow this logic in designing the computer software. Correlation (of wireline logging data) is also based on the large set of subjective rules for pattern recognition that aims to represent human logical processes. Following the foregoing, statistical pattern recognition methods was utilized in this work. This is based on numerical computation procedures (mathematical calculation) in identification of patterns.

The advantages of this approach are that (Nikravesh, et al., 2003; Lim, 2003).

- It can be applied in all depositional environments,
- It can be more helpful for obtaining more reliable correlation results for complex geologic formation
- It has the capability of dealing with probability and uncertainty of data due to fuzziness

The geological information that is meaningful can be derived by selecting, weighting and combining a set of logs which give rise to a set electrofacies that can be correlated with geological facies for a better understanding.

An electrofacies is defined as the set of log responses which characterizes a bed and permits it to be distinguished from others (Serra and Abbott, 1982). In log-analysis applications, an electrofacies is used typically as an indicator of lithology and depositional environment (Wolff and Pelissier-Combescure, 1982; Anxionnaz, Delfiner, and Delhomme, 1990). Electrofacies zonation is based on the attempt to identify clusters of log values coming from levels with similar characteristics. The importance of electrofacies characterization in reservoir description and management has been widely recognized as this kind of data partitioning is to simplify a complex data set into some homogeneous and simple subgroups and to produce a better correlation between dependents and independents within distinct subgroups for further petrophysical properties regression (Lee and Datta-Gupta, 1999). The electrofacies determination procedure which is based on cluster analysis of well logs used in this study is summarized by the flow chart in figure 13.

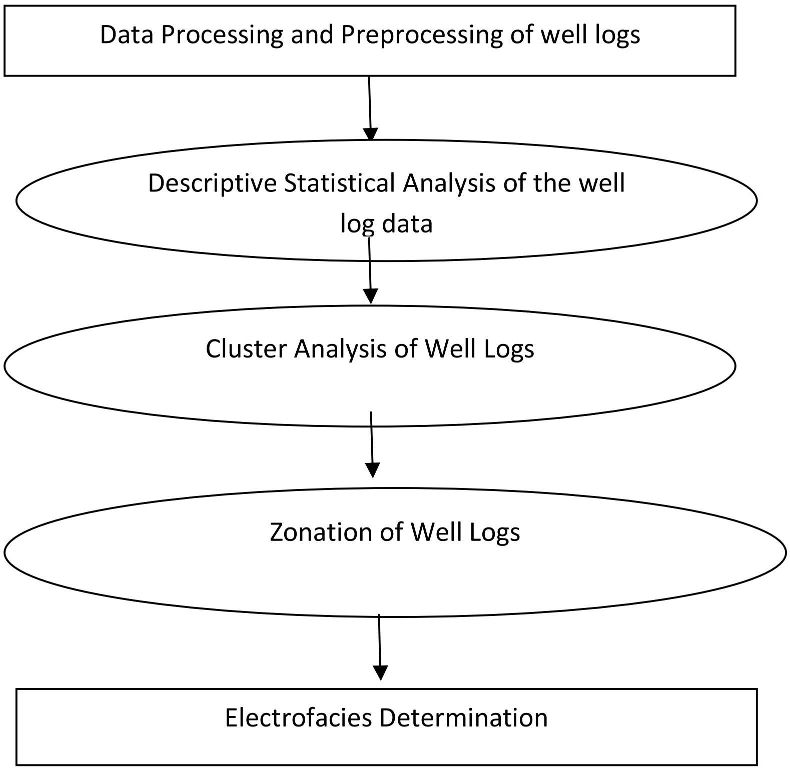


Figure 13. Flow Chart of electrofacies determination

5.2. Cluster Analysis of well logs

The provided well data comprise the readings of gamma radiation, azimuth and dip taken at about every 10cm of the total well depth. This makes the reading too detailed that sediment classification has the tendencies of overlapping. However, samples with similar logs responses need to be identified. This identification is done using a clustering algorithm that can operate on log traces.

Cluster Analysis, also called data segmentation, relates to grouping or segmenting a collection of objects (observations, individuals, cases, or data rows) into subsets or "clusters", such that those within each cluster are more closely related to one another than objects assigned to different clusters. The purpose of cluster analysis is to look for similarities/dissimilarities between data points in order to group them into classes. In multi-dimensional space logs, the distance between data points is a measure of their dissimilarities. Samples with similar log responses will tend to form clusters, separated by areas with a lower density of points. The points are close due to the similarity of log their response. The idea is to ascribe to each level of the well the group or cluster to which it belongs. Statistical theory provides two types of approach (Gnanadesikan, 1977).

One approach is “classification” (normally called “discriminant analysis”). The groups are specified in a lithofacies database and then each depth level is assigned to the correct group by use of an appropriate discriminant function. This method has a merit of being fully automatic as the interpretation work is done once. Meanwhile, correct definition of the database is very important for achieving good results

The second approach which was used here is to determine the clusters or groups from the data in each well. This “clustering” has a merit of letting the data “speak for themselves” and reveals their subtle differences. However, geologic interpretation of the cluster must be repeated each time. There are two major methods of clustering algorithm used in identification of log responses which usually can follow either a hierarchical or relocation strategy (one in which observation are relocated among tentative clusters).

Relocation methods move observations iteratively from one group to another, starting from an initial partition. The number of groups has to be specified in advance and typically does not change during the course of the iteration (Fraley and Raftery 1998).

K-means clustering is the most common relocation method. Conversely, hierarchical clustering which uses some heuristic criteria like single link (nearest neighbour), complete link (farthest neighbour), average link or maximum-likelihood as explained in section 5.2.2.1 was utilized and is described below. This have prove successful in earth sciences application and well log analysis as recorded in the works Delfiner, et al.,(1987); Lim, et al., (1997) and Lim, (2003).

5.2.1. Hierarchical Clustering

This is an approach to clustering based on the representation of data as a hierarchy of clusters nested over set-theoretic inclusion or measured characteristics. It is mainly used as a tool for partitioning. However, the data are not partitioned into a particular cluster in a single step, but a series of *X* partitions takes place, which may run from a single *C* cluster containing all objects to clusters each containing a single object. Thus, when there are *X* cases this involves *X-1* clustering steps or fusions, exemplified into partitions as *X* cluster*, X-1* cluster, *X-2* clusters…..and *Xth* in which all samples forms into one cluster. It can be said that at level *Z*, in the sequence, the number of clusters, $C = X - Z + 1$. Thus, level one corresponds to *X* clusters and level *X* to one (Duda and Hart, 1973). The key components of hierarchical clustering analysis is the repeated calculation of distance measures between objects, and between clusters once objects begin to be grouped into clusters. It is also not limited to a pre-determined number of clusters and can display similarity of samples across a wide range of scale. Hierarchical Clustering is sub-divided into two types- *agglomerative* and *divisive* methods. They construct their hierarchy in the opposite direction possibly yielding different results. (Figure 14)

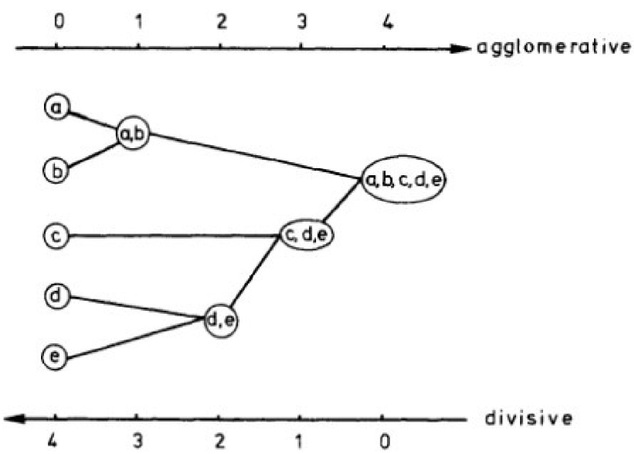


Figure 14 Distinction between Agglomerative and Decisive techniques

5.2.1.1. Agglomerative (bottom up, Clumping) Method

This procedures start with *x* singleton clusters which proceed by series of merging two nearest clusters of the *x* objects into groups at each step. It is particularly common in the natural sciences and will be utilized here.

5.2.1.2. Divisive (top bottom, Splitting)

This procedures start with all of the sample in one cluster which separate/ split a cluster in two distant parts, starting from universal cluster containing all entities.

Hierarchical clustering may be represented by a two dimensional diagram known as dendrogram or clustering tree (Figure 15) which illustrates the fusions or divisions made at each successive stage of the analysis. Figure 15 shows a simple dendrogram of 10 samples, indicating at level one a singleton cluster. At level two, samples X_5 and X_6 are grouped together to form a cluster which stays together at all the subsequent levels. It is important to note that in order to measure the similarity between clusters; the dendogram is usually drawn up to scale to show the similarity between the clusters that are grouped, thereby making the similarity values to be mostly used in determining if the groupings are natural or forced. Level one to eight may be considered natural while between level eight and nine indicates that the clusters are forced due to large reduction in similarity value (Figure 15)

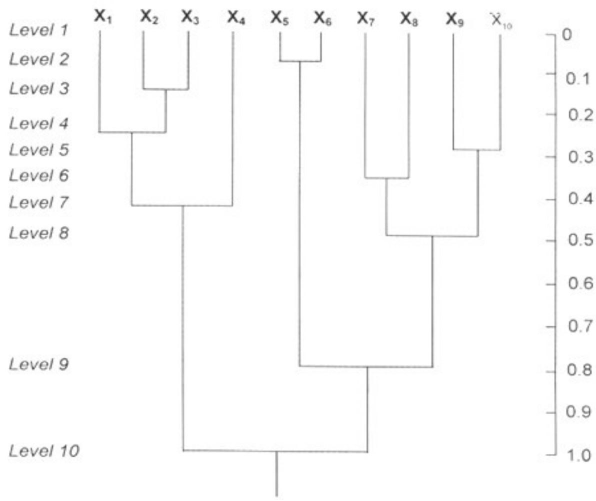


Figure 15 A simple dendrogram for hierarchical clustering

Agglomerative clustering are commonly used than the divisive methods due to its computation simplicity and it can be implemented in R software by `hclust()` and `agnes()`. Section 6 explain the software and the implementations. The steps behind the working principles of the algorithm are shown below, based on the studies of Panigrahi and Sahu, (2004).

1. Read all input patterns
2. Normalize all input patterns by dividing each data point by the maximum value of the corresponding attribute.

   $$Xi = \frac{Xi}{Xi(Max)} \quad (1)$$

3. Assign each item to its own cluster, thus, if there are n samples in set C, there will be n clusters.

   i.e. if $C = \{X_1, X_2, X_3 \ldots\ldots X_n\}$ where $Xi = \{Xi\}$, $I = 1, \ldots\ldots p$ (p is the number of characters in each sample)

   then, $C^{\wedge} = n$ (2)

4. If $C^{\wedge} \leq$ C (no. of clusters required), stop.
5. Find the nearest pair (most similar) of distinct clusters, X_i, and X_j, where $X_i \neq X_j$, whose merger will increase (or decrease) the criterion function as little as possible.
6. Merge X_i and X_j, delete X_j and decrease $C^{\wedge}$ by 1
7. Go to point 4

R software was utilized in this thesis and this automatically computes the above.

5.2.2. Computation

The details of the computation processes using R software, the implementations and the results are discussed below.

Hierarchical clustering function **`hclust()`** is in standard R functions and is available without loading any specific libraries. Hierarchical clustering requires dissimilarities as its input with standard R having functions **`dist()`** to calculate many dissimilarity functions.

Hierarchical agglomerative cluster analysis starts by calculating the distance matrix in the matrix of data. Below are the lists of the common distance functions in R with their respective disadvantages

i. The Euclidean (square) distance

$$d(X, Y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2} \quad (3)$$

The disadvantages is that the application is not scale invariant and not for negative correlation

ii. Maximum, Minowski, Canberra, Manhattan, binary.

iii. Correlation base distance: 1-r. This includes

a. Pearson correlation coefficient (PCC)

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{\sqrt{(\sum_{i=1}^{n} x_i^2 - (\sum_{i=1}^{n} x_i)^2)(\sum_{i=1}^{n} y_i^2 - (\sum_{i=1}^{n} y_i)^2)}} \quad (4)$$

The disadvantage is that it is very sensitive to outliers

b. Spearman correlation coefficient (SCC)

This has the same calculation with Pearson correlation coefficient, although with ranked values.

Cluster analysis can be run in the R-mode especially when seeking relationship among variables and Q-mode. In Q-mode analysis, the distance matrix is a square, symmetric matrix of size n x n that expresses all possible pairwise distance among sample. Hence, a Q-mode analysis is assumed in this work. Euclidean distance is utilized. Equation 3 gives Euclidean (square) distance, *dij,* between points *i* and *j*

The command in R that accomplished this is **`dist`**`(`*data*`),` where *data* is a matrix or dataframe containing the data.

On the complete computation of the distance matrix, a hierarchical cluster analysis can be completed. There are several alternative clustering linkages or methods in the standard function **`hclust.`** The method to be utilized depends heavily on the nature of the data and the expected output. The next section explains all the methods or measures.

5.2.2.1. Similarity Measure

The similarity measurement is very essential to note when deciding the clustering algorithm to be used. The degree of similarity in the clusters and dissimilarity between clusters requires distance measurement. All these commence in a similar manner by fusing two most similar points to a cluster and they differ in the way in which they combine clusters to each other, or new points to existing cluster. In order words, the algorithms differ in the way in which distance is measured between clusters mainly by using two parameters: the distance or likelihood measure (Euclidean, Dice, etc) and the cluster method (Between group linkage, nearest neighbor etc.). Figure 16 illustrates four hierarchical algorithm methods. The explanation below follows the description given by Gelbard *et al*., (2007).

Within groups average: This method calculates the distance between two clusters thereby allowing the cluster with highest average likelihood measure to be united. It applies the likelihood measure to all the samples in the two clusters.

Between groups average: This method calculates the distance between two clusters by applying the likelihood measure to all the samples of one cluster and then comparing it with all the samples of the other cluster. The two clusters with the highest likelihood measure are then united.

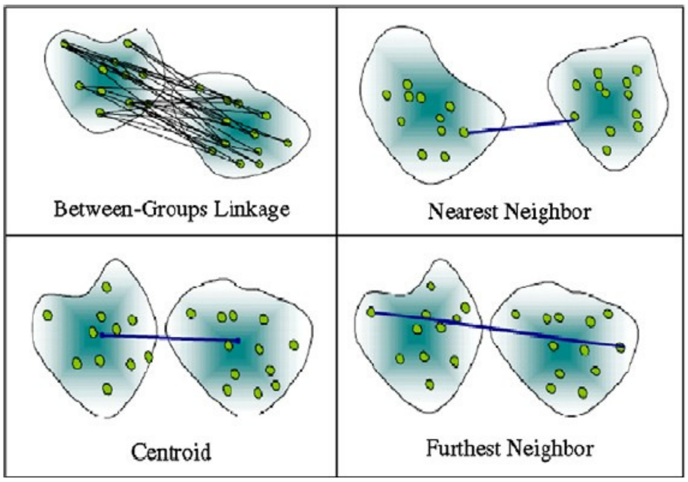


Figure 16 Showing of four hierarchical algorithms methods. Adapted from Gelbard *et al*., (2007)

Nearest neighbor (Single linkage): This method calculates the distance between two clusters by applying the likelihood measure to all the samples of one cluster and then comparing it with all the samples of the other cluster. The two clusters with the highest likelihood measure, from a pair of samples, are united.

Furthest neighbor (Complete Linkage): This method, like the previous methods, calculates the distance between two clusters by applying the likelihood measure to all the samples of one cluster and then comparing it with all the samples of another cluster. For each pair of clusters, the pair with the lowest likelihood measure is taken and two clusters with the highest likelihood measure of those pairs are then united.

Centroid: This method calculates the centroid of each cluster by calculating the mean average properties for all the samples in each cluster. The likelihood measure is then applied to the means clusters and the clusters with the highest likelihood measure between their centroids are united.
Median: This method calculates the median of each cluster. The likelihood measure is applied medians of the clusters and the clusters with the highest median likelihood are then united.
Ward's Method: This method calculates the centroid for each cluster and the square of the likelihood measure of each sample in the cluster and the centroid. The two clusters, which when united have (negative) affect on the sum of likelihood measures, are the clusters that need to be united.

A study of similarity measures as explained above clearly indicates that average linkage method is more suitable for this particular application than the single and complete linkage methods. The original gamma ray log data and the computed difference of the gamma ray logs of each well are considered. Average linkage, used here, is also very useful when clusters are not well separated which has been observed in the present case.
The general syntax/ scripts developed for the computation is printed in Appendix C
Appendix A shows the resulting dendograms of gamma ray in the 21 wellbores from the study area. The results and the implementation are discussed in section 6.

5.3. Electrofacies Classification of Gamma ray well log data using hierarchical clustering

In earth sciences, the hierarchical clustering techniques described above are most widely applied especially in the determination of the electrofacies (Lim et al., 1997), which also corresponds to a cluster in the multivariate space of log data. The clusters can be interpreted as lithofacies, homogenous classes, or similar patterns that exist in the data. For example, shale would typically form a population of data point characterized by high gamma ray while sand will be characterized by a low gamma ray.
An attempt has been made to classify the Athabasca Oil Sand well log data (Original gamma ray data and the computed difference) by hierarchical clustering following the algorithm described above.

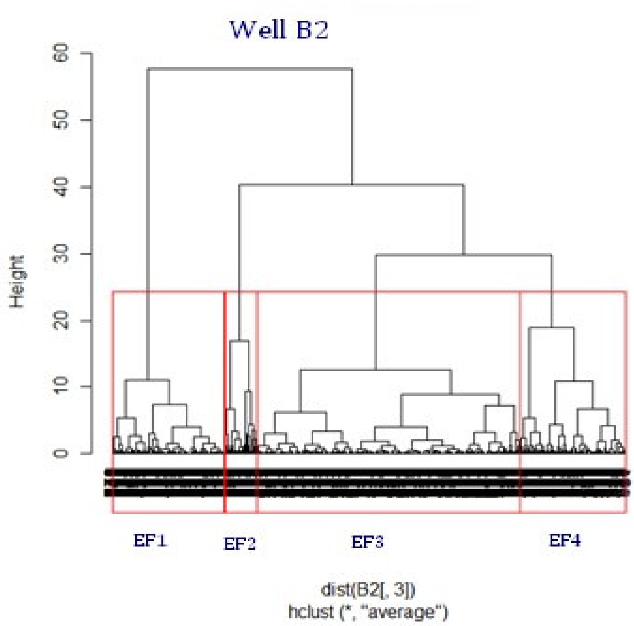


Figure 17 Dendrogram showing clustering scheme resulting in four class. Scale on the left represent multivariate distance and class designations refers to four electrofacies identified within Athabasca Oil sand

The observed gamma ray well log dataset were organized into meaningful structures by the classification algorithm, grouping similar objects to distinguish them from other dissimilar objects on the basis of their measured characteristics. The identification of electrofacies which clearly separate between sand, shale and intermediate sediments following different values of gamma rays logs for each wellbore were recognized. The depths which correspond to each dataset were also noted. The differences in gamma ray log data for each wellbore were computed and the clustering algorithm was also applied to the computed differences. This was done to recognize all the hidden patterns of the gamma ray well log dataset, which ordinarily will not be recognized if only the clustering techniques were done with the original gamma ray dataset. The procedures to achieve the classification of gamma ray well log data are as follows.

i. The datasets of the gamma ray well log (the attributes measurements) were compiled
ii. The clustering algorithm was applied to the similarity matrix as an iterative process to the original gamma ray well log dataset and to the computed difference gamma ray well log dataset
iii. The pairs of object with highest similarities were merged, the matrix was re-computed and the procedures repeat.

All objects are linked together as a hierarchy, shown as a dendrogram. Figure 17 and 18 show the stages of cluster analysis of log data and dendrogram zones for electrofacies classification. It is essential to note that before applying the algorithm, appropriate similarity measures were chosen as described in section 5.2.2.1. All these measures have minimum variance flavor which often give similar results when clusters are compact and well separated. However, if the clusters are close to each other and a different result may be obtained. As stated earlier, the study of similarity measures clearly indicates that average linkage method is more suitable for this particular application where the original and computed difference of gamma ray log data are considered than the single and complete linkage. Therefore, average linkage method was utilized.

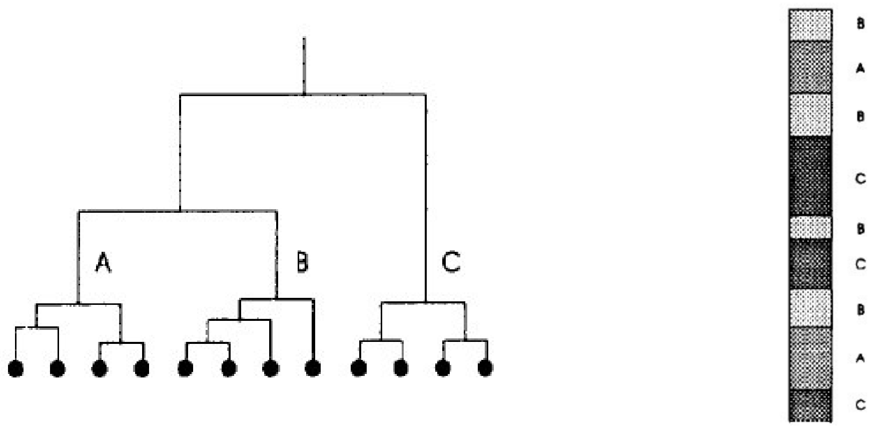
Determining the optimal number of groups in a cluster analysis is crucial. Some objective methods with somewhat arbitrary application have been proposed. These are as follows:

(a) Visually observing for the natural groupings in the data defined by long stem

(b) Definition of clusters at a consistent level of similarity, so that one would draw a line at some chosen level of similarity and all stems that intersect that line would indicate a group.

(c) Definition based on the ratio of within-group variance to total variance

(d) Definition based on the mean of the total mean from all the wellbores.



(A) (B)

Figure 18Stages of Cluster analysis of log data: (a) dendrogram of zones according to hierarchical clustering of zones based on their similarities; (b) classification of zones *(Adapted from Doveton, 1994)*

Sequel to the above, we determine the optimal number of the groups based on the option (b) which supports the assumption that the strength of clustering is indicated by the level of similarity at which elements join a cluster.

Similarity characterization process identified in the above process delineates dominant electrofacies which is referred here as reservoir compartments figure 20. The compartmentalization helps in better understanding of sedimentary structure of the reservoir, where each compartment represents the region with specific sedimentary features that is consistently observed over the locations.

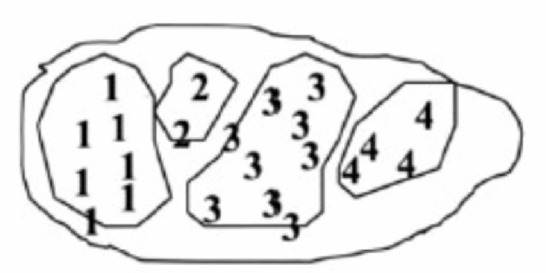


Figure 19 Reservoir compartmentalization

6. IMPLEMENTATION

The details of the methodological framework and the computation processes described in section 5 were implemented using the statistical programming software R. The software provides the clustering functionality that can be applied to a set of well log data.

The well log dataset of 21 wellbores in the depositional environment of the study area as described in section 3 was imported to the software. This section will summarize the programming environment R and discuss the application of the method to the data and results.

R is a programming language for data analysis, manipulation, and graphical representation. It provides the language and environment and packages which includes functions (Venables, et al., 2008). It is a part of GNU project with its source code freely available under the GNU General Public License[2]. Therefore R was chosen for the implementation.

The Packages “sp”, “clusters”, “R.basic”, were utilized for spatial analysis, clustering and graphic visualization.

The subtle variations of lithofacies based on the selection of log data are optimized for characterizing rock types by capturing electrofacies that predicts reservoir complexity. The electrofacies are calibrated using a qualitative interpretation of well logs which helps in filtering a large quantity of data so as to automatically detect potential sandbodies which serve as the reservoir rocks for the bitumen. It is essential to note that the output of the classification will be utilized to delineate prospective areas for in depth geological mapping and interpretation.

6.1. Data Preprocessing and Compilation of Well logs

The well log data was preprocessed into the formats that could be readily used for the implementation in the software. The data comprises well location data and well log data in text files. Well names, surface x-y coordinates and heights are given in the well location file. The log data provided include gamma ray, azimuth and dip which consist of four columns. Each line is a vector of depth, gamma ray, azimuth and dip. The log measurements were taken at approximately every 10cm.

2 www.gnu.org/license

The implementation generates some textual and graphical representation of intermediate and final results which include outputs such as the dendogram for each well (with visualization) and Sand/shale/intermediate sediments electrofacies classification for each well (with visualization).

The dendograms obtained from the aforementioned well log data clusters are presented in Appendix A. For illustration purpose, figure 20 shows the dendogram of a selected three wellbores. The dendograms show the clustering results that depicts four classes in each of the wells

The 21 wellbores were symbolized with an alpha-numeric tags for which they are identified as A1, B2…, U21 representing well 1, 2…and 21 respectively.

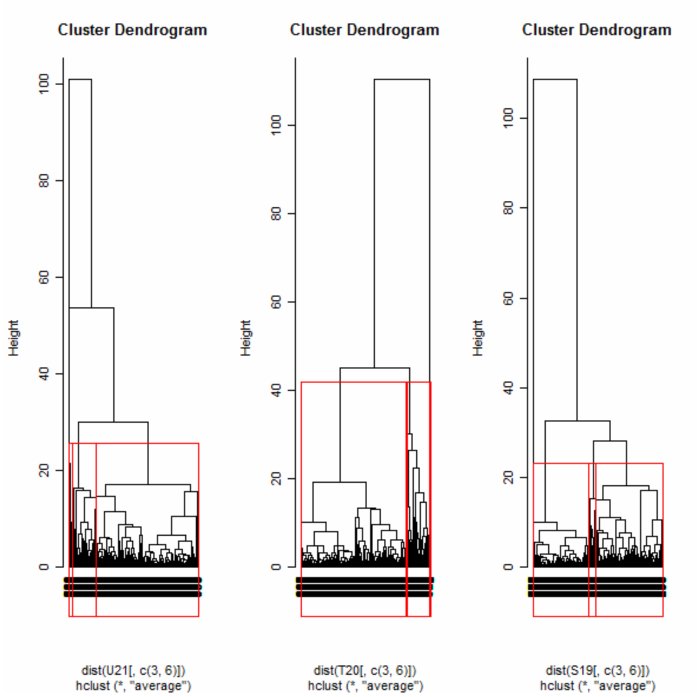


Figure 20 The dendograms of well A1,B2, and C3.

A critical review of the different dendograms generated by all the wellbores reveal the following:

a. The vertical axes of the cluster dendrogram indicate the fusion level with the most similar observations combining first which are at the same level in all dendrograms.
b. At the upper fusion levels, the scales diverge showing the distance among cluster centroids since average linkage was utilized.

c. In all the dendogram in Appendix A, it has been observed that at a similarity value of 10 the number of clusters obtained for the wells is an average of seven clusters.
d. At similarity distance of approximately 80, all the samples are forced to one cluster except in the case of wells F6, G7, and O15 where the value is about 100 while well T20 is about 120.
e. At a similarity distance of approximately 22 (average of all distances), four clusters were obtained in almost all the wellbores. It may be observed here that there are a large number of reductions of clusters in each of the wells. This indicates that the clusters are natural, hence approximately at a distance of 22 four clusters are considered for classification.
f. In some cases where the similarity distance is more or less than the average, the distances were adjusted so that uniformity of four clusters is maintained in all wells.
g. Finally, it may be concluded that in all cases four clusters were chosen for the classification of gamma ray data of the well logs.

The decision to choose four clusters for all the wellbores were based on the nature of the data and the knowledge of the study area's geology, as explained in section 3, where sediments are categorized as Sandstone, Sandy-Shales, Shales and Shale-Sands.

The whole processes were automated by assigning the corresponding parameters to the software. The clustering of the 21 wellbores was carried out automatically, generating 4 clusters in each of the wells and plotting the corresponding dendograms. It took approximately 80 seconds to complete the whole process on a computer processor with AMD Athlon™ 64 X 2 Dual-Core Processor TK-57 1.90GHz, Memory (RAM) 2.00GB and System type 32-bit Operating System. The challenges of this algorithm and the results are described in section 6.3

6.2. Calibration of Electrofacies

The objective of this electrofacies classification was to distinguish between four classes: Sandstone, Sand-Shale, Shale-Sand and Shale. Preferably, the lithology would have been calibrated on core[3] by matching the depth and the rock types with the log data. However, due to unavailability of core samples, the lithology assessments and calibration were based on the qualitative interpretation of well logs over a set of rules. Table 2 depicts north-eastern Alberta

[3] A tabular segment of rocks obtained as a study sample by drilling

well logs suite with their respective measured values, the characteristic rock types and regional stratigraphic unit from where the calibration and interpretation will be based.

Table 2 Stratigraphic Nomenclature and Characteristics of Different Units in Athabasca Wabiskaw-McMurray Succession, Northeastern Alberta. Adapted from Hein and Cotterill (2006)

A	B	C	D	E	F	G	H	I
Age	AGS stratigraphic unit[a]	RGS stratigraphic unit[b]	Gamma ray API units	Neutron porosity %	Density porosity %	Resistivity (ohm-m)	Thickness range (m)	Fissility/Fracture
		Wabiskaw BVF Sand & Shale	60 to 75	up to 45	about 30	20 to 30	5 to 40	none
Middle Albian	Wabiskaw C	Wabiskaw C Sand	60 to 90	up to 30	up to 30	<15	~0.2 to 10	none
		Wabiskaw C Mud	75 to 120	36–45	near 20	5 to 10	0.3 to 5	nonfissile
Early Albian	Wabiskaw D	Wabiskaw D Shale	75 to 90	>30 to >45	27, but variable	2, often < 10	<0.05 to 2	prominent platy
		Wabiskaw D Sand	20 to 30	near 36	near 33	near 100	5 to 9	prominent platy
		Wabiskaw DVF Sand & Shale	30–50, > 70 (shalier)	>30	27, but variable	<10	<0.1 to 2.5	platy in shale
Aptian	Upper McMurray (upper pt)	McMurray A1 Sand	about 75	>30	20 to 30	variable	0.05 to 0.4	
		McMurray A1 Mud	95 to 100	36–45	near 22	variable, >20	0.1 to 0.3	none or octagonal
		McMurray A2 Sand	about 75	>30	20 to 30	variable	0.05 to 0.4	
		McMurray A2 Mud	consistent ~120	36–45	near 22	near 10	1 to 2	none or octagonal
		McMurray B1 Sand	about 75	>30	20 to 30	variable	0.05 to 0.4	
		McMurray B1 Mud	90 to 120	near 45	20 to 30	variable	0.05 to 0.4	none or octagonal
		McMurray B2 Sand	30 to 45	>30	>30	variable	1 to 2	
		McMurray B2 Mud	90 to 120	near 45	near 22	8 to 10	1 to 2	none or octagonal
Aptian	Upper McMurray (lower pt)	McMurray C Channel or McMurray Channel	25 to 45	>30	>30	variable	~10–40 +	none or octagonal
				>30	>30	variable	~10–40 +	none or octagonal
Aptian	Lower McMurray	Undifferentiated		>30	>30	variable	~10–40 +	none or octagonal
Middle	pre-McMurray Sand	Not noted						
Barremian	pre-McMurray Mud	Not noted						

A detail review into the distribution of the original and computed gamma ray values of some selected wells (A1, B2, T20, U21) that are associated with the electrofacies in our case are shown as box plot in Figure 22. The four box plot in each well depicts the data from the clustered well logs were segmented into EF1[4], EF2, EF3 and EF4. The details of the distributions indicate that the electrofacies grouping reflects the nonlinearity between the gamma rays of each clustered well as there was no overlapping of electrofacies in each of the individual well. Outliers are observed as circle in the plots, although it is not conclusive. Hence, the distribution of the data of each electrofacies can be quickly viewed for an informed interpretation. The statistical values are shown in Table 3 this enables the ability to compare between selected wellbores.

[4] EF means Electrofacies

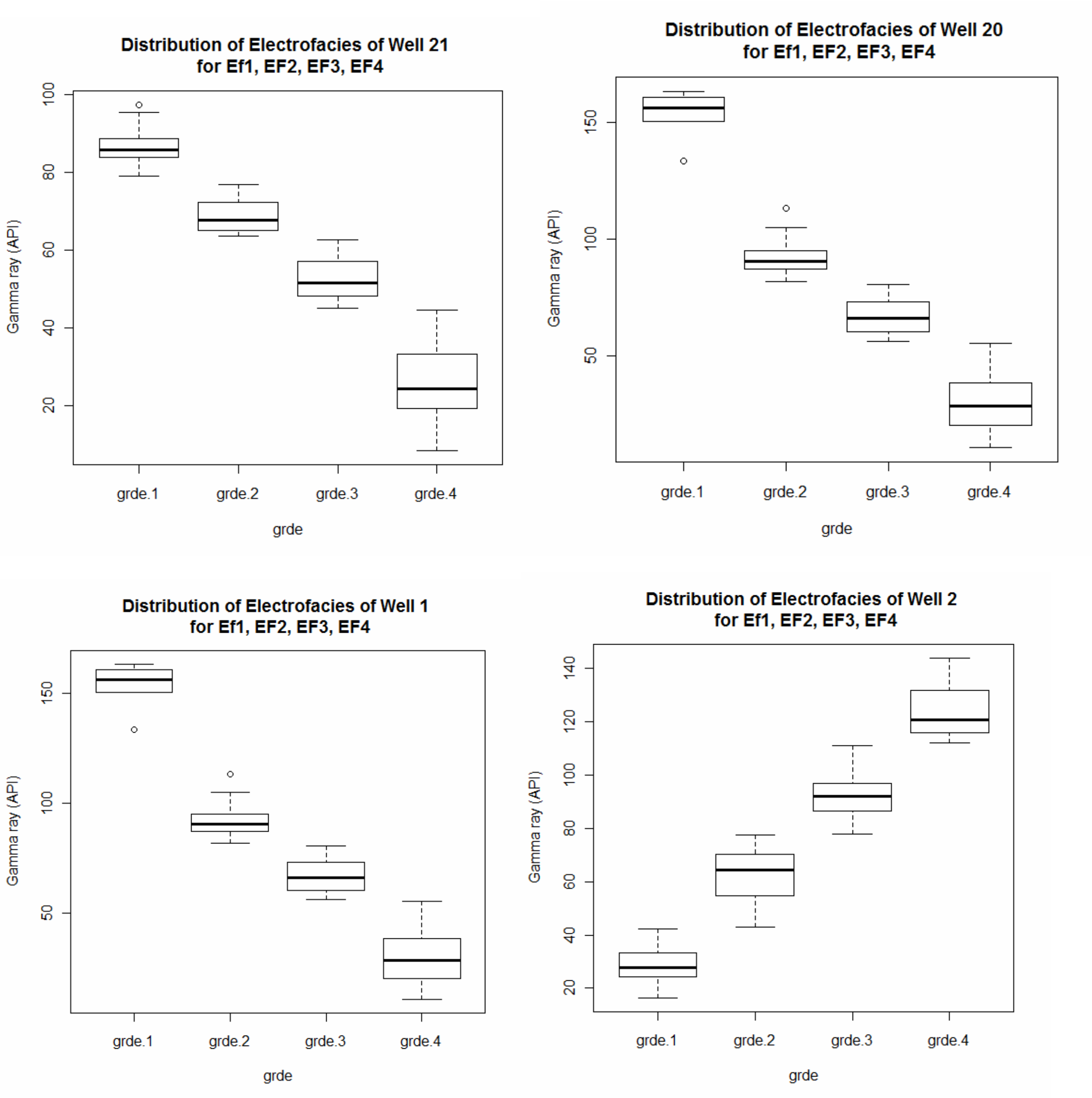


Figure 21 Box plots showing distributions of the gamma ray values within four electrofacies classes determined by clustering on log data. The central box shows the quartile of the data, whiskers indicate range, the median are represented by a line and the outliers are denoted by circle.

Table 3 This shows the statistical values of the computed box plots indicating the values of the lower end of the whisker, the first quartile (25th percentile), second quartile (median=50th percentile), third quartile (75th percentile), and the upper end of the whisker.

Statistical values of Well 21	Statistical values of Well 20
> a<-boxplot(BplotW21, plot=F) > a$stats [,1] [,2] [,3] [,4] [1,] 79.09 63.590 45.050 8.590 [2,] 83.82 65.065 48.370 19.275 [3,] 85.79 67.820 51.580 24.485 [4,] 88.62 72.270 57.285 33.280 [5,] 95.44 76.930 62.750 44.640 > a $stats [,1] [,2] [,3] [,4] [1,] 79.09 63.590 45.050 8.590 [2,] 83.82 65.065 48.370 19.275 [3,] 85.79 67.820 51.580 24.485 [4,] 88.62 72.270 57.285 33.280 [5,] 95.44 76.930 62.750 44.640 $n [1] 26 27 151 612 $conf [,1] [,2] [,3] [,4] [1,] 84.30266 65.62917 50.43372 23.59053 [2,] 87.27735 70.01083 52.72628 25.37947 $out [1] 97.37 $group [1] 1 $names [1] "grde.1" "grde.2" "grde.3" "grde.4"	> b<-boxplot(BplotW20, plot=F) > b$stats [,1] [,2] [,3] [,4] [1,] 150.44 81.650 56.040 10.650 [2,] 150.44 87.130 60.220 20.280 [3,] 156.47 90.630 65.885 28.415 [4,] 160.88 95.015 73.300 38.220 [5,] 163.44 105.020 80.690 55.500 > b $stats [,1] [,2] [,3] [,4] [1,] 150.44 81.650 56.040 10.650 [2,] 150.44 87.130 60.220 20.280 [3,] 156.47 90.630 65.885 28.415 [4,] 160.88 95.015 73.300 38.220 [5,] 163.44 105.020 80.690 55.500 $n [1] 5 55 98 750 $conf [,1] [,2] [,3] [,4] [1,] 149.0931 88.95012 63.79738 27.37998 [2,] 163.8469 92.30987 67.97262 29.45002 $out [1] 133.49 113.39 $group [1] 1 2 $names [1] "grde.1" "grde.2" "grde.3" "grde.4"
Statistical values of Well 1	**Statistical values of Well 2**
> b<-boxplot(BplotW1, plot=F) > b$stats [,1] [,2] [,3] [,4] [1,] 16.430 43.130 77.81 112.21 [2,] 24.350 54.600 86.39 115.98 [3,] 27.805 64.335 92.15 120.88 [4,] 33.440 70.170 97.02 131.68 [5,] 42.120 77.400 110.89 143.90 > b $stats [,1] [,2] [,3] [,4] [1,] 16.430 43.130 77.81 112.21 [2,] 24.350 54.600 86.39 115.98 [3,] 27.805 64.335 92.15 120.88 [4,] 33.440 70.170 97.02 131.68 [5,] 42.120 77.400 110.89 143.90 $n [1] 190 178 444 54 $conf [,1] [,2] [,3] [,4] [1,] 26.76306 62.49111 91.35293 117.5043 [2,] 28.84694 66.17889 92.94708 124.2557 $out numeric(0) $group numeric(0) $names [1] "grde.1" "grde.2" "grde.3" "grde.4"	> c<-boxplot(BplotW2, plot=F) > c$stats [,1] [,2] [,3] [,4] [1,] 16.430 43.130 77.81 112.21 [2,] 24.350 54.600 86.39 115.98 [3,] 27.805 64.335 92.15 120.88 [4,] 33.440 70.170 97.02 131.68 [5,] 42.120 77.400 110.89 143.90 > c $stats [,1] [,2] [,3] [,4] [1,] 16.430 43.130 77.81 112.21 [2,] 24.350 54.600 86.39 115.98 [3,] 27.805 64.335 92.15 120.88 [4,] 33.440 70.170 97.02 131.68 [5,] 42.120 77.400 110.89 143.90 $n [1] 190 178 444 54 $conf [,1] [,2] [,3] [,4] [1,] 26.76306 62.49111 91.35293 117.5043 [2,] 28.84694 66.17889 92.94708 124.2557 $out numeric(0) $group numeric(0) $names [1] "grde.1" "grde.2" "grde.3" "grde.4"

6.2.1. Rule Base

Baker, (1989) proposed some simple rules to recognize lithological zones by visually observing the shape and relationships between different logs such as gamma ray, neutron and density etc. Based on his geologic knowledge of the study area, he constructed a set of rules using a GURU expert system shell. However, neutron porosity and density porosity were among the additional input data with gamma ray data that informed his decision. Nevertheless, this rule can fit in this case study. The rules were useful in recognizing the lithology for log data and it is stated as follows

IF *the gamma ray is less than 50 API*
THEN *the gamma ray is low*

IF *density porosity – neutron porosity is less than 2.5%*
AND *density porosity – neutron porosity is greater than -2.5%*
THEN *porosity difference is equal*

IF gamma ray is low
AND *porosity difference is equal*
THEN *lithology is limestone*

Sequel to the above the distribution of the log data in different electrofacies of different wellbores, Figure 22 reveals that the qualitative interpretation based on the possibilities presented by Baker, (1989); Hein and Cotterill (2006) can be applied to the study data.
In achieving this, the mean of each electrofacies from the clustered gamma ray data in each wellbore were computed. This gives the average value of each electrofacies. The range values which describe the range of the gamma ray distribution in each classification were also computed. Appendix B shows all the values and corresponding ranges for each of the electrofacies respectively. Electrofacies were calibrated based on the mean values, considering the maximum value in a cluster of each group as a factor. The interest as stated earlier is to distinguish between the four lithological classes thereby depicting the sandbodies which serve as reservoir rocks.

In Example 1 (below), six decision rules were constructed for lithology identification. The programme (written in R) was applied when calibrating the electrofacies generated from the clustering performed in each wellbore. It first describes the attributes of the clusters, in which the essential parameters are the mean and maximum values of each cluster. However, the rules would have given preferred results if Bayes classifier based on Bayes formula was implemented in generating the rules.

Bayes classifier has the advantage that the decision is based on the additional information concerning the classification and its conditional probability distributions (Friedman and Kandel, 1999).

Three threshold values were used to assign some rules to encode the lithologic classification. These threshold values are based on the gamma ray values as described in Table 2.

This rule base consists of a number of if/then rules. It has two parts an **IF** (or Condition) part and a **THEN** (or action) parts. The rules are stated as follow

Example 1. Six decision rules for calibrating electrofacies

IF *1. The mean gamma ray is less than **50** API*

THEN *2. The gamma ray is low.*

IF *1. The gamma ray is low and the maximum value in a cluster is less than **45** API*

THEN *2 The gamma ray is **Sand.***

IF *1. The gamma ray is low and the maximum value in a cluster is more than **45** API*

THEN *2. The gamma ray is **Sandy-Shale***

IF *1. The gamma ray is more than **50** API*

THEN *2. The gamma ray is high*

IF *1. The gamma ray is high and the maximum value in a cluster is less than **90** API*

THEN *2. The gamma ray is **Shale-Sand***

IF *1. The gamma ray is high and the maximum value in a cluster is more than **90** API*

THEN *2. The gamma ray is **Shale.***

The above rules were used to calibrate the electrofacies in our study area based on the following

a. The knowledge of the general geology and stratigraphy of the study area.
b. The work of Baker, (1989); Hein and Cotterill (2006) summarized in table 2
c. The intrinsic nature of the gamma ray data.

It is essential to note that a crucial step in the electrofacies calibration is to be certain that the classes defined by the geological interpretation match with the distinct clusters in the multivariate space of the selected logs. However, the challenges posed by inaccurate matching of geological interpretation with the multivariate space of the selected log can be tackled by incorporating other geological analysis like paleontological, petrographic and geochemical analysis. Secondly, considering this case study, the results can be improved with a complete suite of well log data such as Spontaneous Potential (SP), Deep Resistivity (DR), Density (RHO) and Neutron Porosity (NPHI) etc, where different factors based on rock properties and numerical calculations will be considered from there measured characteristics to ascertain a perfect match with the geological interpretation from the well log data. The most important aspect is that the algorithm developed with the above rules can automatically calibrate and classify electrofacies in numerous wells. There are some exceptions which may exist as a result of misclassification in the above set rules. This is because there is a continuum in terms of rock properties between these four classes and this ambiguity is intrinsic in any classification when dealing with continuous phenomena. The problem of misclassification is not addressed in this work, although the likely error of not identifying all the patterns in a clustered wellbore is explained in section 6.3 with the possible solution.

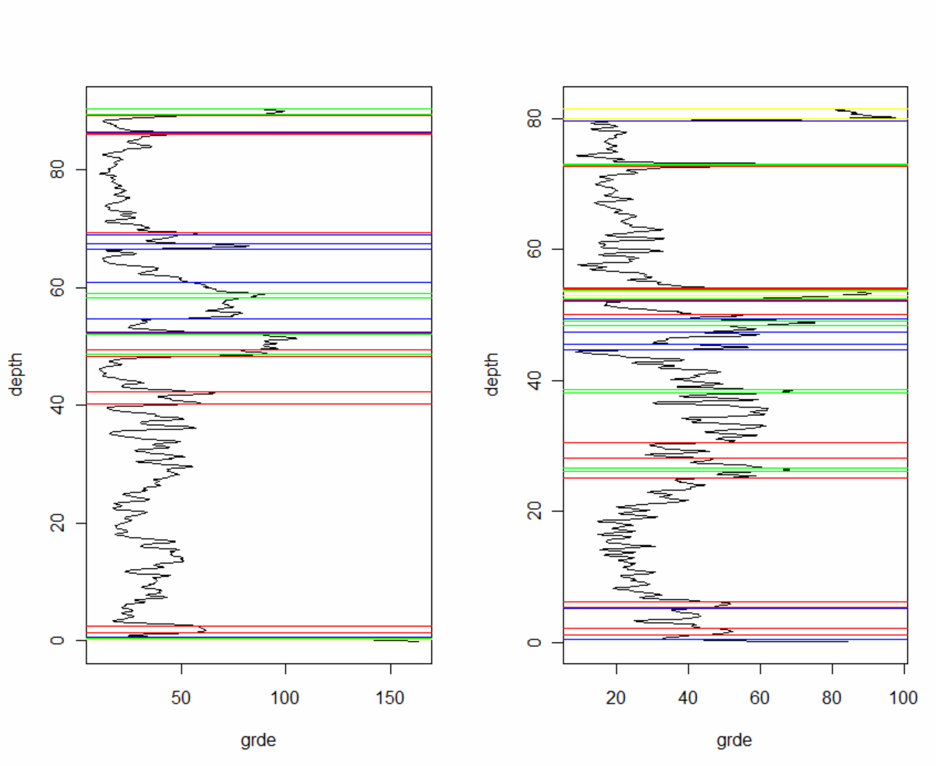
The table below shows an example of the result from the implementation utilizing the above decision rules. This assigns lithology to the clustered well.

Table 4. An example of the result from the implementation using the decision rules

```
> fun=function(x){
+ maxgrde = max(x)
+ meangrde = mean(x)
+ col=c()
+ if ((maxgrde < 45)&&(meangrde<50)){
+   col = rep("sand",length(x))
+ }
+ if ((maxgrde < 45)&&(meangrde<50)){
+   col = rep("sand",length(x))
+ }
+ if ((maxgrde >45)&&(meangrde<50)){
+   col = rep("sand shale",length(x))
+ }
+ if ((maxgrde < 90)&&(meangrde>50)){
+   col = rep("shale sand",length(x))
+ }
+ if ((maxgrde > 90)&&(meangrde>50)){
+   col = rep("shale",length(x))
+ }
+ col
+ }
> fun(w3$grde)
  [1] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [10] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [19] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [28] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [37] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [46] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [55] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [64] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [73] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [82] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
 [91] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[100] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[109] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[118] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[127] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[136] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[145] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
[154] "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale" "shale"
```

A set of wellbores from the study area with log responses typical of the four classes described above were randomly selected for the illustration. Sand, Sand-shale, Shale-sand and Shale were interpreted where a well defined contact was visible based on the set rules. In Figure 22 and 23 the electrofacies profile with the depth interval and the zonation from well log response were defined on the basis of four clusters, as explained above, in two selected wells. The definition depicts the lithoglogy explained in Example 1. The colors used in the interval demarcation are as follows; Yellow, Green, Red, and Blue which represent Electrofacies EF1, EF2, EF3, and EF4 respectively. It was noted that there were less problems in visualization in large stable beds[5] compared to where there were sharp transition between beds especially in thin beds.

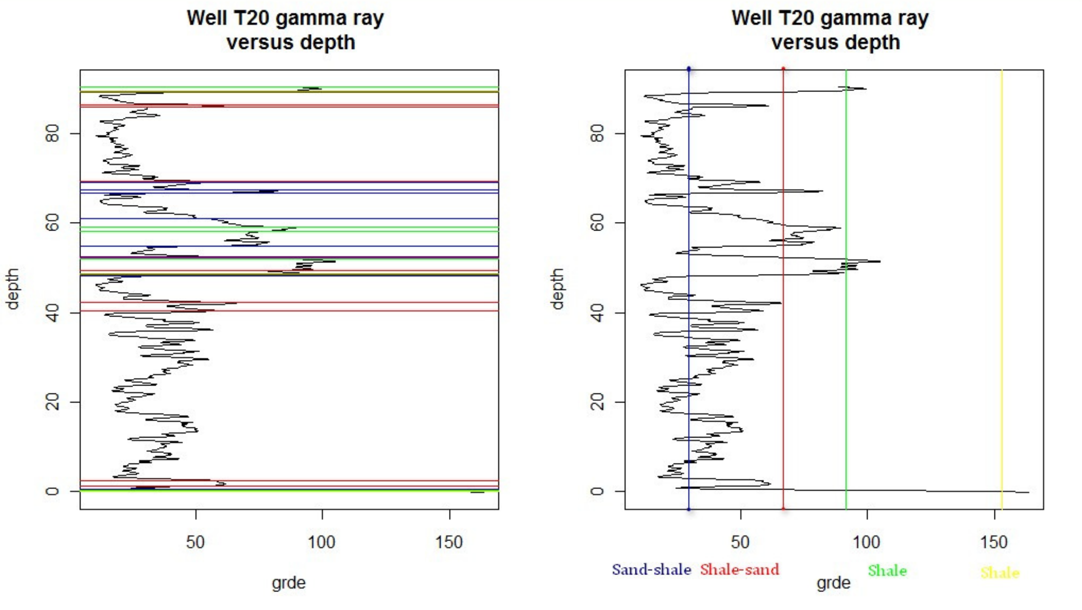
Figure 23b shows the grouping of the lithology identified into segments in the wellbore. The cut-off decision of each segment was based on the mean values, the maximum value in a cluster and the decision rules in Example 1. These have identified three lithology group Sand-shale, Shale-sand Shale in well T20. The representative mean values of the varying electrofacies in each of the wells over the study area is shown in Appendix B



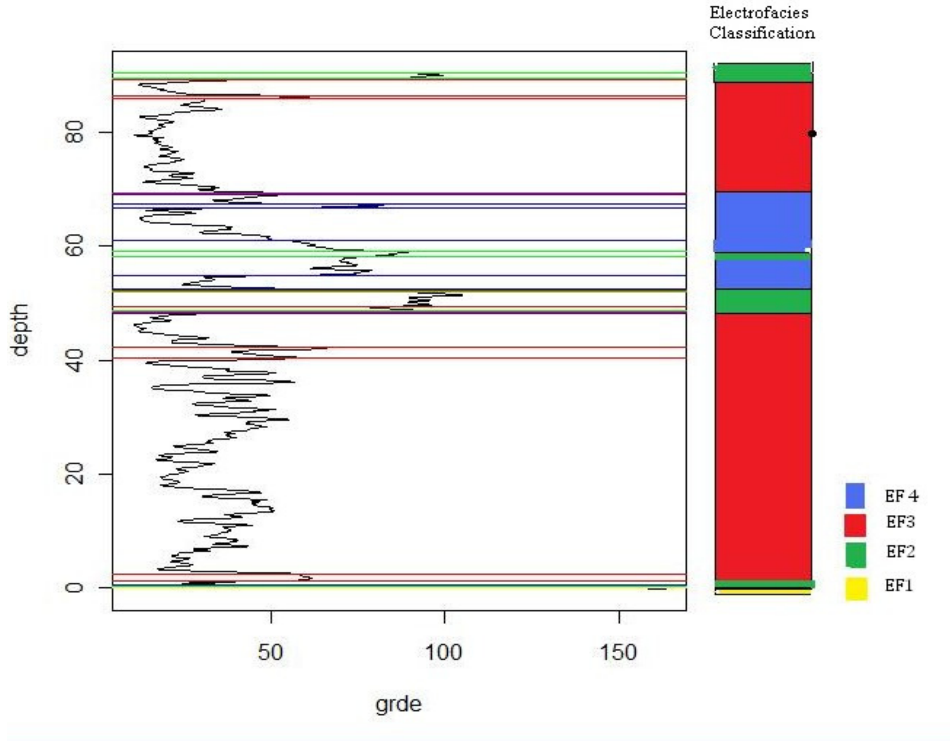
a) b)

[5] A geologic a layer of rock that is generally homogeneous and was deposited more or less continuously without erosion

Figure 22 Well log data with depth intervals for two selected wells. a) Well T20, b) Well U21



a) b)



c)

Figure 23 Profile of four electrofacies at Well T20

6.3. Electrofacies Assignment and Recognition of hidden patterns.

The electrofacies were automatically assigned and calibrated using the set rules by some programmed codes and functions utilizing R programming language to the 21 wellbores in the study area (Table 4). The error of not recognizing the subtle variations in the patterns of a wellbore after applying the clustering algorithm was imperative. A review into figure 24 which depicts the likely structure of the data, reveals that it is possible that the clustering algorithm will ideally recognize the pattern and assign clusters to the values in figure 24a but fail to accurately do the same (match the clusters) in figure 24b because it assumes a different pattern. To tackle this challenge, the differences of the original gamma ray log were computed for all the wellbores. The computations assume this method: If x1, x2, x3, x4…..and so were the original gamma ray values, then the computed difference is x2-x1, x3-x2, x4-x3…….and so. The clustering algorithm was applied to the computed difference. The large circles in figure 24c indicate the likely pattern of the computed differences[6] and the clusters are matched and recognized by the algorithm. This accounts for all subtle variations identified in the pattern of the data.

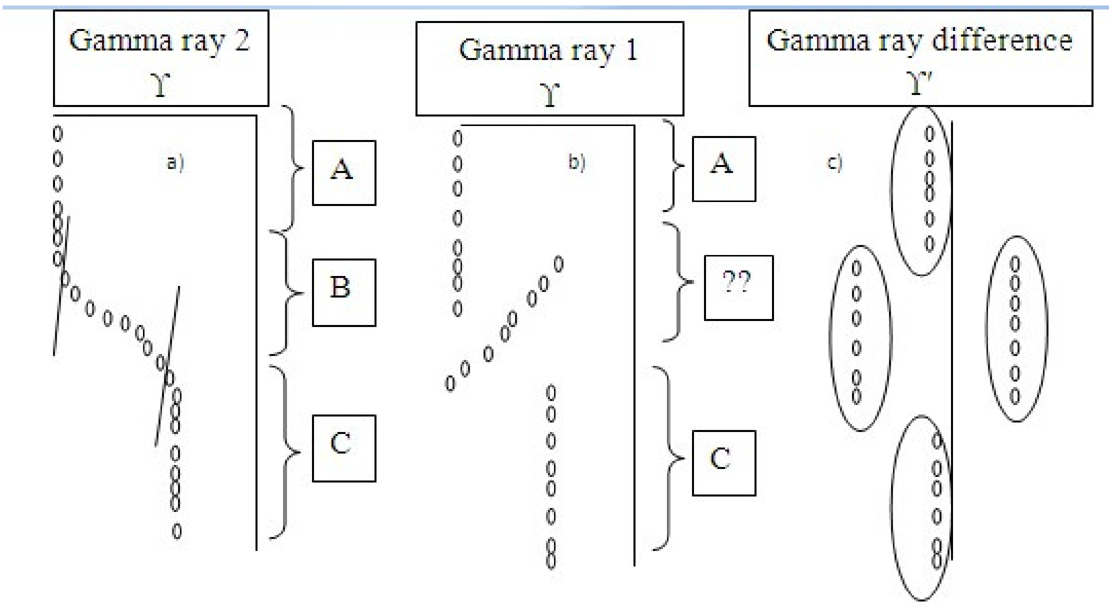


Figure 24 illustrates the possibility of identifying the hidden patterns in the gamma ray data through computing the difference of the gamma ray and applying the clustering algorithm. a) and b) the original gamma ray with diverse pattern c) computed gamma ray difference.

[6] Computed differences of each well have negative and positive values

6.4. Geological evaluation of the clusters and manual interpretation

The above described implementation was carried out on the sample data of the study area. The results were compared with the manual interpretation done by a geologist. This is important in order to evaluate the quality of the various modeling results achieved through automatic interpretation of the wellbores. It is essential to note that the automatic interpretation workflow is divided into two corresponding parts. The clustering of the well logs in the first part and the calibration of the clustered wells in the second part. This calibration assigns the lithology to the clustered electrofacies identified.

The clustering and calibration results were mainly influenced by the optimal number at a level of similarity and the gamma ray threshold values for Sand, Sand-shale, Shale-sand, and Shale respectively.

Firstly, the clustering was generated from the sample data by cutting the hierarchy at different level of similarity in all the wells as explained above until a consistent level was achieved. This was to establish the resulting clusters that were applicable to multiple wells.

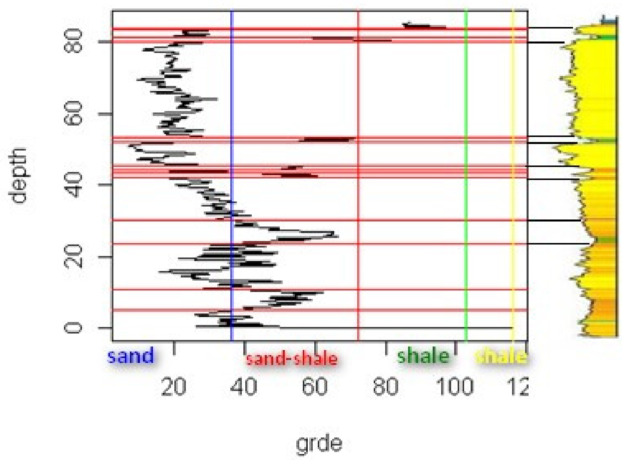
The manual interpretation of the well log data set denotes to define a fixed number of clusters and apply this to all wells in the data set. Therefore, ten and four generated clusters have been tested and the results of the four clusters were found to lead to the classification equal to the manual classification in the wells.

Figure 25 indicates a direct comparison of the manual and the automated classification results for a set of selected wells.

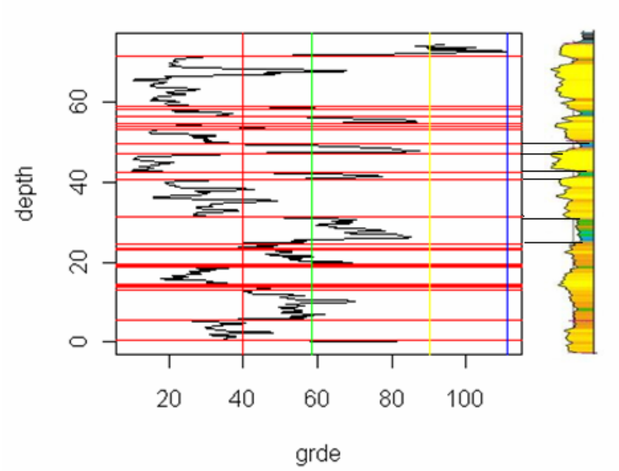
The clusters generated which are described as electrofacies (EF1, EF2, EF3, and EF4) fit perfectly with the manual classification done by a geologist for the wells. Two clusters were chosen for the evaluation of the automatic model, in this case we considered cluster three and four (EF3 and EF4). The idea is to check if the algorithm was able to compute the same solution the geologist found.

In the wells I and II, the automated and the manual interpretation are identical. Thus, the algorithm was able to found the contacts identified in the manual interpretation. In well III, EF3 and EF4 were used to match the manual interpretation. The algorithm identified some contacts correctly. However, there are some errors in proper identifications of all the contacts in the well. This might be due to error during classification (misclassification). In some cases in the wellbores, automatic classification was also identified to be different with the manual interpretation because of the error during classification. Both techniques led to good results to most of the wellbores.

I



II



III

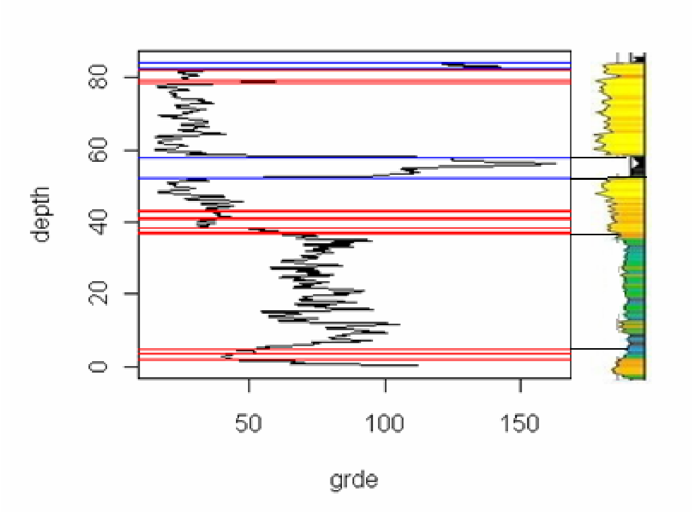


Figure 25. Comparison between automated (left) and manual (right) interpretation of three wells. The horizontal lines mark the identified contacts between the clusters

It is essential to note that the mean and the maximum values of the gamma ray in each cluster of all the wells turned out to be suitable in assigning lithology to the clusters following the developed decision rules. This match the classification of Sand, Sand-shale, Shale-sand, and Shale in the manual interpretation

6.5. Practical Application of electrofacies in exploration.

Electrofacies classification as applied in this study serves as a base in classifying lithology into sand, sand-shale, shale-sand and shale. It identifies potential sandbodies which are the bitumen reservoir rocks in the study area. Its automatic determination tackles the challenges posed by the time consuming manual analysis especially in this study area where the wellbore is more than 250,000. In each stratigraphic sequence, the cumulated thickness of sandstone, sand-shale, shale-sand and shale electrofacies were mapped to identify areas with higher potential of bitumen accumulation. It is important to note that the result of the classifications need to be checked and partly developed further manually as it serve as a first quick assessment of data which gives an avenue to detailed manual interpretation.

Electrofacies classifications form a strong base for correlation. However, due to time constraints, automated correlations are not performed with the result of this electrofacies, but can be implemented on the algorithm developed by Exeler (2009) which explains a topological method in correlating well log data. This author established some connections between sandbodies in this study area with varying parameter such as search distance, search angle around flow direction and slope and sinuosity of the channels. The author developed zonations of the well log data and identified just sand and shale. Hence, incorporating the electrofacies generated which identified Sand, Shale and Intermediate sediments can yield a better correlation and a more informed subsurface model in the study area. Integrating this classification with the trend map of the study area will provide a good basis for selecting prospective areas on which to focus for more detailed work.

7. DISCUSSION AND CONCLUSION

The originality of the methods described in this thesis resides in the automatic classification of stratigraphic column via statistical pattern recognition of the original and computed difference of the gamma ray log data from the wellbores in the study area.

A hierarchical clustering algorithm was implemented in each of the wellbores and these clusters and classifies the wells in four classes that represent the lithologic information of the wells. The likely problems of not identifying the hidden patterns in the data were tackled by computing the gamma ray difference in each of the wellbores. The computed difference of the gamma ray forms a part of the variable that was clustered.

The measurement of similarities in the clusters and dissimilarity between clusters utilized Euclidean (square) distance/likelihood measure and the cluster method used was the Average linkage method. These classes known as electrofacies identify not just sand and shale lithology of the study area but also intermediate sediments. The merits of this approach are as follows:

a. The data are allowed to "speak for themselves" and thereby revealing their subtle differences which are paramount in log response of different rocks.
b. It does not require the pre-specification of the number of clusters. This is important because different geologic environments give varying numbers of classification of log response. Hence, it allows that the numbers of clusters are not pre-defined when applying the algorithm.
c. It has been proved successful in identifying the classification in log response in different geologic environments, evident in the works of Wolff and Pelissier-Combescure, (1982); Delfiner, et al., (1987); Lim, et al., (1997); Lee, et al., (2002) and Lim, (2003). These clusters are easily identified as a hierarchy which is more informative and it gives a single coherent global picture of the data thereby allowing the determination of the desired number of classification.

The method also resides in the use of generated electrofacies as an exploration tool for identifying sandbodies which serve as the bitumen reservoir rocks. As no core was available, the samples were calibrated based on the qualitative interpretation of well logs over a set of decision rules. This application is not meant to be used outright in the field rather it reveals to the expert geologist the general subsurface patterns and classification of litholgy for a more detailed geologic survey. It also saves valuable time in manual analysis.

However, the general challenge/demerit envisaged in the algorithm implemented in this study is the ability of the algorithm to directly address the issue of determining the number of classes within the data i.e. simultaneous determination of the number of clusters and cluster membership in the data.

As explained earlier the optimal number of classes in each wellbore were defined at a consistent level of similarity, by drawing a line at some chosen level of similarity and all stems that intersect that line would indicate a group. In a case of 250,000 wellbore, this approach will be cumbersome in visualizing each of the wellbore and if the classes are fixed to a particular number of classes, a case may also arise that a wellbore will depict either all sand or all shale and the algorithm will generate four classes and classify intermediate sediments.

A univariate classification algorithm was offered as an alternative such as natural breaks, since the key variable of our data is the gamma ray log.

Natural breaks are forms of variance-minimization classification typically uneven, and are selected to differentiate values where significant changes in value occur. The method applied is due to Jenks, as described in Jenks and Caspall (1971). In other words, the method seeks to reduce the variance within classes and maximize the variance between classes. The steps in this classification described by Goodchild and Longley (2006) are as follows

Step 1: The user selects the attribute, x, to be classified and specifies the number of classes required, k

Step 2: A set of k-1 random or uniform values are generated in the range [min{x},max{x}]. These are used as initial class boundaries

Step 3: The mean values for each initial class are computed and the sum of squared deviations of class members from the mean values is computed. The total sum of squared deviations (TSSD) is recorded

Step 4: Individual values in each class are then systematically assigned to adjacent classes by adjusting the class boundaries to see if the TSSD can be reduced. This is an iterative process, which ends when improvement in TSSD falls below a threshold level, i.e. when the within class variance is as small as possible and between class variance is as large as possible. True optimization is not assured. The entire process can be optionally repeated from Step 1 or 2 and TSSD values compared

However, the disadvantages of this method are that:

a. The class ranges are specific to individual data sets
b. The user specifies that number of classes prior to classification. This seems unsuitable in log classifications due to the varying nature of rock log response.
c. Choosing the optimal number of classes is often difficult.

In an attempt to solve the aforementioned challenges in both approaches, the following solutions are proposed

i) The detailed knowledge of the geological settings of the study area helps to understand the general stratigraphic sequence and constitute a strong base for any automatic lithological classification.
ii) Incorporating other geological analysis approach such as paleontological, petrographic and geochemical analysis.
iii) Incorporating a complete well suite data (other geophysical measurement like density, porosity, permeability, spontaneous potential well-log data etc). This will reveal some intrinsic properties of rock materials and form the good base for the algorithm implementation.

In view of the above, Fraley and Raftery (1998) developed a model-based hierarchical clustering algorithm that attempt to address the issue of directly determining the number of groups within the data. This will be a focus on the future research and a comparative analysis is recommended to be made between the model-based hierarchical clustering with the conventional hierarchical clustering algorithm for a more informed classification.

Finally, based on the research hypothesis and question, it can be concluded that the patterns from the well logs are recognizable. The whole process is automated based on a set of decision rules and the results forms the basis of correlation between wells in the study area.
The comparative analysis between the automatic and the manual interpretation results indicated that the algorithm was able to compute exactly the same solution that the geologist found.

8. REFERENCES

Agterbsrg. F.P and Bonhan-Carter, G.F., Eds 1990, Statistical applications in the earth science: Geol. Surv. Can., Paper 8909. 588p

Agterbsrg F.P and Griffith. CM.. ,1991, Computer application in Stratigraphy 1989/1990: a review Computer and Geosciences 17.1105-1118.

Alberta Department of Energy (ADE), 2006, Alberta's oil sands 2006

Annxionnaz, H.; Delfiner, P., and Delhomme, J. P., 1990. Computer-generated corelike descriptions from open-hole logs: *Am. Assoc. Petroleum Geologists Bull., v. 74. no. 4, p. 375-393.*

Baker P.L, 1989, Log Analysis System Incorporating Artificial Intelligence Techniques – Some Examples. Earth Science Department Monash University, Clayton Vic, http://www.publish.csiro.au/?act=view_file&file_id=EG989331.pdf [Accessed 27/10/2010]

Bezdek, J.C., 1981. Pattern Recognition with Fuzzy Objective Function Algorithm. Plenum Press, New York, NY.

Cutler, W.G., 1982, Stratigraphy and sedimentology of the Upper Devonian Grosmont Formation, northern Alberta: Bulletin of Canadian Petroleum Geology, v. 31, p. 282-325.

C&C Reservoirs., 2007, .Reservoir Evaluation Report North America, Athabasca Oil Sands Area, Western Canada Sedimentary Basin, Canada,

Delfiner, P., Peyret, O., and Serra, O., 1987. Automatic Determination of Lithology from Well logs. *SPE Annual Technical Conference and Exhibition,* Houston, TX, September, 16-19, Paper SPE 56658

Duda, R.O and Hart, P.E., 1973, Pattern classification and scene analysis, New York, Wiley-Interscience, 211-236pp

Duda R. O. and Hart P. E.,1996, *Pattern Classification and Scene Analysis,* New York: John Wiley & Sons,(Preliminary Edition)

Elek, I., 1988. Some application of principal component analysis: well-to-well correlation, zonation. *Geobytes, pp.46-55*

Euzen, T., Delamaide, E., Feuchtwanger, T., IFP Technologies, K.D Kingsmith Resource Consultants, 2010, Well Log Cluster Analysis: An Innovative Tool for Unconventional Exploration. Paper CSUG/SPE 137822 presented at the Canadian Unconventional Resources & International Petroleum Conference held in Calgary, Alberta, Canada, 19-21 October 2010.

Flach, P.D., and Mossop, G.D., 1985, Depositional environments of Lower Cretaceous McMurray Formation, Athabasca Oil Sands, Alberta: AAPG Bulletin, v. 69, p. 1195-1207.

Flach, P. D., and Hein, F. J., 2001, Outcrop-core correlation of channel and non-channel facies, McMurray Formation, Fort MacKay area, NE Alberta: Rock the Foundation Conf., Proc. Can. Soc. Petroleum Geologists (Calgary), p. 132–133.

Fraley, C., and Raftery, A., 1998, How Many Clusters? Which Clustering Methods? Answers Via Model-Based Cluster Analysis. *The Computer Journal, Vol, 41, No.8,*

Gelbard, R., Goldman, O. and Spiegler, I., 2007, Investigating diversity of clustering methods: An empirical comparison: *In Data & Knowledge Engineering 63 (2007) 155-166*

Gill, D., Shomrony, A., and Fligelman, H., 1993, Numerical zonation of log suites and logfacies recognition by multivariate clustering. AAPG Bulletin, 77(10):1781{1791, 1993.

Goodchild, M., and Longley, P., 2006, Geospatial Analysis - A Comprehensive Guide. 3rd edition © 2006-2011 de Smith. Available at http://www.spatialanalysisonline.com/output/html/Univariateclassificationschemes.html [Accessed on 22nd February, 2011]

Gnanadesikan, R., 1977 Methods for Statistical Analysis of Multivariate Observations, Wiley & Sons, New York City *pp. 82-120*

Gradstein, F. M., Agterberg, F. P., Brower, J. C,, and Schwarzacher, W. S.. 1985, Quantitative Stratigraphy: D. Reidel Publishing Co., Dordrecht, Holland, 393 p

Hallam, R.J., Hajdo, L.E., Donnelly, J.K., and Baron, R.P., 1989, Thermal recovery of bitumen at Wolf Lake: SPE Reservoir Engineering, May, p. 178-186.

Hassibi, M., Ershaghi, and Aminzaeh, F., 1996. High Resolution Reservoir Heterogeneity Characterization Using Recognition Technology: *Development in Petroleum Science, Vol. 51 (471-497) Elsevier Science B.V*

Hebb D. O., 1949, *The Organization of Behavior*, New York: Wiley Inc.,
Rosenblatt F., 1958 "The Perceptron: A Probabilistic Model For Information Storage and Organization in the Brain" *Psychological Review* vol. 65, pp. 386-408

Hein, F., and Cotterill, D., 2006, The Athabasca Oil Sands—A Regional Geological Perspective, Fort McMurray Area, Alberta, Canada. *Natural Resources Research, Vol. 15, No. 2,*

HILLS L. V. (1974) Oil sands of the future. Canadian. Society. *Petroleum .Geologists Mem. 3.*

Hoyle. I. B., 1986, Computer techniques for the zoning and correlation of well-logs.
Geophysical prospecting, 34(5):648{664, 1986.

Huang Z. and Williamson M.A., 1994, Geological Pattern Recognition and Modelling with A General Regression Neural Network: *Canadian Journal of Exploration Geophysics Vol.30, 1 60-68p*

Jang, J.S.R. and Gulley, N., 1995. Fuzzy Logic Toolbox. The Math Works Inc., Natick, MA.

Jenks, G. F., and Caspall, F. C., 1971. Error on choroplethic maps: Definition, measurement, reductions. Annals of the Association of American Geographers, Volume 61, Number 2, Pages: 217 – 244.

Jolliffe, I. 2005. Principal Component Analysis. Encyclopedia of Statistics in Behavioral Science

Kandel E., Schwartz J., Jessell T., 1991, *Principles of Neural Science,* New Jersey: Prentice Hall Inc.

Kohonen, T., 1997. Self-Organizing Maps, 2nd edn. Springer, Berlin.

Keith, D.A.W., Wightman, D.M., Pemberton, S.G., MacGillivray, J.R., Berizniuk, T., and Berhane, H., 1988a, Fluvial, estuarine and shallow marine sedimentation in the Lower Cretaceous McMurray Formation and Wabiskaw Member (Clearwater Formation), in the central region, Athabasca Oil Sands area, northeastern Alberta: 4th UNITAR/UNDP Conference on Heavy Crude and Tar Sands, paper no. 130, 32 p.

Lee, S. H. Kharghoria, A., and Datta-Gupta, A., 2002. Electrofacies characterization and permeability prediction in carbonate reservoirs: role of multivariate analysis and nonparametric regression. *SPE Annual Technical Conference and Exhibition,* Houston, TX, Oct. 3-6, Paper SPE 56658

Lim, J., Kang, J., Nati, S., and Kim, J., 1997, Multivariate Statistical Analysis for Automatic Electrofacies Determination from Well Log Measurements. SPE 38028 paper presented at SPE Asia Pacific Oil and Gas Conference held in Kuala Lumpur, Malaysia 14-16 April 1997

Lim, J., 2003, Multivariate Statistical Techniques Including PCA and Rule Based Systems for Well log Correlation: *Development in Petroleum Science, Vol. 51 (673-688) Elsevier Science B.V*

McCulloch W. S., Pitts W. 1943, A Logical Calculus of the Ideas Immanent in Nervous Activity *Bulletin of Mathematical Biophysics*, vol. 5, pp. 115-133

Mohaghegh, S., Arefi, R., Ameri, S., Aminiand, K. and Nutter, R., 1996, Petroleum reservoir characterization with the aid of artificial neural networks: *Journal of Petroleum Science and Engineering, 16(1996) 263-274.*

Nandhakumar, N., and Aggarwal, J., 1985, The Artificial Intelligence Approach to PatternRrecognition--a Perspective and an Overview in Pattern recognition. Vol 18, No16 383-389p

Nicholls J. G., Martin A. R., Wallace B. G., 1992, From Neuron to Brain, Sunderland: Sinauer Ass. Inc.

Nikravesh, M., Adams, R., and Levey R., 2003 Soft Computing, Tools for Intelligent Reservoir Characterization and Optimum Well Placement: *Development in Petroleum Science, Vol. 51 (471-497) Elsevier Science B.V*

Nikravesh, M. and Aminzadeh, F., 2003 Soft Computing for Intelligent Reservoir Characterization and Modelling: *Development in Petroleum Science, Vol. 51 (3-48) Elsevier Science B.V*

Norsk Geologisk Forening, 2009 Available at **http://www.geologi.no/data/f/0/16/49/9_2401_0/Geology_of_the_Athabasca_Oil_Sands.pdf**

Panigrahi, D.C. and Sahu, H.B., 2004, Application of hierarchical clustering for classification of coal seams with respect to their proneness to spontaneous heating: *Mining Technology (Trans. Inst. Min. Metall. A) Vol 113 (97-106)*

Pao. Y.H., 1989. Adaptive Pattern Recognition and Neural Networks. Addison-Wesley, New York, NY.

Pavlidis, T., 1977 *Structural Pattern Recognition*. Springer-Verlag, New York

Pitts W., McCulloh W. S., 1947 "How We Know Universals: The Perception of Auditory and Visual Forms" *Bulletin of Mathematical Biophysics*, vol. 9, pp. 127-147

Price, R.A., 1994, Cordilleran tectonics and the evolution of the Western Canada Sedimentary Basin, in Mossop, G.D., and Shetsen, I., compilers, Geological atlas of the Western Canada Sedimentary Basin: Canadian Society of Petroleum Geologists and Alberta Research Council, Calgary, http://www.ags.gov.ab.ca/publications/ATLAS_WWW/ATLAS.shtml. [Accessed 9/11/10]

Schaefer, A., 2005, Klastische Sedimente. Elsevier, Spektrum Akad. Verl., 2005.

Serra. O 1984.Fundamentals of well-log interpretation. Elsevier,

Serra, O. and Abbott, H.T., 1982. The contribution of logging data to sedimentology and stratigraphy. *Soc. Pet. Eng. J., pp. 117-131*

Startzman, Richard A., Texas A and M U.; Kuo, Tsai-Bao, Texas A and M U, 1987, A Rule-Based System for Well Log Correlation: *Journal SPE Formation Evaluation Volume 2, Number 3 311-319pp*

Shanmugam, G., Moiola, R.J., McPherson, J.G. and O'Connell, S., 1988. Comparison of modem Mississippi fan with selected ancient fans. *Trans. Gulf Coast Assoc. Geol. Soc.,* 38: 157-165.

Simaan. M. and Aminzadeh F.. Eds.. 1989 Artificial Intelligient and expert system in petroleum exploration: Advance in Geophysical data processing Vol 3 JAI Press Inc.

Telford, W.M., Geldart, L.P. and Sheriff, R.E., 1990. Applied Geophysics, 2nd edition. Cambridge University Press.

Tipper, J. C.. 1988, Techniques for Quantitative Str'atigraphic Correlation--A Review and Annotated Bibliography: Geological Magazine, v. 125. n. 5, p. 474-494.

Tsai-Bao Kuo. 1986.Well log correlation using artificial intelligence (Rule-based, Expert System, Computer-assisted). PhD thesis,

Venables, W.N., Smith, D.M., and The R Development Core Team., 2008, An Introduction to R:

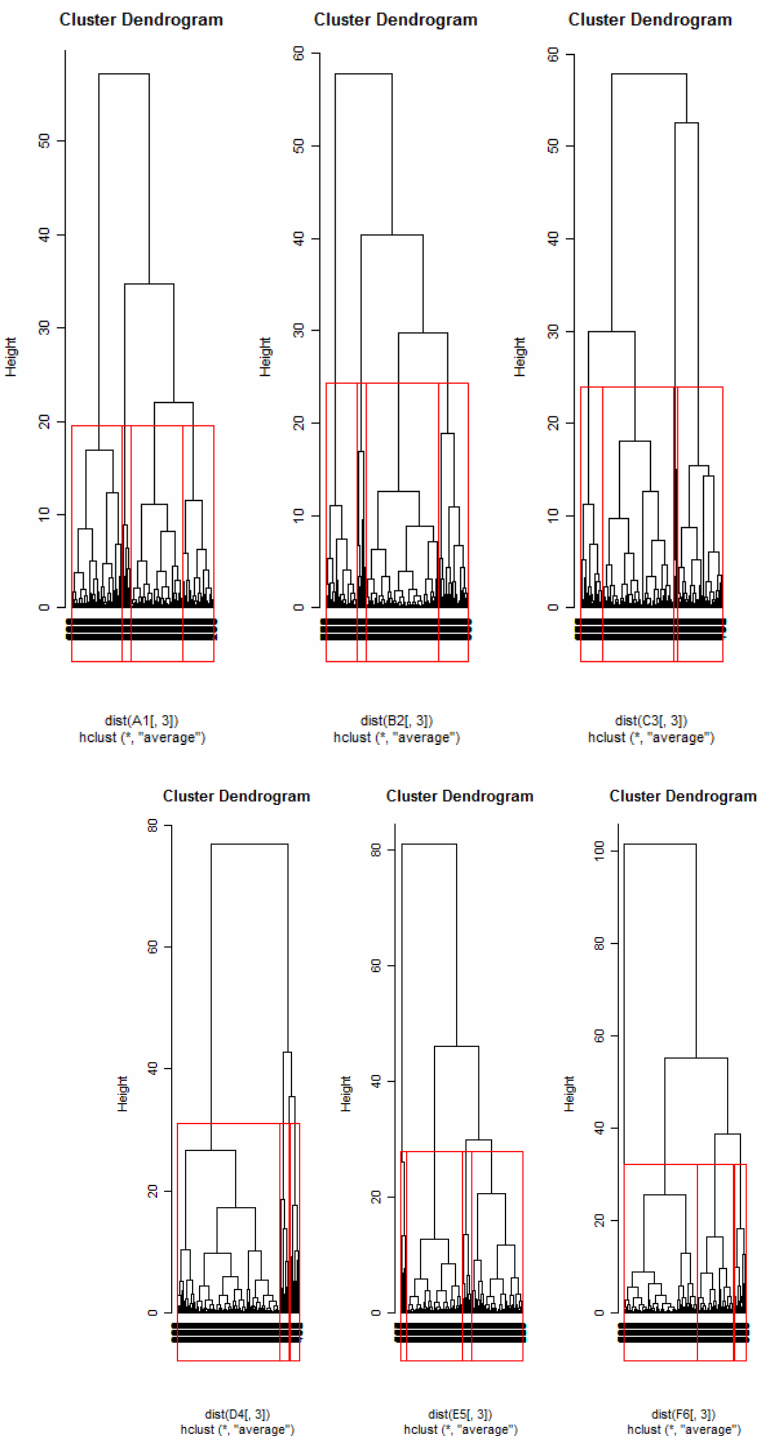
A Programming Environment for Data Analysis and Graphics Version 2.8.1

Wright, G.N., McMechan, M.E., Potter, D.E.G., 1994, Structure and architecture of the Western Canada Sedimentary Basin, in Mossop, G.D., and Shetsen, I., compilers, Geological atlas of the Western Canada Sedimentary Basin: Canadian Society of Petroleum Geologists and Alberta Research Council, Calgary, http://www.ags.gov.ab.ca/publications/ATLAS_WWW/ATLAS.shtml. [Accessed 10th January, 2011]

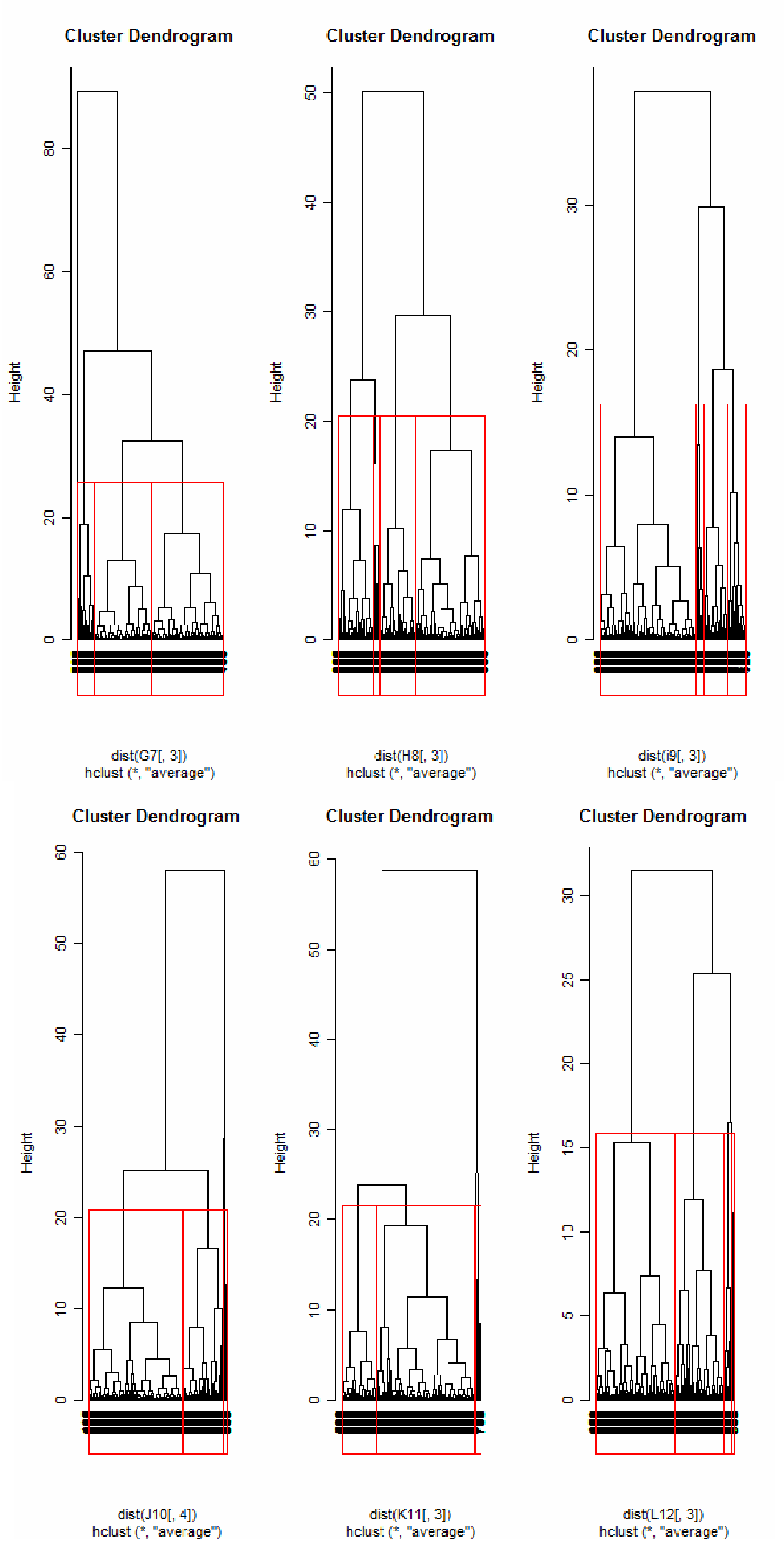
Wu, X., and Nyland, E., 1987, Automated stratigraphic interpretation of well-log data*: J, Geophysis,* **52(12), 1665-1675**

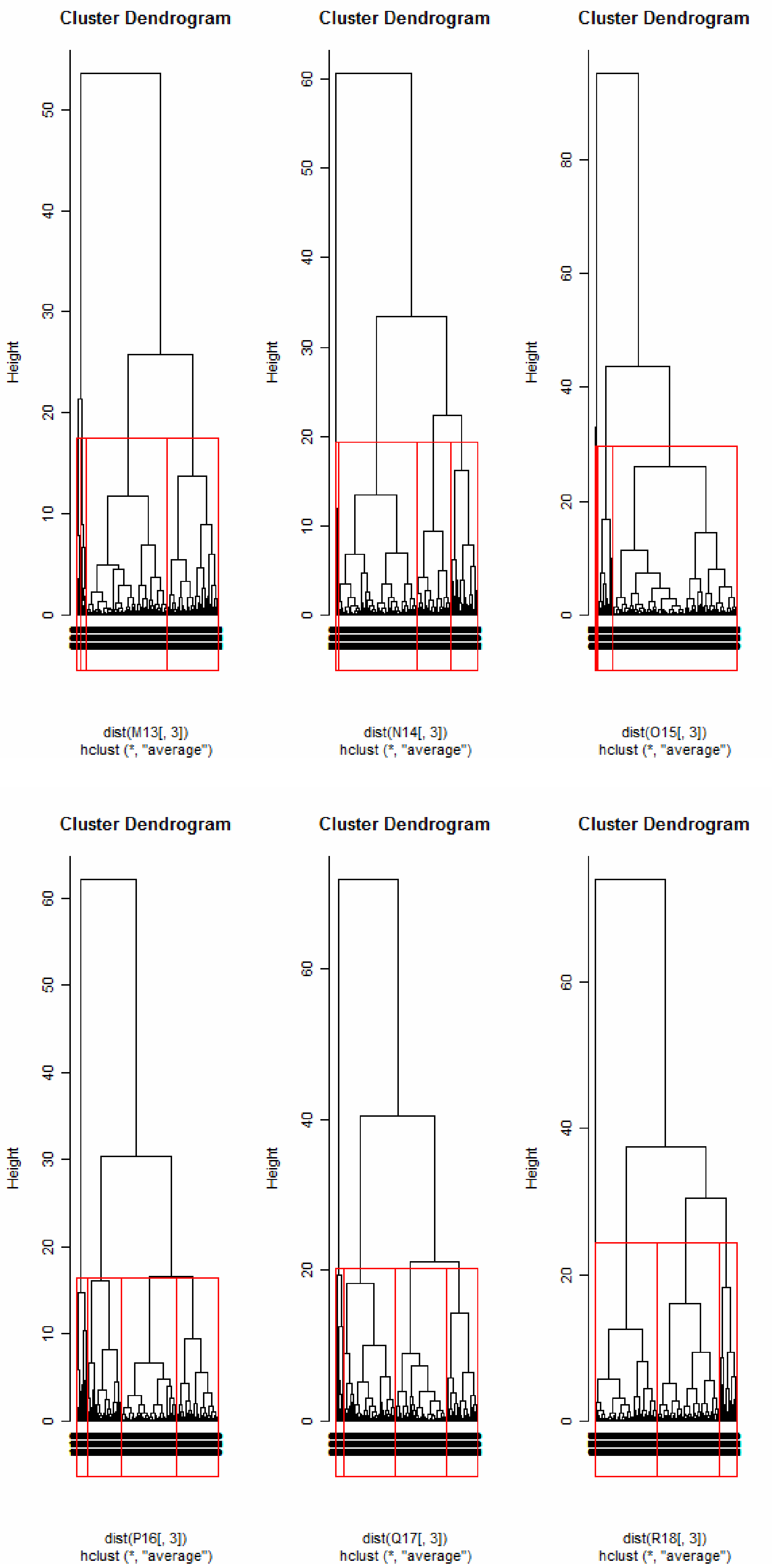
Wolff', M., and Pelissier-Combescure, J., 1982. Faciolog--automatic electrofacies determination, in Transactions, 23rd annual logging symposium, paper FF, Society of Professional Well Log Analysts, Houston, 23 p.

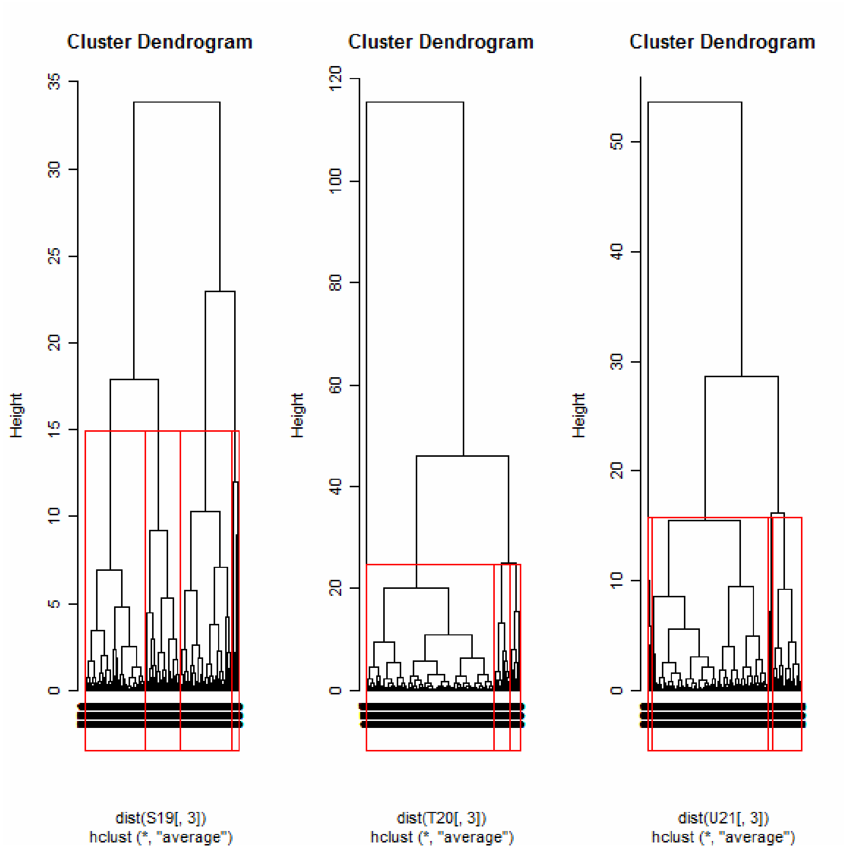
APPENDIX A



Cluster Dendrogram
Height
dist(A1[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(B2[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(C3[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(D4[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(E5[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(F6[, 3])
hclust (*, "average")


Cluster Dendrogram
Height
dist(G7[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(H8[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(i9[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(J10[, 4])
hclust (*, "average")
Cluster Dendrogram
Height
dist(K11[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(L12[, 3])
hclust (*, "average")


Cluster Dendrogram
Height
dist(M13[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(N14[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(O15[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(P16[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(Q17[, 3])
hclust (*, "average")
Cluster Dendrogram
Height
dist(R18[, 3])
hclust (*, "average")


Cluster Dendrogram
Cluster Dendrogram
Cluster Dendrogram
Height
Height
Height
dist(S19[, 3])
hclust (*, "average")
dist(T20[, 3])
hclust (*, "average")
dist(U21[, 3])
hclust (*, "average")

APPENDIX B

Well U21					
			Cluster Number	Average Gr Value (Mean)	Min & Max (Range)
			1	86.5573	79.09 97.37
			2	68.85333	63.59 76.93
			3	52.71947	45.05 62.75
			4	26.51753	8.59 44.64
Well T20					
			1	152.944	133.49 163.44
			2	91.72036	81.65 113.39
			3	66.7397	56.04 80.69
			4	29.77443	10.65 55.50
Well S19					
			1	74.98778	67.53 90.89
			2	52.01346	40.83 66.67
			3	32.17957	22.53 40.51
			4	14.25871	4.23 22.29
Well R18					
			1	48.84613	34.57 65.33
			2	17.82185	6.23 34.02
			3	79.3171	65.70 104.65
			4	112.748	109.83 115.72
Well Q17					
			1	120.8937	102.52 143.23
			2	74.07846	61.15 98.03
			3	46.75461	36.18 60.50
			4	25.56304	13.82 35.79

Well P16					
			1	120.8937	102.52 143.23
			2	74.07846	61.15 98.03
			3	46.75461	36.18 60.50
			4	25.56304	13.82 35.79
Well O15					
			1	142.502	126.35 156.99
			2	109.4333	99.59 120.35
			3	73.95608	62.90 96.07
			4	30.36205	7.63 62.34
Well N14					
			1	71.7705	58.96 90.26
			4	49.33184	40.1 58.5
			3	25.73789	9.89 39.80
			4	100.8473	92.07 111.10
Well M13					
			1	48.47233	36.70 66.99
			2	22.65931	10.74 36.45
			3	73.53321	67.93 81.48
			4	94.9075	84.16 103.72
Well L12					
			1	88.04286	81.41 96.02
			2	71.57667	66.97 79.63
			3	50.31506	38.37 65.94
			4	23.23514	8.54 37.99
Well K11					
			1	39.73314	23.32 66.55
			2	15.83508	7.00 23.21
			3	85.29667	72.31 96.68
			4	110.4311	100.00 120.01
Well J10					
			1	116.09	116.09 116.09

			2	87.50579	77.08 102.90
			3	48.35233	36.81 72.68
			4	23.22054	6.69 36.61
Well I9					
			1	23.64029	8.95 40.44
			2	48.66513	40.84 57.24
			3	67.30424	57.92 78.12
			4	86.83297	79.38 101.49
Well H8					
			1	103.9209	92.82 122.02
			2	80.12142	66.26 92.17
			3	53.96439	43.44 65.77
			4	24.25281	9.12 42.80
Well G7					
			1	88.57747	74.38 111.66
			2	55.90752	37.81 73.42
			3	23.45217	11.75 37.42
			4	136.4933	133.10 141.75
Well F6					
			1	120.2275	105.73 146.19
			2	81.51292	65.60 104.11
			3	35.93195	16.95 65.22
			4	159.132	154.55 164.64
Well E5					
			1	100.0767	89.65 112.70
			2	70.32242	45.96 89.25
			3	29.32306	15.32 45.47
			4	134.1986	116.06 162.63
Well D4					
			1	134.8376	114.41 157.03
			2	95.04333	77.39 112.04
			3	40.12844	12.47 74.59
			4	170.326	163.30 174.99
Well C3					
			1	91.30409	75.04 116.49
			2	60.26241	49.62 72.69

			3	30.35025	7.69 49.14
			4	143.8575	125.72 152.60
Well B2					
			1	28.859	16.43 42.12
			2	62.22983	43.13 77.40
			3	91.96489	77.81 110.89
			4	123.8133	112.21 143.90
Well A1					
			1	32.04702	14.63 55.64
			2	72.13055	56.44 81.86
			3	94.19606	82.29 108.70
			4	120.6119	111.20 130.17

Appendix C

`hclust`(*dmat,* **method** = "average"), where *dmat* is an object created by **`dist()`.** The optional command method gives the method used for clustering.

To plot the resulting dendogram, the following syntax was utilized.

plot(*clust,* **labels** = *NULL*, **hang** = *-1*), where *clust* is an object created by **`hclust()`.** The command **label** permits the specification of a vector containing names for the value being clustered whereas the command **hang** specifies where the labels are to be placed. A negative value aligns the labels just below a distance value of zero and this is mostly preferred option