AN EMPIRICAL CORRELATION OF PERMEABILITY FOR THE KHUFF FORMATION

BY

Muhammad Hasan

A Thesis Presented to the
DEANSHIP OF GRADUATE STUDIES

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE

In

PETROLEUM ENGINEERING

December 2017

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN- 31261, SAUDI ARABIA

DEANSHIP OF GRADUATE STUDIES

This thesis, written by **MUHAMMAD HASAN** under the direction of his thesis advisor and approved by his thesis committee, has been presented and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN PETROLEUM ENGINEERING.**

Dr. Sidqi A. Abu-Khamsin
(Advisor)

Dr. Dhafer Abdullah Al Shehri
Department Chairman

Dr. Michael Fraim
(Member)

Dr. Salam A. Zummo
Dean of Graduate Studies

Dr. Abdulazeez Abdulraheem
(Member)

8/1/12

Date

Dedication

I dedicate my work to my parents, sisters and family who have been supporting and encouraging me in all my endeavors.

ACKNOWLEDGEMENTS

Thanks to Almighty Allah for giving me strength to complete this thesis. Some heartfelt thanks goes to my family for their endless support, patience and sound advice at every stage of my Master's program.

I express my sincere gratitude to my advisor, Dr. Sidqi Abu-Khamsin for generously sharing his wealth of knowledge and experience. I would also like to thank my committee member Dr. Abdulazeez Abdulraheem for dedicating tireless efforts towards this study. Special thanks goes to my committee member Dr. Michael Fraim for his insightful comments and discussions.

Special thanks are due to the Department of Petroleum Engineering at King Fahd University of Petroleum & Minerals, including Faculty, Staff and Fellow Graduate Students who contributed directly or indirectly to the accomplishments of this work.

I would like to acknowledge the efforts of Mr. Muhammadin for assisting me during the days of petro-physics lab work, Mr. Abdulsamed Iddris, Mr. Ali and Mr. Musa for providing any possible help during the lab work, Dr. Mohamed Mahmoud for providing cores, Dr. David Abiola Obembe for his constant support. I am extremely thankful to Mr. Najmudeen Sibaweihi for his support with tech log and matlab programs.

Thanks to my friends Talha Mirza Baiq, Arqam Muqtadir, Sarim jamal, Zeeshan Tariq, Ahmed Sadeed, Wasay Raza, Shams Kalam, Muzammil Hussain Ramay, Sarmad Zafar

Khan, Danish Hashmat for their support and guidance, without which this thesis would be incomplete.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

FCM	:	Fuzzy c-means algorithm
KM	:	K-means algorithm
EM	:	Expectation maximization algorithm
AI	:	Artificial Intelligence
KC	:	Kozeny and Carman
NMR	:	Nuclear magnetic resonance
FFI	:	Free flow index
BVI	:	Bulk volume irreducible
FZI	:	Flow-zone indicator
RMSE	:	Root means square error
AAD	:	Average absolute difference

ABSTRACT

Full Name : Muhammad Hasan

Thesis Title : An Empirical Correlation of Permeability for the Khuff formation.

Major Field : Petroleum Engineering

Date of Degree : December 2017

Carbonate reservoirs usually have quite a complicated texture with a high degree of heterogeneity. They can have large aerial extent but with very low porosities. Some strata have low matrix permeability, while others have natural fractures. For this reason, well log evaluation techniques applicable to sandstone reservoirs usually fail in carbonate reservoirs. In the Middle East, the economic significance of oil and gas production from carbonate reservoirs is huge, especially due to gigantic fields present in the region.

This research focuses on the Khuff formation of the Middle East which has one of the largest accumulation of gas in the world. The major mineral constituents of Khuff formation are dolomite, limestone and anhydrite. It has been known that some minerals significantly affect permeability; however, no general trend or rule of thumb appears to exist for carbonate rocks. While permeability in clastic rocks can be described by a model based on mineral content in the form of Herron's correlation, developing a universal model for carbonate rock permeability based on mineral content is not possible due to different heterogeneity mechanisms present from one formation to another. Nevertheless, a local field correlation calibrated with well logs is possible. This research presents an initial

attempt to discern the impact of mineralogy and porosity on permeability of the Khuff formation.

A data set that belongs to Khuff-B and Khuff-C strata was obtained from one Saudi Arabian reservoir. The data set has 537 points from several wells and includes well log porosity, bulk density, hole diameter, and volume fractions of dolomite, limestone and anhydrite. The set also included the corresponding core permeability values. Initially, a thorough quality analysis was carried after which the data was filtered. Supervised clustering produced different trends and clusters based on mineral ratio. Symbolic and non-linear regression techniques as well as neural networks were then employed to explore possible relations between the parameters. Non-linear regression produced a rigorous mathematical model that showed reasonable accuracy. The model consists of five terms of input parameters linked with different coefficients and exponents. On the other hand, neural networks were employed to model the heterogeneity of the system.

The developed mathematical model returned reasonable results with squared correlation coefficient (R^2) of 0.6 for the filtered data. Neural networks presented good results as well with R^2 of 0.73 for training and 0.71 for testing. The developed mathematical model also showed reasonable match with a separate set of Arab-D data.

ملخص الرسالة

الاسم الكامل : محمد حسن

عنوان الرسالة : إنشاء علاقة تجريبية لحساب نفاذية طبقة الخف

التخصص : هندسة النفط

تاريخ الدرجة العلمية : 2017

عادة ما تكون للخزانات الكربونية طبيعة معقدة للغاية و على درجة عالية من اللا تجانس. إذ يمكن أن تمتد على مساحات واسعة ولكن بمساميات منخفضة جدا. و قد يكون لبعض الطبقات نفاذية صخرية منخفضة بينما لبعضها الآخر شقوق طبيعية. لهذا السبب تفشل عادة تقنيات تقييم سجلات الآبار الصالحة لمكامن الحجر الرملي في المكامن الكربونية. و تمتاز الشرق الأوسط بمكامن كربونية ضخمة مما يجعل الجدوى الاقتصادية لإنتاج الزيت والغاز منها كبيرة.

يركز هذا البحث على طبقة الخف في الشرق الأوسط و التي تحتوي على واحدة من أكبر تراكمات الغاز في العالم. و المكونات المعدنية الرئيسية لطبقة الخف هي الدولومايت والحجر الجيري والأنهيدرايت. ومن المعروف أن بعض المعادن تؤثر بشكل واضح على نفاذية الصخور الرملية، إلا أنه لا يبدو أن هنالك قاعدة رئيسية او إتجاه عام للصخور الكربونية. فبينما يمكن وصف نفاذية الصخور الترابية على أساس نموذج يعتمد على المحتوى المعدني بصيغة علاقة هيرون التجريبية، فإن تطوير نموذج عام للصخور الكربونية على أساس المحتوى المعدني غير ممكن بسبب إختلاف آليات عدم التجانس من طبقة لأخرى. إلا أن تطوير علاقة محلية لحقل محدد معايرة بسجلات الآبار قد يكون ممكنا. و يقدم هذا البحث محاولة مبدئية للتعرف على تأثير المحتوى المعدني والمسامية على نفاذية طبقة الخف.

تم الحصول على قاعدة بيانات تنتمي إلى طبقتي الخف ب و الخف ج من أحد المكامن السعودية. و تحتوي القاعدة على 537 نقطة بيانية من عدة آبار مختلفة و تشمل قراءات سجلات الآبار للمسامية و الكثافة الكلية و قطر حفرة البئر و النسبة الحجمية لكل من الدولومايت و الحجر الجيري و الأنهيدرايت. كما إحتوت القاعدة على القيم المقابلة لنفاذية الأقوار. تم في البداية إجراء تحليل شامل لجودة البيانات ثم تمت بعد ذلك تصفية البيانات. و لقد أنتجت تقنية التجميع

الخاضعة للرقابة نزعات ومجموعات مختلفة على أساس نسبة المعادن. ثم استخدمت تقنيات النكوص الرمزي و النكوص اللا خطي إضافة إلى الشبكات العصبية لإستكشاف العلاقات الممكنة بين المتغيرات. و لقد أنتج النكوص اللا خطي نموذج رياضي شامل و ذو دقة معقولة. يحتوي النموذج على خمسة حدود من المدخلات مربوطة بأسس و معاملات مختلفة. و في المقابل، تم تطبيق تقنية الشبكات العصبية لنمذجة عدم التجانس في النظام.

أخرج النموذج المطور نتائج معقولة بمعامل علاقة (R^2) بقيمة 0.6 للبيانات المصفاة. كما أنتج نموذج الشبكات العصبية نتائج جيدة أيضا بمعامل علاقة (R^2) بقيمة 0.73 لبيانات التدريب و 0.71 لبيانات الإختبار. كذلك أظهر النموذج تطابقا جيدا مع قاعدة بيانات أخرى من طبقة العرب د.

CHAPTER 1

INTRODUCTION

1.1 Carbonate Rock Characteristics

Clastic rocks are formed usually from sediments that go through transportation, deposition and lithification, or compaction and eventually cementation to become a solid form of rock. Carbonate rocks, on the other hand, can be very different from clastic rocks and usually have quite complicated nature. They are developed from biogenic sediments, which are formed by biological activity, like reef building and seafloor accumulation of skeletal organisms. Other types are formed by water evaporation from shallow onshore basins or by precipitation from seawater. Furthermore, carbonate rock sediments are rarely transported to long distances like sediments of clastic rocks.

Clastic rocks are mostly shales and sandstones that have a variety of particles and minerals, which include clay minerals, feldspar, quartz, remnants of animals or plants and fragments of preexisting rocks. On the other hand, carbonate rocks have a small mineral group - mainly dolomite and calcite. Some additional minerals that can also exist in carbonate rocks are glauconite and phosphate; secondary minerals contain clay minerals, ankerite, anhydrite, quartz, pyrite, siderite and chert [1].

Due to the differences mentioned above, a classification system for carbonate and clastic rocks has been established. Clastic rocks are usually differentiated by size and grain composition, while carbonate rocks are distinguished by factors like pore or grain types, depositional texture, diagenesis or rock fabric. A famous classification of carbonate rocks is provided by Dunham [2], which is shown in Fig. 1-1.

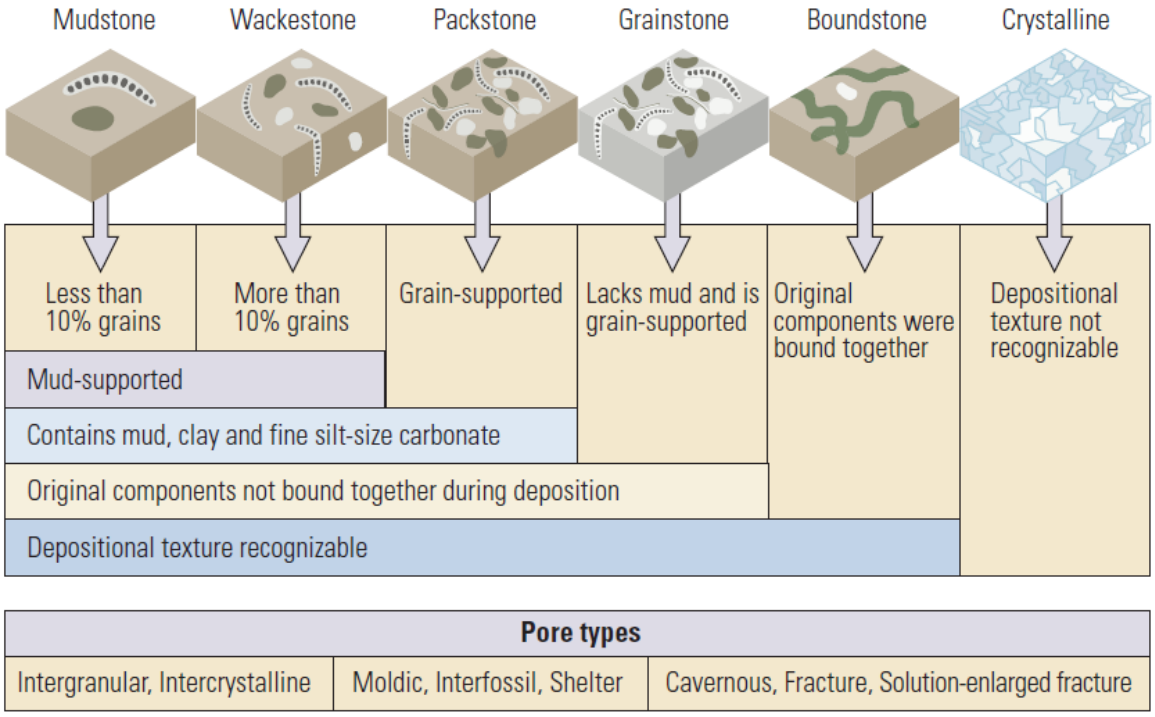


Figure 1-1: Carbonate Rock Classification as presented by Dunham and Lucia [2]

During diagenetic processes in carbonate rocks, pore space and permeability can be significantly modified. As carbonate rocks are extremely vulnerable to dissolution, new pore spaces are created after grain dissolution, while large vugs and caves are produced by dissolution along bedding planes and fractures. It can also produce pyrite or anhydrite fills. Change in mineralogy is not usually witnessed in clastic rocks. On the other hand, diagenetic activity in carbonates usually involves dolomitization in which the original

aragonite or calcite ($CaCO_3$) are replaced with dolomite ($CaMg(CO_3)$) by replacing calcium (Ca) by magnesium (Mg). Hydrocarbon recovery is usually enhanced by dolomitization due to enhancement of porosity since the ionic volume of Mg is considerably smaller than that of Ca resulting in greater porosity in dolomite. Although, carbonate and clastic rocks typically go through burial, compaction and cementation, of carbonate sediments comprise substantial quantities of the minerals that are metastable like magnesium calcite and aragonite; calcite itself is willingly dissolved and re-precipitated by migrating aqueous pore fluids. Thus, carbonate formations are expected to experience dissolution, recrystallization and mineralogical replacement. The controlling factors of these are temperature, pressure and pore-fluid chemistry that cause massive heterogeneity in carbonate formations. Stylolites are formed by deep burial dissolution process which can totally destroy reservoir permeability.

There is also an uneven distribution of pores in carbonate formations with variety of pore sizes and types. The pore system in clastic rocks is predominantly of intergranular structure, while some are uniformly distributed through the rock matrix. Carbonate formations also possess intergranular pores. As a primary pore type, intragranular porosity can also be present or usually developed when grains, like shell fragments, are partially dissolved. The shape of dissolved shell fragments is usually preserved by moldic porosity. In the carbonate Arab formation of the Arabian Peninsula, researchers have shown that there are two pore types: intergranular and intragranular. Therefore, carbonate reservoirs can be gigantic yet the pore spaces within the rock may be microscopic, and fluids can flow easily through fractures yet matrix permeability may be very small.

Carbonate rocks have a huge tendency for dissolution. For example, at the seafloor, as carbonic acid is formed by the reaction between carbon dioxide and water, dissolution may result in notable karst topography, that includes caves, sinkholes and intricate drainage patterns expressed on the surface of the earth as “disappearing” streams in active karst systems. In actual reservoirs, seismic imaging reveals collapse caves as karst systems. Wells drilled into these karst systems have huge productivity.

1.2 Carbonate Rocks in the Gulf Region

The most economically important carbonate formations in the Gulf Region are the Khuff (upper Permian-lower Triassic) and the Arab (upper Jurassic). Both formations house major petroleum reservoirs in the Arabian/Persian Gulf region with the Khuff producing gas almost exclusively – with the exception of Yibal field in Oman, which has significant oil production - and the Arab producing mainly oil. The Khuff formation contains the world’s largest gas reserves in South Pars (Iran) and North field (Qatar) dome with approximately 1500 TCF of recoverable reserves while the Arab formation houses the world’s largest oil accumulation: the Ghawar field in Saudi Arabia, with approximately 120 billion barrels of recoverable reserves. A comparison of the evolution of both the rocks over a period of time is presented in Table 1-1.

Table 1-1: A comparison of Evolution of Khuff and Arab-D reservoir rocks [3]

Factors	Khuff	Arab
Burial diagenesis	More chemical compaction and associated cementation caused by greater thermal exposure	Shallower burial depths correspond to greater preservation of primary porosity
Depositional setting	Extensive, poorly circulated, very low-relief shelf	Well-circulated conditions nearer to margins facing deep intracratonic basins
Dominant lithologies	Grainstones having relatively fine grain size and mudstones	Grainstones and grain-dominated packstones
Primary mineralogy	Aragonite	Mainly calcite
Anhydrite cement	Extensive anhydrite plugging of entire zones	Localized anhydrite cementation
Dolomite	Major proportion of reservoirs, finely crystalline	Minor to moderate proportion of reservoirs, medium to coarsely crystalline
Dominant pore types	Moldic and intercrystalline	Primary intergranular, intrafossil, and microporosity
Petroleum phase	Gas	Oil

As discussed above, carbonate rocks have extreme ranges of porosity and permeability. Since, we cannot develop a correlation for every carbonate rock present across the world, we intend to focus on the Khuff formation as the subject of this study. This is due to availability of data required to derive a correlation of permeability based on mineralogical and petrophysical parameters. Further details regarding the Khuff formation is given in the following section [3].

1.3 The Khuff Formation

The Khuff formation of eastern Arabia is 3.7 million km^2 in area, which makes it one of the largest carbonate ramps in Earth's history. It is of late Permian early Triassic age and is roughly 250 million years old. The Khuff formation of Arabia is present as the Kangan and Dalan formations of Iran, the Chia Zairi formation of Iraq and the Bih Hagi and lower Ghail formations in outcrops of the Musandam area of the eastern United Arab Emirates and northern Oman. The Khuff thickness increases from zero, where siliciclastic facies pinch out in central Saudi Arabia, to more than 400 m (1300 ft) in Ghawar field,

eastern Saudi Arabia, to 800 m (2600 ft) in the North field, Qatar, to nearly 1000 m (3300 ft) in the eastern United Arab Emirates [4].

The Khuff reservoirs possess an average porosity of less than 12% and display weak porosity-permeability correlation. Since the Khuff reservoir is deeper than the Arab reservoir, it depicts an exposure to higher temperature and pressure resulting in severe dolomitization. This also causes the Khuff reservoirs to be mainly gas bearing. The Khuff is part of a Paleozoic petroleum system sourced mainly from Lower Silurian hot shales in the gas window.

Burial depth is of utmost importance since it translates into significant alteration of Khuff rock. The Khuff strata were deposited on an extensive (epeiric), very low-relief shelf, which was sheltered from the open ocean by a reefal barrier. The Khuff reservoirs have a layer-cake geometry and consist mainly of interbedded mudstones and relatively fine-grained grainstones. Due to time of aragonite seas during the late Permian, the Khuff deposits are composed of less stable mineralogy. As a result, the Khuff grainstones were susceptible to intense eogenetic dissolution and cementation, commonly leading to the porosity inversion for which these strata are notorious, whereby volumes that were initially solid (grains) become pores, and the initial intergranular pores become solid (cement).

The Khuff has gone under greater dolomitization and quite significant calcium sulfate cementation. In the United Arab Emirates, for example, the Khuff formation is 75–85% dolomitized, and the Khuff dolomite tends to be finely crystalline. The Khuff strata in both South Pars and Ghawar fields are predominantly dolostone, with limestone intervals chiefly confined to the most open-marine grain shoal facies. These dolostones formed in

evaporitic sabkha and shallow reflux settings, although evidence of dolomite recrystallization and cementation during burial is abundant. The combined result of the above depositional and eogenetic factors is that the Khuff lithologies tend to be dominated by moldic pores in limestones and by very fine inter crystalline pores in dolostones, as well as having extensive anhydrite-cemented barrier zones. These tendencies are expected to result in lower permeability for a given porosity and poorer overall connectivity in the Khuff strata.

Khuff reservoirs almost exclusively contain gas. Only Yibal field has significant oil production. The Khuff is part of a Paleozoic petroleum system sourced mainly from Lower Silurian hot shales in the gas window.

1.4 Permeability Prediction

There are different logs that reveal the desired petrophysical parameters for a particular section of the wellbore. Famous logs for the determination of porosity are neutron, sonic and density logs. Bulk density is determined with help of density logs based on gamma ray logs. Lithology is determined by gamma ray and self/spontaneous potential logs. Caliper log determines the borehole diameter. Nuclear magnetic resonance (NMR) logging gives the pore size distribution along with porosity. This can be observed that despite many advancements in well logging operations, accurate matrix permeability determination is still a challenge until a well testing operation is carried out. For this reason, there always exist a need for a correlation that predicts permeability within a reasonable accuracy. Many researchers discussed in chapter 2 have developed matrix permeability correlations based

on different input parameters. This study focusses on log porosity, bulk density and mineral content for the development of permeability correlation.

Due to heterogeneity in carbonate formations, researchers believe that porosity and permeability relationships cannot be determined without the comprehension of different petrophysical classes within a particular formation. Evaluation methods that are effective in sandstone reservoirs usually are non-effective in carbonate reservoirs due to diagenesis of the pore structure. These differences obscure hydrocarbon recovery as well as reservoir evaluation. Therefore, there is a need of a rigorous and accurate predictor of matrix permeability that could incorporate the mineralogical variations of a given carbonate formation as well as the pore size distribution from the NMR log.

After extensive literature survey, presented in the second chapter, it has been observed that there are very few correlations of matrix permeability available for carbonate formations. This essentially is due to heterogeneity offered by carbonate formations. Despite some available correlations, there are none which correlate permeability with mineral content of carbonate reservoirs. The is a general trend of diagenesis from limestone to dolomite with mineral content. This study focusses on data from Khuff formation. In this study, an attempt to correlate permeability with mineral content has been carried out.

In this study, core and well-log data has been utilized to achieve the desired results. Core permeability was obtained by laboratory measurement. The corresponding log data includes porosity, bulk density, hole diameter (caliper log), and the rock mineral content of limestone, dolomite and anhydrite.

1.5 Thesis Organization

This thesis is organized in seven chapter, first being the introduction. Chapter 2 of this thesis provides a literature survey of all attempts at deriving permeability correlations. Chapter 3 states the problem and defines the objectives of this study. Workflow for this study is also discussed in the end. Chapter 4 provides the comprehensive data analysis and possible interpretations based on various established tools. Chapter 5 discusses the attempts to identify hidden patterns and unique characteristics within the data with the help of clustering. In chapter 6, development of mathematical correlation is discussed in detail. The overall approach and steps are presented. In the end, chapter 7 presents conclusions and recommendations of this study.

CHAPTER 2

LITERATURE REVIEW

There are many permeability correlations available for carbonate reservoirs in the literature. Some are developed with percolation and fractal concepts, some are based on pore and grain properties, while others are based on well log parameters. All correlations available in literature are presented in this chapter.

2.1 Permeability correlations developed based on the pore and grain properties

Kozeny and Carman

Kozeny and Carman (KC) [5][6] derived an equation of permeability based on the specific surface area (ratio of pore surface area to grain volume) and tortuosity $(L_a/L)^2$ of the sample (rock). The equation is given by:

$$k = \frac{\phi^3}{f\tau A_g^2(1-\phi^2)} \tag{2.1}$$

Where,

f : empirically measured constant

τ : tortuosity $(L_a/L)^2$, ratio

A_g : specific surface area, 1/m

ϕ = porosity, fraction

L_a = length of actual flow path followed by fluid, mm

L = capillary tube length, mm

Amaefule et al. and Altunbay et al.

A modification by Amaefule et al. and Altunbay et al. [7][8] of Kozeny and Carman equation after Archie relationship ($F=a\phi^{-m}$) now defines permeability in terms of effective pore radius and formation factor. The equation is given by:

$$k = \frac{r_{eff}^2}{8F} \quad (2.2)$$

Where,

r_{eff} = effective pore radius of capillary tube, μm^2

F = formation factor ($F=a\phi^{-m}$), constant [9]

Krumbein and Monk

Krumbein and Monk [10][11] presented a permeability correlation in terms of the geometric mean of grain diameter, standard deviation of the grain diameter. Equation is given by:

$$k = 760 D_g^2 \exp(-1.3\sigma_D) \quad (2.3)$$

Where,

K = permability in darcies

D_g = geometric mean of grain diameter, mm^2

σ_D = standard deviation of grain diameter, mm

Berg

Berg [10] presented a permeability correlation in terms of the median grain diameter. Equation is given by:

$$k = 80.8\phi^{5.1}D^2e^{-1.385\,p} \quad (2.4)$$

Where,

K = permability, darcy

D = median grain diameter, mm

ϕ = porosity, %

P = sorting term, constant

Van Baaren

Van Baaren [10] presented a permeability correlation in terms of the dominant grain size. The equation is given by:

$$k = 10\phi^{3.64+m}{D_d}^2C^{-3.64} \quad (2.5)$$

Where,

K = permeability, darcy

D_d = dominant grain size, μm

C = sorting constant {ranges from 0.7(well sorted)-1(poorly sorted)}

ϕ = porosity, %

m = cementation exponent, constant

Coates and Denoo

Coates and Denoo [10] defined permeability in terms of the irreducible water saturation. The equation is given by:

$$k = [\frac{100\phi_e{}^2(1-S_{wi})}{S_{wi}}]^2 \quad (2.6)$$

Where,

S_{wi} = irreducible water saturation, fraction

ϕ_e = effective porosity, fraction

Sen et al.

Sen et al. [10][12] presented permeability correlation in terms of the volume to surface ratio, exchange cation molarity, proton NMR decay constant. The equation is given by:

$$k = 10^{6.59}(\phi^m V_p/_S)^{2.08} \quad (2.7)$$

$$k = 10^{2.65}(\phi^m/_{Q_v})^{2.11} \quad (2.8)$$

$$k = 10^{-0.1}(\phi^m T_1)^{2.15} \quad (2.9)$$

Where,

k = permeability, mD

V_p = pore volume, μm^3

Q_v = exchange cation molarity per unit volume {meq (million equivalent)/ml}

ϕ = porosity, fraction

T_1 = proton NMR decay constant, milliseconds

S = surface area, μm^2

m = conductivity exponent

Swanson

Swanson [10][13] defines permeability in terms of maximum capillary pressure and mercury saturation in percent. The equation is given by:

$$k = a[\frac{S_b}{P_c}]^c_{max} \quad (2.10)$$

Where,

P_c = capillary pressure

S_b = mercury saturation as percent of bulk volume

Further he also presented permeability in terms of logarithmic mean of the distribution T_2 and NMR porosity. Equation is given by:

$$k = 4.6\phi^4 {T_{2lm}}^2 \quad (2.11)$$

Where,

T_{2lm} = logarithmic mean of relaxation time

ϕ = porosity

Coates

Coates [10] defined permeability in terms of FFI (free flow index) and BVI (bulk volume irreducible). The equation is given by:

$$k = [\frac{\phi}{C}]^2[\frac{FFI}{BVI}] \quad (2.12)$$

Where,

ϕ = porosity

C = constant

FFI and BVI are determined from NMR.

Quintero et al.

Quintero et al.[10][14] presented a permeability correlation in terms of T_2 distribution. The equation is given by:

$$k = Cpf 4.6 \phi^4 {T_{2lm}}^2 \quad (2.13)$$

Where,

T_{2lm} = logarithmic mean of relaxation time

ϕ = porosity

C_{pf} = permeability factor constant (0.4 for carbonates)

2.2 Permeability Correlations based on the Flow-Zone Indicator

The flow-zone indicator (FZI) is a term used to define the flow zones based on the surface area and tortuosity. Flow Zone Indicator is a unique and useful value to quantify the flow character of a reservoir and one that offers a relationship between petrophysical properties at small-scale, such as core plugs, and large-scale, such as well bore level [15]. A general equation relating the permeability to the FZI was derived by (Amaeful et al, Altunbay et al) [7] as given by Eq. 2.14.

$$k = 1014\ (FZI)^2[\frac{\phi^3{}_e}{(1-\phi_e)^2}] \quad (2.14)$$

Where,

ϕ_e = effective porosity

Several methods to calculate the FZI for a given reservoir were presented. Amaefule et al. [7] defined FZI as:

FZI = RQI/ ϕ_z

Where,

RQI: Reservoir quality index given by:

$$RQI\ (\mu m)\ =\ 0.0314\sqrt{\frac{k\ (mD)}{\Phi}} \quad (2.15)$$

ϕ_z: normalized porosity given by:

$$\phi_z = \frac{\phi}{1-\phi} \quad (2.16)$$

Ohen et al. presented the following equation:

$$FZI = [\frac{b(1-NMRS_{WR}}{1+a(NMRS_{WR}-1)}]^{1/c} \quad (2.17)$$

Where,

$NMRS_{WR}$ = NMR measured water saturation

a,b,c =Constants

Altunbay et al. defined the FZI based on resistivities, gamma ray and grain density as:

$$FZI = a + bRt + cRxo + dGR + eGD + fRtRxo + gRtGR + hRtGD + jRxoGR + kRxoGD + iGRGD + mRt2 + nRxo2 + pGR2 + rGD2 \quad (2.18)$$

Where,

R_t = true resistivity

R_{xo} = resistivity of flushed zone

GR = spectral gamma ray log value in API

GD = grain density

2.3 Permeability Correlations Developed with Percolation and Fractal Concepts

Katz and Thompson

Katz and Thompson [16][17] developed permeability correlation based on some characteristic length of pore space (l_c), conductivity of water (σ_o) and conductivity of water saturated rock (σ).

$$k = \frac{1}{226} l_c^2 \frac{\sigma}{\sigma o} \quad (2.19)$$

Where,

l_c: This is some characteristic pore space which is a subset of all the pores. It is argued by the author that permeability is governed by the smallest pores. This subset contains smallest connected pores within the larger pore spaces. It can be calculated from the pressure at the inflection point on a capillary pressure curve by converting it to diameter.

σ_o = conductivity of water

σ = conductivity of water saturated rock

Mavko and Nur

Mavko and Nur [10] developed a permeability correlation based on the particle size and threshold porosity.

$$k \, \alpha \, c(\phi - \phi_c)^3 d^2 \quad (2.20)$$

Where,

ϕ = porosity

ϕ_c = threshold (critical) porosity

d = particle size

Martys et al.

Martys et al. [16][18] developed a permeability correlation based on the specific surface area and the threshold porosity;

$$k = \frac{2\phi_2^{\ *}}{s^2}(\phi_1 - \phi_1^{\ c})^f \quad (2.21)$$

$$k = k_o(\phi_1 - \phi_1^{\ c})^g \quad (2.22)$$

Where,

g = 4 (constant)

f = 4.2 (constant)

ϕ_1 = porosity

ϕ_2 = 1-ϕ_1

$\phi_1^{\ c}$ = critical porosity at which pore space first percolates or threshold porosity

s = specific surface area (pore surface area to the total volume of the sample)

k_o = permeability bound for bed of poly dispersed spheres

Pape et al.

Pape et al. [16][19] developed the following permeability correlation;

$$k = A\phi + B\phi^m + C(10\phi)^n \quad (2.23)$$

Where**,**

ϕ = porosity

m: Archie's constant

n: m + 2/[c_1(3-D)], 0: 39<c_1<1

c_1, D: Fractal dimensions

Muller and Mccauley

Muller and Mccauley [16][20] developed a following permeability correlation based on the fractal dimensions;

$$k\ \alpha\ \phi^{4-D/D} \tag{2.24}$$

Where,

D= median grain diameter

ϕ = porosity

Wong

Wong [16][21] developed a following permeability correlation based on the formation factor, characteristic grain size (l_g) and characteristic throat size (l_t);

$$k = \frac{c_1 l_g^2}{F} \tag{2.25}$$

$$k = \frac{c_2 l_t^2}{F} \tag{2.26}$$

Where,

F = formation factor

l_g = characteristic grain size

l_t = characteristic throat size

Hansen and Skjeltorp

Hansen and Skjeltorp [16][22] developed the following permeability correlation based on the characteristic length, euclidean dimensions and fractal dimension;

$$k_r = c^1 \phi \left(\frac{l_1}{l_n}\right)^{2[(E_S - D_S) + (D_v - E_v)]} \quad (2.27)$$

Where,

ϕ = porosity

E_s and E_v = Euclidean surface and volume dimension

D_v and D_s = fractal volume and surface dimension

C = rock dependent constant

Kr = rock permeability

Garrison et al (1993)

Garrison et al developed a following permeability correlation based on the apparent surface fractal dimension and area shape factor;

$$k_r = 10^{-1.7-0.6D_{s\,(control)}+113\left(S_{a(contol)}\right)} \quad (2.28)$$

Where,

D_s = apparent surface fractal dimension

S_a = area shape factor

2.4 Permeability Correlations Derived with Well Log Parameters

Tixier

Tixier et. al. [16][23] developed a following permeability correlation based on the resistivity gradient, resistivity of samples saturated with water, densities of formation water and oil;

$$k = C\left[\frac{2.3}{R_o(\rho_w-\rho_0)}\frac{\Delta R}{\Delta D}\right]^2 \quad (2.29)$$

Where,

$\Delta R/\Delta D$ = resistivity gradient

R_o = resistivity of samples saturated with water

$\rho_{w,o}$ = densities of formation water and oil

Coates and Dummannoir

Coates et. al. and Dummannoir et. al. [16] developed a following permeability correlation based on the water resistivity, true resistivity, and cementation and tortuosity factor:

$$k = [\frac{C\ \phi^{2w}}{w^4(\frac{R_w}{R_t})}]^2 \quad (2.30)$$

Where,

R_w = water resistivity

R_t = true resistivity

w = cementation and tortuosity factor

ϕ = porosity

Yao and Holditch

Yao et. al. and Holditch et. al. [24] developed a following permeability correlation based on the gamma ray index, deep induction, shallow resistivity:

$$k = \frac{\phi^{e1}\ (1-I_{GR})^{e2}\ R_{ild}^{e3}}{(\frac{R_{ild}}{R_{sfl}})^{e4}} \quad (2.31)$$

Where,

I_{GR} = gamma ray index

R_{ild} = deep induction

R_{sfl} = shallow resistivity

Saner et al.

Saner et al. [25] developed the following permeability correlation based on the formation factor:

$$log(k) = 7.04 - 4.19(F) \quad (2.32)$$

Where,

F = formation factor

Mohaghegh et al.

Mohaghegh et al. [26] developed a following permeability correlation based on the gamma ray index, bulk density, deep induction:

$$k = 126.5 + 0.0011\gamma - 50.3\rho_D + 0.0625\, I_D \quad (2.33)$$

Where,

γ = gamma ray index

ρ_D = bulk density

I_D = deep induction

Xue et al.

Xue et al. [27] developed a following permeability correlation based on the sonic travel time, density porosity, gamma ray index, resistivity ratio:

$$\log(k) = 0.151\Delta t - 0.019\phi_d - 0.0392GR + 0.0222RR - 7.7 \quad (2.34)$$

Where,

Δt = sonic travel time

ϕ_d = density porosity

GR = gamma ray index

RR = resistivity ratio

2.5 Permeability Correlation based on Rock Fabric

A correlation is presented by Lucia and Jennings [28][29] which is based on the petrophysical classes. The band of petrophysical classes were defined by rock fabric number which was incorporated in the permeability porosity relationship. The equation is given by:

$$\log(k) = \left(A - Blog(rfn)\right) + \left(\left(C - Dlog(rfn)\right)\log\left(\phi_{ip}\right)\right) \quad (2.35)$$

Where,

A=9.7982

B=12.0838

C=8.6711

D=8.2965

Rfn= rock fabric number (determined by plotting data k vs ϕ_{ip} on a developed chart by Lucia and Jennings, Fig. 2-1)

ϕ_{ip} = fractional interparticle porosity

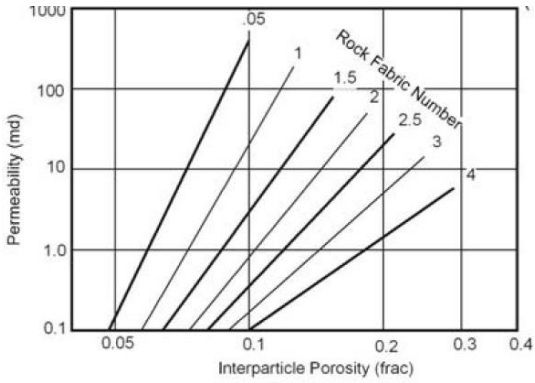


Figure 2-1: Plot of Permeability Vs Interparticle porosity to determine rock fabric number[29].

It is observed that none of the above correlations contain mineral content or composition of the rock as an input. There are various correlations available for siliclastic rocks that are based on mineral content but none for carbonate rocks. One such correlation was presented by Herron.

2.6 Herron's Correlation

Herron et. al. [30] presented a correlation for clastic rocks based on mineralogy, determined from geochemical well logging, and porosity to estimate the intrinsic permeability. Correlation is given by:

$$k = A_f \frac{\phi^3}{(1-\phi^2)} \exp(\Sigma BiMi) \qquad (2.36)$$

Where,

Af = textural maturity based on feldspar content

$$A_f = 4.9 + 2F_{max}$$

Mi = mineral content

Bi = constant for each mineral (table 2)

ϕ = porosity

Table 2-1: Constant for each Minerals for Eq. 2.36

Clays		**Cements**		**Framework Minerals**	
Kaolinite	-4.5	**Calcite**	-2.5	**Quartz**	0.1
Illite	-5.5			**Feldspar**	1.0
Smectite	-7.5				

2.7 Artificial Intelligence Utilization for Correlations

Many researchers are using artificial intelligence to capture the heterogeneities offered by natural systems. Neural networks, Fuzzy logics, Support verctor machines etc. are employed to develop correlations. Al-anazi et al. [31] employed support vector machine for petrophysical classification as well as to correlate well-log parameters with permeability. Mohaghegh et al. [26] compared statistical and artificial neural network to estimate permeability from well-log data. He demonstrated after comparison that accuracies from neural network are unmatched to any other technique. Fang et. al. [32] discusses the application of neural networks to a pattern recognition problem in ecology: the determination of lithology from well-logs. They demonstrated the application of neural network for the determination of lithologies (limestone, dolomite, sandstone, shale, sandy and dolomitic limestones, sandy dolomite, and shale sandstone) from selected well logs. Saggaf et al. [30] and Bhatt et al. [31] both applied neural networks to estimate reservoir

parameters. There are countless other researchers who have solved their complex systems with the help of Artificial Intelligence. Due to heterogeneity in carbonate system, we also intend to apply Artificial Neural Network to formulate an empirical correlation.

CHAPTER 3

PROBLEM STATEMENT

3.1 KNOWLEDGE GAP

Recent advances in well logging techniques allow the determination of the mineral content of a rock formation along its depth. Service companies offer various tools that work on

Elemental Capture Spectroscopy to achieve that. Since some minerals are known to significantly affect permeability, it would be very useful to develop a permeability correlation for carbonate rocks based solely on their mineral composition - as measured by well logs - as well as other well log data. But since different carbonate formations have gone through different transformations during depositional, diagenetic and tectonic activities, heterogeneities exhibited by carbonate rocks are not easy to capture even with conventional well logs and seismic attributes. That's why a universal matrix permeability correlation for all carbonate rock is far-fetched. Nevertheless, a correlation developed specifically for a given carbonate formation would be reasonable if the mineral content can trend the diagenesis of the rock type. This research presents an initial attempt to discern the impact of mineralogy on the permeability of the Khuff formation.

3.2 OBJECTIVES

The objective of this research is to develop a correlation for the prediction of permeability of the Khuff formation based on petrophysical and mineralogical data obtained from well logs. The correlation will allow the estimation of matrix permeability for the Khuff formation only in Saudi Arabia.

3.3 Research Methodology

Data available for Khuff formation will be used to develop the empirical correlation for permeability based on mineral content. Currently we have 537 core sample data. Filtration process will be employed to remove outliers and meaningless data points. Using data analysis techniques, data will be divided into different petrophysical classes based on which different rock flow units will be identified. Based on these zones, any statistical analysis tool like non-linear regression, artificial neural network etc. will be used to develop the correlation.

Once the correlation is developed, it will be tested on the data available in literature for the Khuff formation in the Arabian Gulf region.

The roadmap of this study is summarized in the following flowchart:

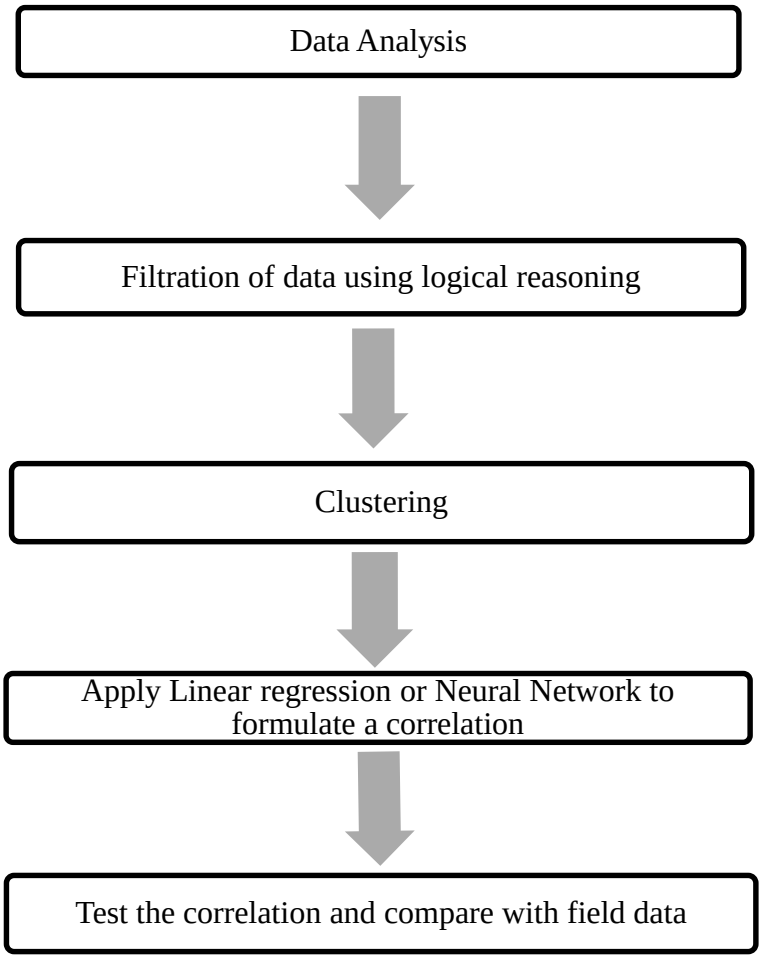


Figure 3-1: Work flow for this study

CHAPTER 4

DATA ANALYSIS

This chapter presents details of data analysis and handling. Initially the theory and underlying concepts of data analysis tools are given. This is followed by an interpretation of the different data analysis tools. This includes crossplots of permeability versus available parameters or variables as well as the results of analysis by the concept of hydraulic flow units.

4.1 Theory

In the data analysis process, several statistical parameters were used such as mean, mode, median, standard deviation, kurtosis, skewness, standard error, range etc. A brief description of these parameters is provided in this section.

4.1.1 Measure of Central Tendency

A measure of central tendency can be explained as the tendency of the set of data points to be around a point on a straight line or the point around which data is aligned. The first and foremost measure of central tendency is the mean or average. The median and mode are also indicators of central tendency of a data set.

Mean

The mean or average is the term generally used for an arithmetic mean of a data set defined by the following formula:

$$\bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad (4.1)$$

Where,

$\bar{x}$ = mean,

$\sum$xi = sum of all points in the data set,

N = number of points in the data set.

Some other forms of mean include harmonic mean and the geometric mean. The arithmetic mean describes the center of a set of data points.

Median

The median is defined as the value for a set of data points that divides the data into two equal halves when data is arranged in an ascending order. The median value is such that half of the data is exactly above the median value as its below. This value is used to define the central tendency of data sets which are not evenly or uniformly distributed such that they are either positively or negatively skewed. For example, a data set that is positively skewed would show a higher value of mean, which may mislead the interpretation process as demonstrated by Fig. 4-2. For a data set having equal mean and median values is

uniformly distributed as shown in Fig. 4-1. Hence, it can be deduced that the mean is influenced by a few very high or very low value points in the data set, which will not in turn fully represent the central tendency of the data set. This anomaly makes the median a better measure of central tendency in certain cases.

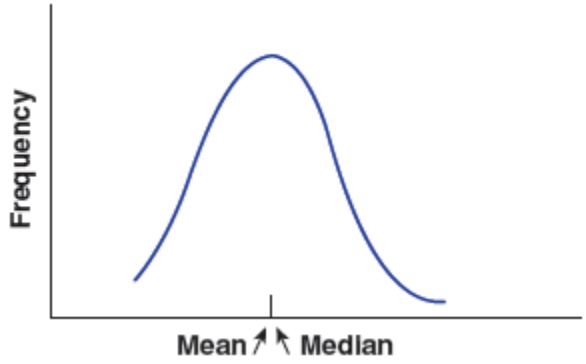


Figure 4-1: Normal distribution will show similar Mean and Median value

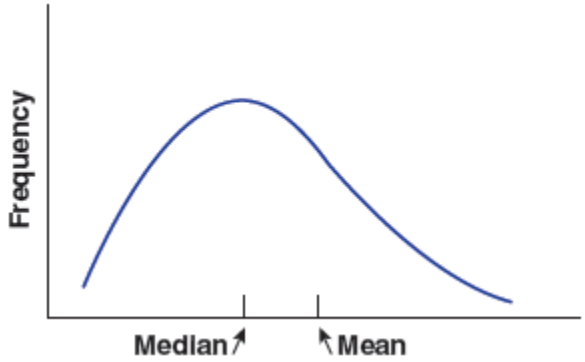


Figure 4-2: Positively skewed data will show Mean higher than Median

Mode

The mode can be defined as the most frequently occurring number in a set of data. The mode is generally not a very good measure of central tendency but works best in a classification problem.

Usually, the arithmetic mean is considered to be the best measure of central tendency for a given data set. But it can be used in conjunction with other measures of central tendency to attain a better understanding of the given data set.

4.1.2 Measures of Variation

A measure of variation deals with the dispersion of a data set around a center value. Most commonly used measures of variation include range, standard deviation and variance.

Measures of variation help understand how wide or deep a data set is distributed around a central value while measures of central tendency help understand how data points are clustered about a center value.

Range

Range can be defined as the difference between the highest and lowest data points within a given data set. It describes how far the data set is distributed from a center value. Two data sets could have the same mean and median but their measures of dispersion from the center value could be different. So, for a preliminary measure of variation, range could be an indicator of dispersion but for data sets of complicated nature, it might not be too

useful. It is because of the dependence of range on highest and lowest value, which can affect the range dramatically due to a single aberrant data point.

Standard Deviation

Standard deviation may be defined as the amount with which each data set varies from a center value. The formula to measure standard deviation is given by:

$$\sigma = \sqrt{\sum \frac{(x_i - \bar{x})^2}{N}} \quad (4.2)$$

Where,

σ = standard deviation of a data set,

$\bar{x}$ = mean,

xi = i'th data point in the data set,

N = number of points in the data set.

Standard deviation is used widely in inferential statistics as the representation of measure of variation.

Variance

Variance may be defined as the average of the square of the deviations for a set of data from their mean value. It is given by formula:

$$a^2 = \sum \frac{(x_i - \bar{x})^2}{N} \quad (4.3)$$

Where,

a^2 = variance of a data set,

$\bar{x}$ = mean,

xi = i'th data point in the data set,

N = number of points in the data set.

One of the conceptual drawback of this statistical parameter is that it leaves the unit of measurement to be squared. Due to this reason, standard deviation is widely used as a measure of variation in inferential as well as descriptive statistics.

4.1.3 Standard Error

Standard error can be defined as a statistical term that can measure the accuracy with which a sample represents a population. It can also be described as the deviation of mean of a sample from the mean of a population. This difference is simply known as standard error. Smaller standard error depicts more representative sample from a population. Small standard error depicts larger sample size as both are inversely proportional to each other. Hence, data having smaller standard error is considered to be more representative of the true mean and high standard error depicts certain irregularities in the data. It can be simply calculated by dividing square root of sample size with standard deviation.

SE = σ/√ (sample size) (4.4)

4.1.4 Skewness

Skewness can be defined as the degree of symmetry or lack of symmetry for a given data set. It can be obtained through following formula:

$$a^3 = \sum \frac{(x_i - \bar{x})^3}{N\sigma^3} \quad (4.5)$$

Where,

a^3 = skewness of a data set,

σ = standard deviation of a data set,

$\bar{x}$ = mean,

xi = i'th data point in the data set,

N = number of points in the data set.

A symmetrical data has the skewness value equal to 0. This can be demonstrated through Fig. 4-3.

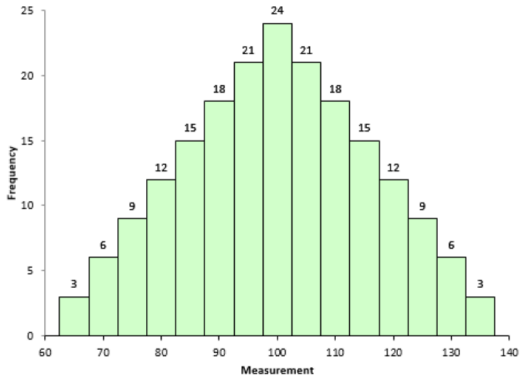


Figure 4-3: Symmetrical Data having skewness value equal to 0.

If right handed tail is longer than left handed tail, then data is said to be positively skewed as demonstrated by the Fig. 4-4.

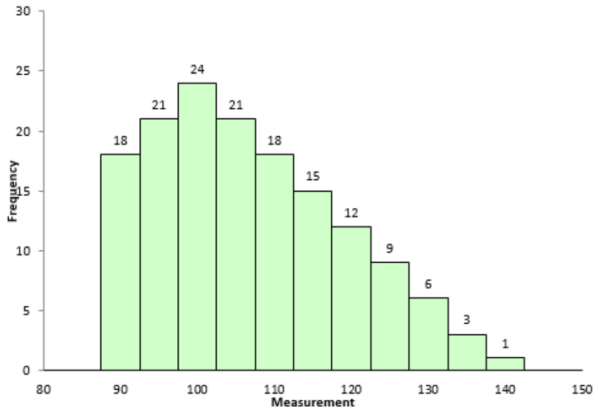


Figure 4-4: Positively skewed data having right handed tail longer than left handed tail.

If left handed tail is longer than right handed tail, then data is said to be negatively skewed as demonstrated by the Fig. 4-5.

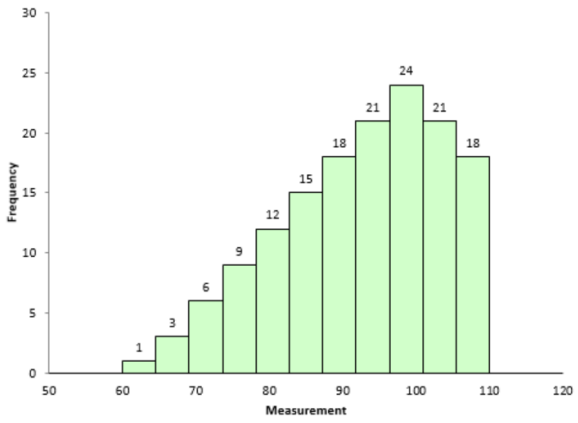


Figure 4-5: Negatively skewed data having left handed tail longer than right handed tail.

If skewness has a value in the range from -0.5 to 0.5 then data set is fairly symmetrical. If skewness has value from -0.5 to -1 or from 0.5 to 1 then data set is moderately skewed. Skewness values beyond -1 and 1 indicate highly skewed data sets.

4.1.5 Kurtosis

Kurtosis can be defined as the degree of peakedness or flatness of a data set. This can be calculated through following formula:

$$a^4 = \sum \frac{(x_i - \bar{x})^3}{N\sigma^4} \tag{4.6}$$

Where,

a^4 = kurtosis of a data set,

σ = standard deviation of a data set,

$\bar{x}$ = mean,

xi = i’th data point in the data set,

N = number of points in the data set.

Kurtosis showing higher values depicts heavier tails as compared to lower values showing lighter tails. Kurtosis for the data set of Fig. 4-6 is negative.

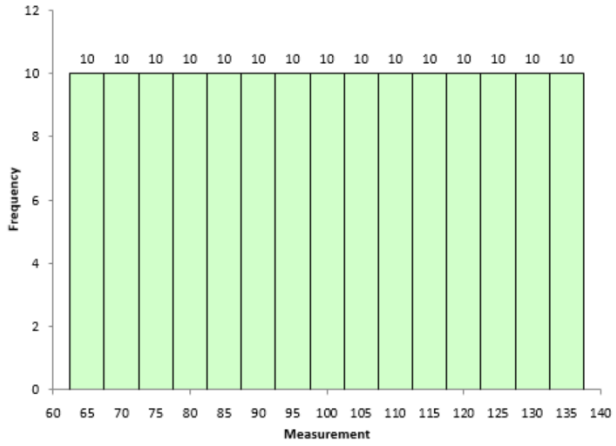


Figure 4-6: Example of negative kurtosis value

Kurtosis for the data set of Fig. 4-7 is positive.

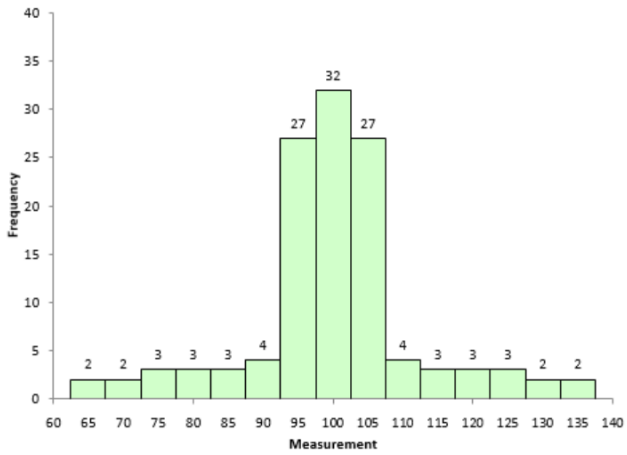


Figure 4-7: Example of positive kurtosis value

4.2 Preliminary Analysis

The data utilized for this study belongs to Khuff B and Khuff -C formations within eastern Saudi Arabia. Core and well log data was obtained from few wells completed in the Khuff region. The total number of data points are 537. The available data includes core permeability, log porosity, bulk density, rock content of limestone, dolomite and anhydrite in terms of volumetric fractions, caliper log etc.

Based on the statistical parameters mentioned above, a preliminary analysis was carried out on the available data. The main objectives of preliminary analysis are to prepare the data for further analysis by describing the key features of the data and summarizing the results. Quantitative and qualitative approaches used to describe the data include measure of variation, measure of dispersion, skewness, kurtosis, range etc. Summary of the preliminary data analysis is provided in Table 4-1.

Table 4-1: Summary of Data Analysis

	Permeability, K (mD)	Porosity, ϕ (fraction)	Density, ρ (g/cc)	Limestone Content, ML (%)	Dolomite Content, MD(%)	Anhydrite Content, MA (%)
Mean	67.20	0.14	2.69	38.27	46.72	15.05
Standard Error	8.21	0.00	0.02	1.29	1.34	0.45
Median	4.95	0.13	2.80	42.60	40.10	14.84
Mode	0.10	0.21	2.87	0.02	6.50	6.50
Standard Deviation	196.47	0.07	0.40	30.79	32.02	10.74
Sample Variance	38601	0.01	0.16	947.85	1025.33	115.43
Kurtosis	19.03	-1.14	258.28	-1.35	-1.32	0.33
Skewness	4.27	0.06	13.07	0.13	0.35	0.63
Range	1340.49	0.28	8.31	97.80	99.86	54.74
Minimum	0.01	0.00	2.20	0.01	0.13	0.00
Maximum	1340.50	0.28	10.51	97.80	99.99	54.74

4.2.1 Core Permeability

Visual representation of data is provided in Fig. 4-8 where permeability is arranged in an ascending order and plotted against sample no. A histogram based on Freedman-Diaconis rule is also displayed in the Fig. 4-8 [35]. This rule is employed in all histograms in this Chapter. Permeability has an arithmetic mean of 67.2 mD and a standard error of 8.21, which is quite high for a data set. A high standard error in a data set depicts certain irregularities in the data, which can be seen through histogram as well. As confirmed by histogram, permeability data is positively skewed with a skewness value of 4.27 mD, which explains the major chunk of values below 10 mD. Permeability is highly skewed with kurtosis value around 19.03 mD, which explains the sharp peaks. Difference of mean and median also confirms the concentration of data to lower values. Data has high and lows of 0.01mD and 1340 mD, respectively, with a very high standard deviation of 196.47 mD. This very large permeabilities could be the result of cracked or fractured core samples.

Lower values, on the other hand, could indicate very tight cores where permeability measurement was not possible.

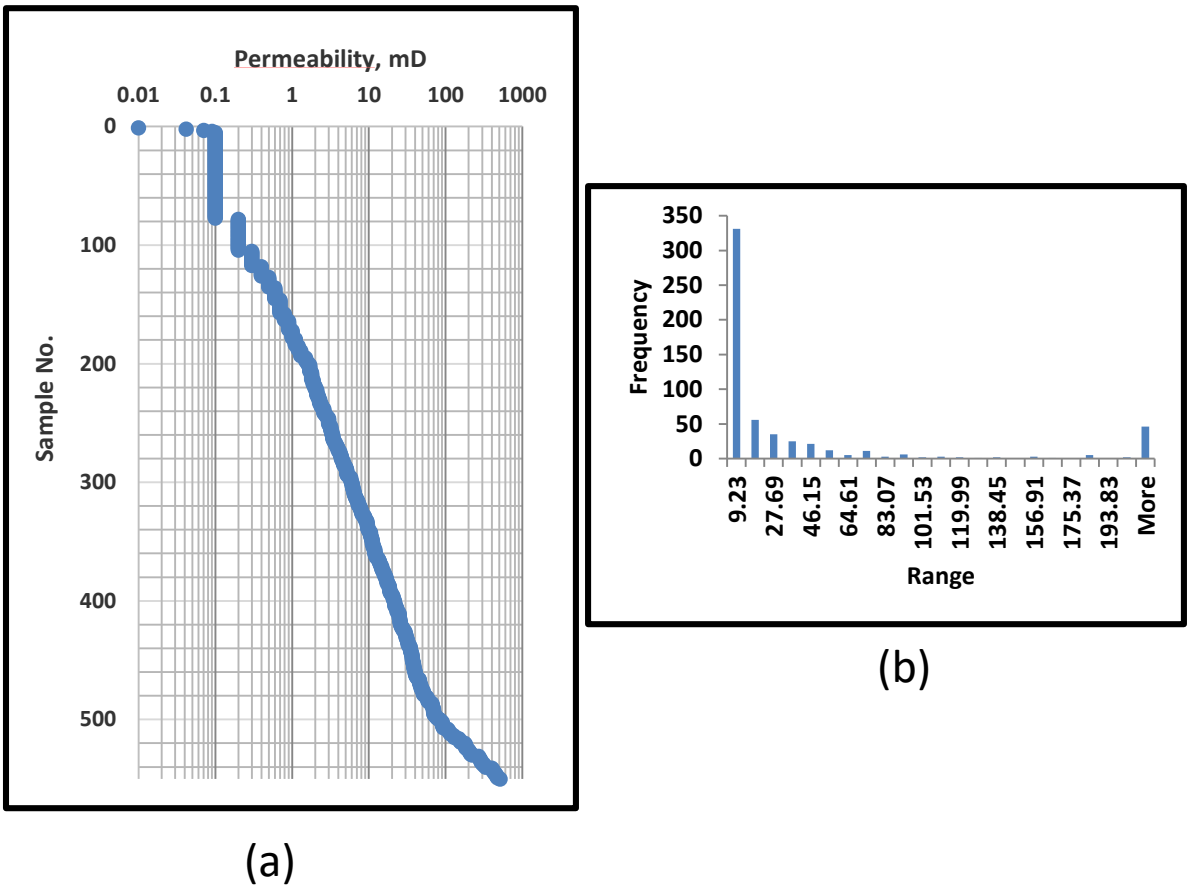


Figure 4-8: (a) Permeability (ascending order) Vs Sample No. (b) Histogram of permeability data based on Freedman-Diaconis rule.

4.2.2 Porosity

Visual representation of porosity data is provided in Fig. 4-9 where porosity is plotted against sample number as well as a histogram. Porosity has arithmetic mean equal to 0.14 fraction having standard error of 0.00. Standard error being zero depicts true mean. It can also be confirmed by the median which is very close to mean. Standard deviation and variance is low as well which depicts that data is not scattered as was in the case of permeability. Skewness value of 0.06 fraction depicts fairly symmetrical data having a kurtosis of -1.14 giving the impression of lighter tails. Graphical representation also confirms this but the histogram shows two peaks, which could be an indication of dual porosity system within the data. This will be further demonstrated in the clustering section. High and lows of porosity is 0.28 and 0, respectively. Mean porosity of this data set is also equivalent to mean porosity of the Khuff formation reported in literature [3].

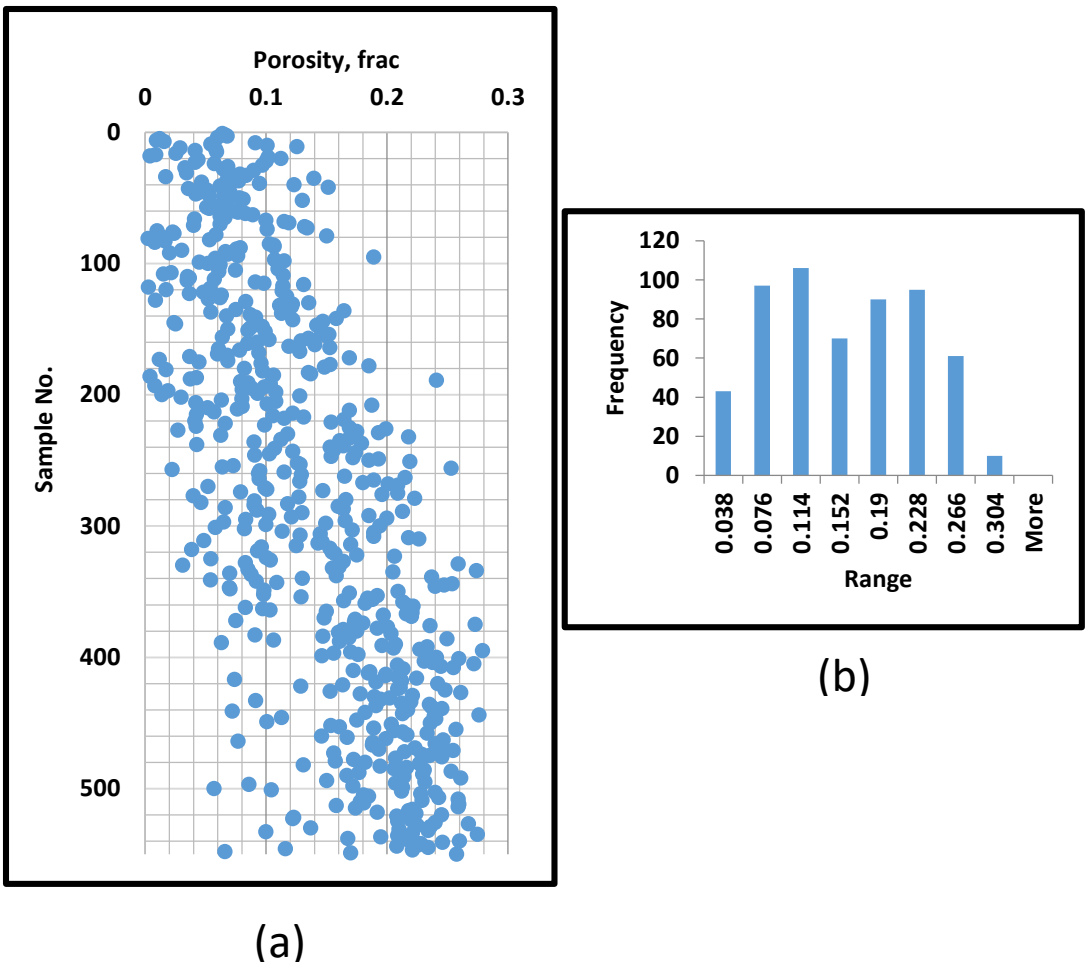


Figure 4-9: (a) Porosity Vs Sample No. (b) Histogram of porosity data based on Freedman-Diaconis rule.

4.2.3 Bulk Density

Visual representation of data is provided in Fig. 4-10 where bulk density is plotted against sample number as well as a histogram. The mean of bulk density is 2.69 g/cc but due to concentration of data in dolomitic range having density 2.87g/cc, median is 2.8 g/cc. This is also confirmed by mode which is around 2.87. Graphical representation and

histogram also confirms the same. Standard deviation of the data set is 0.4 g/cc showing the concentration of data within very small range. Data is highly skewed in the positive direction having very high kurtosis value showing a heavy tail. Range of the data is highly misleading as it's not possible to have a bulk density value to be 10.51 g/cc. This shows the existence of outliers which need to be removed for further analysis.

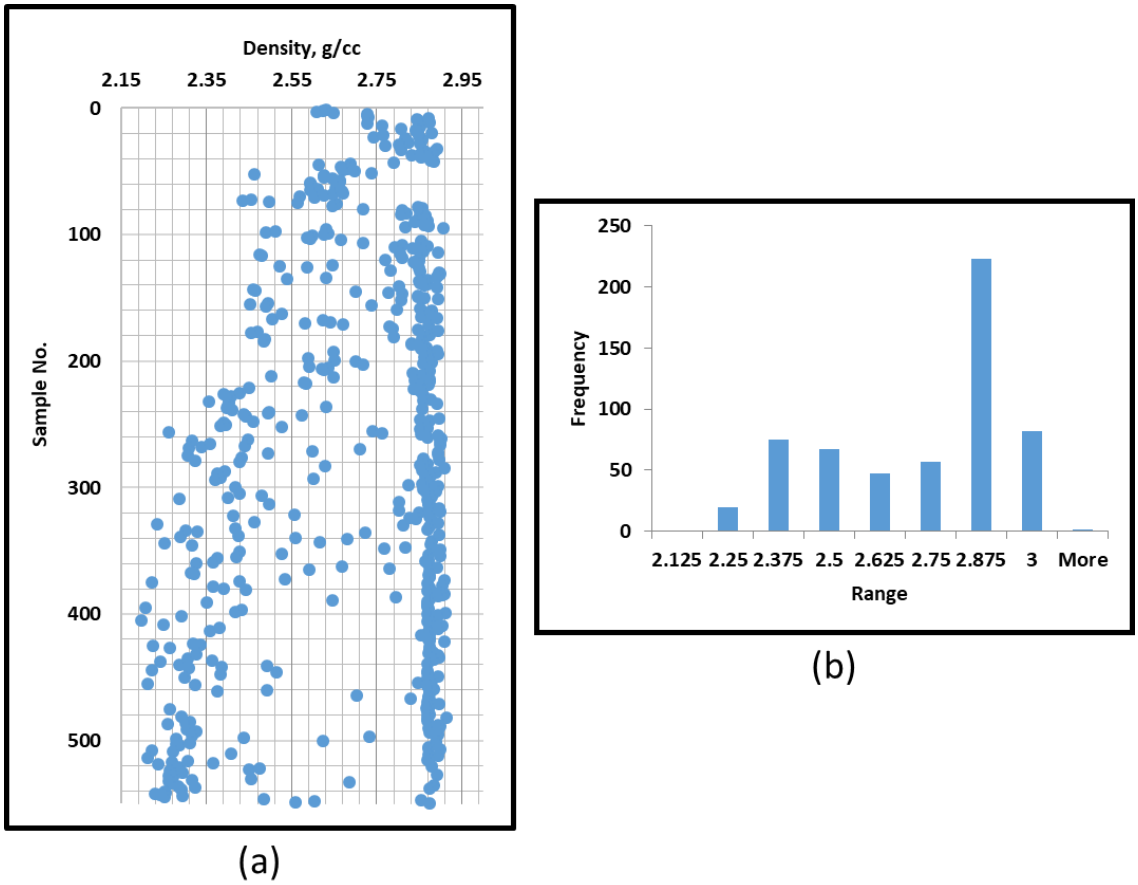


Figure 4-10: (a) Density Vs Sample No. (b) Histogram of density data based on Freedman-Diaconis rule.

4.2.4 Mineral Content

Mineralogical composition of the Khuff log data show limestone, dolomite and anhydrite. Graphical representation along with histograms of each mineral content is provided in Fig. 4-11. Fig. 4-11 shows mineral content vs sample no. Green color represents anhydrite content, which has high and low of 54.74% and 0 %, respectively, with mean around 15%. Anhydrite content is positively skewed as shown in Fig. 4-12. Red color represents dolomite content which has high and low of 99.99% and 0.13 % respectively with mean around 46.7%. It shows highly dolomitized rock. Dolomite content is also positively skewed having negative kurtosis showing almost uniform peaks in Fig. 4-12. Standard deviation for the dolomite content is high due to extreme values of dolomite content away from mean. Blue color represents limestone content with a mean of 38.27% and a high and low of 97.8% and 0.01%, respectively. Like dolomite, the standard deviation of the limestone content is high. Limestone content is positively skewed but fairly symmetrical as shown in Fig. 4-12.

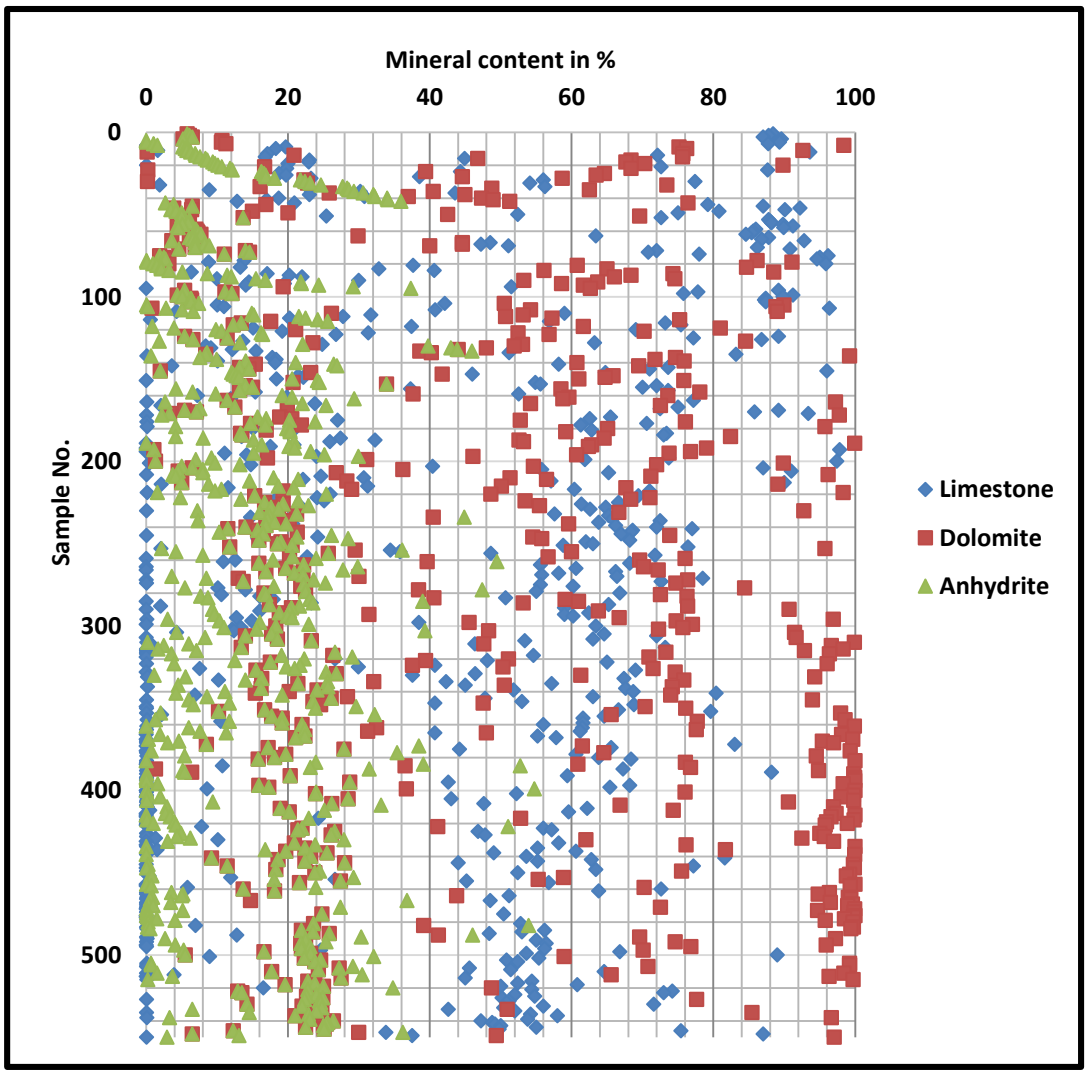


Figure 4-11: Mineral content Vs Sample No.

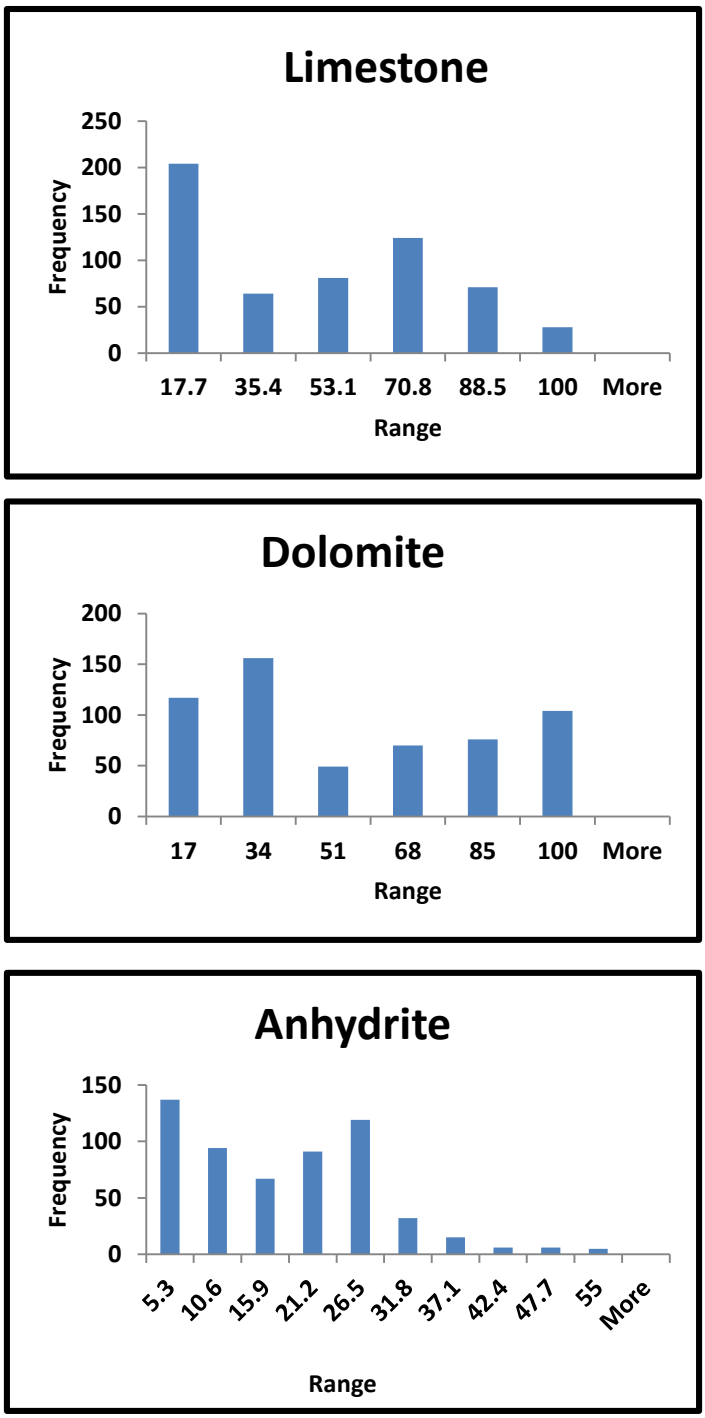


Figure 4-12: Histogram of mineral content based on Freedman-Diaconis rule.

4.3 Crossplots and Correlation Coefficients

After the preliminary data analysis, the crossplot is another method to visualize petrophysical data. A crossplot is a two-dimensional plot between two variables. It is a clever approach which can reveal formational features better than any log vs depth or log vs sample plots. Crossplots help in identifying trends and clusters and help understand distribution of data with respect to each other. Usually, these plots are linear but could be logarithmic based on the variable being plotted. For permeability, logarithmic scale is generally used. Crossplots of permeability with each other parameter are presented later in this section.

Correlation coefficient is a method to quantify closeness or inter-dependence of two variables over each other. It varies from 1 to -1. Correlation coefficient of 0 indicates no relationship. Positive correlation coefficient means directly proportional and negative means indirectly proportional relationship. Correlation coefficients for the available data are presented in Table 4-2 and are graphically represented in Fig. 4-14. It has been noticed that like every conventional permeability problem, logarithmic value of permeability returns better correlation coefficients with other parameters. This has prompted to plot log of permeability in every crossplot discussed in this section.

Table 4-2: Correlation Coefficient matrix

	Permeability, K (mD)	Log k	Porosity, ϕ (fraction)	Density, ρ (g/cc)	Limestone Content, ML (%)	Dolomite Content, MD(%)	Anhydrite Content, MA (%)
Permeability, K (mD)	1.00						
Log k	0.59	1.00					
Porosity, ϕ (fraction)	0.32	0.70	1.00				
Density, ρ (g/cc)	-0.17	-0.20	-0.22	1.00			
Limestone Content, ML (%)	0.04	-0.21	-0.24	-0.36	1.00		
Dolomite Content, MD(%)	-0.08	0.14	0.17	0.39	-0.94	1.00	
Anhydrite Content, MA (%)	0.12	0.18	0.21	-0.05	-0.05	-0.28	1.00

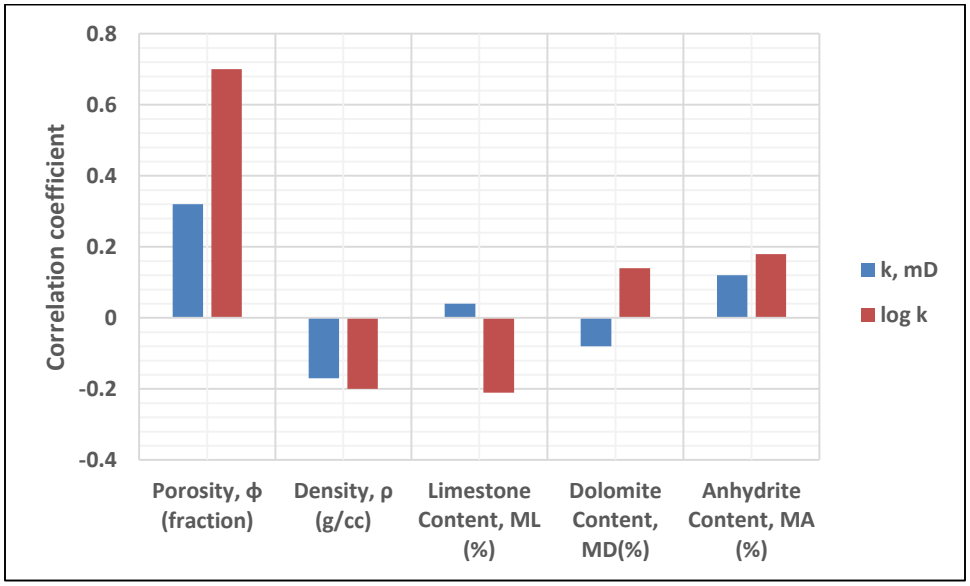


Figure 4-13: Comparison of Correlation Coefficients between permeability in mD and logarithmic value of permeability.

4.3.1 Permeability Vs Porosity

A crossplot of permeability vs porosity is presented in Fig. 4-15. The plot shows permeability to be directly proportional to porosity. Correlation coefficient is 0.7, which, as expected, indicates a strong relationship between both the parameters. Lower permeability values show unrealistic behavior and represent lowest reading of the measurement method. For example, permeability value of 0.1 mD remains constant despite huge variation across porosity axis. As indicated in section 4.2.1 there are certain irregularities in the data set. This will be further analyzed in the discussion on hydraulic flow units. The data contains some very high permeability values, which could be vugular/cracked cores but the bulk of data falls under 100 mD.

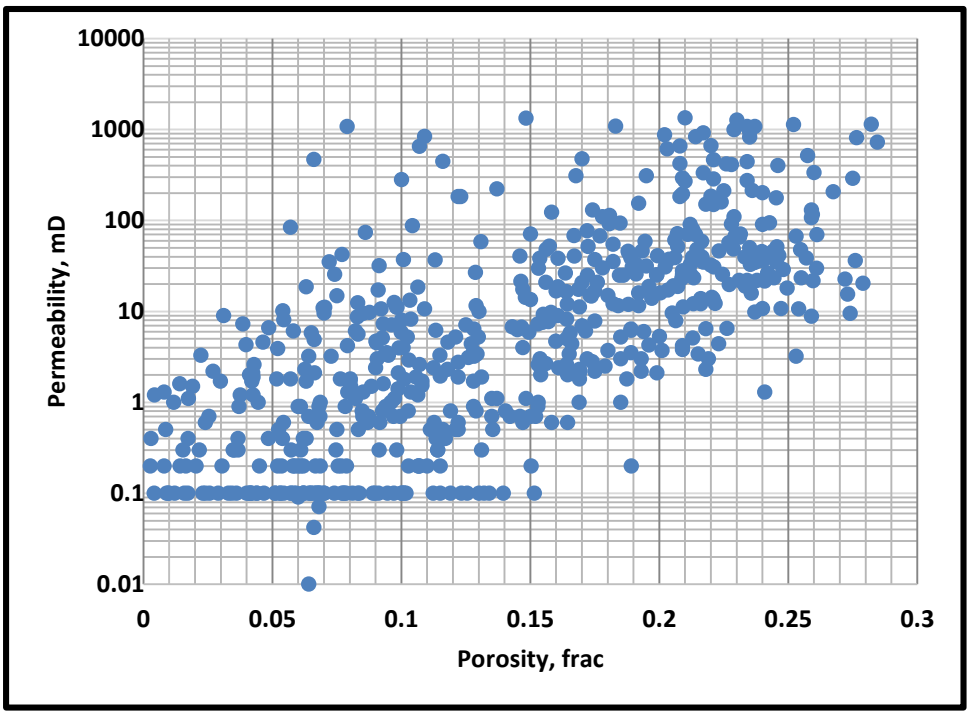


Figure 4-14: Crossplot of permeability vs porosity showing a directly proportional behavior.

4.3.2 Permeability vs Bulk Density

Crossplot of permeability vs bulk density is shown in Fig. 4-16. It is observed that there are two major groups of points in this crossplot. The first group shows permeability to be indirectly proportional to bulk density. The second group shows that with a small variation in bulk density, the permeability changes dramatically. Correlation coefficient is coming out to be -0.2, which can explain the first group but cannot explain the second. It looks like as if there is no correlation between permeability and bulk density for the second group. This matter will be further analyzed in the section on data filtration.

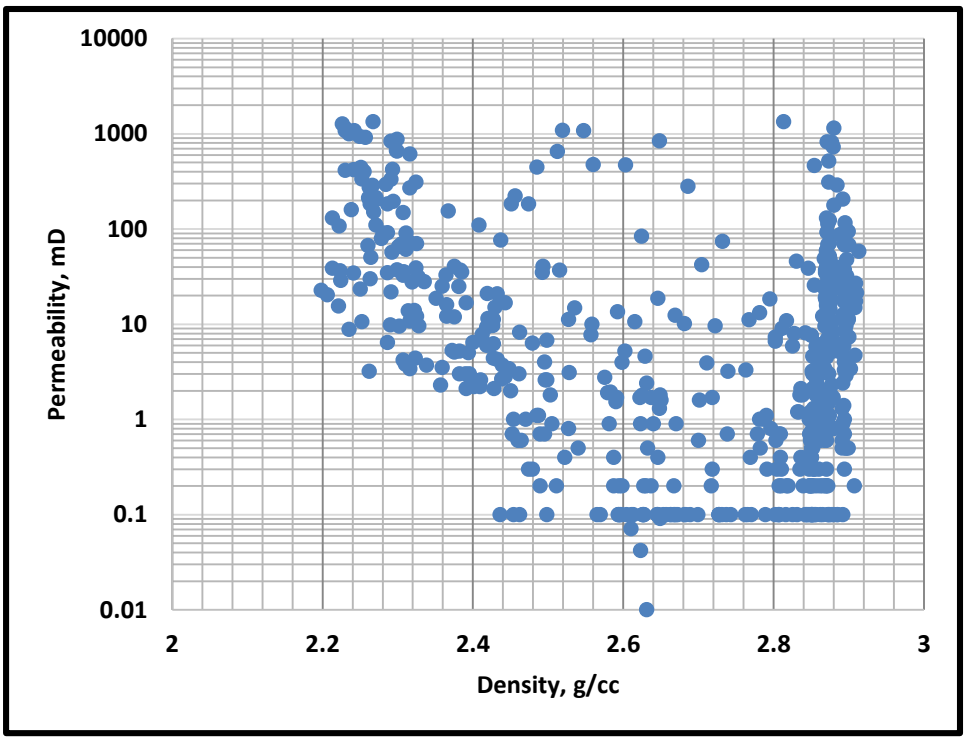


Figure 4-15: Crossplot of permeability vs density showing two trends.

4.3.3 Permeability Vs Limestone Content

As this study is intended to correlate permeability with mineral content, it was necessary to analyze crossplots of permeability with various minerals present in the rock even though the correlation coefficient came out to be -0.21, which indicates an indirect relationship between both parameters. However, the crossplot (Fig. 4-17) reveals that different trends exist within this data. Data points with 50-80% limestone content show an indirect relationship, which is confirmed by the correlation coefficients as well. But data points with very low or negligible limestone content shows no behavior. The rest of the data points are scattered all over.

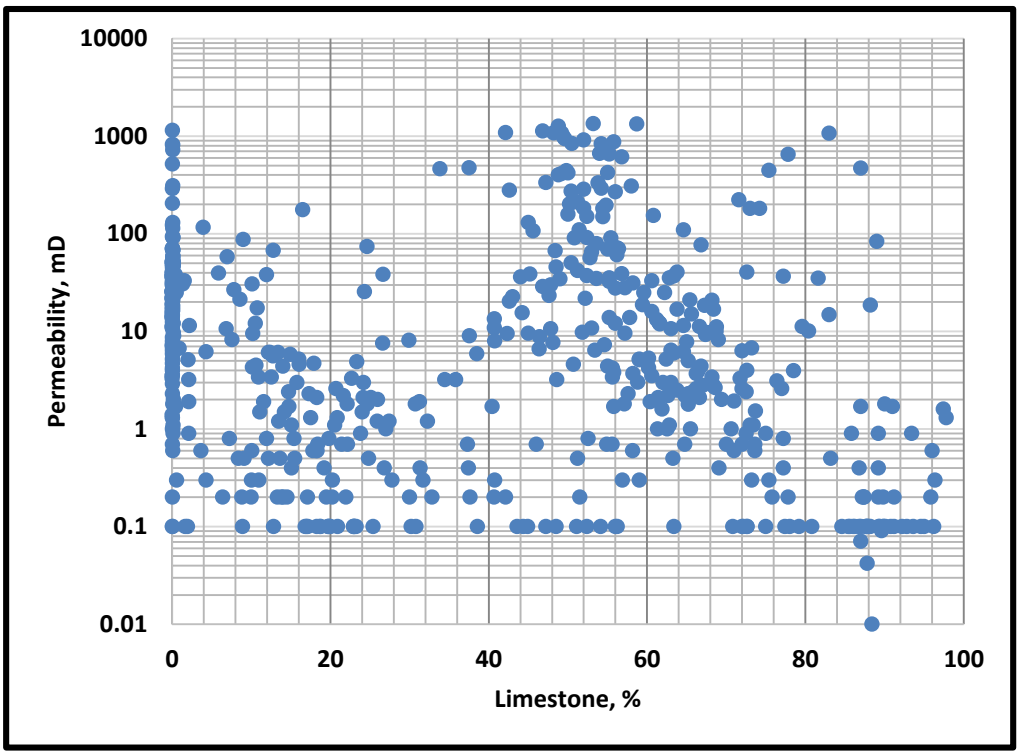


Figure 4-16: Crossplot of permeability vs limestone content

4.3.4 Permeability Vs Dolomite Content

A crossplot of permeability vs. dolomite content is given in Fig. 4-18. The correlation coefficient between the parameters is 0.14, which shows very weak correlation. But the positive correlation coefficient shows they both are directly proportional to each other. Only data points with 10-30% dolomite content depict this trend; the rest of the points are scattered and show no direct relationship.

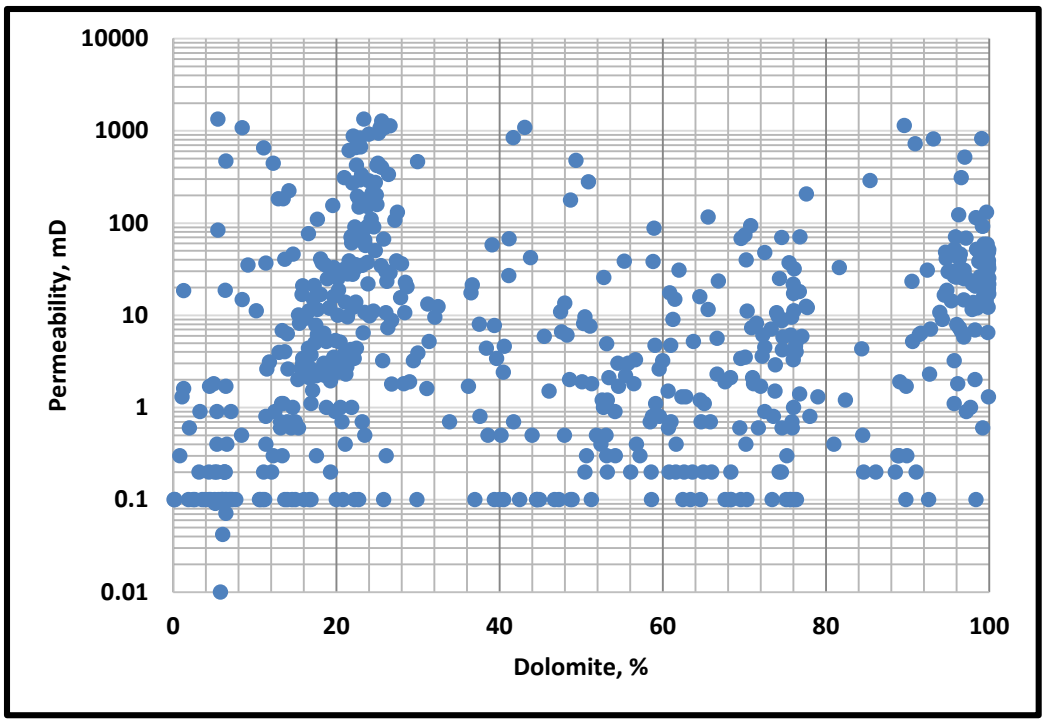


Figure 4-17: Crossplot of permeability vs Dolomite content

4.3.5 Permeability Vs Anhydrite Content

Crossplot of permeability vs anhydrite content is shown in Fig. 4-19. Correlation coefficient between both the parameters came out to be 0.21, which reveals a direct relationship between the parameters. This is contrary to the literature as anhydrite is considered as a plugging material in carbonate formations and it usually destroys the permeability. But the nature of this data cannot be overlooked as it shows a different trend. There must be underlying reasons, which should be further investigated. But due to the limitation of information regarding the data provided, this matter couldn't be further investigated. Despite the scattered data, the higher anhydrite content shows direct relationship.

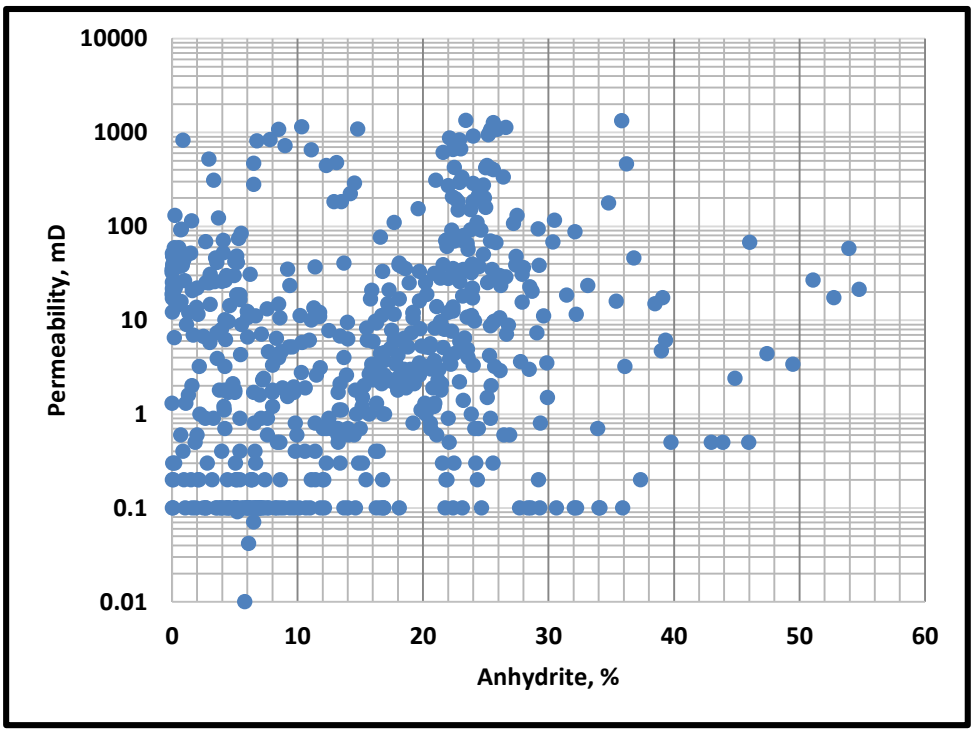


Figure 4-18: Crossplot of permeability vs Anhydrite content

4.4 Variation of Permeability with Mineral Content

After preliminary data analysis and crossplots, it has been established that there is a strong correlation between permeability and porosity. But crossplots of permeability and mineral content show weak correlation. This study is essentially conducted to correlate permeability with mineral content. To further investigate this aspect of the problem, the variation of mineral content is plotted on a crossplot of permeability vs porosity for each mineral.

The dolomite content of each data point is indicated on permeability vs porosity crossplot as shown in Fig. 4-20. Dolomite content is divided into 10 intervals of 10% difference each. Data is scattered yet it can be observed that higher dolomite content

concentrate in the higher permeability and porosity region on the crossplot. However, there is no clear demarcation that can simplify the problem at hand.

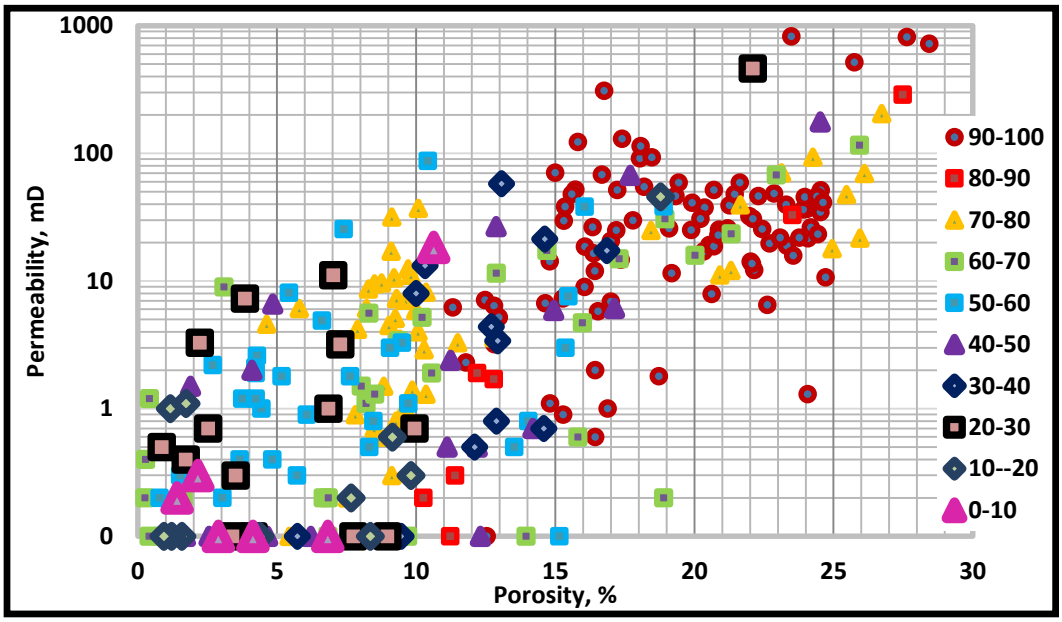


Figure 4-19: Variation of Dolomite content on permeability vs porosity crossplot.

Similarly, the variation of limestone content on permeability vs porosity crossplot is presented in Fig. 4-21. Data points with negligible limestone content occupy the region of high permeability and porosity. Rest of the data is scattered but it can be deduced that permeability is indirectly proportional to limestone content, which was also verified by the crossplot of permeability and limestone.

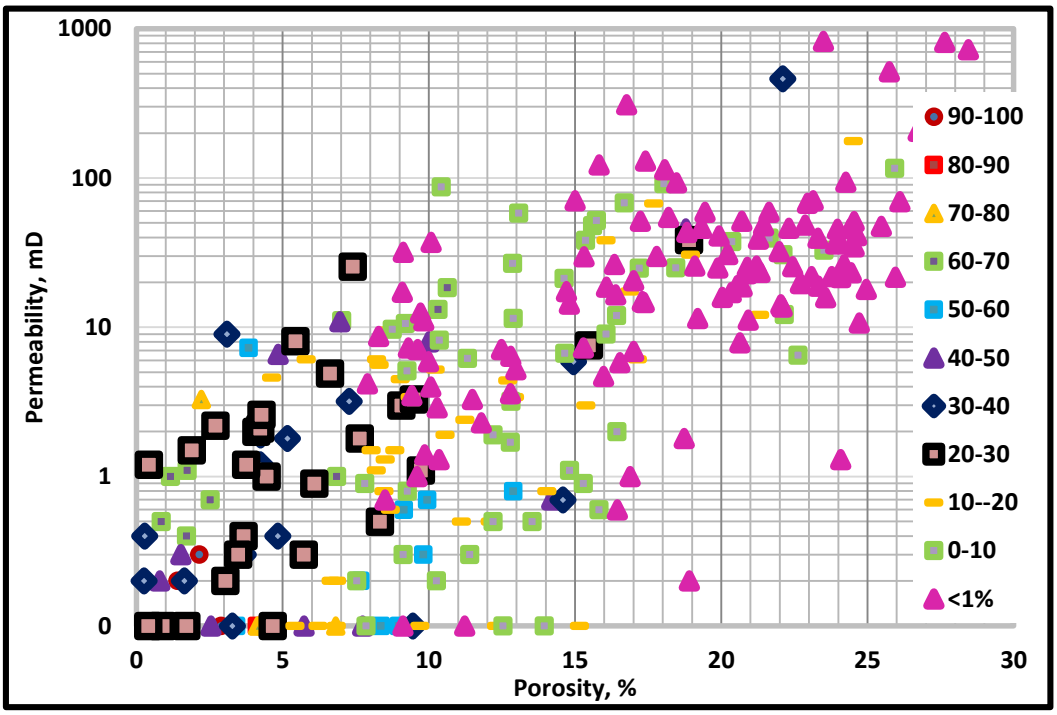


Figure 4-20: Variation of Limestone content on permeability vs porosity crossplot

The anhydrite content was also analyzed as shown in Fig. 4-22. No clear trend can be observed from this plot as the data is scattered and it is difficult to discern any conclusion.

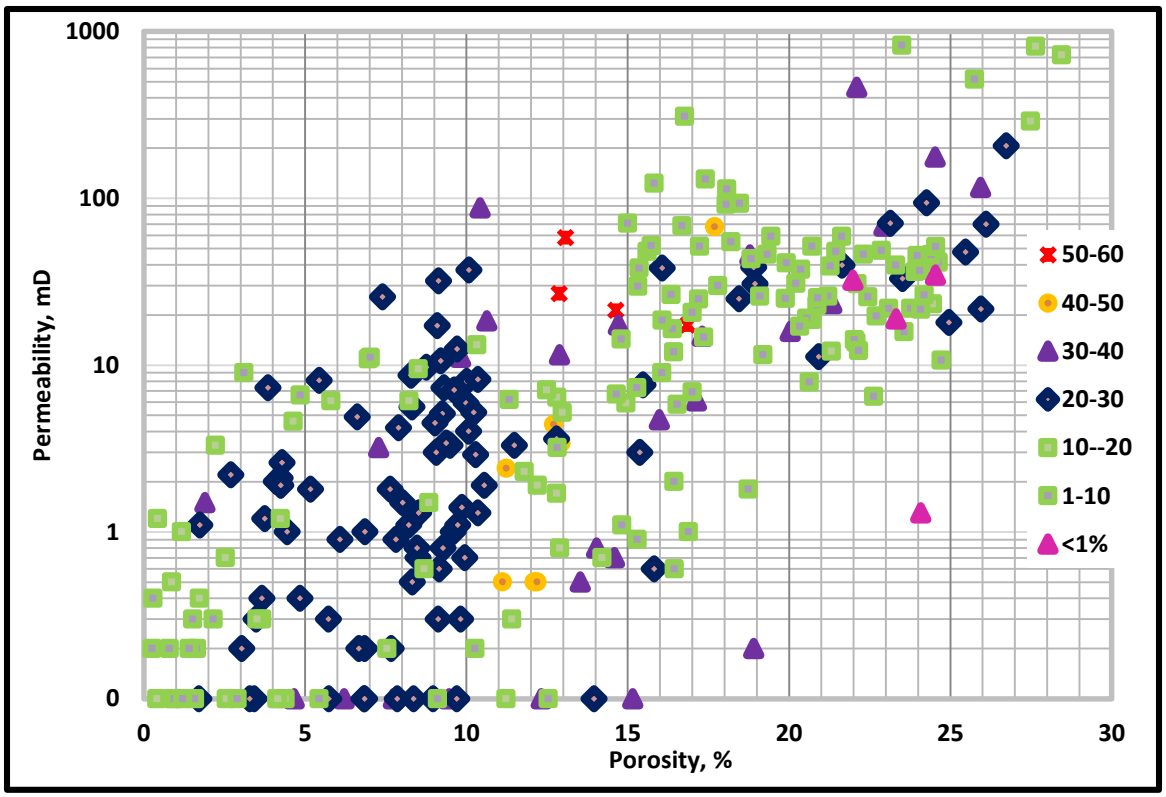


Figure 4-21: Variation of Anhydrite content on permeability vs porosity crossplot

4.5 Hydraulic Flow Units

An elementary volume of reservoir rock which has similar geological and petrophysical properties can be identifies as a hydraulic flow unit [36]. According to Ebanks [37], a mappable portion of the reservoir within which the geological and petrophysical properties that affect the flow of fluid are consistent and predictably different from the properties of other reservoir rock volume is known as hydraulic flow units.

To evaluate hydraulic flow units, reservoir quality index (RQI), normalized porosity (ϕ_z) and flow zone indicator (FZI) need to be determined. These parameters can be calculated by the following equations:

$$RQI\ (\mu m)\ =\ 0.0314\sqrt{\frac{k\ (mD)}{\Phi}} \tag{4.7}$$

$$\phi_z = \frac{\phi}{1-\phi} \tag{4.8}$$

$$\log\ (\text{RQI}) = \log\ (\phi_z) + \log\ (\text{FZI}) \tag{4.9}$$

Where,

K = permeability, mD

ϕ = porosity, fraction

As shown by the line equation, log-log plot of RQI vs ϕz will have a unit slope with FZI as an intercept as shown in Fig. 4-23. Data points having similar FZI will lie on the straight line of slope 1. This FZI will help distinguish one hydraulic flow unit from the other.

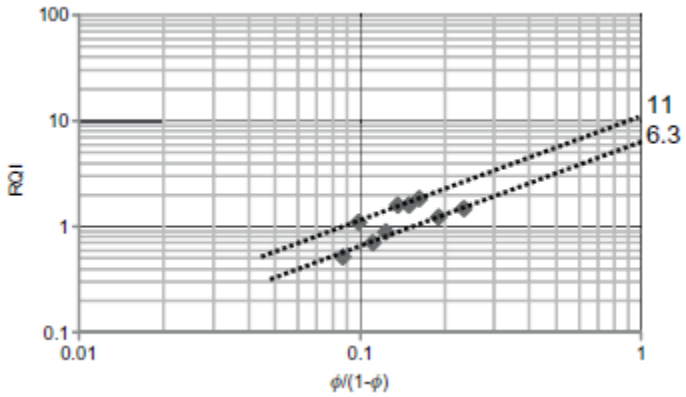


Figure 4-22:Unit slope lines to identify inherent hydraulic flow units.

The log-log plot of RQI vs ϕ_z for the given data is shown in Fig. 4-24. As can be observed that points are not aligned and segregated along the unit slope straight line. This proves the complexity of the data at hand and the need to find out a different approach to differentiate the hydraulic flow units. The straight-line present at the lower end of the plot is due to assigned permeability values of 0.1 mD, which are misleading and need to be removed from the data for further analysis. This step is carried out in the data filtration process discussed next.

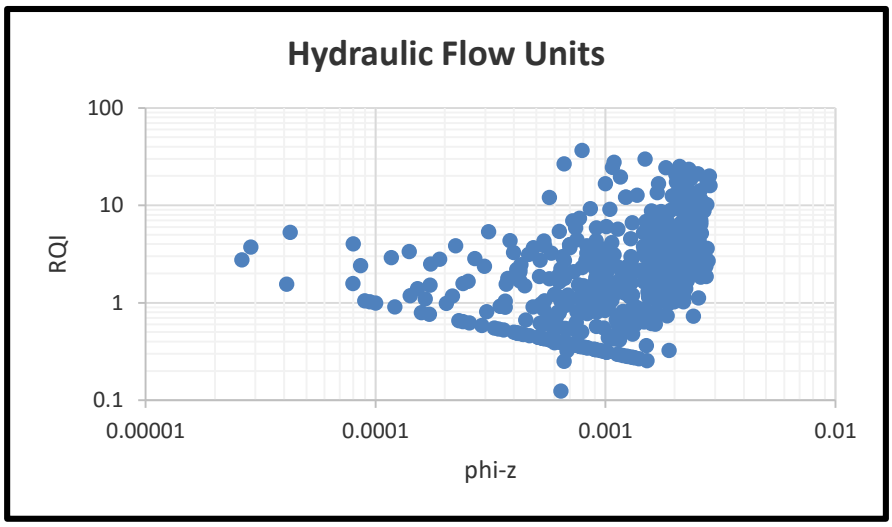


Figure 4-23: log-log plot of RQI vs ϕz to identify hydraulic flow units.

4.6 Data Filtration

Data filtration is one of the most important part of statistical data analysis. There are many reasons to filter data to improve its quality and the results such as removal of some discrepancies that can be present. One of such discrepancies could be presence of klinkers, which are data points that have incorporated errors while data gathering such as equipment failure or equipment limitations. Such data points should be discarded and additional data points should be collected if feasible. Some researchers argue that if we insist to keep such data points then they would destroy the spirit of the statistical investigation process [38]. The other type of discrepancy could be outliers. Outliers are legitimately gathered data points that have a deviant nature, which is not at all due to any equipment failure or similar causes. Outliers are not easily discarded like klinkers, and removing outliers may introduce certain limitations to the intended results.

4.6.1 Logical Data Filtering

In this step, klinkers are removed. Data points with 0.1 mD permeability are considered klinkers because the original data report states that cores, which didn't have any permeability or were very tight were assigned 0.1 mD permeability. Also, data points with larger than 1 D permeability, which most likely indicates vugular/fractured cores, were also removed. Lastly, washout zones where the hole diameter is excessively large as revealed by the caliper log were identified and data points corresponding to those zones were removed as well.

4.6.2 Filtering based on density

After removing the klinkers several regression techniques were applied to the data to formulate a correlation. All attempts to correlate permeability with mineral content failed to return any meaningful results. This was construed as either there was no correlation between the parameters or the data needed further investigation. In section 4.3.2, where a crossplot of permeability vs bulk density was discussed, two trends were witnessed and it was pointed that the second trend deserved further investigation. The second trend show permeability changing quite drastically over a very small range of density. This trend corresponds to core samples with high dolomitization.

The average grain density of a rock is given by:

$$\bar{\rho}_g = \sum v_i \rho_{gi} \tag{4.10}$$

Where,

v_i = volume fraction of i'th mineral

ρ_{gi} = density of i'th mineral, g/cc

For this case,

i= limestone, dolomite and anhydrite

hence, average grain density will be given by

$$\rho_g = M_L\rho_L + M_D\rho_D + M_A\rho_A \tag{4.11}$$

Where, M_L , M_D and M_A are the fractional mineral content of limestone, dolomite and anhydrite, respectively, while ρ_L, ρ_D and ρ_A are densities of limestone, dolomite and anhydrite, respectively, in g/cc.

Using Eq. 4.11, average grain density was calculated. Since, the density provided by the data is bulk density, the grain density should only match bulk density when porosity is equal to zero. The absolute percentage error between the average grain density (calculated) and bulk density (given) was then computed and plotted against porosity in Fig. 4-25. Conventionally, the error should be a strong function of porosity and should increase as the porosity increases because greater the porosity, the more offset would be average grain

density from bulk density. When absolute percentage error was plotted against porosity Eq. 4.12, it broke down the data into two clusters. One showed absolute percentage error increasing with the increment in porosity as expected. The other showed that the absolute percentage error is zero irrespective of porosity variation. This essentially means that for this cluster, the average grain density is equivalent to bulk density, which is not possible. This is depicted in Fig. 4-25.

$$relative\ difference\ (\%)\ = \frac{\rho_g - \rho_B}{\rho_g} * 100 \qquad (4.12)$$

$$relative\ difference\ (\%)\ \alpha\ \phi$$

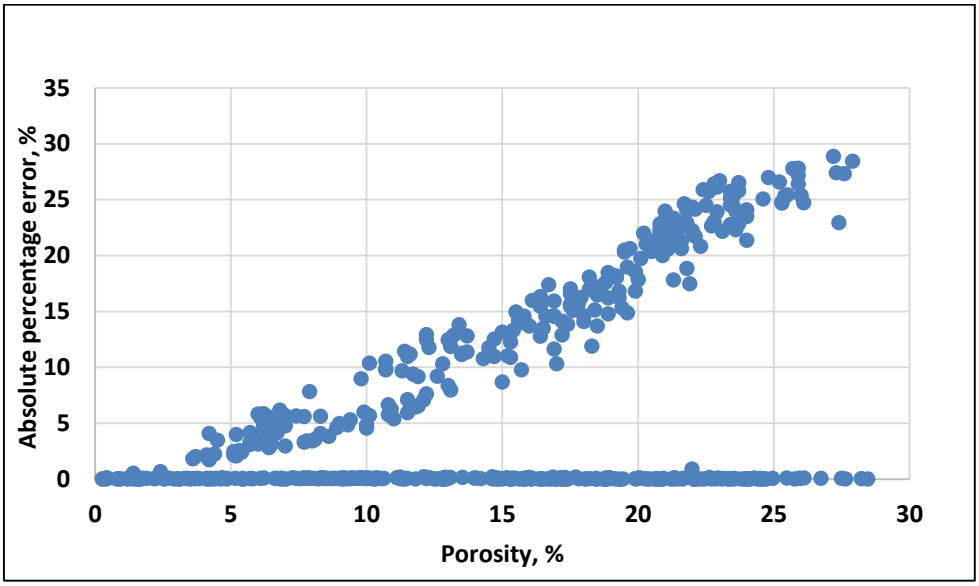


Figure 4-24: Average Absolute Percentage of Error (between grain density and bulk density) Vs Porosity.

Thus, the data points in this cluster were deemed erroneous and were removed from the data set. The remaining number of data points, after the application of filtration based

on density, is 226. This proved to be a breakthrough in this study and clustering and regression techniques applied afterwards showed good results.

CHAPTER 5

CLUSTERING

In the previous chapter, the data was analyzed and filtered. In this chapter, an attempt is made to identify hidden patterns and unique characteristics within the data. For this purpose, fuzzy C-means (FCM) clustering, K-means (KM) clustering algorithm and expectation maximization (EM) algorithm were employed. These are powerful machine learning tools. Then, clustering is carried out with the help of graphical representation of data [39].

Clustering may be defined as an organizing tool which can help divide data into groups of identical nature. It divides data in such a fashion that data point belonging to one group are similar to points in the same group as compared to points in any other group. The major types of clustering method include supervised and unsupervised clustering. **Supervised clustering** is a type of machine learning process where training is carried out for every input with corresponding target. The major application of this type of machine learning is classification and regression analysis. **Unsupervised clustering**, on the other hand, is a type of machine learning process where training is carried out with a set of input parameters only. Major applications of unsupervised learning are identification of structures and different relationships within the data [40].

5.1 Unsupervised Clustering

As described above, unsupervised learning is carried out when only input data is available with no corresponding output variables. Main objective for unsupervised learning

is to model and identify underlying structures and distribution in the set of data in order to learn more about the data. For this purpose, algorithms based on unsupervised learning themselves devise to discover and present the unique structures within the data [41]. The nature of this study is also such that we have parameters yet we don't have any scheme for data clustering. This is the reason to apply unsupervised clustering techniques to our problem. There are many researchers who have demonstrated the use of cluster analysis for the classification of rock masses, electro-facies, textural properties etc. [42][43]

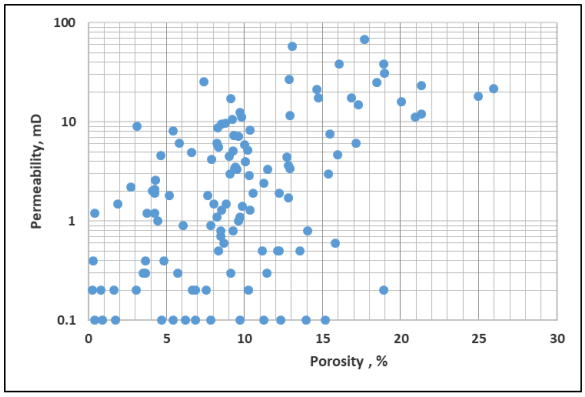
5.1.1 Fuzzy C-Means Clustering (FCM)

Fuzzy c-means (FCM) is a data clustering technique in which a dataset is grouped into n clusters with every data point in the dataset belonging to every cluster to a certain degree. For example, a certain data point that lies close to the center of a cluster will have a high degree of belonging or membership to that cluster and another data point that lies far away from the center of a cluster will have a low degree of belonging or membership to that cluster. It starts with an initial guess for the cluster centers, which are intended to mark the mean location of each cluster. The initial guess for these cluster centers is most likely incorrect. Next, function assigns every data point a membership grade for each cluster. By iteratively updating the cluster centers and the membership grades for each data point, function iteratively moves the cluster centers to the right location within a data set. This iteration is based on minimizing an objective function that represents the distance from any given data point to a cluster center weighted by that data point's membership grade. For further in-depth study, references [44], [45], [38] and [39] can be inspected.

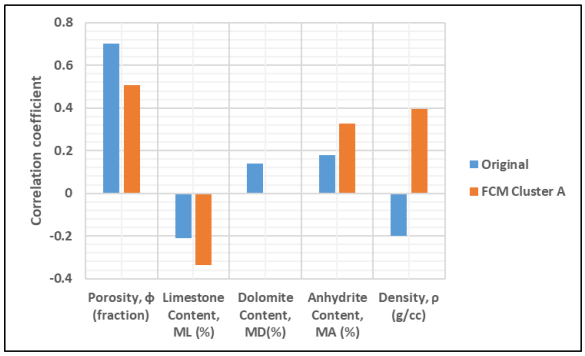
When applied to the filtered data set, FCM returned three clusters singling out some outliers. Characteristics of these clusters are presented in the following discussion.

Cluster A

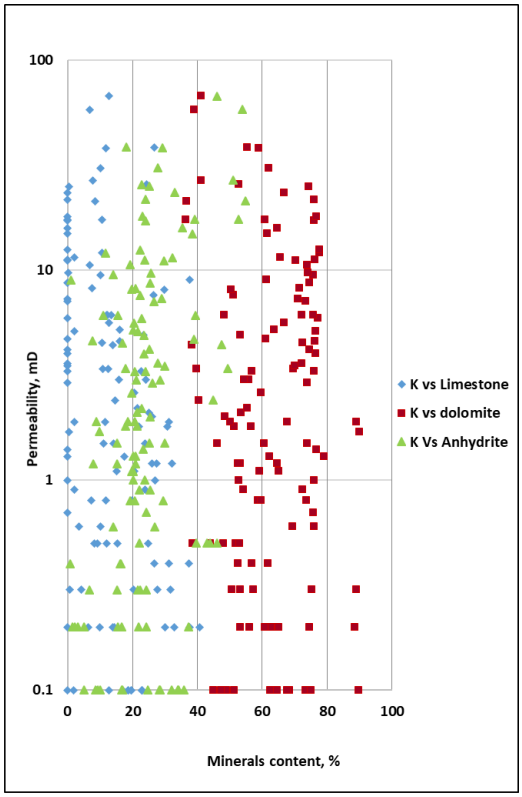
In this cluster, algorithm segregated data points having less than 40% limestone and greater than 40% dolomite content. Correlation coefficients of porosity and dolomite with log of permeability have been reduced while of limestone, anhydrite and density have been increased. All these observations are presented in Fig. 5-1. It can also be noticed that direction of proportionality in case of density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.



(a)



(b)

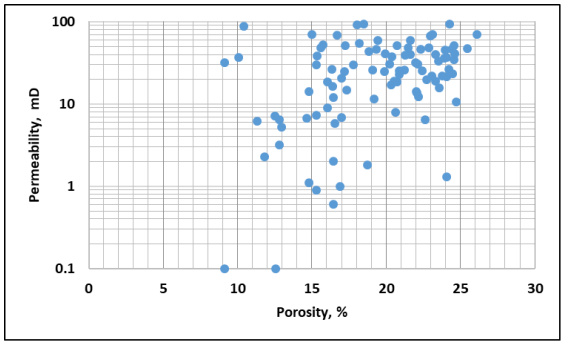


(c)

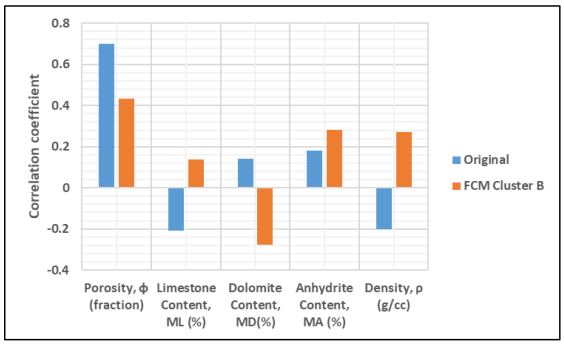
Figure 5-1: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and FCM Cluster A. (c) Permeability Vs Mineral Content for the FCM Cluster A.

Cluster B

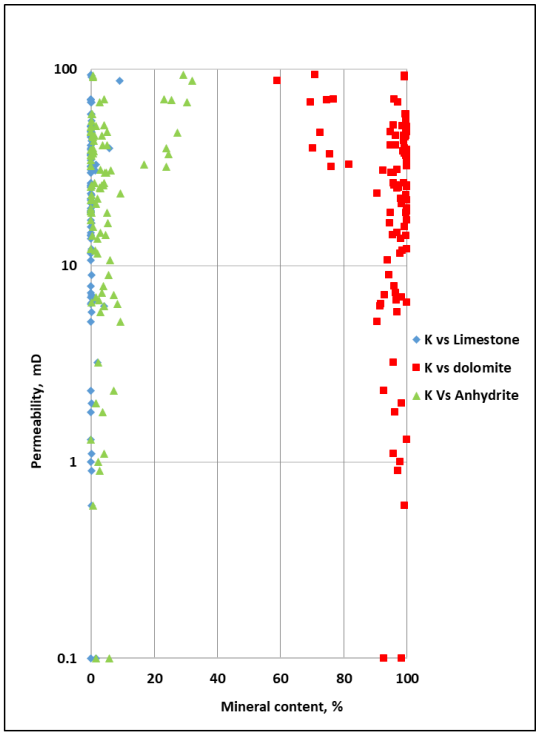
In this cluster, algorithm segregated data points having negligible or 0% limestone content. Correlation coefficients of porosity and limestone with log of permeability were reduced while of dolomite, anhydrite and density were increased. All these observations are presented in Fig. 5-2. It can also be noticed that direction of proportionality in case of limestone, dolomite and density have been totally reversed. Permeability vs porosity crossplot shows scattered data points.



(a)



(b)

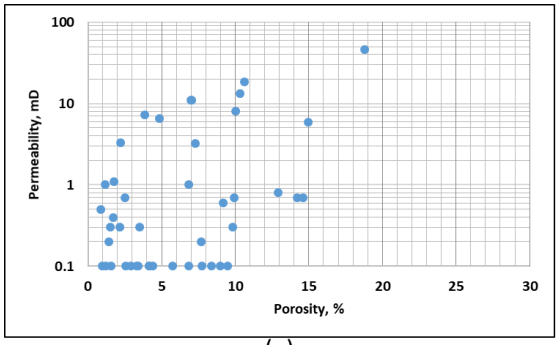


(c)

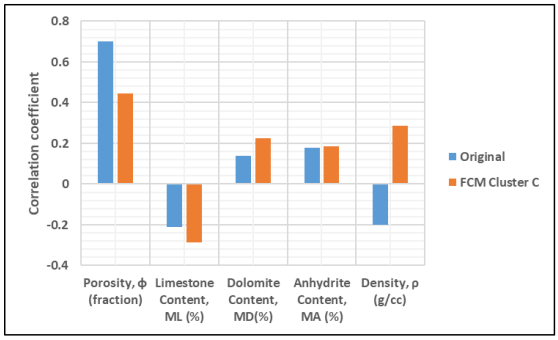
Figure 5-2: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and FCM Cluster B. (c) Permeability Vs Mineral Content for the FCM Cluster B.

Cluster C

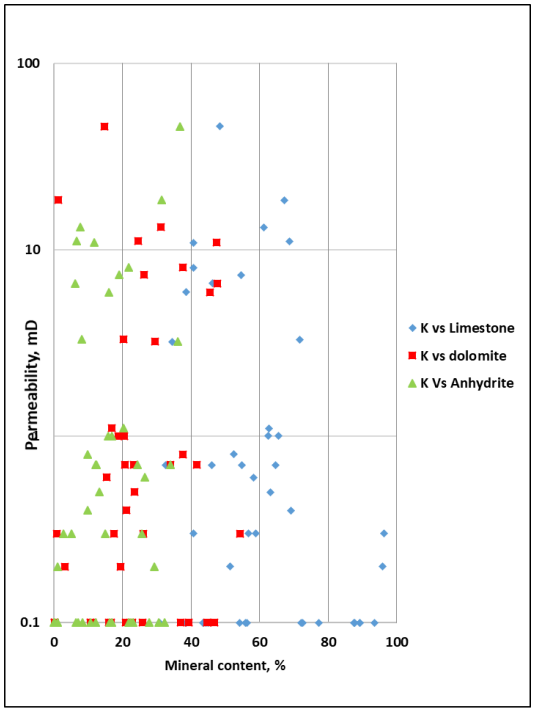
This cluster is a reflection of cluster A. In this cluster, algorithm segregated data points having greater than 40% limestone and less than 40% dolomite content. Correlation coefficients of porosity with log of permeability have been reduced while of dolomite, limestone and density have been increased. For Anhydrite, correlation coefficient is almost same. All these observations are presented in Fig. 5-3. It can also be noticed that direction of proportionality in case of density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.



(a)



(b)



(c)

Figure 5-3: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and FCM Cluster C. (c) Permeability Vs Mineral Content for the FCM Cluster C.

Application of FCM algorithm improved the correlation coefficients between dependent and some independent variables yet it was unable to identify very clear patterns within the system. Due to this reason, clustering of data based on other algorithms was pursued.

5.1.2 K-means Algorithm

K-means (MacQueen, 1967)[48] is one of the simplest unsupervised learning algorithms that solve the well-known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster. These centroids should be placed in a crafty way because different locations cause different results. So, the better choice is to place them as far away from each other as possible. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early groupage is done. At this point we need to re-calculate k new centroids as barycenters (mean of vertices of a triangle or any irregular shape) of the clusters resulting from the previous step. After we have these k new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop is then generated. As a result of this loop we

may notice that the k centroids change their location step by step until no more changes are seen. In other words, centroids do not move any more. For further in-depth reading, reference [49] can be consulted.

Like FCM algorithm, K-means algorithm returned three clusters singling out some outliers. Characteristics of these clusters are presented in the following discussion.

Cluster A

This cluster resembles cluster B of FCM algorithm. In this cluster, algorithm segregated the data points having negligible or 0% limestone content with dolomite greater than 80%. Correlation coefficients of the porosity and limestone with log of permeability have been reduced while of dolomite and density have been increased. For anhydrite, correlation coefficient remains same. All these observations are presented in Fig. 5-4. It can also be noticed that direction of proportionality in case of dolomite and density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.

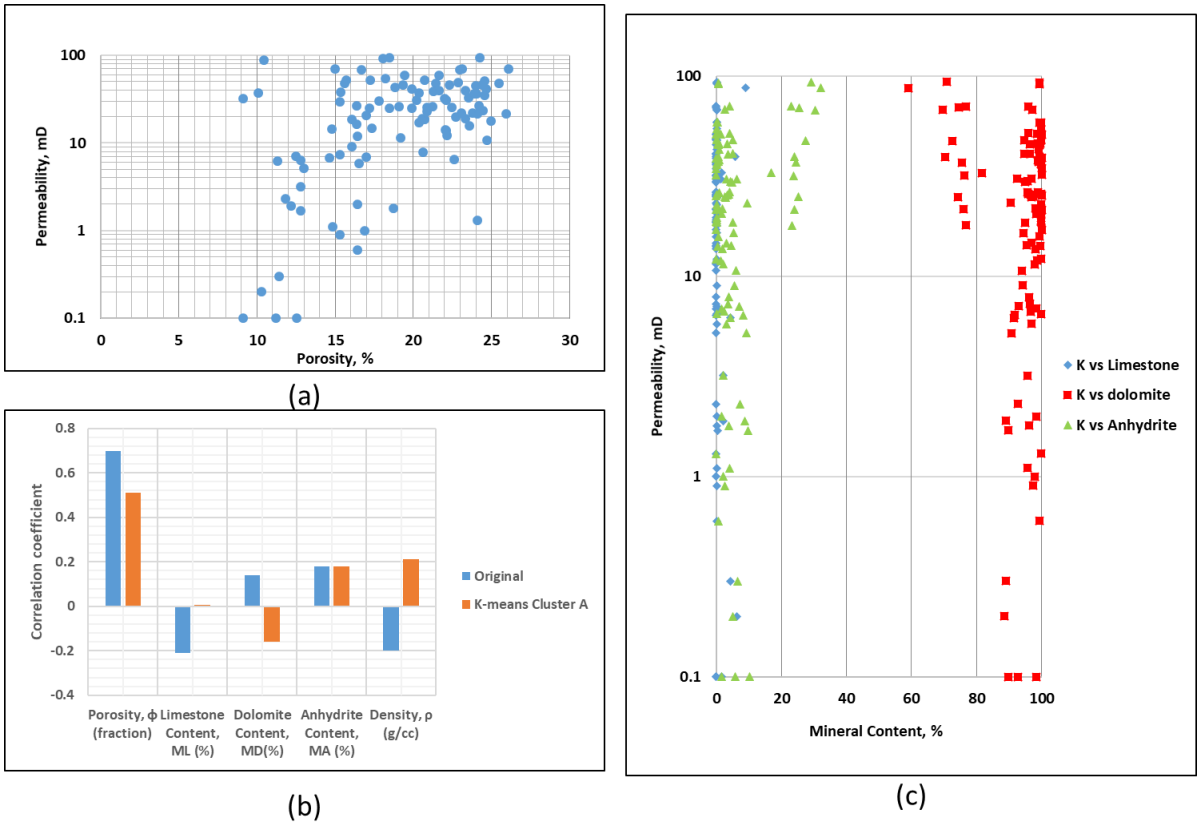


Figure 5-4: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and K-means Cluster A. (c) Permeability Vs Mineral Content for the K-means Cluster A.

Cluster B

This cluster resembles cluster C of FCM algorithm with better correlation coefficients. In this cluster, algorithm segregated data points having greater than 40% limestone and less than 40% dolomite content. Correlation coefficients of porosity with log of permeability have been reduced while of limestone, dolomite, anhydrite and density have been increased. All these observations are presented in Fig. 5-5. It can also be noticed that

direction of proportionality in case of density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.

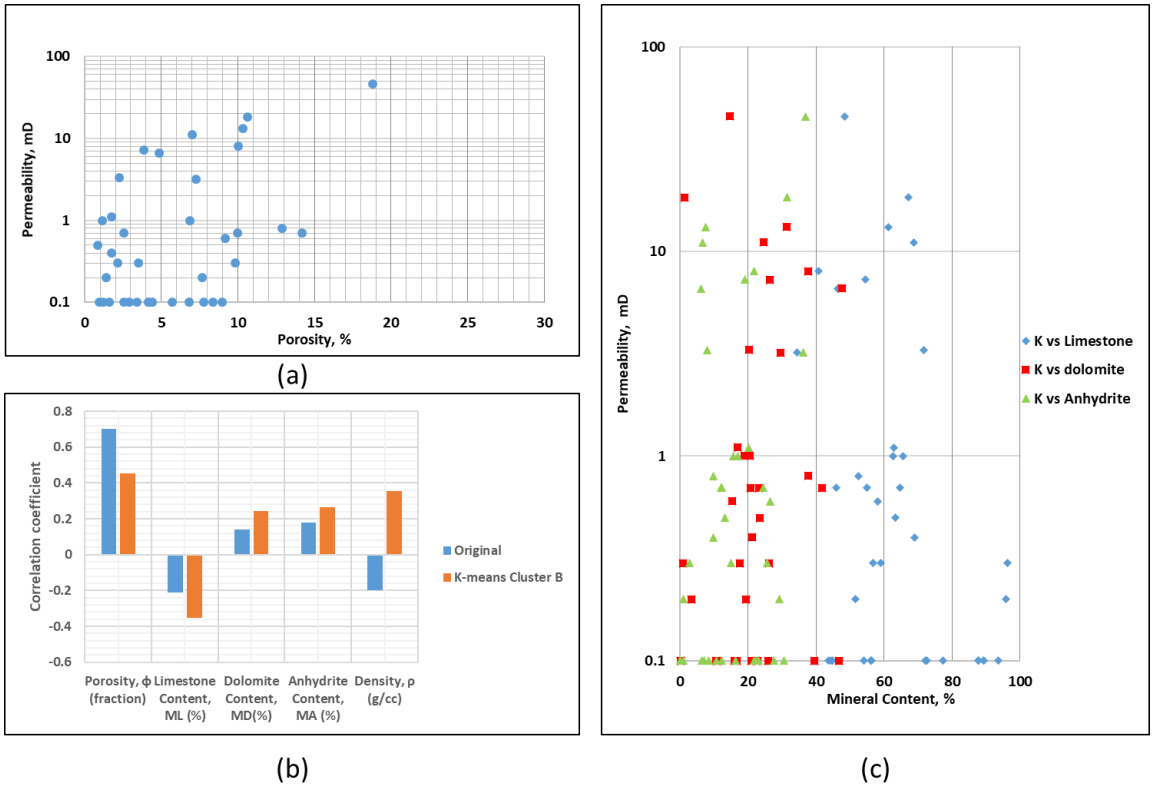


Figure 5-5: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and K-means Cluster B. (c) Permeability Vs Mineral Content for the K-means Cluster B.

Cluster C

This cluster resembles cluster A of FCM algorithm with almost similar correlation coefficients. In this cluster, algorithm segregated data points having less than 40% limestone and greater than 40% dolomite content. Correlation coefficients of porosity and dolomite with log of permeability have been reduced while of limestone, anhydrite and

density have been increased. All these observations are presented in Fig. 5-6. It can also be noticed that direction of proportionality in case of density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.

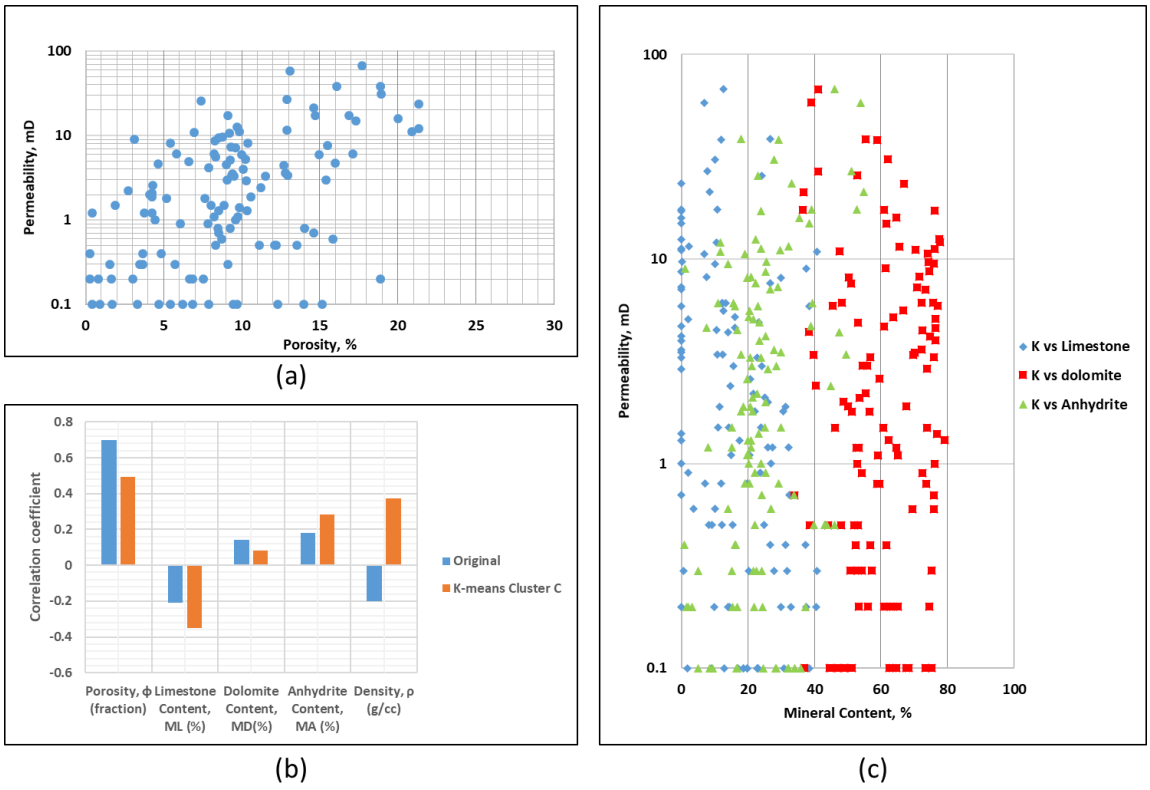


Figure 5-6: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and K-means Cluster C. (c) Permeability Vs Mineral Content for the K-means Cluster C.

It can be deduced that K-means and FCM algorithm returned almost similar clusters. This is an indication of inherent patterns that exist with the system which were identified by both algorithms employed. Further analysis was carried out based on these clusters which will be discussed in the next chapter.

5.1.3 EM (Expectation-Maximization) Algorithm

Expectation-maximization (EM) algorithm is an iterative process to find maximum likelihood or maximum a posteriori (MAP) estimates parameters in statistical models where the models depend on unobserved latent (hidden) variables. EM algorithm is based on an expectation (E) step and a maximization (M) step. E and M steps are carried out alternatively. Expectation (E) step creates a function for the expectation of the log-likelihood evaluated using the current estimate for the parameters. While maximization (M) step computes parameters maximizing the expected log-likelihood found on the E step. These parameter-estimates are then used to determine the distribution of the latent variables in the next E step. A detailed background and theory of EM algorithm can be found in documents presented by references [50], [51], [52], [53], [54].

EM algorithm returned five clusters singling out some outliers. Characteristics of these clusters are presented in the following discussion:

Cluster A

In this cluster, algorithm segregated data points randomly and a pattern was not found. Correlation coefficients of porosity and limestone with log of permeability have been reduced while of dolomite, anhydrite and density have been increased. All these observations are presented in Fig. 5-7. It can also be noticed that direction of proportionality in case of anhydrite has been totally reversed. Permeability vs porosity crossplot shows scattered data points. This cluster contains very few data points.

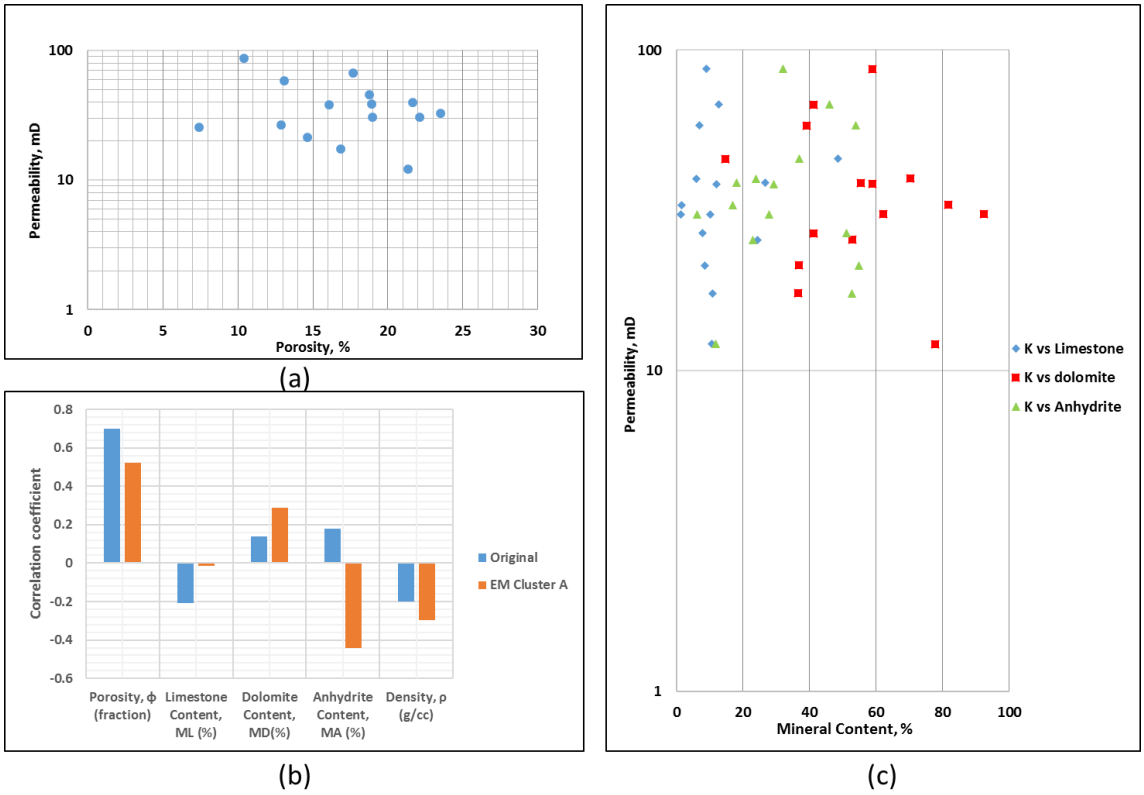


Figure 5-7: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and EM Cluster A. (c) Permeability Vs Mineral Content for the EM Cluster A.

Cluster B

In this cluster, algorithm segregated data points having almost similar mineral content with few exceptions. Correlation coefficients of porosity and limestone with log of permeability have been reduced while of dolomite and anhydrite have been increased. Density has almost similar correlation coefficient. All these observations are presented in Fig. 5-8. It can also be noticed that direction of proportionality in case of dolomite and density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.

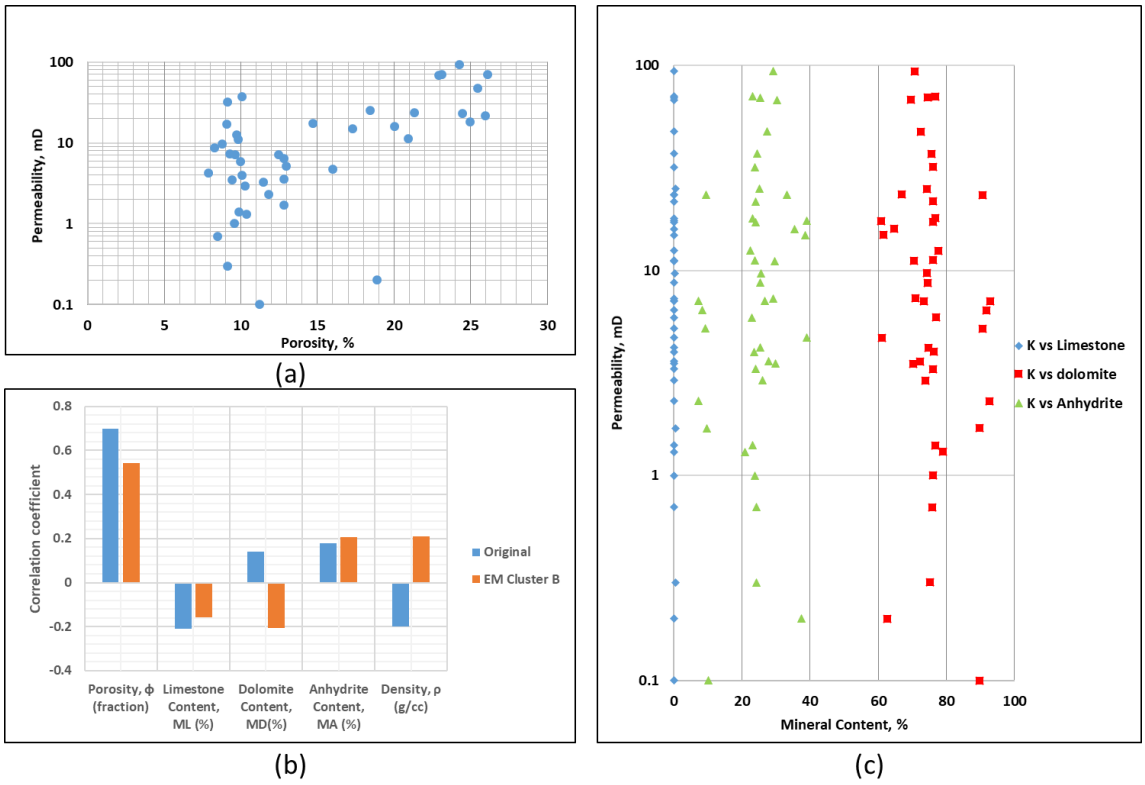


Figure 5-8: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and EM Cluster B. (c) Permeability Vs Mineral Content for the EM Cluster B.

Cluster C

In this cluster, algorithm segregated data points having greater than 40% Limestone and less than 40% Dolomite content. It is similar to cluster B of K-means algorithm and cluster C of FCM algorithm. But this cluster has lower number of data points and has chosen permeability values either less than or close to 10 mD. Correlation coefficient of porosity and anhydrite with log of permeability has been reduced while of dolomite, density and limestone has been increased. All these observations are presented in Fig. 5-9.

It can also be noticed that direction of proportionality in case of density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.

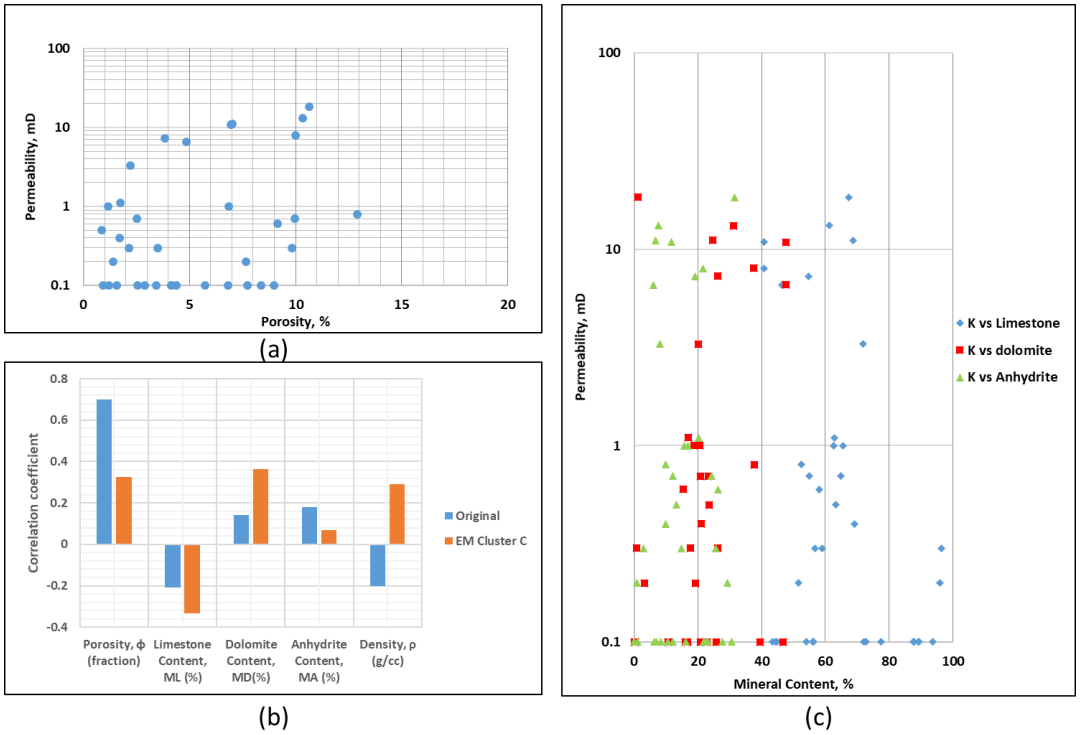


Figure 5-9: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and EM Cluster C. (c) Permeability Vs Mineral Content for the EM Cluster C.

Cluster D

In this cluster, algorithm segregated data points having greater than 40% dolomite and less than 40% limestone content. It is similar to cluster C of K-means algorithm and cluster A of FCM algorithm. But this cluster has fewer data points and has chosen permeability values either less than or close to 10 mD. Correlation coefficient of porosity, dolomite, limestone, anhydrite and density with log of permeability has been reduced. All these

observations are presented in Fig. 5-10. It can also be noticed that direction of proportionality in case of density has been totally reversed. Permeability vs porosity crossplot shows scattered data points.

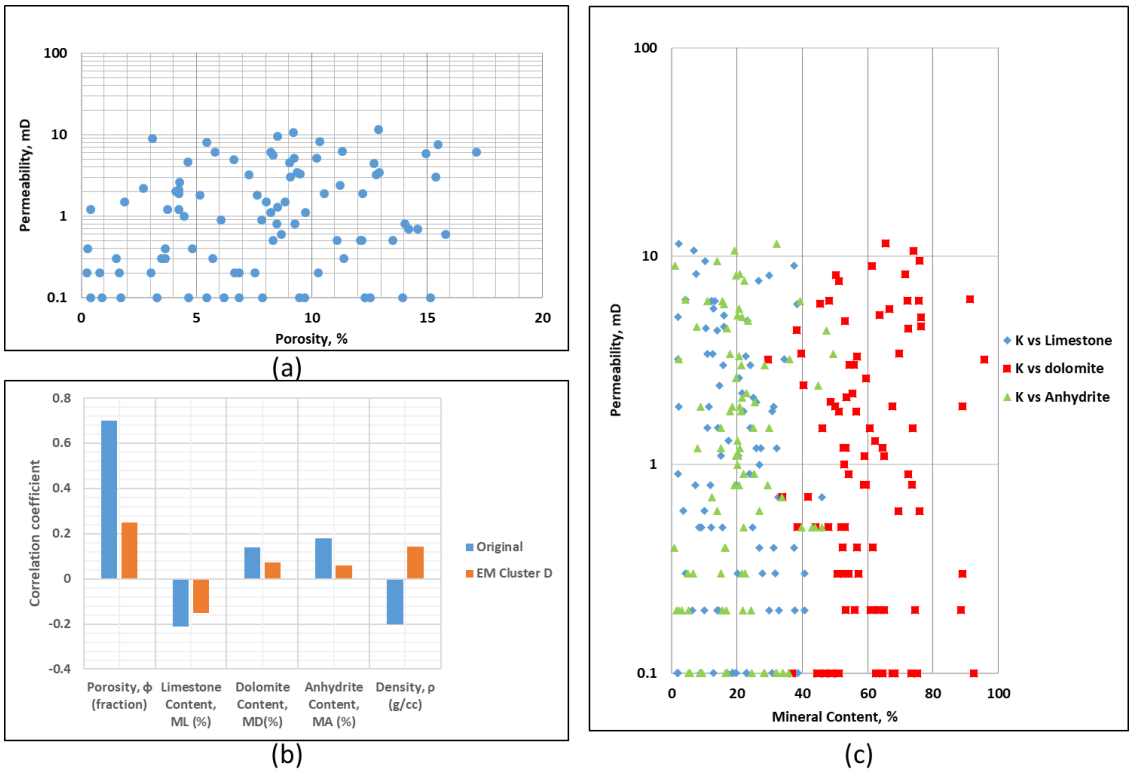


Figure 5-10: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and EM Cluster D. (c) Permeability Vs Mineral Content for the EM Cluster D.

Cluster E

In this cluster, algorithm segregated data points having almost similar values of each mineral i.e. 0% limestone and anhydrite while 100% dolomite. Correlation coefficient of porosity, dolomite, limestone, anhydrite and density with log of permeability has been reduced. All these observations are presented in Fig. 5-11. Permeability vs porosity crossplot shows a cluster of data points.

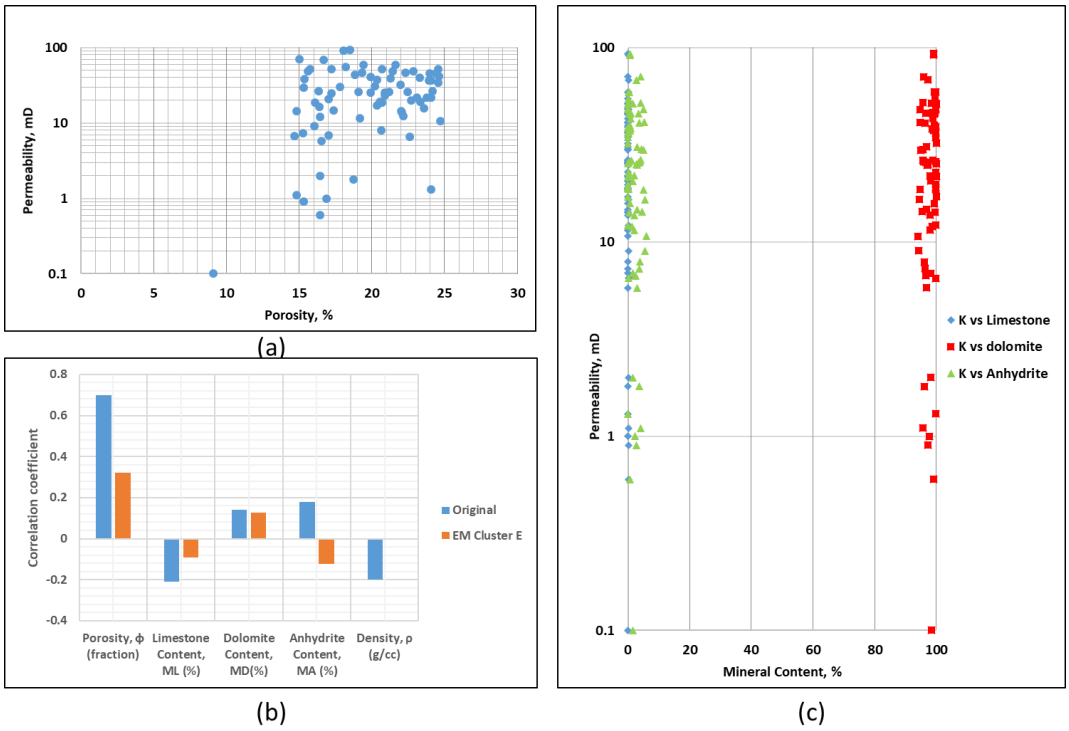


Figure 5-11: (a) Permeability Vs Porosity crossplot. (b) Comparison of correlation coefficients between original (all data) and EM Cluster E. (c) Permeability Vs Mineral Content for the EM Cluster E.

EM algorithm returned clusters mostly having lower correlation coefficients which is of prime importance to this problem. Due to this reason, further analysis of the clusters obtained by EM algorithm was not pursued.

5.2 Clustering Based on Mineral Ratio

After clustering process, different regression analysis techniques were applied to develop the correlation. All such attempts did not return the desired results, which are discussed in more details in chapter 6. This has prompted to further analysis of the data, which eventually discovered the discrepancy discussed in section 4.6.2. After removal of the data points showing 0% relative error with respect to porosity, a detailed analysis of remaining data was carried out. Fortunately, two subsets were identified.

Subsets that came into existence after many combinations tested on the data were based on the ratio of dolomite content to anhydrite content. It was discovered that a major subset of the filtered data contains equal dolomite and anhydrite contents, which can be observed in Fig. 5-12. Based on this observation, a subset of filtered data having MD/MA=1 was taken as subset 1 and rest of the points as subset 2. After the segregation of data on these grounds, data trends started making sense and eventually resulted in a correlation. These details are discussed in chapter 6.

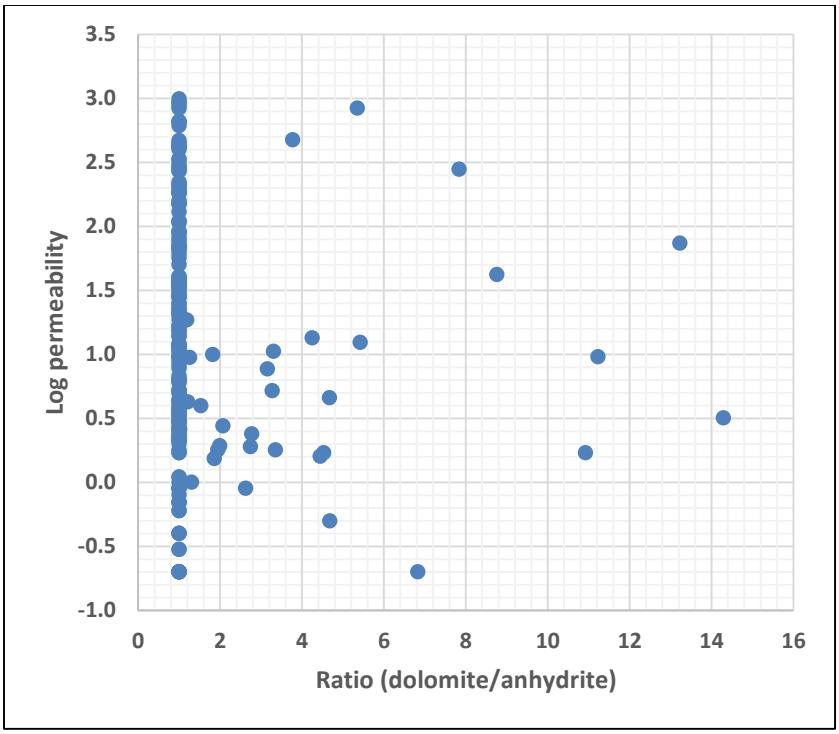


Figure 5-12: A plot of log k Vs ratio MD/MA. It is evident that major chunk of data has ratio = 1.

CHAPTER 6

DEVELOPMENT AND TESTING OF THE MATHEMATICAL CORRELATION

In this chapter, development of mathematical correlation is discussed in detail. The overall approach and steps are discussed. In the end, the testing of the developed correlation is also carried out.

Regression analysis is one technique of modeling the relationship between one variable and another set of variables. The relationship is expressed as an equation that predicts the response variable “the dependent variable” from a function of regressor variables “the independent variables” and parameters. Empirical equation or correlation analysis measures the strength of such relationship by means of statistical analysis.

6.1 Symbolic Regression

A regression technique that allows to search for a mathematical expression to fit the data in terms of simplicity and accuracy is known as **Symbolic Regression**. For this purpose, no model is provided prior to the algorithm. Algorithm itself searches for a mathematical expression with the help of building blocks such as mathematical operators, constants, analytic functions and state variables. Genetic algorithm is employed to combine the individual building blocks to generate a mathematical expression representative of the data. Through this approach, data reveals intrinsic patterns and relationships within it without any structure of model imposed by a human being. To ensure the model to predict

the patterns within the data accurately, error metrics along with special complexity measure are employed by the algorithms[55].

As other (conventional) regression techniques require a pre-specified model structure to optimize its parameters, symbolic regression without any imposition of pre-specified assumptions, seek the model from the data. Hence, it seeks to discover both model parameters and structures within the data. Due to such a large space to search, data could be fit into many models. This needs to be controlled by limiting the set of input building blocks for a possible mathematical model based on known underlying principles of the system being investigated.

In this study, to investigate the possible mathematical models using symbolic regression, Eureqa has been used. Eureqa is an AI (Artificial Intelleigence) powered search engine which employs evolutionary search to discover mathematical models for the dataset in their simplest form. Evolutionary search is a machine learning process that generate random equations for the dataset. Some of them are useful which will eventually pave a way to various other equations until a desired result is achieved [56], [38].

6.1.1 Application of Eureqa in Search of a Mathematical Model:

Eureqa takes dependent as well as independent variables as an input. It needs a specification of dependent variables in terms of independent variables. Function parameters are provided with the selection of mathematical operators to be used to formulate a model. For example, an input could be log k = f (ϕ, ρ, ML, MD, MA) or any

different combination. With this, formula building blocks are selected and program is started. It will continue to search in the space provided by the user until a desired mathematical model is generated.

As we know that FCM, K-means and EM algorithm were employed to develop clusters to improve correlation coefficients. FCM and K-means returned almost similar clusters. While EM algorithm returned clusters, which were not good enough as far as problem at hand is considered. So, the clusters developed by FCM algorithm was taken as the basis to apply symbolic regression. It contained three clusters A, B and C details of which can be found from section 5.1.1.1 to 5.1.1.3.

6.1.2 FCM Cluster a Mathematical Model:

Cluster A belongs to the data set which contains less than 40 % limestone and greater than 40% dolomite. This data set was provided to Eureqa for the development of a mathematical model. After several attempts, the best correlation that it could come up is given in Eq. 6.1. R^2 for this model is 0.44.

$$log(\boldsymbol{k}) = b*\boldsymbol{\phi}/erf((c*\boldsymbol{\phi})^{(e*\boldsymbol{\phi})} - \boldsymbol{\phi}*cos((f*\boldsymbol{\phi}^2)^g)) + h*\Delta\boldsymbol{\rho}/erf((i*\Delta\boldsymbol{\rho})^{(j*\boldsymbol{r})} - \Delta\boldsymbol{\rho}*cos((m*\Delta\boldsymbol{\rho}^2)^n)) + o*\boldsymbol{ML}/erf((q*\boldsymbol{ML})^{(s*\boldsymbol{ML})} - \boldsymbol{ML}*cos((t*\boldsymbol{ML}^2)^u)) + v*\boldsymbol{MD}/erf((w*\boldsymbol{MD})^{(x*\boldsymbol{MD})} - \boldsymbol{MD}*cos((y*\boldsymbol{MD}^2)^z)) + A*\boldsymbol{MA}/erf((B*\boldsymbol{MA})^{(C*\boldsymbol{MA})} - \boldsymbol{MA}*cos((D*\boldsymbol{MA}^2)^E)) - F \quad (6.1)$$

Where,

k = permeability, mD

ϕ = porosity, %

$\Delta\rho$ = normalized density, g/cc

ML = limestone content, %

MD = dolomite content, %

MA = anhydrite content, %

Rest of the symbols are coefficients and exponents.

Observed (measured) vs predicted for the mathematic model (6.1) is given in Fig. 6-1. The complexity of the Eq. 6.1 makes it almost impractical. Further, observed vs measured are also not aligned to the unit slope line.

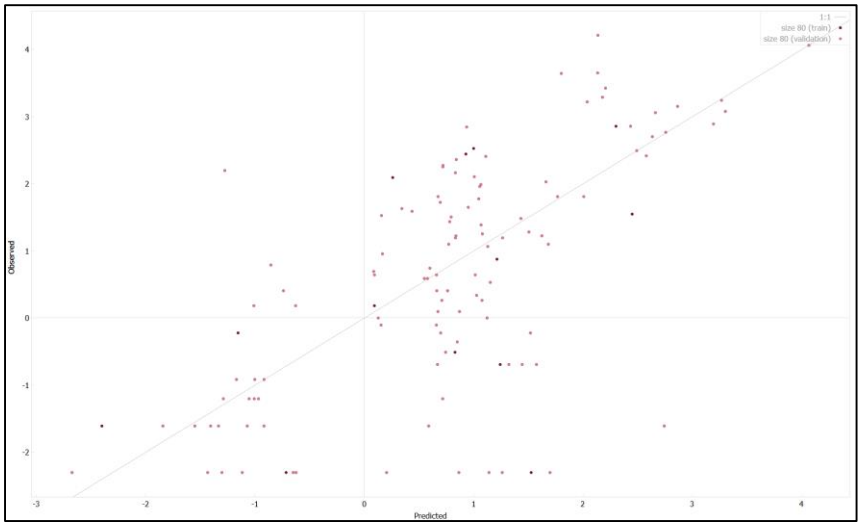


Figure 6-1: Observed Vs Measured of FCM Cluster A for the Mathematical Model developed by Symbolic regression

6.1.3 FCM Cluster B Mathematical Model:

Cluster B belongs to the data set that predominantly contains dolomite with no limestone. This data set was provided to Eureqa for the development of a mathematical model. After several attempts, the best correlation that it could come up with is given in Eq. 6.2. R^2 for this model is 0.42.

log(***k***) = logistic((b + c****ϕ*** + tan(e + f****ϕ***) - tan(***ϕ***) - tan(g****ϕ***) - tan(h****ϕ***) - ***ϕ****tan(i****ϕ***) - j*tan(m****ϕ***))/(***ϕ*** - n)) + logistic((o + q****Δρ*** + tan(s + t****Δρ***) - tan(***Δρ***) - tan(u****Δρ***) - tan(v****Δρ***) - ***Δρ****tan(w****Δρ***) - x*tan(y****Δρ***))/(***Δρ*** - z)) + logistic((A + B****ML*** + tan(C + D****ML***) - tan(***ML***) - tan(E****ML***) - tan(F****ML***) - ***ML****tan(G****ML***) - H*tan(I****ML***))/(***ML*** - J)) + logistic((K + L****MD*** + tan(M + N****MD***) - tan(***MD***) - tan(O****MD***) - tan(P****MD***) -

MD*tan(Q****MD***) - R*tan(S****MD***))/(***MD*** - T)) + logistic((U + V****MA*** + tan(W + X****MA***) - tan(***MA***) - tan(Y****MA***) - tan(Z****MA***) - ***MA****tan(b****MA***) - c*tan(e****MA***))/(***MA*** - f)) (6.2)

Where, same definition of parameters is used as in Eq. 6.1.

Observed (measured) vs predicted for the mathematic model (6.2) is given in Fig. 6-2. The complexity of the Eq. 6.2 also makes it almost impractical. Further, observed vs measured are also not properly aligned to the unit slope line.

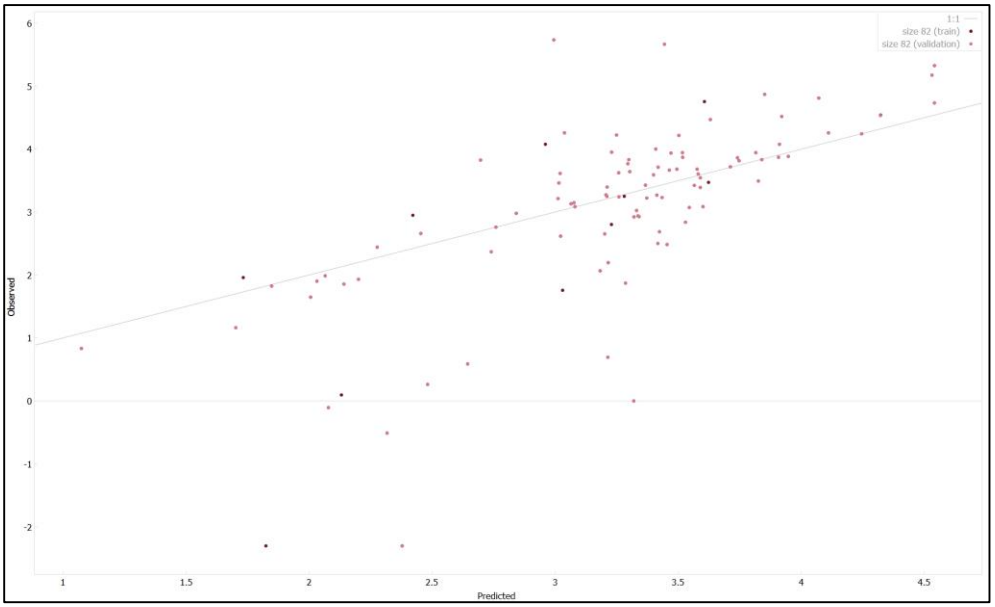


Figure 6-2: Observed Vs Measured of FCM Cluster B for the Mathematical Model developed by Symbolic regression

6.1.4 FCM Cluster C Mathematical Model:

Cluster C belongs to the data set which contains greater than 40 % limestone and less than 40% dolomite. This data set was provided to Eureqa for the development of a mathematical model. After several attempts, the best correlation that it could come up is given in Eq. (6.3). R^2 for this model is 0.48.

$$\log(\boldsymbol{k}) = \cos(b + \boldsymbol{\phi}^2 + c*\boldsymbol{\phi}/\mathrm{erf}(e - \tan(f*\boldsymbol{\phi})) + g*\boldsymbol{\phi}/(h*\boldsymbol{\phi} + \boldsymbol{\phi}^2*\sin(i*\boldsymbol{\phi})^2 - \boldsymbol{\phi}*\tan(j*\boldsymbol{\phi}) - \mathrm{factorial}(m - n*\boldsymbol{\phi}))) + \cos(o + \boldsymbol{\Delta\rho}^2 + q*\boldsymbol{\Delta\rho}/\mathrm{erf}(s - \tan(t*\boldsymbol{\Delta\rho})) + u*\boldsymbol{\Delta\rho}/(v*\boldsymbol{\Delta\rho} + \boldsymbol{\Delta\rho}^2*\sin(w*\boldsymbol{\Delta\rho})^2 - \boldsymbol{\Delta\rho}*\tan(x*\boldsymbol{\Delta\rho}) - \mathrm{factorial}(y - z*\boldsymbol{\Delta\rho}))) + \cos(A + \boldsymbol{ML}^2 + B*\boldsymbol{ML}/\mathrm{erf}(C - \tan(D*\boldsymbol{ML})) + E*\boldsymbol{ML}/(F*\boldsymbol{ML} + \boldsymbol{ML}^2*\sin(G*\boldsymbol{ML})^2 - \boldsymbol{ML}*\tan(H*\boldsymbol{ML}) - \mathrm{factorial}(I - J*\boldsymbol{ML}))) + \cos(K + \boldsymbol{MD}^2 + L*\boldsymbol{MD}/\mathrm{erf}(M - \tan(N*\boldsymbol{MD})) + O*\boldsymbol{MD}/(P*\boldsymbol{MD} + \boldsymbol{MD}^2*\sin(Q*\boldsymbol{MD})^2 - \boldsymbol{MD}*\tan(R*\boldsymbol{MD}) - \mathrm{factorial}(S - T*\boldsymbol{MD}))) + \cos(U + \boldsymbol{MA}^2 + V*\boldsymbol{MA}/\mathrm{erf}(W - \tan(X*\boldsymbol{MA})) + Y*\boldsymbol{MA}/(Z*\boldsymbol{MA} + \boldsymbol{MA}^2*\sin(b*\boldsymbol{MA})^2 - \boldsymbol{MA}*\tan(c*\boldsymbol{MA}) - \mathrm{factorial}(e - f*\boldsymbol{MA}))) \quad (6.3)$$

Where, same definition of parameters is used as in Eq. 6.1.

Observed (measured) vs predicted for the mathematic model (6.3) is given in Fig. 6-3. The complexity of the Eq. 6.3 also makes it almost impractical. Further, observed vs measured are also not properly aligned to the unit slope line.

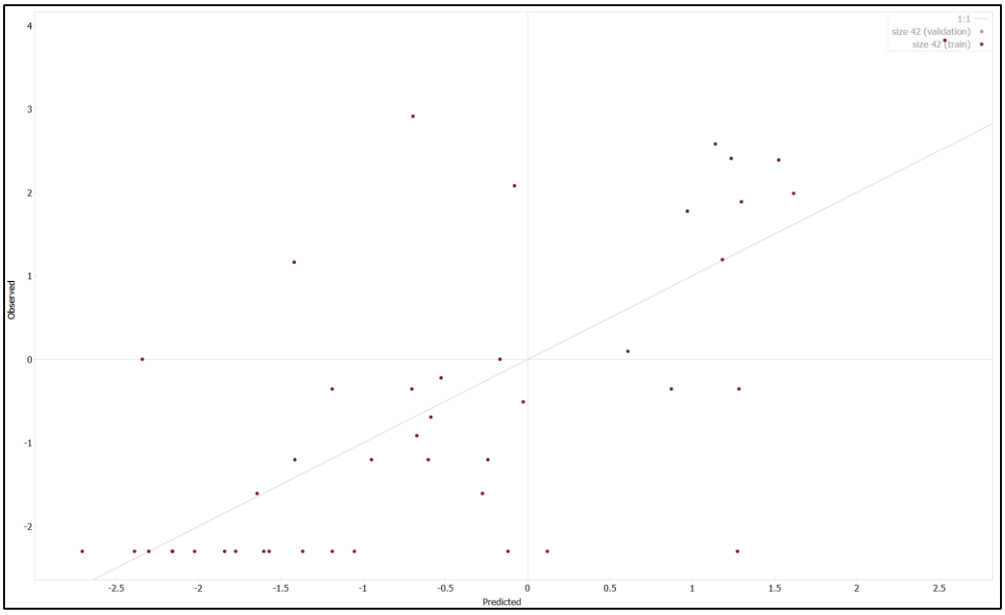


Figure 6-3: Observed Vs Measured of FCM Cluster C for the Mathematical Model developed by Symbolic regression

Due to such complex equations and low R^2, the clusters developed by machine learning algorithms were abandoned and new clustering was carried out. New clustering resulted in two clusters which are explained in section 5.2.

Data of new clusters was investigated and the trends were identified which prompted to use clustering based on mineral ratio for further analysis. Non-linear regression was applied to the new clustering approach and much better results were obtained. This is discussed in detail in the following section.

6.2 Non-linear regression

Regression analysis in which one or more independent variables are modeled to fit a dataset by a function which is a nonlinear combination of parameters is known as non-linear regression[57]. It is a powerful tool to analyze the data of considerable complexity.

A nonlinear regression model can be written as:

$$Y_n = f(\boldsymbol{x}_n, \boldsymbol{\theta}) + Z_n \tag{6.4}$$

Where, *f* is the expectation function and *xn* is a vector of associated regressor variables or independent variables for the nth case. In this model, the random variable *Yn*, which represents the response for case n, n = 1, 2, . . ., N, has a deterministic part and a stochastic part. The deterministic part, *(xn, . . ., xn)* θ, depends upon the parameters θ. The stochastic part, represented by the random variable *Zn*, is a disturbance which perturbs the response for that case. For nonlinear models, at least one of the derivatives of the expectation function with respect to the parameters depends on at least one of the parameters [58].

When analyzing a particular set of data, we consider the vectors *xn*, n = 1, 2, . . ., N, as fixed and concentrate on the dependence of the expected responses on θ. We create the N-vector $\eta(\theta)$ with nth element:

$$\eta_n(\boldsymbol{\theta}) = f(\boldsymbol{x}_n, \boldsymbol{\theta}) n = 1, \ldots, N \tag{6.5}$$

and write the nonlinear regression model as:

$$\boldsymbol{Y} = \boldsymbol{\eta}(\boldsymbol{\theta}) + \boldsymbol{Z} \tag{6.6}$$

with $\boldsymbol{Z}$ assumed to have a spherical normal distribution. That is,

$$\mathrm{E}[\boldsymbol{Z}] = \mathbf{0}$$

Where E is the expectation function or equivalently it can be written as $E(Y)= \eta(\theta)$.

If Z is normally distributed then:

$$\mathrm{Var}(\boldsymbol{Z}) = \mathrm{E}[\boldsymbol{Z}\boldsymbol{Z}^{\mathrm{T}}] = \sigma^2 \boldsymbol{I}$$

where I is an N×N identity matrix.

6.2.1 Development of Correlation:

In this study, Excel and Matlab programming was employed to carry out non-linear regression. Excel has certain linear and non-linear regression models that can be used to fit the trends of the data set. Matlab can be used to fit any mathematical model without the pre-requisite of built in models.

As discussed previously, after symbolic regression failed with the clustering carried out by machine learning algorithms, there were many non-linear regression models and possible combinations tried with the help of SPSS (a statistical analysis software to develop correlations) and Matlab. None of them returned any desired results. This led us to the re-evaluation of data set which resulted in new clustering of the data.

These clusters were further analyzed and proportionality of individual parameters with permeability was checked. This was carried out to come up with a basic mathematical model that can be further manipulated to provide desired results. The data was divided into three clusters. First cluster was the 'Filtered Data'. This cluster consists of 226 data points. Second cluster was a subset of the first cluster in which dolomite and anhydrite contents are equal. This cluster contains 174 data points. Third cluster was also a subset of the first cluster in which dolomite and anhydrite contents are not equal. This cluster contains 52 data points.

Plots of each parameter or variable with permeability are provided in the following sections for the first cluster which is 'Filtered Data'.

6.2.1.1 Relationship of log k with porosity:

Figure 6-4 shows a crossplot of logk versus porosity. This figure reveals logk is directly proportional to porosity and can be fit by a linear equation: $\log k = 0.086\phi - 0.3584$ with $R^2 = 0.31$. This is a weak correlation but direction of proportionality has been identified which is:

$$log\ k\ \alpha\ \phi \tag{6.7}$$

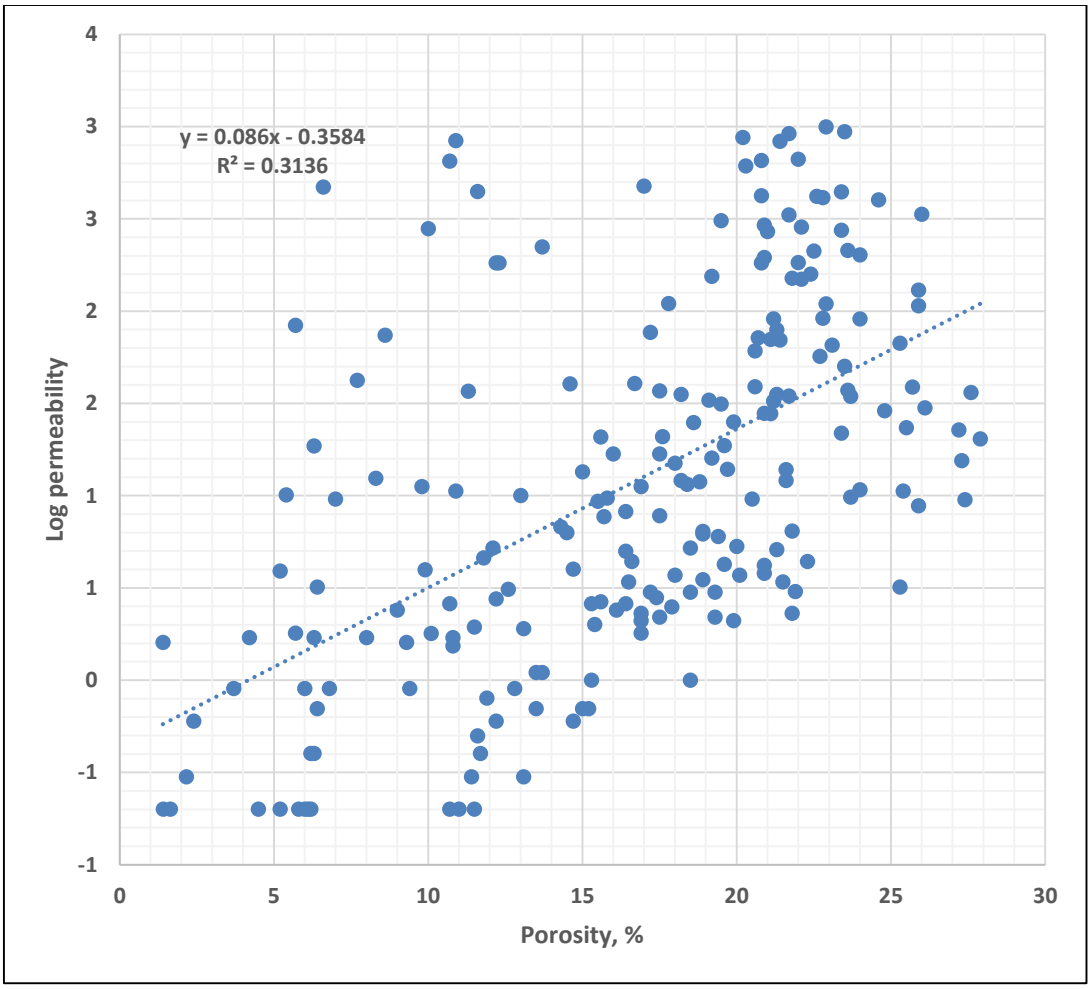


Figure 6-4: Crossplot of Log k vs porosity showing both are directly proportional to each other.

6.2.1.2 Relationship of log k with density:

Figure 6-5 shows a crossplot of logk versus density. The figure shows logk to be inversely proportional to porosity and can be fit by a linear equation: logk = -3.81ρ +10.326 with R^2 = 0.33. This is a weak correlation but direction of proportionality has been identified which is:

$$log\ k\ \alpha\ 1/\rho \tag{6.8}$$

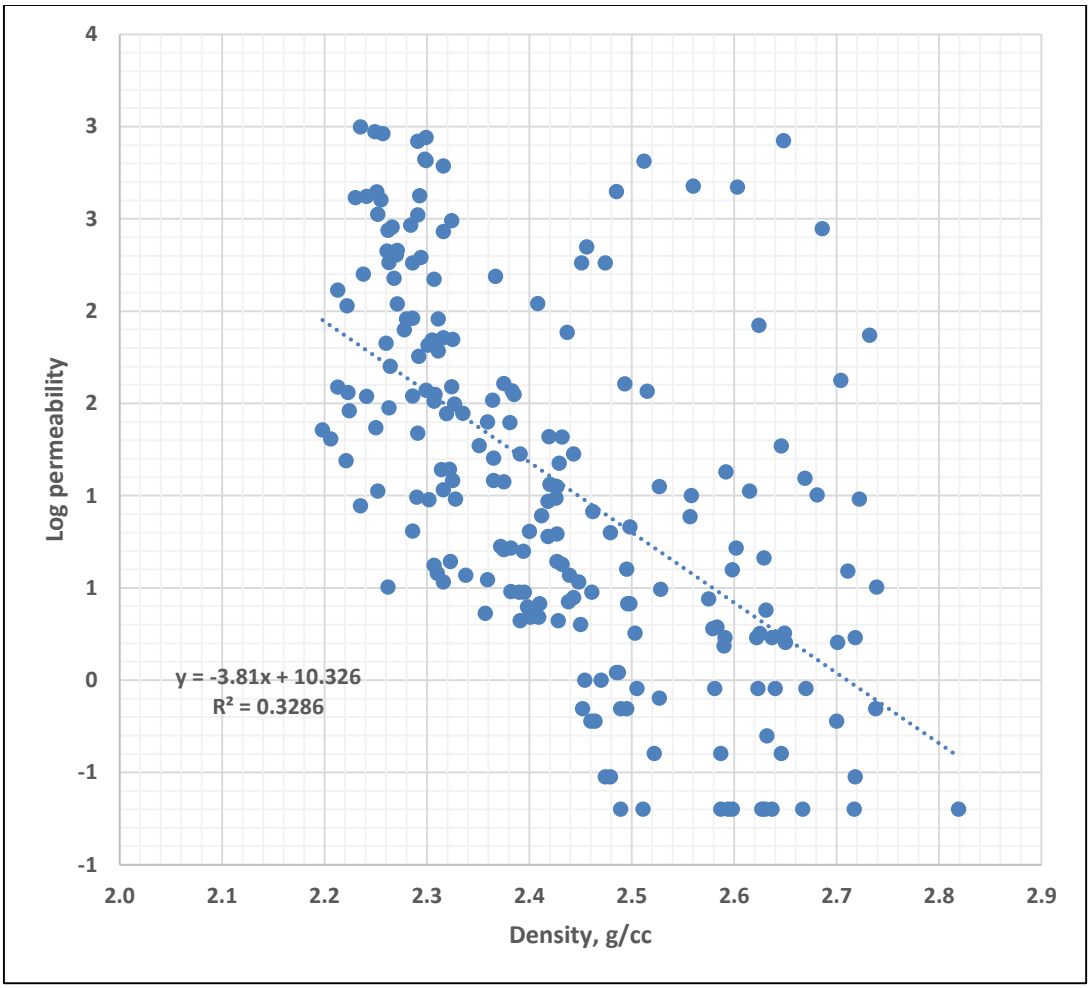


Figure 6-5: Crossplot of Log k vs density showing both are indirectly proportional to each other.

6.2.1.3 Relationship of log k with limestone content:

Figure 6-6 shows a crossplot of logk versus limestone content. As it can be observed that logk is inversely proportional to limestone content and can be fit by a linear equation: logk = -0.0362ML +3.32 with R^2 = 0.25. This is a weak correlation but direction of proportionality has been identified which is:

$$log\ k\ \alpha\ 1/ML \tag{6.9}$$

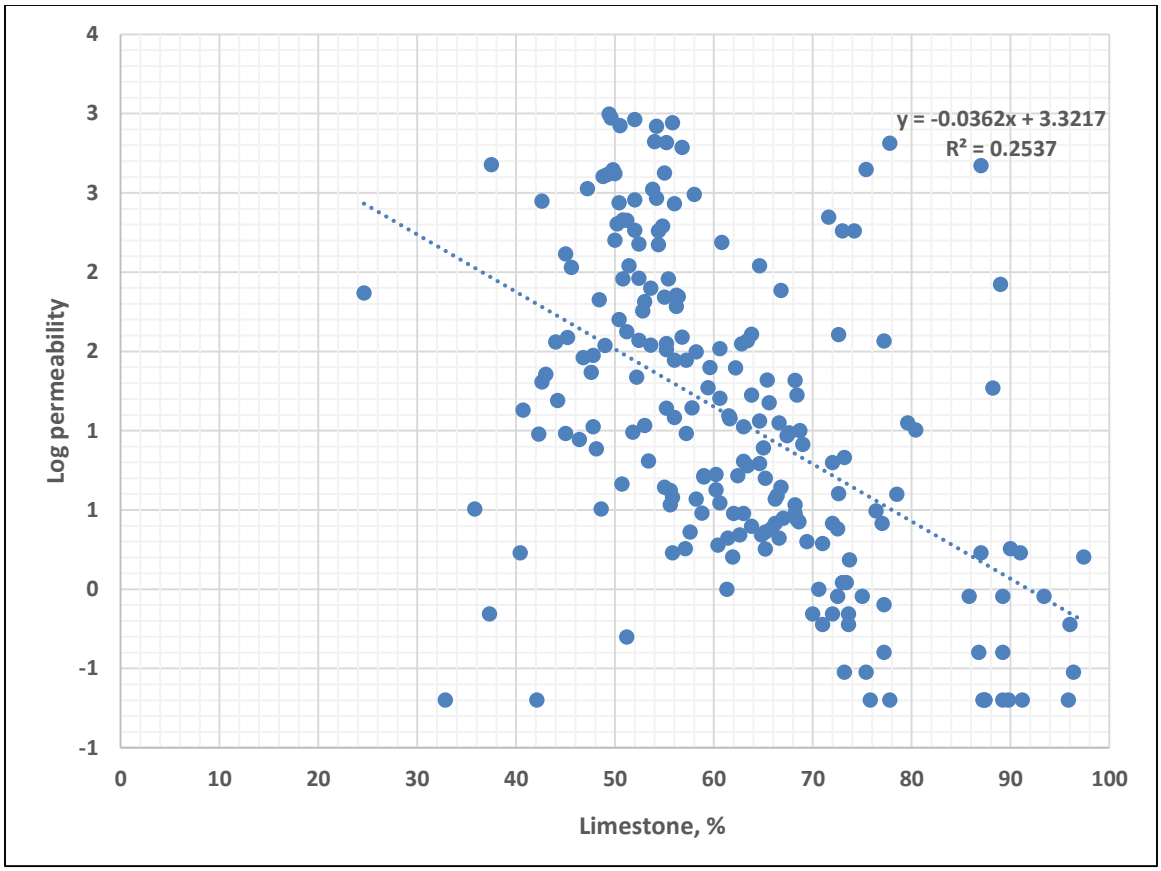


Figure 6-6: Crossplot of Log k vs limestone content showing both are indirectly proportional to each other.

6.2.1.4 Relationship of log k with dolomite content:

Figure 6-7 shows a crossplot of logk versus dolomite content. As it can be observed that logk is directly proportional to dolomite content and can be fit by a linear equation of logk = 0.0227MD +0.59 having R^2 = 0.06. This is a very weak correlation but direction of proportionality has been identified which is:

$$log\ k\ \alpha\ MD \tag{6.10}$$

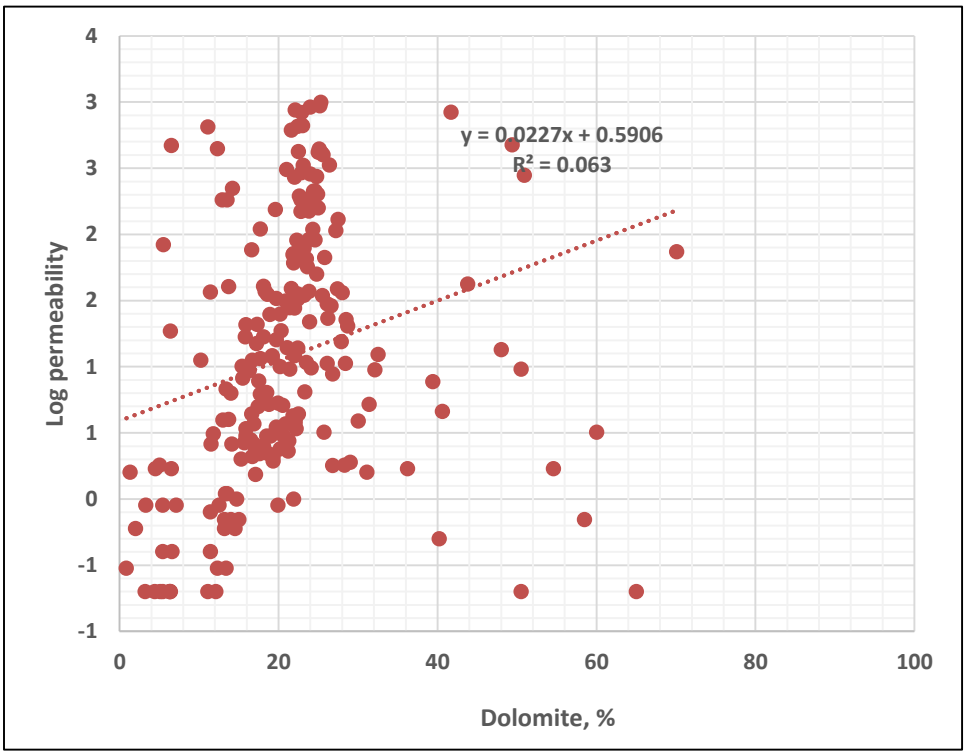


Figure 6-7: Crossplot of Log k vs dolomite content showing both are directly proportional to each other.

6.2.1.5 Relationship of log k with anhydrite content:

Figure 6-8 shows a crossplot of logk versus anhydrite content. As it can be observed that logk is directly proportional to anhydrite content and can be fit by a linear equation of logk = 0.07MA-0.2453 having R^2 = 0.33. This is contrary to literature where anhydrite is considered to be a plugging material. This is a weak correlation but direction of proportionality has been identified which is:

$$Log\ k\ \alpha\ MA \tag{6.11}$$

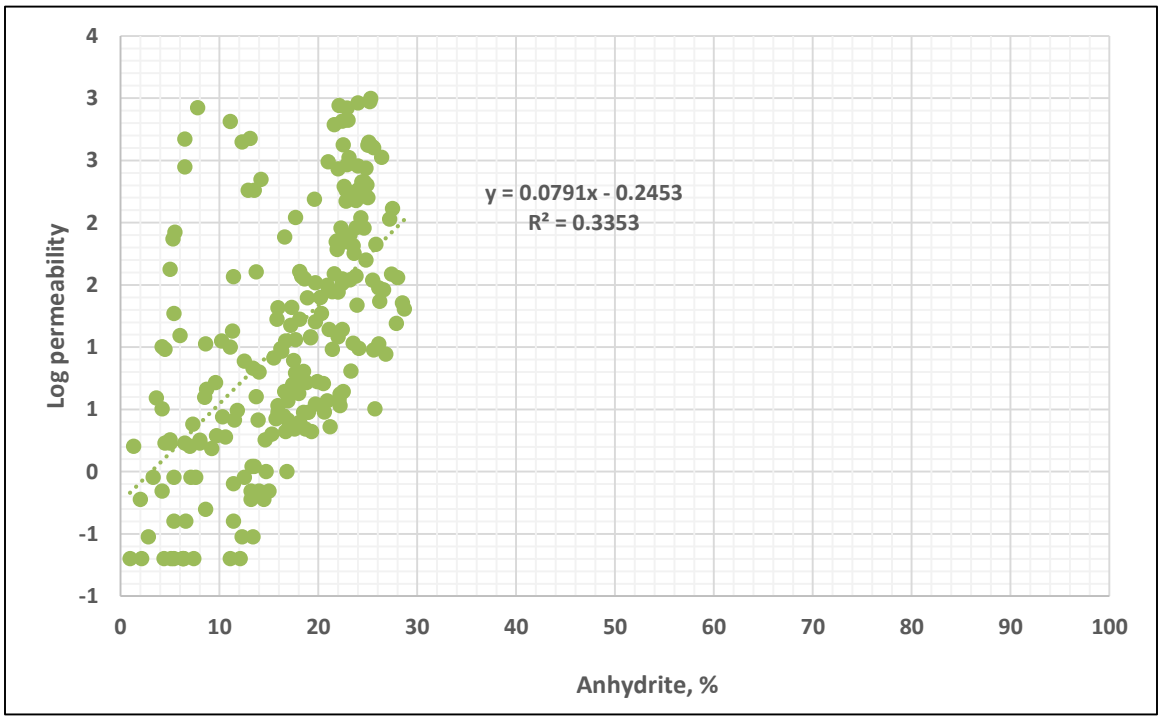


Figure 6-8: Crossplot of Log k vs anhydrite content showing both are directly proportional to each other.

6.2.1.5 Basic Mathematical Model:

Among the Eq.s (6.7), (6.8), (6.9), (6.10) and (6.11), dolomite has the weakest correlation with logk. Combining all the equation will yield:

$$log\ k\ \alpha\ \frac{\phi * MD * MA}{ML * \rho} \tag{6.12}$$

which leads to:

$$log\ k\ = (constant) * \frac{\phi * MD * MA}{ML * \rho} \tag{6.13}$$

This was the basic structure for the mathematical model that was pursued.

6.2.1.6 Mathematical Model for the filtered data:

The basic mathematical model discussed in section 6.2.1.5 was taken as an initial model for the regression. R^2 and root mean squared error was used to define the accuracy of the model. During optimization, it was discovered that a more rigorous mathematical model may be developed using the basic model defined by Eq. 6.14. The optimized mathematical model which worked best for this data set came out to be:

$$log\ k\ = \{(c1) * \frac{\Phi^{e1} * MD^{e2} * MA^{e3}}{ML^{e4} * \rho^{e5}}\} + \{(c2) * \frac{\Phi^{e6} * MD^{e7}}{ML^{e8} * \rho^{e9}}\} + \{(c3) * \frac{\Phi^{e10}}{ML^{e11} * \rho^{e12}}\} + \{(c4) * \frac{\Phi^{e13}}{\rho^{e14}}\} + \{(c5) * \Phi^{e15}\} - 1 \quad (6.14)$$

Where,

k = permeability, mD

ϕ = porosity, fraction

ρ = density, g/cc

ML = limestone content, fraction

MD = dolomite content, fraction

MA = anhydrite content, fraction

e = exponents

c = coefficients

Using Matlab, non-linear regression analysis was applied to the mathematical model of Eq. 6.14. After an iterative process, the optimized exponents and coefficients returned by the program are provided in Table 6-1. R^2 for the developed mathematical model was 0.6 with a root mean squared error (RMSE) of 0.575.

Table 6-1 reveals that the contribution of the ratio of porosity and density in Eq. 6.14 is insignificant as c4 is almost negligible. Removing the negligible term, the final form of Eq. 6.14 for this data set was thus modified to:

$$log\ k\ = \{(145.67) * \frac{\phi^{0.30} * MD^{0.24} * MA^{0.19}}{ML^{0.12} * \rho^{2.71}}\} + \{(11.48) * \frac{\phi^{2.12} * MD^{1.25}}{ML^{-0.12} * \rho^{-1.04}}\} + \{(40.87) * \frac{\phi^{1.43}}{ML^{-0.85} * \rho^{1.79}}\} + \{(-7.24) * \phi^{1.44}\} - 1 \quad (6.15)$$

Table 6-1: Coefficients and exponents of mathematical model given by equation 6.14 for filtered data

Coefficients		**Exponents**	
c1	145.67	***e1***	0.30
c2	11.48	***e2***	0.24
c3	40.87	***e3***	0.19
c4	3.11e-12	***e4***	0.12
c5	-7.24	***e5***	2.71
		e6	2.12
		e7	1.25
		e8	-0.12
		e9	-1.04
		e10	1.43
		e11	-0.85
		e12	1.79
		e13	1.97

		e14	-28.22
		e15	1.44

Measured vs predicted plot for this model is given by Fig. 6-9. Measured vs predicted data points are aligned along a line of slope 1. R^2 of the correlation is 0.6.

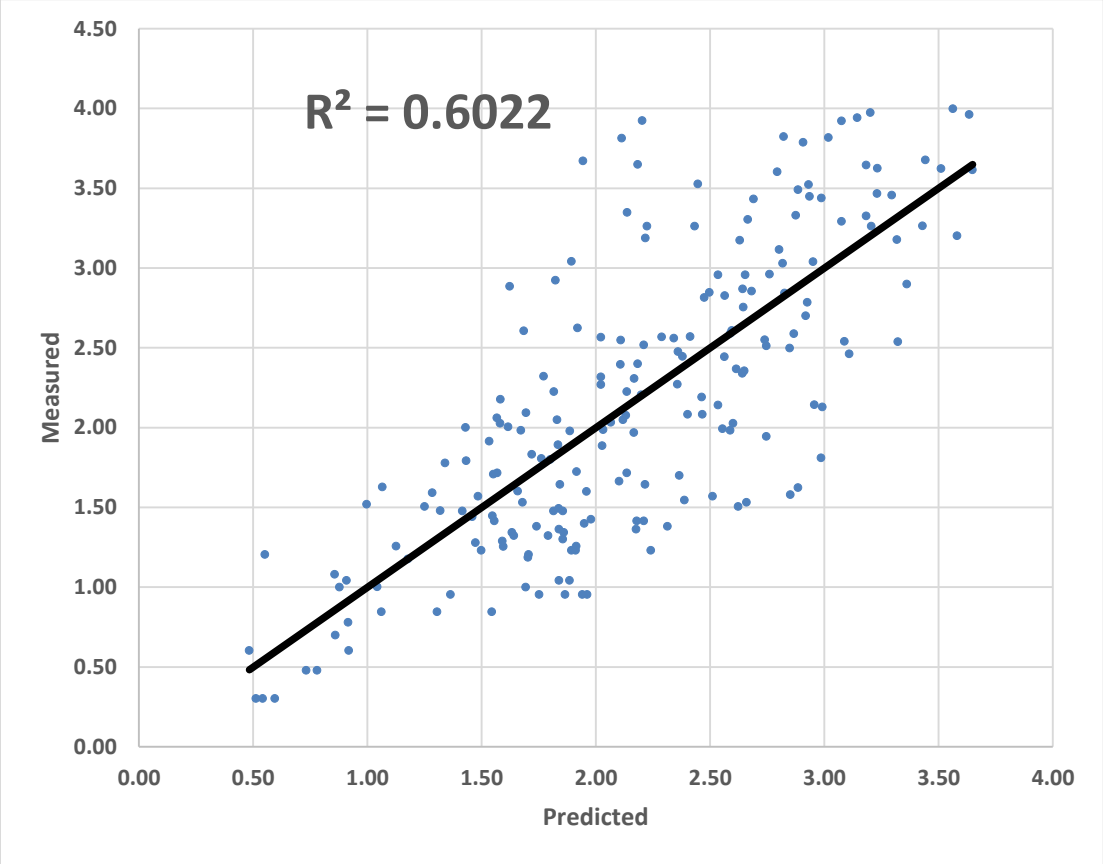


Figure 6-9: Measure vs predicted (logk +1) plot for the case of filtered data

6.2.1.7 Mathematical Model for Subset 1 (MD=MA):

Mathematical model given by Eq. 6.14 was taken as the initial model for the matlab program for non-linear regression analysis of subset 1. After an iterative process, the optimized exponents and coefficients which are provided in Table 6-2 were returned by the program. R^2 for the developed mathematical model was improved to 0.669 having root mean squared error (RMSE) to be 0.481.

The final form of Eq. 6.14 for this data set came out to be:

$$log\ k\ = \{(13.7) * \frac{\Phi^{-2.47} * MD^{3.02} * MA^{3.02}}{ML^{-15.35} * \rho^{-11.16}}\} + \{(-3.71) * \frac{\Phi^{-0.45} * MD^{-0.01}}{ML^{5.03} * \rho^{3.15}}\} + \{(2.27) * \frac{\Phi^{-0.51}}{ML^{5.17} * \rho^{2.78}}\} + \{(817.21) * \frac{\Phi^{0.46}}{\rho^{7.19}}\} + \{(-0.16) * \Phi^{0.52}\} - 1 \quad (6.16)$$

Since we know that, for this case:

$$MD = MA = \frac{1-ML}{2}$$

Replacing MD and MA in terms of ML in equation and solving lead to:

$$log\ k\ = \{(0.21) * \frac{\Phi^{-2.47} * (1-ML)^{6.04}}{ML^{-15.35} * \rho^{-11.16}}\} + \{(-3.73) * \frac{\Phi^{-0.45} * (1-ML)^{-0.01}}{ML^{5.03} * \rho^{3.15}}\} + \{(2.27) * \frac{\Phi^{-0.51}}{ML^{5.17} * \rho^{2.78}}\} + \{(817.21) * \frac{\Phi^{0.46}}{\rho^{7.19}}\} + \{(-0.16) * \Phi^{0.52}\} - 1 \quad (6.17)$$

Table 6-2: Coefficients and exponents of mathematical model given by Eq. 6.14 for Subset 1

Coefficients		**Exponents**	
c1	13.7	***e1***	-2.47
c2	-3.71	***e2***	3.02
c3	2.27	***e3***	3.02

c4	817.21	***e4***	-15.35
c5	-0.16	***e5***	-11.16
		e6	-0.45
		e7	-0.01
		e8	5.03
		e9	3.15
		e10	-0.51
		e11	5.17
		e12	2.78
		e13	0.46
		e14	7.19
		e15	0.52

Measured vs predicted plot of subset 1 for this model is shown in Fig. 6-10. Measured vs predicted data points are aligned along a line of slope 1. R^2 of the correlation is 0.67.

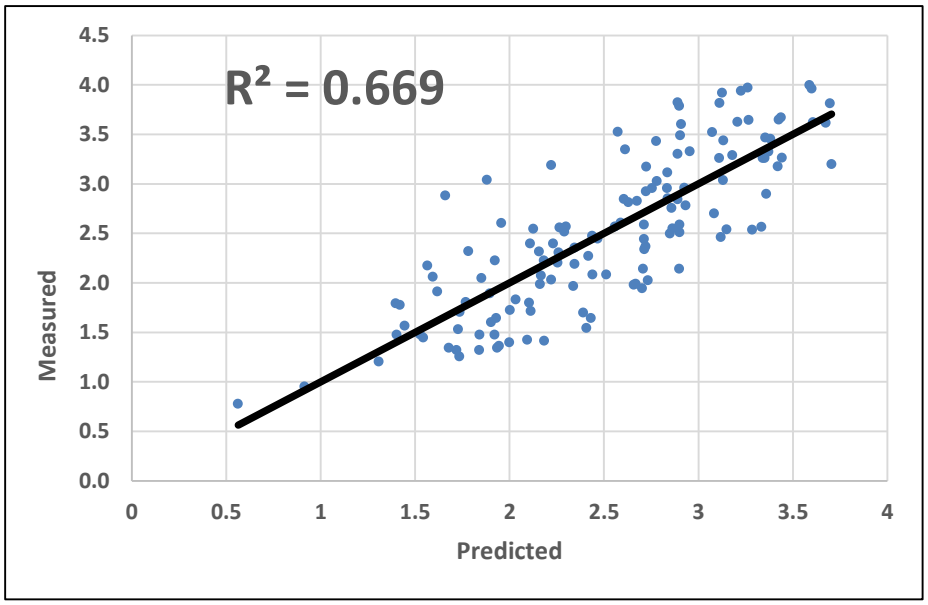


Figure 6-10: Measure vs predicted (logk + 1) plot for the case of Subset 1

6.2.1.8 Mathematical Model for Subset 2 (MD ≠ MA):

Mathematical model given by Eq. 6.14 was taken as the initial model for the matlab program for non-linear regression analysis of subset 2. After an iterative process, the optimized exponents and coefficients which are provided in Table 6-3 were returned by the program. R^2 for the developed mathematical model was improved to 0.71 having root mean squared error (RMSE) to be 0.44. As witnessed in the case of subset 1, new clustering has improved R^2 of the model.

Table 6-3 reveals that the ratio of porosity and density in Eq. 6.14 is insignificant as the 4th coefficient is almost negligible. The final form of Eq. 6.14 for this data set is coming out to be:

$$log\ k\ = \{(749.42) * \frac{\Phi^{0.07} * MD^{0.26} * MA^{0.51}}{ML^{0.18} * \rho^{4.75}}\} + \{(9.91) * \frac{\Phi^{1.85} * MD^{1.03}}{ML^{0.19} * \rho^{0.02}}\} + \{(23.19) * \frac{\Phi^{1.59}}{ML^{-0.45} * \rho^{0.68}}\} + \{(-11.18) * \Phi^{1.57}\} - 1 \quad (6.17)$$

Table 6-3: Coefficients and exponents of mathematical model given by Eq. 6.14 for Subset 2

Coefficients		**Exponents**	
c1	749.42	***e1***	0.07
c2	9.91	***e2***	0.26
c3	23.19	***e3***	0.51
c4	3.018e-10	***e4***	0.18
c5	-11.18	***e5***	4.75
		e6	1.85
		e7	1.03
		e8	0.19
		e9	0.02
		e10	1.59
		e11	-0.45
		e12	0.68
		e13	1.46
		e14	-21.56
		e15	1.57

Measured vs predicted plot of subset 2 for this model is given by Fig. 6-11. Measured vs predicted data points are aligned along a line of slope 1. R^2 of the correlation is 0.71.

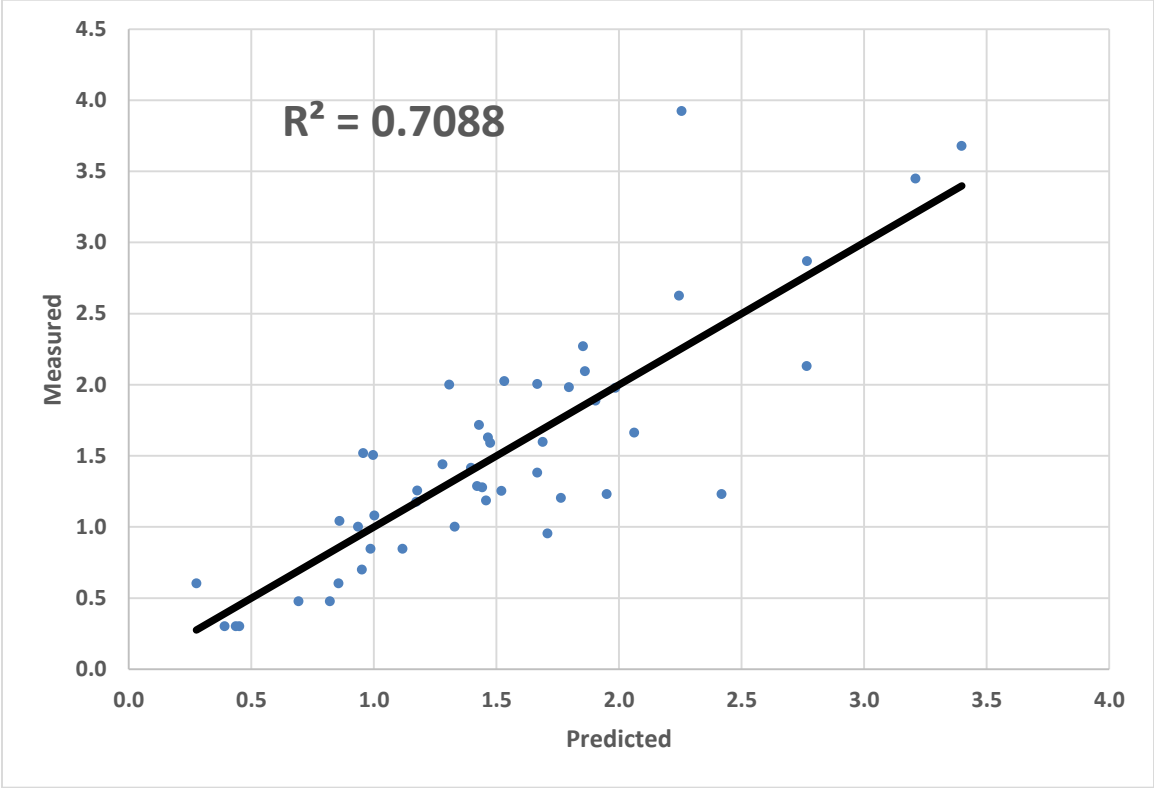


Figure 6-11: Measure vs predicted (logk + 1) plot for the case of Subset 2

6.2.2 Power law regression for subset 2 (MD = MA):

For the development of mathematical model presented in the previous section, an extensive investigation to identify trends with in the data set was carried out. During this investigation, various combinations of parameters were plotted against logarithmic of permeability. Among these combinations, a power law expression was also tried. This power law expression was able to fit the crossplot of (log k +1) vs ratio of limestone content and porosity. One is added to logarithmic of permeability to remove negative values. This plot is given in Fig. 6-12.

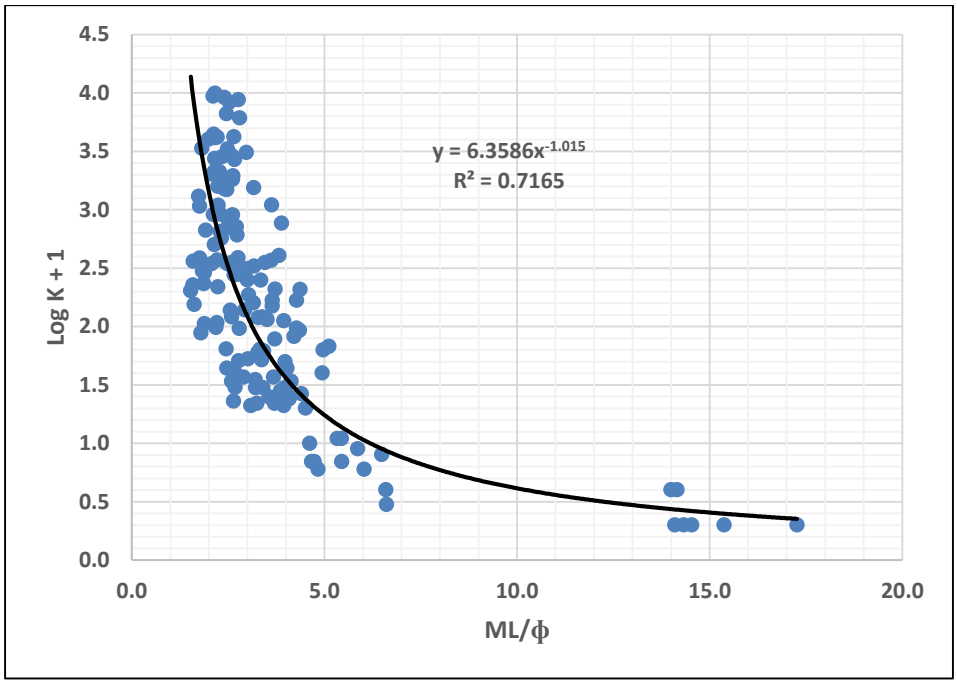


Figure 6-12: log k +1 vs ML/ϕ plot for Subset 1

R^2 for the fit is given as 0.72. This fitting gave one more equation for subset 1 (MD = MA) which is given as:

$$log\ k = 6.36 * (\frac{ML}{\phi})^{-1.02}\text{-}1 \qquad (6.17)$$

Measured vs predicted for the above-mentioned model is given by:

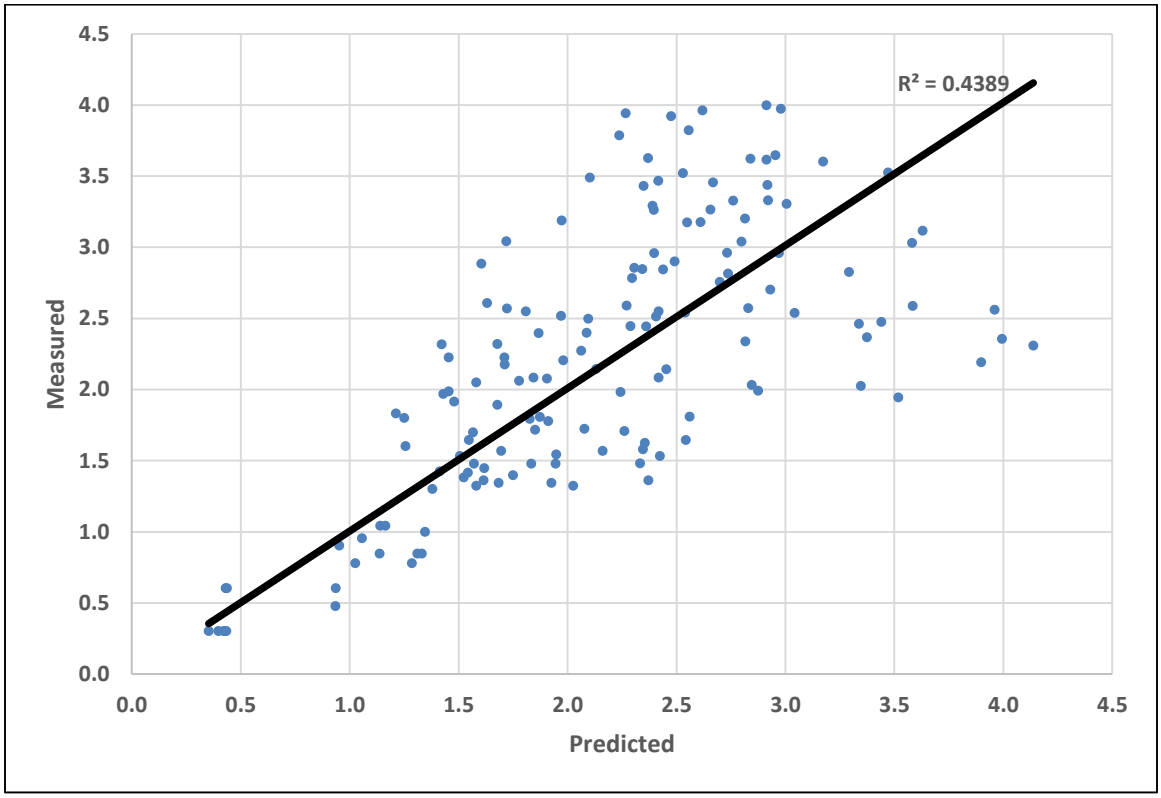


Figure 6-13: Measure vs predicted (logk + 1) plot for the case of Subset 1

This model is not good for prediction as the data is concentrated between ML/ϕ = 2 and ML/ϕ = 4. This is incorporating error, which is eventually evident in the plot of measured vs predicted in Fig. 6-13.

6.3 Artificial intelligence:

Artificial Neural network is the powerful statistical tool to recognize and classify complex patterns and system which human brain cannot do [53]. This technique is inspired from biological neurons that are found in human brain [59]. Many researchers applied neural network techniques in petroleum application especially in rock mechanics (Abdulraheem et al., 2009, and Tariq et al., 2016b).

The neural network models are structured on three basic components, namely; learning algorithm, transfer function and neurons architecture [62]. The network model comprises of at-least three layers, input layer, hidden layer and output layer Fig. (6-14). Each layer connects with other layers with the help of weights. The network performance is solely based on the adjustment of weights between these layers [63].

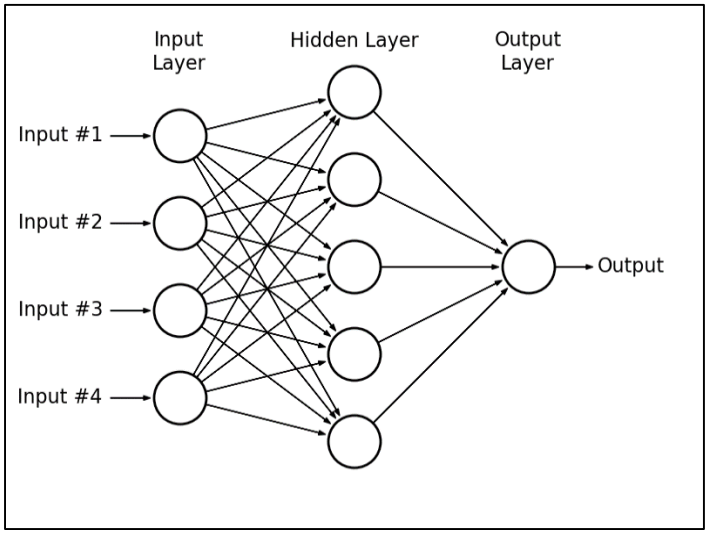


Figure 6-14: Basic structure of ANN systems

6.3.1 ANN Architecture for Permeability Correlation Development:

A back propagation neural network algorithm was implemented to model permeability. The proposed ANN model was based on five input parameters, namely, bulk density, neutron porosity, mineralogical contents such as; Limestone, dolomite and anhydrite, with one hidden layer (Fig. 6-16). The number of neurons in the hidden layer were varied between 5 and 20. The optimum number of neurons were found to be 20. Tan-sigmoidal type activation function was used as a transfer function between input and hidden layer and linear type activation function was used between hidden and output layer. Levenberg-Marquardt back propagation algorithm was selected as the training algorithm to obtain the weights and biases. This optimum combination (Table 6-4) of ANN parameters to predict permeability was selected based on highest correlation coefficient between measured and predicted data both during training and testing phases.

Table 6-4: Optimized parameters for the ANN model

Parameter	Value
Learning function	trainlm
Transfer function	tansig
Number of hidden layers	1
Number of neurons	20

To avoid the model to stuck on local minima 10,000 realizations were performed with the initialization of different weights and biases during training and cross-validation

phases of the modeling. After training, the weights and biases from the optimum model were extracted which are given in (Table 6-6).

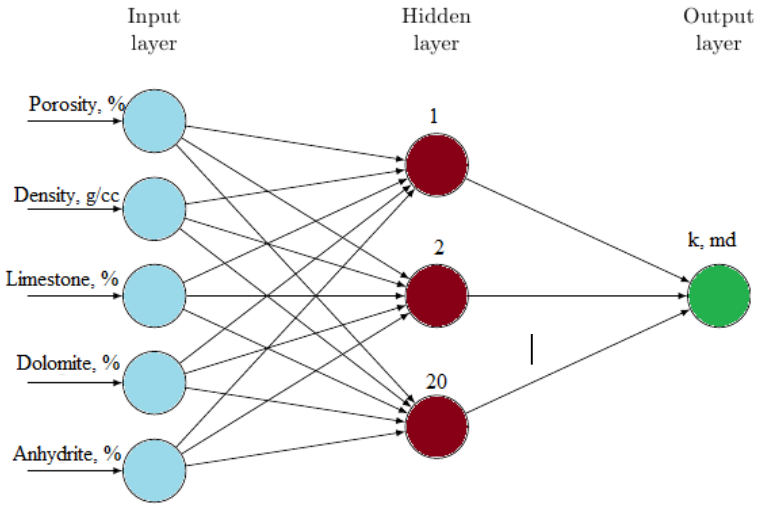


Figure 6-15: ANN architecture for proposed correlation

6.3.2 Empirical Correlation:

All the data that goes in to the neural network model is first normalized between [-1 1] by using two-point slope form given by Eq. 6.18 and 6.19.

$$\frac{Y-Ymin}{Ymax-Ymin} = \frac{X-Xmin}{Xmax-Xmin} \tag{6.18}$$

$$Y = \frac{X-Xmin}{Xmax-Xmin}(Ymax - Ymin) + Ymin \tag{6.19}$$

Where,

Y = Input Parameter in normalized form

Y_{max} = 1

Y_{min} = -1

X_{min} = Input Data maximum value.

X_{max} = Input Data minimum value

X = Input Parameter normalized form

$$Y = \frac{X - Xmin}{Xmax - Xmin} * 2 - 1 \quad (6.20)$$

6.3.3 Proposed Permeability Model

The empirical correlation made from the weights of ANN model for permeability in normalized form is given below:

$$K_n = \left[\sum_{i=1}^{N} w_{2_i}\left(\frac{2}{1+e^{-2\left(w_{1_{i,1}}\phi_n + w_{1_{i,2}}\rho_n + w_{1_{i,3}}Ls_n + w_{1_{i,3}}Dol_n + w_{1_{i,3}}Anh_n + b_{1_i}\right)}} - 1\right)\right] + b_2 \quad (6.21)$$

Where, the w_1, w_2, b_1 and b_2 are the weight matrix of input layer, weight vector of hidden layer, bias vector of input layer and bias of hidden layer respectively. All these values are presented in Table 6-6.

The output obtained by using the above correlation will be in normalized form that must be de-normalized using the following equation:

$$K = 10^{(0.4854K_n - 0.5145)} \quad (6.22)$$

Table 6-5-Weights and Biases Matrix for Permeability Correlation

$\mathbf{w_1}$					$\mathbf{w_2}$	$\mathbf{b_1}$	$\mathbf{b_2}$
-1.731	-0.623	-0.468	1.072	1.513	-0.438	2.584	0.540
0.307	0.397	1.425	-0.519	-1.264	-0.560	2.010	
-0.265	-2.157	0.725	0.784	-1.989	0.639	1.486	
-1.312	-0.601	-1.491	-0.625	-0.118	1.152	2.653	
-1.413	0.923	1.893	-0.501	-0.290	1.211	1.912	
2.882	3.015	-2.157	-0.385	0.499	-0.905	-0.661	
-2.625	1.165	-0.756	0.771	-0.151	-0.822	0.422	
-2.522	-1.081	-1.474	-1.485	0.020	-0.780	0.660	
-0.492	1.319	1.188	0.802	1.837	-1.119	1.527	
2.512	-1.388	-0.101	0.395	1.788	0.905	-0.319	
0.267	0.008	-0.061	0.534	2.018	1.745	0.642	
-1.377	1.384	-1.309	-0.482	-1.459	2.393	-1.036	
0.896	-0.750	0.822	1.319	2.346	0.688	-0.433	
-1.353	-1.416	2.138	-0.753	0.512	1.013	-1.479	
1.636	0.887	-0.318	-0.913	-0.674	0.813	1.635	
2.748	-1.004	-0.753	1.303	0.596	1.252	1.841	
0.079	1.278	1.393	1.190	1.277	-0.540	1.930	
1.032	0.792	1.336	1.420	0.083	0.098	2.409	
-1.433	0.910	-0.954	2.112	-0.343	0.813	-2.496	
1.606	-1.708	2.673	-0.926	0.861	1.007	3.265	

6.3.4 Results and Discussion

A MATLAB based program was developed to train the model with Artificial Neural Network (ANN) having optimized parameters in Table 6-4. The model can be easily re-generated using the mentioned optimized parameters. Weights and biases of the trained model are presented in Table 6-6. Using these weights and biases, a simple worksheet can also be developed to estimate permeability.

Results can be broadly categorized into training & testing parts. The Khuff data is divided into training data set which is taken to be 70% (194) of the total data. While unseen data which is 30% (83) is used to test the developed model. Correlation coefficient, average absolute difference (AAD) and root mean square error (RMSE) are used to evaluate the accuracy of developed model. Appendix can be referred for the definition of these utilized errors.

6.3.5 Training Results

As mentioned above, 70% data points were taken from the original data set to train the model. These data points were further divided by the MATLAB program to train and test the system. The correlation coefficient of the estimated permeability with the measured permeability data is found to be 0.86. While, the crossplot of measured vs predicted permeability has coefficient of determination R^2 of 0.7308. As shown in Fig. 6-16, the proposed correlation fit the measured values of permeability in log scale quite well. Further, the measured vs predicted values are aligned around 45-degree line showing a good prediction by the proposed correlation. The root mean square error and average absolute difference of the training set of data are obtained to be 0.04 and 0.08 respectively.

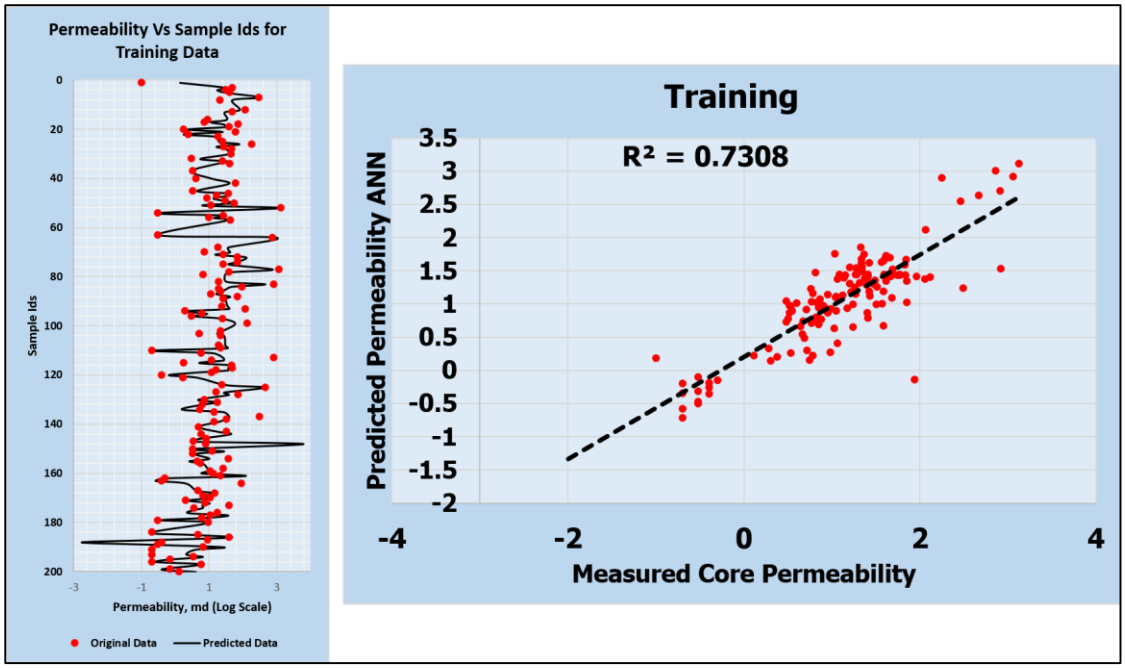


Figure 6-16: Results obtained for training set of data (a) Comparison of Measured permeability values against predicted permeability values on a plot of Sample ids vs permeability (log scale) (b) Crossplot of Measured Vs Predicted permeability values.

6.3.6 Testing

For testing, 30% unseen data points were taken from the original data set to test the proposed correlation. The correlation coefficient of the estimated permeability with the measured permeability data is 0.85. While, the crossplot of measured vs predicted permeability has R^2 equal to 0.72. As shown in Fig. 6-18, the proposed correlation fit the test with minor deviations which confirms the validity of the proposed correlation. Measured vs predicted values are also aligned around 45-degree line showing a good prediction by the proposed correlation. The root means square error (RMSE) and average

absolute difference(AAD) of the testing set of data are found to be 0.03 and 0.04 respectively.

The reason of R^2 being low for the measured vs predicted permeability could be the approach that we followed for the formulation of this correlation. Carbonate formations, due to the incorporation of many variations over a considerable period, depends on various other factors that have not been considered. Those factors may include but not limited to texture, rock fabric, pore size distribution etc. This could be the way forward to link this correlation with such factors making it into a more robust estimator of permeability with high certainty.

6.3.7 How to use the proposed Permeability correlation?

Proposed correlation can be easily used by developing a worksheet using weights and biases given in Table 6-6 or a MATLAB program may be developed. Steps that can be followed to solve the equation are as follows:

1. Input parameters are required to be normalized between the range of [-1 1] with the help of Eq. 6.18 and 6.19.
2. Now, Eq. 6.21 can be employed to estimate permeability in a normalized form using weights and biases presented in Table 6-6. The sequence of input parameters should be in the order; bulk density (g/cm3), neutron porosity (%), limestone content (%), dolomite content (%) and anhydrite content (%).
3. The value obtained from Eq. 6.21 should be de-normalized using Eq. 6.22.

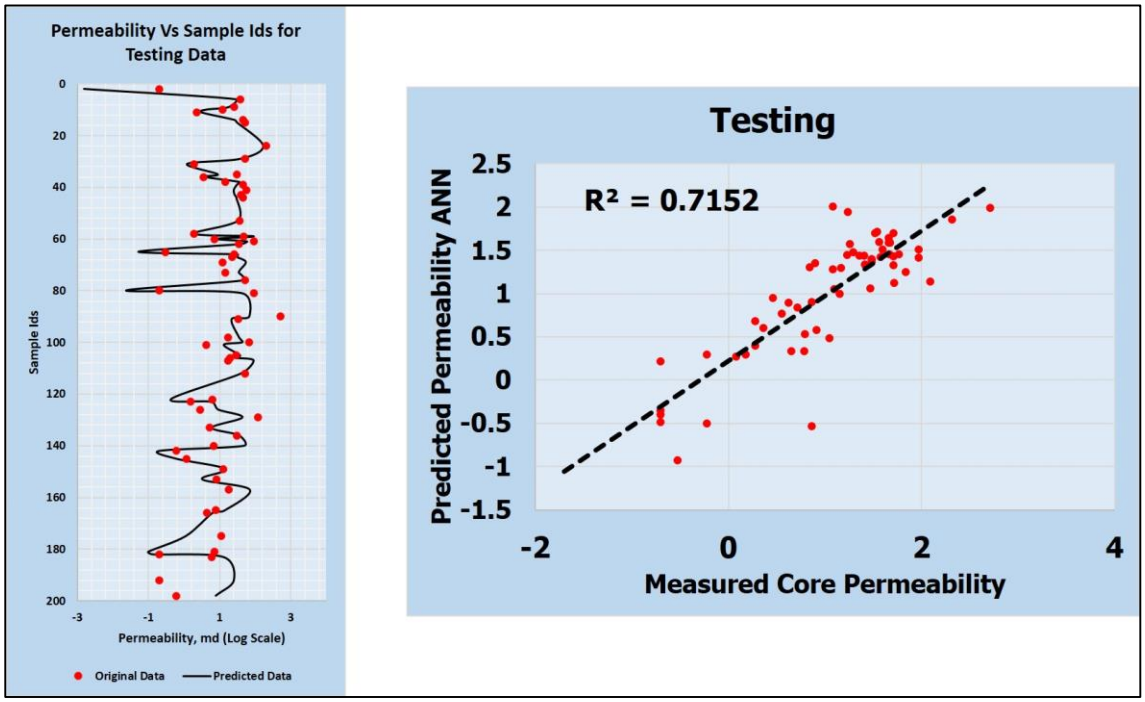


Figure 6-17: Results obtained for testing set of data (a) Comparison of Measured permeability values against predicted permeability values on a plot of Sample ids vs permeability (log scale) (b) Crossplot of Measured Vs Predicted permeability values.

6.4 Validation with Khuff-C Data

For the development of any new mathematical model, validation is an integral part. Validation is carried out either with unseen data or data of similar nature from a different field. To validate the model developed in this study, data of Khuff formation was obtained from another field. Unfortunately, comprehensive field data was not available. The data which was made available belonged to a particular section of Khuff-C from one well. The data reported core permeability, core porosity and grain density as well as well logs

including neutron porosity, bulk density, gamma ray and resistivity logs. Simulated mineral content (from Tech log) was also available. The available section of formation did not show any anhydrite content. Only dolomite and limestone were present. This information did not match with the kind of training data employed for this study. Nevertheless, an attempt was carried out to predict permeability with the mathematical model of filtered data (Eq. 6.14). After certain adjustments of exponents and coefficients, the model returned results within reasonable framework. This can be observed by Fig. 6-18. Estimated and measured values are within the same envelop. The mathematical model is given by Eq. 6.22.

$$log\ k\ = \{(3592.3) * \frac{\phi^3 * MD^{-0.45}}{ML^{-3.53} * \rho^{-1.07}}\} + \{(-3.23) * \frac{\phi^3 * MD^{-0.34}}{ML^{0.44} * \rho^{-0.75}}\} + \{(0.45) * \frac{\phi^{3.14}}{ML^{0.72} * \rho^{-1.91}}\} + \{(6.29) * \phi^{2.92}\} - 1 \quad (6.22)$$

It can be deduced that for a different field, the coefficients and exponents need to be optimized by taking in few points for training and then prediction can be carried out. The results returned some appreciable results but not very accurate. It could be due to many factors but most importantly the diagenetical activities undergone by each formation is very different.

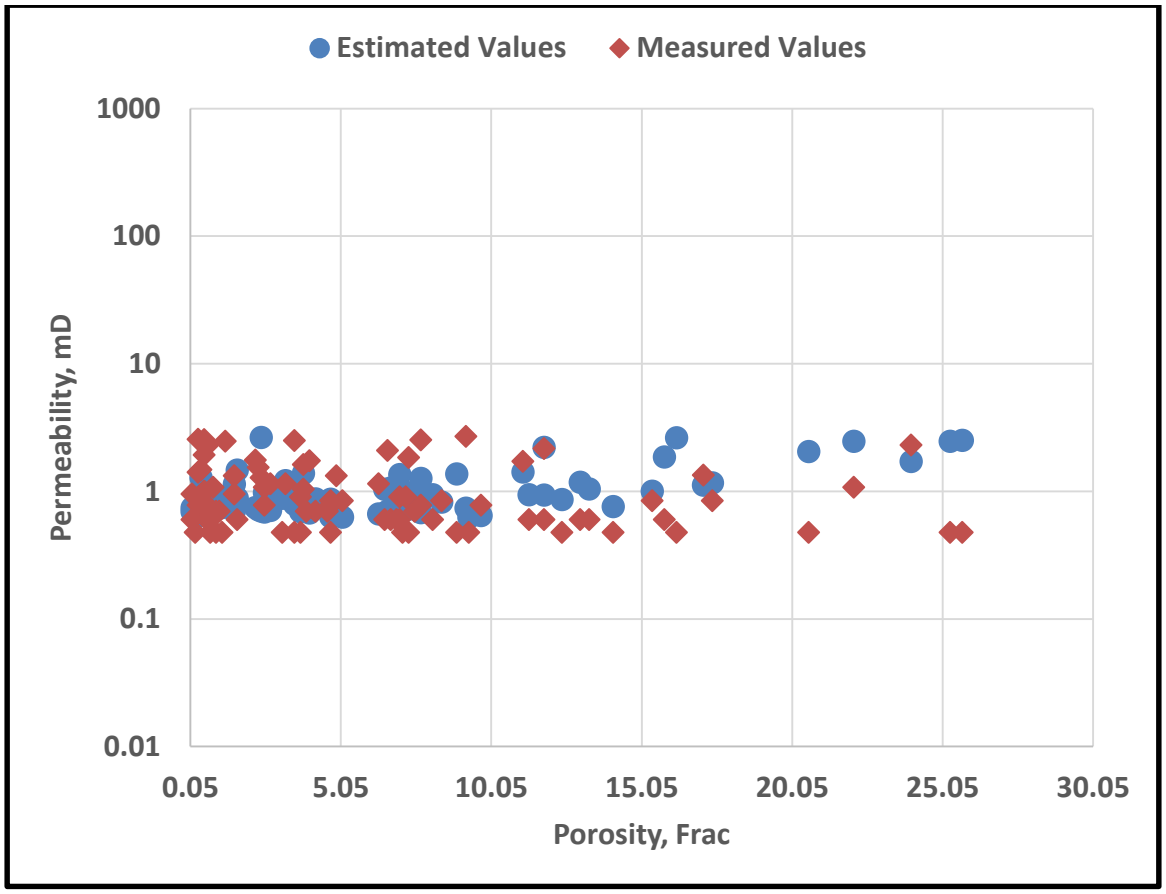


Figure 6-18: Comparison of Permeability vs Porosity crossplot for the data set of Khuff formation from another field

6.5 Validation with Arab-D Data

Another data set was used to check the accuracy of the developed model. This data had been obtained from one well completed in the Arab-D formation. The Arab D is also a well-known formation producing oil in the Arabian Peninsula. The data set contains permeability and well logs showing bulk density, neutron porosity and mineral content

estimated through log photoelectric effect. A crossplot of permeability and porosity are shown in Fig. 6-19.

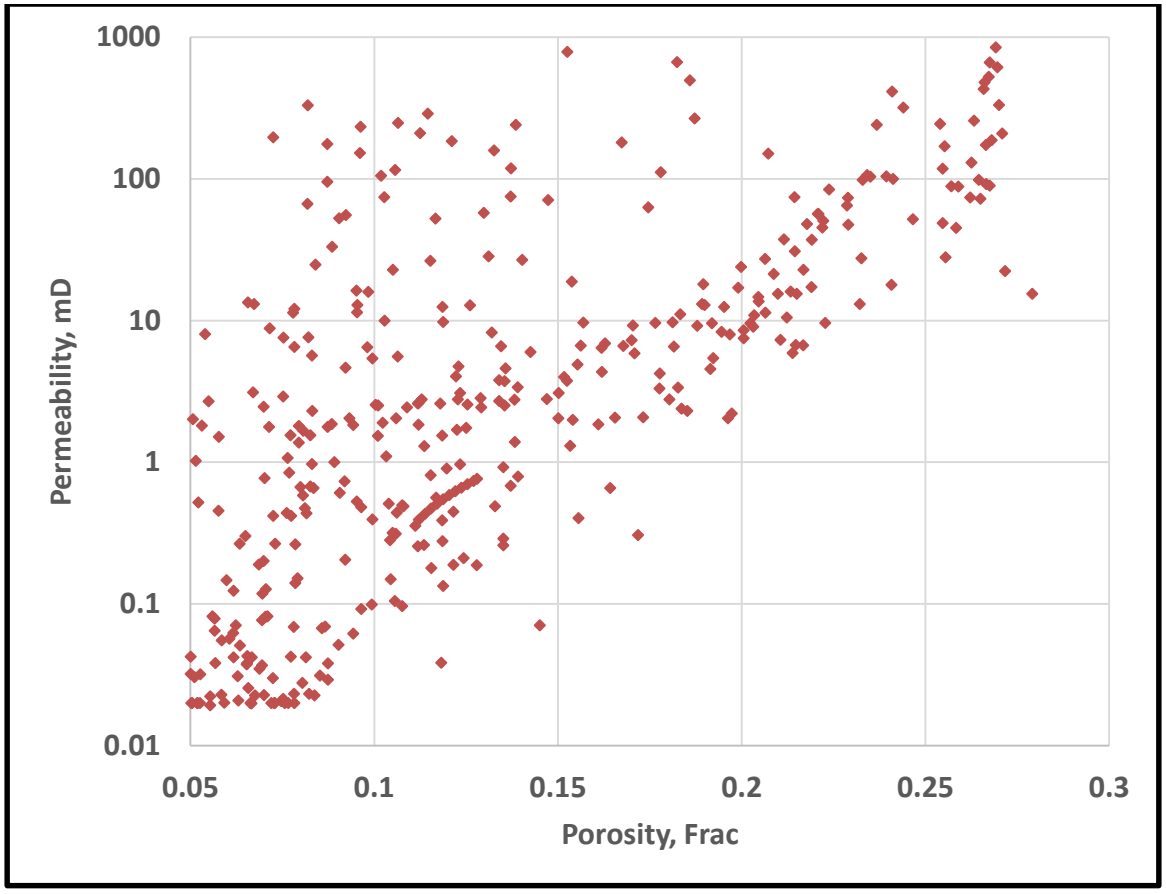


Figure 6-19: Crossplot of measured permeability vs porosity for Arab-D formation.

With slight adjustments of coefficients and exponents, the mathematical model given by Eq. 6-15 was applied to the Arab-D data. The predicted values are shown in Fig. 6-20, where they show a similar trend as the actual values.

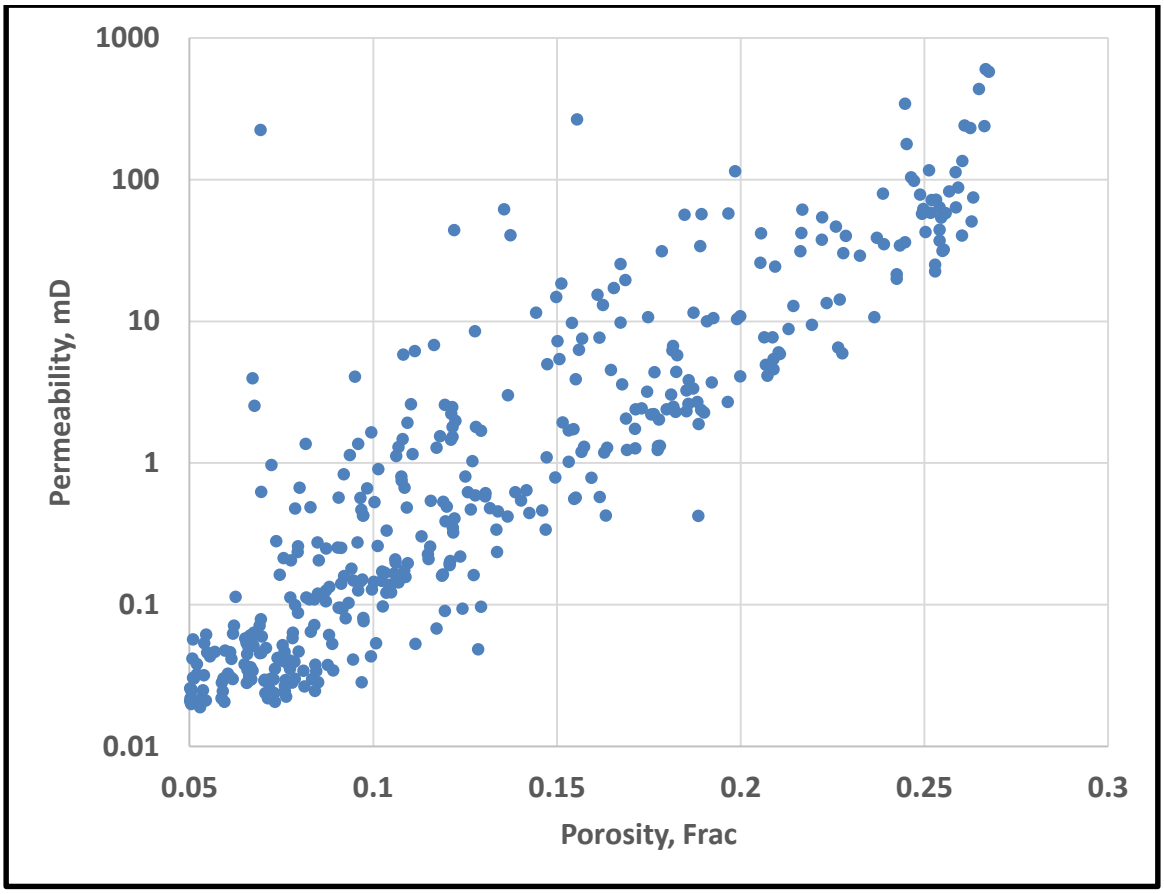


Figure 6-20: Crossplot of estimated permeability vs porosity for Arab-D formation.

When both plots were superimposed over each other, as shown in Fig.6-21, the close match between the two sets becomes evident. This exercise shows that the developed model could be applicable to other carbonate formations with a reasonable prediction power.

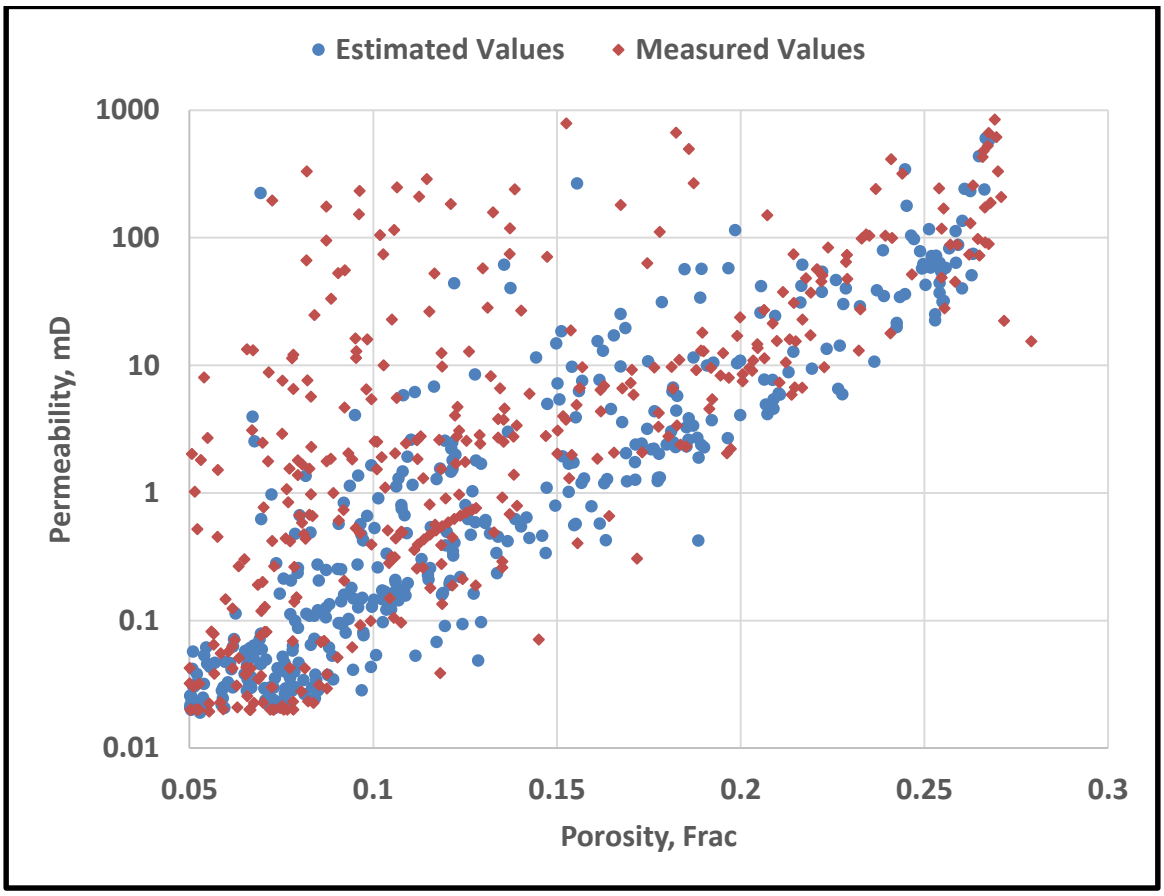


Figure 6-21: Comparison of Permeability vs Porosity crossplot for the data set of Arab-D formation

CHAPTER 7

CONCLUSIONS AND RECOMMENDATIONS

This research work was conducted to develop a correlation that estimates matrix permeability from porosity, bulk density and the rock's mineral content provided by well logs.

7.1 Conclusions

Conclusions drawn from this study are:

1- A new empirical correlation has been developed to predict the permeability for the Khuff formation based on core and log data from eastern Saudi Arabia.
2- The correlation was developed using various techniques such as symbolic regression, neural network and non-linear regression. Neural network and non-linear regression presented reasonable results.
3- The correlation can be applied to other Khuff reservoirs provided that training data is available.
4- The correlation was tested on Arab-D data and showed reasonable accuracy.

7.2 Recommendations

Recommendations for the future work are:

1. R^2 for the developed correlation is low which shows that for accurate permeability determination for carbonates, a parameter that defines diagenesis should be incorporated.
2. NMR log was not available for this data set. Including such data can provide pore- size distribution that would capture the rock texture and provide better accuracy.
3. Validation did not return satisfactory results that shows an in-depth analysis of deviation and possible inclusion of that within the mathematical model.
4. The new correlation should also be tested and tuned for more fields.

References

[1] B. Montaron, “Carbonate Evolution,” *Oil Gas Middle East*, no. August, pp. 26–32, 2008.

[2] R. J. Dunham, “Classification of Carbonate Rocks According to Depositional Textures,” vol. 38, pp. 108–121, 1962.

[3] S. N. Ehrenberg, P. H. Nadeau, and A. A. M. Agrawi, “A comparison of Khuff and Arab reservoir potential throughout the Middle East,” *Am. Assoc. Pet. Geol. Bull.*, vol. 91, no. 3, pp. 275–286, 2007.

[4] A. S. Alsharhan and A. E. M. Nairn, “The late Permian carbonates (Khuff Formation) in the western Arabian Gulf: Its hydrocarbon parameters and paleogeographical aspects,” *Carbonates and Evaporites*, vol. 9, no. 2. pp. 132–142, 1994.

[5] J. Kozeny, “ber kapillare Leitung des Wassers im Boden, Sitz. der Wien,” *Akad. der Wissenschaften*, p. 136, 1927.

[6] Carman, “Fluid Flow Through Granular Beds,” *Trans. - Inst. Chem. Eng.*, vol. 15, 1937.

[7] J. O. Amaefule, M. Altunbay, D. Tiab, D. G. Kersey, and D. K. Keelan, “Enhanced reservoir description: using core and log data to identify hydraulic (flow) units and predict permeability in uncored intervals/wells,” *SPE 68th Annu. Tech. Conf. Exhib. Soc. Pet. Eng.*, no. c, pp. 205–220, 1993.

[8] and H. M. T. M. Altunbay, D. Georgi, “Permeability prediction for carbonate,” *Spe*. 1997.

[9] H. A. Nooruddin and M. E. Hossain, “Modified Kozeny–Carmen correlation for enhanced hydraulic flow unit characterization,” *J. Pet. Sci. Eng.*, vol. 80, no. 1, pp. 107–115, Dec. 2011.

[10] T. Babadagli and S. Al-Salmi, “A Review of Permeability-Prediction Methods for Carbonate Reservoirs Using Well-Log Data,” *SPE Reserv. Eval. Eng.*, vol. 7, no. 2, pp. 8–10, 2004.

[11] W. C. Krumbein and G. D. Monk, “Permeability as a Function of the Size Parameters of Unconsolidated Sand,” *Trans. AIME*, vol. 151, no. 1, pp. 153–163, Dec. 1943.

[12] P. N. Sen, C. Straley, W. E. Kenyon, and M. S. Whittingham, “Surface-to-volume ratio, charge density, nuclear magnetic relaxation, and permeability in clay-bearing sandstones,” *GEOPHYSICS*, vol. 55, no. 1, pp. 61–69, Jan. 1990.

[13] P. H. Nelson, “Permeability-porosity relationships in sedimentary rocks,” *The Log Analyst*, vol. 35, no. 3. pp. 38–62, 1994.

[14] L. Quintero, A. Boyd, A. Gyllensten, and F. El-Wazeer, “Comparison of Permeability from NMR and Production Analysis in Carbonate Reservoirs,” 1999.

[15] T. Chandra, “Permeability estimation using flow zone indicator from Well log data.”

[16] T. Babadagli and S. Al-Salmi, “A Review of Permeability-Prediction Methods for Carbonate Reservoirs Using Well-Log Data,” *SPE Reserv. Eval. Eng.*, vol. 7, no. 2, pp. 75–88, Apr. 2004.

[17] A. H. Thompson, A. J. Katz, and C. E. Krohn, “The microgeometry and transport properties of sedimentary rock,” *Adv. Phys.*, vol. 36, no. 5, pp. 625–694, Jan. 1987.

[18] Martys, “Martys et al., 1991.pdf,” 1991.

[19] H. Pape, C. Clauser, and J. Iffland, “Variation of Permeability with Porosity in Sandstone Diagenesis Interpreted with a Fractal Pore Space Model,” *Pure appl. Geophys*, vol. 157, pp. 603–619, 2000.

[20] J. Muller and J. L. McCauley, “Implication of fractal geometry for fluid flow properties of sedimentary rocks,” *Transp. Porous Media*, vol. 8, no. 2, pp. 133–147, Jun. 1992.

[21] P. M. Wong, “PERMEABILITY PREDICTION FROM WELL LOGS USING AN IMPROVED WINDOWING TECHNIQUE,” *J. Pet. Geol.*, vol. 22, no. 2, pp. 215–226, Apr. 1999.

[22] J. P. Hansen and A. T. Skjeltorp, “Fractal pore space and rock permeability implications,” *Phys. Rev. B*, vol. 38, no. 4, pp. 2635–2638, Aug. 1988.

[23] M. P. Tixier, “Evaluation of Permeability from Resistivity Gradient on Electric Logs,” vol. 17, pp. 68–73, 1949.

[24] C. Y. Yao, S. A. Holditch, and T. A&m, “Estimating Permeability Profiles Using Core and Log Data.”

[25] S. Saner, M. Kissami, and S. Al Nufaili, “Estimation of Permeability From Well Logs Using Resistivity and Saturation Data,” 1997.

[26] S. Mohaghegh, B. Balan, and S. Ameri, “Permeability Determination From Well Log Data.”

[27] G. Xue, “Optimal Transformations for Multiple Regression: Application to Permeability Estimation From Well Logs,” 1997.

[28] F. J. Lucia, “Rock-Fabric / Petrophysical Classification of Carbonate Pore Space for Reservoir Characterization 1,” vol. 9, no. 9, pp. 1275–1300, 1995.

[29] J. Jennings and F. Lucia, “Predicting permeability from well logs in carbonates with a link to geology for interwell permeability mapping,” *SPE Reserv. Eval. Eng.*, vol. 6, no. 4, pp. 215–226, 2003.

[30] M. M. Herron, “Mineralogy from Geochemical Well Logging,” *Clays Clay Miner.*, vol. 34, no. 2, pp. 204–213, 1986.

[31] A. F. Al-anazi, I. D. Gates, and J. Azaiez, “Innovative Data-Driven Permeability Prediction in a Heterogeneous Reservoir,” in *EUROPEC/EAGE Conference and Exhibition*, 2009.

[32] and D. A. S. American Association of Petroleum Geologists., J. H. FANG , C. L.SAMUEL J. ROGERS (2), J. H. FANG (2), C. L. KARR (3), “Determination of Lithology from Well Logs using Neural Netwrok,” *Am. Assoc. Pet. Geol. Bull.*, vol. 76, no. 5, pp. 731–739, 1992.

[33] M. M. Saggaf, M. Nafi Toksoz, and H. M. Mustafa, “ESTIMATION OF RESERVOIR PROPERTIES FROM SEISMIC DATA BY SMOOTH NEURAL NETWORKS.”

[34] Alpana Bhatt, “Reservoir properties from well logs using neural networks,” 2002.

[35] D. Freedman and P. Diaconis, “On the histogram as a density estimator:L 2 theory,” *Zeitschrift f*

Wahrscheinlichkeitstheorie und Verwandte Gebiete, vol. 57, no. 4, pp. 453–476, Dec. 1981.

[36] TohidNejadGhaffarBorhani and SeyedHosseinEmadi, "ApplicationofHydraulicFlowUnitsandIntelligent SystemsforPermeabilityPredictioninaCarbonate Reservoir."

[37] W. J. Ebanks, "Flow Unit Concept--Integrated Approach to Reservoir Description for Engineering Projects, by W. J. Ebanks, Jr.; #91038 (2010)," *Journal*, 1987.

[38] W. R. Blischke, M. Rezaul Karim, and D. N. Prabhakar Murthy, "Preliminary Data Analysis," Springer, London, 2011, pp. 159–189.

[39] "Cluster Analysis: Basic Concepts and Algorithms."

[40] G. Gan, C. Ma, and J. Wu, *Data Clustering: Theory, Algorithms, and Applications*, vol. 20. 2007.

[41] jason Brownlee, "Supervised and Unsupervised Machine Learning Algorithms - Machine Learning Mastery." [Online]. Available: https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/. [Accessed: 28-Nov-2017].

[42] M. C. Kitzig, A. Kepic, and D. T. Kieu, "Testing cluster analysis on combined petrophysical and geochemical data for rock mass classification," *Explor. Geophys.*, vol. 48, no. 3, p. 344, 2017.

[43] T. Euzen, "Well Log Cluster Analysis and Electrofacies Classification: A Probabilistic Approach for Integrating Log with Mineralogical Data*," 2014.

[44] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters," *J. Cybern.*, vol. 3, no. 3, pp. 32–57, Jan. 1973.

[45] J. C. Bezdek and J. C., *Pattern recognition with fuzzy objective function algorithms*. Plenum Press, 1981.

[46] F. Zhao and L. Le, "Fuzzy C-means Clustering for 3D Seismic Parameters Processing," pp. 47–50, 2009.

[47] M. Angel and R. Canelon, "Fuzzy Clustering Based Models Applied To Petroleum Processes."

[48] J. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 1967, pp. 281–297.

[49] "Clustering - K-means." [Online]. Available: https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html. [Accessed: 28-Nov-2017].

[50] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 39. WileyRoyal Statistical Society, pp. 1–38, 1977.

[51] "The EM Algorithm for Gaussian Mixtures Finite Mixture Models."

[52] E. Weinstein, "Expectation- Maximization Algorithm and Applications," 2006.

[53] A. del. Castillo, A. Santoyo, and E. García-Valladares, “A new void fraction correlation inferred from artificial neural networks for modeling two-phase flows in geothermal wells,” *Comput. Geosci.*, vol. 41, pp. 25–39, Apr. 2012.

[54] M. R. Gupta and Y. Chen, “Theory and Use of the EM Algorithm,” *Signal Processing*, vol. 4, no. 3, pp. 223–296, 2011.

[55] E. J. Vladislavleva, G. F. Smits, and D. Den Hertog, “Order of Non-linearity as a Complexity Measure for Models generated by Symbolic Regression via Pareto Genetic Programming.”

[56] D. Keohane, “Nutonian - At the Cutting Edge of Technology, Science, and Data Analysis,” *www.venturefizz.com*.

[57] A. Ruckstuhl, “Introduction to Nonlinear Regression,” 2010.

[58] D. G. Bates,Douglas M. and Watts, “Nonlinear Regression Analysis and Its Applications,” *John Wiley Sons, Inc.*, no. Second.

[59] A. Graves, M. Liwicki, S. Fernandez, R. Bertolami, H. Bunke, and J. Schmidhuber, “A Novel Connectionist System for Unconstrained Handwriting Recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, May 2009.

[60] A. Abdulraheem, M. Ahmed, A. Vantala, and T. Parvez, “Prediction of Rock Mechanical Parameters for Hydrocarbon Reservoirs Using Different Artificial Intelligence Techniques,” in *SPE Saudi Arabia Section Technical Symposium*, 2009.

[61] Z. Tariq, S. Elkatatny, M. Mahmoud, and A. Abdulraheem, “A New Artificial Intelligence Based Empirical Correlation to Predict Sonic Travel Time,” in *International Petroleum Technology Conference*, 2016.

[62] R. P. Lippman and R. P. Lippman, “An Intrduction to Computing with Neural Nets,” in *IEEE ASSP Magazine*, 1987, no. April, pp. 4–22.

[63] S. Mohaghegh, “Neural Network: What It Can Do for Petroleum Engineers,” *J. Pet. Technol.*, vol. 47, no. 1, pp. 42–42, Jan. 1995.

Appendix

Formula used in this manuscript are given as below:

Average Absolute Difference (AAD)

$$AAD = \frac{1}{N}\sum_{n=1}^{N}\left|k_e - \overline{k_m}\right|$$

Where,

N = number of data points

k_e = estimated permeability

k_m = measured permeability

Root mean Square Error (RMSE)

$$RMSE = \sqrt{\frac{\sum(k_e - k_m)^2}{n}}$$

Where,

n = number of data points

k_e = estimated permeability

k_a = measured permeability

Vitae

Name:	Muhammad Hasan
Date of Birth:	May 31, 1990
Permanent Address:	34, Zulfiqar Street No. 4, Phase 8, DHA, Near Marina Club, Karachi, Pakistan.
Email Address:	smh.hasan@gmail.com

Publications:

An Anomalous Productivity Model for Naturally Fractured Shale Gas Reservoirs. **SPE-188033-MS**

A Mathematical Model for Transient Testing of Naturally Fractures Shale Gas Reservoirs. **SPE-188058**

Enhanced Gas Recovery (EGR) Methods and Production Enhancement Techniques for Shale & Tight Gas Reservoirs. **SPE-188090-MS**

Educational Qualification:

MS (Petroleum Engineering) **December, 2017**

King Fahd University of Petroleum & Minerals,

Dhahran, Saudi Arabia.

B.E. (Petroleum Engineering) **December, 2012**

NED University of Engineering and Technology,

Karachi, Pakistan.