

**AES/PE/12-02 Mining a Massive Reservoir Engineering Database
for Determinants of Recovery Efficiency**

24-01-2012 Nathalie Marie-Clair Chün

Title : Mining a Massive Reservoir Engineering Database for Determinants of Recovery Efficiency

Author(s) : Nathalie Marie-Clair Chün

Date : January 2012

Professor(s) : Cor van Kruijsdijk

Supervisor(s) : Al Tucker

TA Report number : AES/PE/12-02

Postal Address : Section for Petroleum Engineering
Department of Applied Earth Sciences
Delft University of Technology
P.O. Box 5028
The Netherlands

Telephone : (31) 15 2781328 (secretary)

Telefax : (31) 15 2781189

Copyright ©2012 Section for Petroleum Engineering

All rights reserved.

*No parts of this publication may be reproduced,
Stored in a retrieval system, or transmitted,
In any form or by any means, electronic,
Mechanical, photocopying, recording, or otherwise,
Without the prior written permission of the
Section for Petroleum Engineering*

Acknowledgements

It is a pleasure for me to thank the persons and institutions that have contributed to the completion of this thesis project.

Firstly, I would like to thank my parents for their love and support throughout all my studies in University.

I am very grateful to Prof.ir. C.P.J.W. van Kruijsdijk, chair person of the thesis committee. I would like to thank him for his unsurpassed knowledge, guidance and support that contributed to the fulfillment of the thesis requirements.

Special thanks are owed to Al Tucker for his time and dedication to his role as my supervisor and mentor. I enjoyed and benefitted greatly from the weekly conversations we had throughout the year. His enthusiasm, encouragement and industry experience have been invaluable on both professional as personal level, for which I am extremely grateful.

I also thank Prof.dr. S.M. Luthi and Prof.dr. K-H.A.A. Wolf for their time and willingness to serve on the thesis committee.

I gratefully acknowledge that this thesis project would not have been possible without the use of proprietary data and other resources provided by Shell. I value that I was granted with the unique opportunity to complete my thesis project with them.

Finally, I am grateful for the time I spent at the TU Delft in the preparation of a Master of Science in Petroleum Engineering. The warm and familiar environment of the Geotechnology department has been beneficial for me in pursuing my degree.

Abstract

A global and industry wide dataset was data mined for determinants of recovery efficiency. Understanding of the factors that are driving variance in reservoir performance is essential for benchmarking current performance and for the screening of new opportunities. The following insights in the origin of variance in reservoir performance could be extracted from this analysis.

Global trends for recovery factor with drive mechanism, reservoir type, geological age, lithology and depositional environment were extracted through subgroup analysis. Other property trends, such as porosity with depth and geological age, were found to be basin specific.

The internal structure of the database and correlations was revealed through principal component analysis. Relative importance of the predictor variables was determined using automatic multivariate linear regression. It was found that the predominant variables include: API gravity, permeability and reservoir temperature.

Additional data was identified through combination of literature review, dimensional- and statistical analysis. The following variables are suggested: dip angle, flow rate, fractional water cut, and pressure drop. Furthermore continuous scales for heterogeneity and fracture intensity, especially for carbonate reservoirs are suggested. To express the confidence level for each reservoir in the database, categorical variables for maturity and data quality are proposed. This research forms the basis for future data mining of the dataset and further improvement of the TQ EUR TOOL in which the data is stored. In a wider context this report presents a high level overview of observations on reservoir performance based on actual reservoirs worldwide rather than laboratory data or theory.

Table of contents

Acknowledgements	II
Abstract	III
List of Tables	VI
List of Figures	VII
Nomenclature	IX
1. Introduction	1
2. Literature review	2
2.1. Prediction of recovery factor	2
2.2. Statistical Analysis	3
2.2.1. <i>Statistical studies</i>	3
2.2.2. <i>Well spacing</i>	4
2.2.3. <i>Equations for predicting recovery efficiency</i>	5
2.2.4. <i>C&C database</i>	7
2.2.5. <i>TORIS and GASIS databases</i>	8
2.2.6. <i>Sample size</i>	8
2.3. Conclusion	10
3. Theory- Reservoir engineering principles	11
3.1. Drive mechanisms	11
3.1.1. <i>Water drive</i>	11
3.1.2. <i>Gas cap drive</i>	12
3.1.3. <i>Solution gas drive</i>	12
3.1.4. <i>Secondary recovery</i>	12
3.2. Fluid displacement	13
3.2.1. <i>Displacement efficiency</i>	13
3.2.2. <i>Mobility</i>	13
3.3. EOR – mobilizing residual oil	15
3.3.1. <i>Thermal recovery</i>	15
3.3.2. <i>Chemical flooding</i>	15
3.3.3. <i>Miscible gas injection</i>	16
3.3.4. <i>Outlook</i>	16
4. Methods	17
4.1. Input data	17
4.1.1. <i>Economic bias</i>	17
4.1.2. <i>Data uncertainty</i>	19
4.1.3. <i>Calculated variables</i>	20
4.2. Workflow	22
4.3. Preprocessing	23
4.3.1. <i>Feature selection</i>	23
4.3.2. <i>Data quality</i>	23
4.4. Analyzing	26

4.4.1.	<i>Univariate analysis</i>	26
4.4.2.	<i>Subgroup analysis</i>	26
4.4.3.	<i>Regression</i>	27
4.4.4.	<i>Principal component analysis</i>	28
5.	Results and Discussion	30
5.1.	Statistical overview of the data	30
5.1.1.	<i>Recovery efficiency</i>	30
5.2.	Subgroup analysis	33
5.2.1.	<i>Drive mechanisms</i>	33
5.2.2.	<i>Lithology</i>	38
5.2.3.	<i>Aquifer strength, lithology and fracture intensity</i>	40
5.2.4.	<i>Depositional environment</i>	43
5.2.5.	<i>Sedimentary basins, geological age and depth</i>	46
5.3.	Multivariate analysis results	50
5.4.	Principal component analysis	50
5.4.1.	<i>Correlations</i>	50
5.4.2.	<i>Components</i>	51
5.4.3.	<i>Loadings and principal scores</i>	51
5.4.4.	<i>Deepwater reservoirs</i>	55
5.4.5.	<i>Performance of gas cap reservoirs</i>	57
5.4.6.	<i>EOR methods</i>	58
5.5.	Multi linear regression	59
5.5.1.	<i>Subgroups and sample size</i>	59
5.5.2.	<i>Most important variables</i>	59
5.5.3.	<i>Expansion to entire dataset</i>	60
5.5.4.	<i>Predict missing variables</i>	61
6.	Additional data	63
6.1.	Literature study	63
6.2.	Dimensional analysis	63
6.3.	Statistical analysis	64
6.4.	Formation water properties	65
7.	Conclusions and Recommendations	66
7.1.	Conclusions	66
7.2.	Recommendations	67
	References	68
	Appendix 1.	70

List of Tables

Table 2.1:	Summary of determinants fo recovery factor according to literature.	9
Table 4.1:	Different measurement methods encountered in the tool.	19
Table 5.1:	Ranges of independent variables. The first three columns give the minimum, median and maximum for the all reservoirs. The ranges are also presented for clastic and carbonate reservoirs separately. Please note that there are (current and ultimate) recovery factors higher than 1. This is the effect of pessimistic estimates of in-place volumes.	31
Table 5.2:	Natural drive mechanism versus secondary recovery method.	35
Table 5.3:	Relative measure of fracture intensity in the TQ EUR TOOL.	40
Table 5.4:	Clastic reservoirs in the dataset per geological age versus depositional environment.	49
Table 5.5:	Carbonate reservoirs in the dataset per geological age versus depositional environment.	49
Table 5.6:	Correlation matrix. The darker the cell, the stronger the correlation.	50
Table 5.7:	Left: Loadings of input variables on the components. Right: Loadings on the rotated solution. Rotation VARIMAX maximizes high loadings and minimizes low loadings per component.	52
Table 5.8:	Comparison of mean values of key parameters for deepwater, offshore and onshore fields in our dataset. The deepwater reservoirs are characterized by low average well densities but high porosity and permeability.	55
Table 5.9:	Percentage of variance in oil recovery factor accounted for by the two most important variables of component 1. Note the change of order for all the reservoirs.	60

List of Figures

Figure 4.1:	Number of oil reservoirs per era. Note that the younger reservoirs are relatively more common. However, the distribution of oil reservoirs over geologic time is not the same for all the regions.	18
Figure 4.2:	Porosity versus depth trend per location.	18
Figure 4.3:	Workflow showing the sequential steps of the data mining process and the resulting output for each step. Modified after (Sharma, 2010).	22
Figure 4.4:	Number of missing values (blue) and zeroes (red) per variable. Exp. STOIP, Ultimate Recovery Factor, Permeability and Initial oil saturation have values for all 1203 reservoirs.	24
Figure 4.5:	Analysis of Variance (ANOVA) flowchart.	27
Figure 4.6:	'Scree', derived from the Old Norwegian word: 'skriða', is the term for the rock debris covering a sloping cliff. Source: http://en.wikipedia.org/wiki/File:Yamnuska_bottom_cliff.jpg	29
Figure 5.1:	Top: Frequency distribution of oil recovery factor. Normal distribution is fitted over the distribution. Bottom: Normal Q-Q plot of the oil recovery factor. The straight line indicates a normal distribution and the dots represent the data. Both plots indicate that there are more reservoirs with low recovery factors than if it would be a normal distribution.	32
Figure 5.2:	Box plot of Current Recovery Efficiency versus drive mechanisms. Five basic statistics can be obtained from the box plot: the median, represented by the bar inside the box, the upper quartile and the lower quartile, represented by respectively the upper and lower outline of the box, the minimum and the maximum value. Between 1.5 and 3 times the box length we find the outliers that are represented by the circles. Not present in this plot are values further than three box lengths away which would be flagged as extreme values and represented by asterisks.	34
Figure 5.3:	Ultimate Recovery Efficiency versus drive mechanisms.	35
Figure 5.4:	Microscopic Sweep Efficiency versus Natural Drive mechanism. The microscopic sweep efficiency is significantly lower for solution gas drive reservoirs. This is caused by the fact that this group is characterized by higher residual oil saturations.	35
Figure 5.5:	Macroscopic Sweep Efficiency versus Natural Drive Mechanism.	37
Figure 5.6:	Initial oil saturation versus lithology.	39
Figure 5.7:	Final oil saturation versus lithology.	39
Figure 5.8:	Ultimate recovery factor for clastic reservoirs as a function of aquifer strength.	41
Figure 5.9:	Ultimate recovery factor for carbonate reservoirs as a function of aquifer strength.	41
Figure 5.10:	Ultimate recovery efficiency for carbonate reservoirs with no fractures and low fracture intensity. The plot shows that for these reservoirs the aquifer strength has a positive influence on recovery efficiency.	42
Figure 5.11:	Ultimate recovery efficiency for carbonate reservoirs with high fracture intensity. For highly fracture reservoirs the aquifer strength has a negative impact on the recovery efficiency.	42
Figure 5.12:	Net to gross ratio versus depositional environment.	44
Figure 5.13:	Unrecovered mobile oil versus depositional environment. Note that the unrecovered mobile becomes negative in some cases. This is unphysical and is caused by either too low estimates of microscopic sweep efficiency, which is a common problem in the dataset due to uncertainties around the origin of the 'Final So' value. See chapter 5, paragraph	

	5.2.1 'Drive mechanisms. It could also be the result of too high estimates of ultimate recovery, caused by pessimistic in-place estimates.	44
Figure 5.14:	EUR per well versus depositional environment. There appears to be a relation between reservoir heterogeneity and EUR per well. Only Deep Marine reservoirs have a significantly higher EUR per well. This is due to the fact that all Deep Marine reservoirs are located in current day deepwater. The economic bias allows only for reservoirs with high EUR in this economically difficult location.	45
Figure 5.15:	Porosity versus geological age. A potential trend of decreasing porosity with age seems to be violated by Paleozoic reservoirs. More detailed inspection of the data revealed that most of those reservoirs are located in the same basin. Also, the Jurassic and Devonian reservoirs are restricted to the North Sea basin.	47
Figure 5.16:	Porosity versus Geological Age for reservoirs in the North Sea Basin.	48
Figure 5.17:	Porosity versus depth trend per region. Note that no trend for the Americas region could be obtained. This is because the reservoirs in the Gulf of Mexico follow a different trend than the other reservoirs.	48
Figure 5.18:	Scree plot. This plot can be used to determine the required amount of components by determining the 'elbow' of the graph, here at the elbow sits at 5 components. Alternatively, one can select the number of components by selecting ones with the eigenvalues of each component. In this case 4 have eigenvalues greater than 1. If the eigenvalue is less than 1, a component does not explain more variance than a single variable itself would.	52
Figure 5.19:	Combined loadings (black) and principal scores (red) plot for all reservoirs. The first principal component has heavy loadings on fluid properties. Reservoirs scoring high on this component are light type oils. The second component has heavy loadings on rock properties but also on water depth. Reservoirs scoring high on this component form a distinct cluster and could be identified as deepwater reservoirs.	53
Figure 5.20:	Principal scores of reservoirs with markers indicating: location, geological age, lithology and drive mechanism. The principal components represent the same variables as shown in Figure 5.19.	54
Figure 5.21:	Combined loadings (black) and principal scores (red) plot for all Non-Deepwater reservoirs.	56
Figure 5.22:	Principal scores plot with markers indicating reservoir type. The principal components represent the same variables as shown in Figure 5.19. To increase visibility, the axes are cut at PC scores of 3. Note that the heavy oils form a coherent group at the bottom left of the cloud. Under saturated reservoirs plot over the entire range, even outside the plotted are.	57
Figure 5.23:	Principal scores with markers indicating EOR method. The principal components represent the same variables as shown in Figure 5.19. The plot shows that the first principal component, representing fluid properties, is most important in determining which method is applied. Also note that in this plot we can recognize the heavy oils from Figure 5.22 as reservoirs on steam flooding.	58
Figure 5.24:	Standardized weights for predictors of recovery efficiency. In red, the significant variables according to t-test with tolerance level of 0.10.	60

Nomenclature

Reservoir engineering symbols and definitions

API	Oil gravity, °API
Boi	Oil Formation Volume Factor at initial reservoir pressure, reservoir barrels per stock tank barrels RB/STB
D	Datum depth, m
Dw	Water depth, m
EOR	Enhanced Oil Recovery
Exp. STOIPP	P50 estimate of STOIPP
GIIP	Gas in place at initial reservoir pressure
Hc	Column height, m
IOR	Improved Oil Recovery
k	Rock permeability, milliDarcies
NET	Net pay thickness, m
STOIPP	Stock Tank Oil in place at initial reservoir pressure
P	Initial reservoir reservoir pressure, psi
Pb	Bubble point pressure, psi
GOR	Initial gas-oil ratio, standard cubic feet per stock tank barrel, SCF/STB
Oil RF%	Estimated Ultimate Oil Recovery Factor, %
Swi	Initial water saturation
Soi	Initial oil saturation
Sor	Residual oil saturation
T	Initial reservoir temperature, °C
Well	Well density, wells/km ²
λ_g	Gas mobility
λ_o	Oil mobility
λ_w	Water mobility
ϕ	Porosity, fraction
μ or μ	Oil viscosity, cP
μ_g	Gas viscosity, cP
μ_o	Oil viscosity, cP
μ_w	Water viscosity, cP

Statistical symbols and definitions

ANOVA	Analysis of Variance, statistical method to compare the mean value of rational subgroups.
PCA	Principal Component Analysis
Q-Q plot	'Q' stands for Quantile. A probability plot to compare two distributions, in this case the data and the normal distribution, by plotting their quantiles against each other.
TQ EUR TOOL	Top Quartile Estimated Ulitmate Recovery Tool. Tool in which the analyzed data set is stored.

1. Introduction

Traditionally, one of the most fundamental tasks of the reservoir engineer has been to predict the reservoir's performance. Reservoir simulation is a very powerful tool to predict recovery of a single field for numerous of development options. But, for decision taking and portfolio management these numbers need to be put into perspective. To benchmark current performance and for the screening of new opportunities analogues databases are employed. However, benchmarking recovery factor is only useful if we understand the factors that are driving variance in reservoir performance. Only a handful of studies paid attention to determinants for recovery factor and their datasets are geographically limited. This study aims to put reservoir performance in a global context using a massive industry wide reservoir engineering database. The research was structured along three lines, namely; Statistical analysis, Reservoir Engineering principles and Additional data.

This three-way approach is also reflected in the structure of this report. In the introduction both the previous statistical studies on reservoir engineering datasets, chapter 2 'Literature review', and the reservoir engineering background, chapter 3 'Theory- Reservoir Engineering Principles', will be discussed. The middle part consists of the results from the statistical analysis of the database. The results are discussed through reservoir engineering principles. The most important findings of the analysis are summarized in the conclusion in chapter 7 'Conclusion and Recommendations' paragraph 7.1. Readers particularly interested in the suggested improvements to the TQ EUR TOOL will find a detailed discussion of recommended additional data in chapter 6 'Additional data'. A summary of the suggested additional data and proposed further research is presented in chapter 7 'Conclusion and Recommendations' paragraph 7.2 'Recommendations'.

2. Literature review

A thorough literature review has been conducted in order to get a better view on the current status of research related to TQEUR and to help define the project direction. The prime source of information was the online paper database onepetro.org. Via this website several studies published by the Society of Petroleum Engineers (SPE) and the American Petroleum Institute (API) have been obtained. After a short introduction to the different approaches for predicting reservoir performance, a discussion of previous statistical studies of reservoir engineering databases will be presented. Finally, the key learning points are summarized in the conclusion.

2.1. Prediction of recovery factor

When studying the literature it is striking that a lot of papers describing reservoir behavior appeared around 1940's. Wyckhoff, 1940, explains that before this time, tasks of the petroleum engineer were more focused on optimizing drilling and production. With a growing understanding of the reservoir, one of the most important tasks of the reservoir engineer became to predict the reservoir's performance. The earliest approach to tackle this problem was to classify the reservoir according to its properties and compare it to analogs. This method remains popular because it is easy, quick and can be done even with very limited data. Nevertheless, quantifying the accuracy and the validity of the analog is problematic and varies from case to case. Analog data is therefore almost always used in addition to more sophisticated methods.

In addition to the analog studies, empirical relationships were formulated and used for estimations and predictions. Many empirical relationships turned out to be fairly good approximations of the reality and are still broadly used throughout the industry. However, driven by advancements in technology with time the emphasis shifted towards the development of more complex theoretical and analytical solutions to describe the underlying physical phenomena mathematically. Analytical methods often incorporate a lot of assumptions and therefore miss out on essential details. A well known example is the material balance method, which neglects the reservoir's heterogeneity. As a result, the oil-in-place estimates for heterogeneous reservoirs are invariably too low (van Everdingen, 1980).

Some of the concerns related with the analytical solutions could be overcome by numerically modeling of the reservoir. Due to increasing computer capacity reservoir simulators became very popular from the 1980s onwards. Whereas, the use of analytical methods remains of interest not only in the academic world but is commonly used for pre-screening in the industry too.

In the 1970s, the first major Enhanced Oil Recovery (EOR) projects were launched and in the early 80s the earliest field results from EOR pilots became available. This is reflected in literature by a

renewed interest to determine parameters that influence recovery efficiency with special focus on those playing a key role in EOR.

Nowadays planning of the EOR development programs relies heavily on reservoir simulators and simulation plays a great role in the development of new fields and re-development of mature assets. Despite the ever increasing computing power, robust and reliable reservoir modeling remains an expensive and time-consuming exercise. Additionally, reservoir simulation requires many input parameters and the accuracy of simulation results are limited by uncertainties in the input data. Continuous technical improvement of the measurements devices increased the quality of the input data. In spite of that, sometimes there is not enough data available for example in the exploratory state of the field life or when an oil property is sold from one company to another. In those cases, quick screening of opportunities based on very limited data is required. Hence the industries need for and renewed interest in methods that allow fast screening without reservoir modeling. For this reason and also due to the increased capacity that allows keeping large databases there is a diversion to the use of data mining techniques in reservoir engineering. It is most likely that in the end a combination of all these methods and the comparison between them may yield the best results in the prediction of recovery factors.

2.2. Statistical Analysis

2.2.1. Statistical studies

One of the aims of this thesis is to find determinants for recovery efficiency by applying statistical analysis to the database. Few studies like this have been carried out for the reason that this type of study consumes a lot of time and thorough data mining requires an extensive and reliable database which are commonly not available to the public. Most of the studies focus on a specific rational subgroup for which they determine what the controlling factors are. Rational subgroups represent variation that is inherent to the process, which makes it easier to distinguish special-cause variation. Furthermore, the subgroups are very useful for benchmarking, one of the aims of the TQ EUR Tool database.

On the other hand, when the chosen subgroups are too specific one is left with too few samples to draw statistically significant conclusions on observed trends. On top of that, little attention appears to be given to prove whether the reasonable subgroups are indeed (statistically) different. This leaves an interesting opportunity for this study: to test the validity of the subgroups from literature on the TQ EUR database and determine the importance of certain groups in the portfolio. Themes of subgroups encountered are: drive mechanisms, lithology and fluid type. Additionally, we will search for unknown/unexpected subgroups through the process of data clustering.

2.2.2. *Well spacing*

Craze and Buckley, 1945, were the first to analyze an API reservoir engineering database containing approximately 100 reservoirs and their respective parameters. More specifically the objective of their research was to study the effect of well spacing on the recovery factor. A relationship like this is very valuable, as it can be used as an argument to justify infill drilling. However, they have not found evidence to support a correlation between well spacing and recovery efficiency and based on the results of their studies they propose that there might not be such a relationship. Furthermore, they concluded that the following factors are predominant in determining the oil recoveries; oil viscosity, reservoir pressure decline (%) and formation permeability. The effects of porosity, connate water saturation and shrinkage were eliminated from the analysis. Since the data points of reservoirs still show some scatter after evaluation and elimination of the effects of oil viscosity, reservoir pressure decline and formation permeability it is assumed that there are more factors that exert influence on the recovery efficiency.

In the discussion of this paper it is noted that even though the data does not support a correlation between well spacing and recovery efficiency, the outcomes of the study are not sufficient to entirely rule out the possibility either. Other critics include that the approach is not satisfying to detect the impact of well spacing if this effect would be minor. The author believes that the research of Craze and Buckley was too much focused on either proving or rejecting the hypothetical relationship between well spacing and recovery efficiency. The resulting biased approach is not useful for drawing firm conclusions about the impact and importance of other parameters on the recovery factor.

Also the study of Vietti, 1945, tried to obtain a relation between well spacing and recovery and similarly failed in finding one. In their opinion using intra-field parameters to obtain a relation for total field recovery yields erroneous results. Van Everdingen, 1980, presented another study of the correlation between well spacing and recovery factor as motivation for infill drilling and combined water flooding/infill drilling programs. Based on the results of his study he too concluded that it was impossible to find such a relationship. He suggests that the failure of his and previous attempts might be caused by disregarding reservoir heterogeneity. His study reveals that water flooding is recommended to maintain the pressure above bubble point in depletion type reservoirs, but that with low well densities water flooding becomes highly ineffective.

The question whether well spacing has an influence on the ultimate recovery remains unsolved as none of the studies succeeded in finding a relationship and none of them found strong enough evidence to reject the hypothesis entirely either. A different dimension is added to the discussion by including horizontal wells, since they allow completely different drainage patterns.

2.2.3. Equations for predicting recovery efficiency

Guthrie and Greenberger, 1955, studied 73 water drive reservoirs from the same list published by Craze and Buckley. For their analysis they selected ten parameters and constructed what is called a correlation matrix. A correlation matrix determines the correlation coefficients of the variables and is used to calculate the weights used for a linear regression model. The advantage of this method is that it is purely mathematical and there is no bias towards any of the selected parameters. On the other hand, selecting the right parameters for the analysis is therefore essential to the success of the outcome. Based on their research they proposed the following equation.

- *Guthrie and Greenberger, 1955, equation for the recovery factor of water drive reservoirs:*

$$\text{Oil RF\%} = 0.272 \log(k) + 0.256 S_{wc} - 0.136 \log(\mu_{oi}) - 1.538\phi - 0.0003h + 0.114 \quad \text{Equation 1}$$

Where,

- k = Permeability, md
- S_{wc} = Connate water saturation
- μ_{oi} = Initial oil viscosity, cp
- Φ = Porosity
- h = Net-pay thickness, ft

In addition they proved that many parameters are inter-dependent and that there is a high correlation between depth and formation volume factor, followed by correlations between formation volume factor and log oil viscosity and between porosity and permeability. Guthrie & Greenberger recommend using the average of all log permeabilities instead of the logarithm of the average permeability. Because the impact of permeability on the recovery factor is so large, this will improve the accuracy of the correlation significantly. For future research they suggest to include a measure for shale content of the sand, net-pay thickness of the reservoir, porosity and connate water saturation.

The API Bulletin D14: “A Statistical Study of Recovery Efficiency” presents another statistical analysis of determinants for ultimate recovery carried out by a subcommittee of the API under supervision of J.J. Arps in 1967. The committee concluded that water drive and solution gas drive are fundamentally different and therefore came up with two equations. For water drive reservoirs the following equation was established.

- *Arps, 1967, equation for the recovery factor of water drive reservoirs:*

$$\text{Oil RF\%} = 0.549 \left[\left(\frac{\phi(1-S_{wc})}{B_{oi}} \right)^{0.0422} \left(\frac{k\mu_{wi}}{\mu_{oi}} \right)^{0.0770} (S_{wc})^{-0.1903} \left(\frac{P_i}{P_a} \right)^{-0.2159} \right] \quad \text{Equation 2}$$

Where,

- Φ = Porosity
- S_{wc} = Connate water saturation
- B_{oi} = Initial formation volume factor
- k = Permeability, md
- μ_{wi} = Initial water viscosity, cp
- μ_{oi} = Initial oil viscosity, cp
- P_i = Initial pressure, psia
- P_a = Abandonment pressure, psia

The data set of 70 water drive reservoirs only included sands and sandstones. Another equation was derived for solution-gas-drive reservoirs below the bubble point. The data set of 80 reservoirs included both sandstones and carbonates.

- *Arps, 1967, equation for the recovery factor of solution gas drive reservoirs:*

$$\text{Oil RF\%} = 0.418 \left[\left(\frac{\phi(1-S_{wc})}{B_{ob}} \right)^{0.1611} \left(\frac{k}{\mu_{ob}} \right)^{0.0979} (S_{wc})^{0.3772} \left(\frac{P_b}{P_a} \right)^{0.1741} \right] \quad \text{Equation 3}$$

Where,

- P_b = Bubble point pressure, psia
- B_{ob} = Formation volume factor at P_b
- μ_{ob} = Oil viscosity at P_b , cp

When we compare the equations of Arps to that of Guthrie & Greenberger the observation can be made that both equations depend on the same parameters, though with different weights. This could either mean that the controlling variables for both mechanisms are indeed the same or that there was a bias in selecting parameters for the analysis. We also note that where Guthrie & Greenberger, 1955, included net pay thickness in the equation, Arps, 1967, left it out and instead added pressure drop. The differences in the weights of the parameters are the effect of biases in the input data which makes it hard to transfer the equations from one data set to another. This was also shown by Sharma, 2010. Hence, the value of these studies is that they show which parameters are the most influential rather than that they present a universal equation to predict recovery efficiency.

A key difference is that the signs for water saturation (S_{wc}) and porosity (ϕ) differ. The signs in the equation of Guthrie & Greenberger are counter intuitive. This is most likely the effect of the interrelationship between the independent variables that multi linear regression does not account for. Therefore, the weight that a variable gets does not represent a physical relation anymore but

rather a correction to the weights of related variables. In this case the water saturation can be a correction to the weights of porosity. Because Guthrie & Greenberger took a more mathematical approach, their correlation yielded a good fit for their data set but seems unphysical. This highlights the need for integrating both statistical analysis and reservoir engineering principles to obtain meaningful results.

An unsuccessful attempt to develop an equation for recovery factor was carried out by Gulstad, 1995. His work explains that statistical methods might be inadequate for describing the complex nature of oilfield characteristics; additionally he argues that uncertainties in the input data could mask the underlying relationships. Gulstad agrees with van Everdingen, 1980, that neglecting the heterogeneous nature of the reservoirs might be responsible for the fact no valid correlation could be obtained. It is interesting to notice that reservoir heterogeneity was also not accounted for by the work of Craze & Buckley, Guthrie & Greenberger and Arps.

2.2.4. C&C database

The C&C Reservoirs' database is a commercial digital analog system containing approximately thousand producing reservoirs worldwide. Qing Sun, 2003, limited his study to 250 carbonate reservoirs of the C&C database. It is interesting to note that this research is aiming to classify reservoirs according to their characteristics rather than to build a statistical model for the prediction of recovery factors. Qing Sun concludes that the factors influencing the recovery factor in order of importance are as follows: fluid type, pore/fracture network, reservoir heterogeneity, drive mechanism and wettability. Whether carbonate reservoirs behave like conventional reservoirs is strongly dependent on these factors.

The influence of fractures on the recovery efficiency is also highlighted by the research of Allan, 2003. Based on the study of 100 fractured fields within the C&C database, Allan proposes to further divide fractured reservoirs in four groups that differ in the sense whether storage and flow paths are provided by the matrix, by the fractures or by a combination of both. In reservoirs where the matrix porosity is low, the fracture network has a great influence on the recovery efficiency. Due to the fact that many fractures extent into the aquifer, the strength of the aquifer is a critical factor for the reservoir's performance. In fractured reservoirs with higher porosities, fluid and reservoir properties such as API gravity and matrix permeability are also strong determinants for recovery efficiency. None of these types of reservoirs have been produced to final depletion without application of EOR techniques.

2.2.5. TORIS and GASIS databases

In Sharma, 2010, two datasets were used for the classification and estimation of ultimate recovery: the Tertiary Oil Recovery Information System (TORIS) for oil reservoirs and the Gas Information System (GASIS) for gas reservoirs. In the study the correlations of Guthrie & Greenberger, 1955, and Arps et al., 1967, were applied to 24 reservoirs from the TORIS database. However, both correlations turned out to be incapable of accurate prediction of the recovery factor for the tested oil reservoirs. One can argue whether or not this sample size is large enough to draw statistically significant conclusions.

In addition, Sharma proposed an alternative approach to formulate new regression models for both oil- and gas reservoirs. In case of the oil reservoirs, cluster and principal component analysis was needed to introduce orthogonality to the dataset and improve the accuracy of the multivariate linear regression model. Introduction of orthogonality also allowed the use of the naïve Bayesian approach to construct likelihood functions. A linear regression model was fit on the data for gas reservoirs and the model was successfully tested for robustness on tight gas reservoirs within the dataset. Based on the fact that it was easier to obtain a good fit for gas reservoirs than for oil reservoirs the assumption was made that behavior of gas reservoirs is generally less complex. Besides the conclusions based on the data sets studies, the paper also provides a general work flow for data mining. The recommended approach includes data preprocessing.

2.2.6. Sample size

It is vital to understand that because of outlier removal and segregation in subgroups the size of the used data sets is often significantly reduced. When statistically analyzing data it is essential to have enough data points, since the larger the dataset the better the correlation will be. Guthrie & Greenberger mentioned that for the method that involves a correlation matrix the more factors are included in the correlation, the larger the sample size to test it on should be. A great advantage of the TQ EUR TOOL is therefore its size. The database includes a total of 1193 reservoirs and contains a variety of drive mechanisms and rock types. Also the TORIS and GASIS datasets have a large size, respectively 1381 oil reservoirs and 19.220 gas reservoirs. However, they are geographically concentrated as all of these reservoirs are located in the U.S. In the study of Sharma, 2010, only 95 reservoirs of the TORIS data set were selected for further investigation because the data fields of the other reservoirs were incomplete.

Table 2.1: Summary of determinants fo recovery factor according to literature.

Study	Craze& Buckley (1945)	Guthrie& Greenberger (1955)	Arps (1967)	Arps (1967)	Sharma (2010)	Sharma (2010)
<i>Subgroup</i>	Gas solution drive and water drive sands and sandstones	Water drive, Sands and sandstones	Solution gas drive, All rock types	Water drive, Sands and sandstones	Oil reservoirs TORIS database	Gas reservoirs GASIS database
<i>Uncontrollable parameters</i>						
Oil viscosity	*	*	*	*	*	
Water viscosity				*		
Permeability	*	*	*	*	*	
Porosity		*	*	*	*	*
Water saturation		*	*	*	*	*
Oil saturation					*	
Initial reservoir pressure				*	*	*
Bubble point pressure			*			
Net-pay thickness		*			*	*
Oil formation volume factor			*	*	*	
API gravity					*	*
Reservoir heterogeneity					*	
Depositional environment					*	
Structural compartmentalization					*	
Dip angle					*	
Depth					*	*
Water depth						*
Temperature						*
Gas-In-Place						*
<i>Controllable parameters</i>						
Pressure decline	*					
Abandonment pressure			*	*		
No injectors					*	
No producers					*	
Well spacing					*	
Well density					*	
Production rate					*	
Cumulative Gas Production						*

2.3. Conclusion

To date, few studies that have used a statistical approach have resulted in a successful regression model for the prediction of recovery efficiency. Generally such studies apply to a specific subset of reservoirs and/or locations. Qing Sun 2003 and Allen 2003, for example only provided a hierarchical ranking indicating which of the reservoir rock and fluid properties exert the largest influence on recovery efficiency.

Literature reveals that important predictors for recovery efficiency are as follows: oil viscosity, permeability, porosity, (connate) water saturation, (initial) reservoir pressure, thickness, oil formation volume factor, water viscosity. See Table 2.1. It is interesting to note that all studies that included permeability also described oil viscosity, together forming the mobility factor, as determinants for recovery factor. The same observation can be made for porosity and water saturation.

Guthrie & Greenberger, 1955, and Arps, 1967, presented a statistically derived equation to calculate recovery factor. However, Sharma recently showed that applying the correlation to a database other than the one it was designed for yields poor results. Sharma proposed a new correlation and hereby used cluster- and principal component analysis. Where it was hard to obtain a solid correlation for oil recovery a good fit was reached for gas reservoirs.

Moreover, Sharma showed that both well spacing and reservoir heterogeneity play a role in the ultimate recovery of oil reservoirs. Reservoir heterogeneity was neglected by most of the previous studies, but is assumed to be predominant in determining the recovery factor. It is proposed that disregarding the heterogeneous nature of reservoirs might misleadingly have led to the conclusions that well density has no influence on the recovery factor. Finding a correlation between well spacing and recovery efficiency has been of interest for a long time as such a relationship would provide a great motivation for infill drilling.

Sofar, combining the statistical analysis with reservoir engineering principles has been restricted to parameter selection. It is assumed that including more sophisticated reservoir engineering knowledge and reasonable subgroups in the statistical analysis can lead to less variance of the predictor variables and therefore possibly to better correlations.

3. Theory- Reservoir engineering principles

The previous chapter, literature research, discussed statistical studies of recovery factor. Now we will focus on the background theory. Understanding the reservoir engineering principles will help to identify what the determinants for recovery factor are. In this chapter we first discuss the different type of drive mechanisms and their main production characteristics. Subsequently a more detailed explanation of the displacement process will be given. Finally, EOR methods will be discussed in the light of the discussed reservoir engineering principles.

3.1. Drive mechanisms

The energy that allows for oil production namely; a pressure gradient under which oil flows from the reservoir rock to the well bore is provided by what is called the reservoir drive mechanism. Six types of primary production drive mechanisms have been identified: 1) water drive, 2) gas cap drive, 3) dissolved-gas drive, 4) gravity drainage, 5) compaction drive and 6) combination drive. Early identification of the drive mechanism is essential to optimize field development and ensure high recovery efficiency. In this report we will focus on the three main primary recovery mechanisms: water drive, gas cap drive and solution gas drive which together account for 65% of the reservoirs. Additionally, we will discuss the reservoirs on water injection (25.5%), gas injection (2%) and other enhanced oil recovery methods (7%).

3.1.1. Water drive

Water drive is the most abundant and most efficient natural drive mechanism in oil reservoirs. The production is characterized by the maintenance of constant reservoir pressure, retention of well productivity and low gas oil ratios. Recovery factor ranges from 35 up to 75 (Ahmed, 2010). Conditions that promote high recoveries from water drive reservoirs are thick oil columns, good communication with the aquifer, high permeabilities, uniform and continuous sands, fairly low oil viscosities (favorable mobility ratio).

Water drive reservoirs can be further subdivided into bottom water and edge water reservoirs. Edge water drive is mainly horizontal parallel to the bedding and bottom water drive vertical. A key challenge for most edge water drive reservoirs is the risk of gravity tonguing. As the water tongue under runs the oil, early water break through occurs while parts of the reservoir remain unswept.

Although the displacement can be stable in some cases, if the advancement angle is large with a maximum of 90 degrees. Most reservoirs are operated at a rate of fluid withdrawal which is too high for stable displacement as economic considerations demand high production rates. Hence, there is often a risk of water coning especially when perforations are placed too close to the oil-water-contact. The efficiency of bottom water drive is mainly a function of vertical permeability. Since

vertical permeability is usually less than the permeability parallel to the bedding, the permeability (in the direction of the flow) of bottom water drive reservoirs is on average lower compared to that of edge water reservoirs.

3.1.2. Gas cap drive

In gas cap drive reservoirs the energy is provided by the expansion of free gas that is accumulated in the gas cap. Reservoir pressure slightly declines upon production and wells can produce at a more or less steady rate. The gas-oil-ratio will be increasing slightly but remains low, except for the wells close to the gas-oil-contact. For this reason, the wells should be perforated as far from the gas oil contact as possible to avoid high gas production. Also, oil withdrawal should be restricted as preservation of gas cap should be insured. When managed properly, gas cap drive reservoirs can be very efficient and ultimate recovery ranges from 20 up to 40% (Ahmed, 2010). Segregation is one of the main factors influencing the recovery efficiency. Ineffective segregation will lead to a hybrid form of dissolved gas and gas-cap drive and yields higher GOR than normal gas cap drive. Causes for ineffective segregation are flat structures, low permeabilities or re-pressurization through gas injection in oil zones.

3.1.3. Solution gas drive

The principle behind solution gas drive is that a reduction in reservoir pressure leads to the liberation of associated gas from the oil which in turn provides pressure support. The depletion is characterized by a rapid drop in both reservoir pressure and well potential, whilst gas-oil ratio initially rises and then steadily declines. Characteristic conditions for this type of drive are: flat structures, absence of gas cap or aquifer and high production rates. Because of the large viscosity difference between oil and gas, the gas is far more mobile and will be depleted before the oil. Hence, the risk that the drive energy available from the release and expansion dissolved gas might be exhausted before all oil is produced. As a result the expected recovery factor is relatively low and ranges from as low as 5 to as high as 30 % (Ahmed, 2010).

3.1.4. Secondary recovery

When natural drive mechanisms are not sufficient to drive the oil to the wellbore, external fluids either water or gas can be brought into the reservoir to maintain pressure and displace the oil. The most common secondary recovery method is water flooding, which besides the purpose of enhancing production can be an elegant solution to dispose undesired production water. The limit of secondary recovery efficiencies is determined by quantities of injection fluid that are being produced and the operational costs.

3.2. Fluid displacement

Displacement of oil is controlled by the interaction between three fundamental forces, namely viscous-, capillary- and gravity forces. For higher flow rates it is assumed that the displacement process is mainly dominated by viscous forces and is proportional to permeability. For lower flow rates capillary- and gravity forces will become more important.

3.2.1. Displacement efficiency

In the fluid displacement process the overall recovery efficiency is a product of macroscopic displacement efficiency and microscopic displacement efficiency. Microscopic displacement is a measure for the reduction of oil saturation in the parts of the reservoir that have been swept. The efficiency of the microscopic displacement process is influenced by wettability and interfacial tension.

Macroscopic displacement or volumetric sweep efficiency describes how well the displacement fluid is in contact with oil bearing parts. Macroscopic displacement is affected by the density difference of the fluids, heterogeneities of rock matrix and mobility-ratio of the displacement fluid over the displaced fluid.

Macroscopic displacement itself can be divided into vertical and areal sweep efficiency. Vertical sweep is dominated by permeability heterogeneity in the vertical direction. At the locations of the wells, good data on the vertical sweep efficiency is available. Areal heterogeneity is determined by more factors than permeability such as the areal distribution of reservoir thickness, porosity, fluid saturations, fractures and faults. Variation in permeability of the layers that is continuous over a significant distance (from injector to producer) will promote early breakthrough in high permeable streaks leading to more water required to sweep the low permeability layers.

Methods for prediction of the water/oil displacement efficiencies have been developed. A good overview of methods for performance prediction of waterflooding is given by (Thomas, 2007). However, most of the fundamental methods are based on mathematical analysis of linear segments of stratified reservoirs and homogeneous distribution of the reservoir properties. In reality, geology and injector/producer patterns are more complex.

3.2.2. Mobility

Displacement efficiency is strongly affected by the viscosity and relative permeabilities of the rock to the fluids. For single phase flow permeability is a property of the rock irrespectively of the fluid. On the other hand, in two phase flow the relative permeability of the rock with respect to a particular fluid is dependent on the saturation of this fluid. Below the residual-oil saturation, S_{or} , it becomes

impossible to mobilize the remaining oil and reduce the saturation any further because the relative permeability will be zero.

Mobility (λ) is the ratio of the effective permeability of a fluid to the fluid viscosity. Hence we define:

$$\text{Oil mobility} \quad \lambda_o = \frac{k_o}{\mu_o} = \frac{kk_{ro}}{\mu_o} \quad \text{Equation 4}$$

$$\text{Water mobility} \quad \lambda_w = \frac{k_w}{\mu_w} = \frac{kk_{rw}}{\mu_w} \quad \text{Equation 5}$$

$$\text{Gas mobility} \quad \lambda_g = \frac{k_g}{\mu_g} = \frac{kk_{rg}}{\mu_g} \quad \text{Equation 6}$$

Defined as such, the both the fluid mobility and the mobility ratio is a strong function fluid saturation. The mobility ratio (M) is defined as the ratio of the mobility of the displacing fluid over that of the displaced fluid. A mobility ratio larger than one implicates that the displacement fluid is more mobile than the oil and this is unfavorable. In a heterogenous porous medium frontal instabilities may lead to viscous fingering and leaving parts of the reservoir unswept. Different definitions of mobility ratio appear in the literature, the most commonly used definition uses endpoint relative permeabilities and is defined as:

$$\text{End-point mobility ratio} \quad M = \frac{\mu_o k_{rw,or}}{\mu_o k_{ro,cw}} \quad \text{Equation 7}$$

However, Hagoort, 1974 found that this definition was over estimating the risk of unfavorable displacement and proposed the use of the shock front mobility ratio. The shock front mobility ration is always smaller than the end-point mobility ratio and is defined as:

$$\text{Shock front mobility ratio} \quad Ms = \frac{\frac{k_{rw}(Sf)}{\mu_w} + \frac{k_{ro}(Sf)}{\mu_o}}{\frac{k_{ro,cw}}{\mu_o}} \quad \text{Equation 8}$$

For a wider discussion on which of the definitions for mobility ratio should be favoured under certain conditions, readers are referred to Kumar, 2008.

3.3. EOR – mobilizing residual oil

Enhanced oil recovery (EOR) aims to improve ultimate recovery by either increasing the stability of the macroscopic displacement process or increasing the microscopic sweep efficiency by reduction of the residual oil saturation. Based on the approach to reach this goal, the different methods can be divided into three categories: thermal recovery, chemical flooding and miscible gas injection.

3.3.1. Thermal recovery

Thermal recovery factor is a large contributor to the production of heavy oil and enables to produce from fields that would otherwise be unproducible. In thermal recovery steam is injected in the reservoir to heat up the oil. The production is improved by viscosity reduction of heavy oil components and vaporizing light components and the steam also displaces the oil towards the production wells. A disadvantage is that the process of steam injection is highly water and energy intensive. Because of wellbore heat losses, steam flooding is more cost efficient in shallow reservoirs.

3.3.2. Chemical flooding

In chemical EOR, surface active agents are used to improve recovery efficiency. The following chemicals can be added to the injection water:

- Polymers that increase the injection water viscosity
- Surfactants that reduce the interfacial tension and
- Alkaline salts that alter the reservoir rock wettability, reduce adsorption of surfactants and produce soap from crude

Success or failure is strongly dependent on selection of the appropriate chemicals and the optimal concentration, which has to be determined on a case by case basis. A key challenge is to control the loss of the chemicals, as they are being adsorbed at the reservoir rock's surface. Additionally, research is directed towards finding chemicals that are both environmentally friendly and cost efficient.

Recently, the industry has also shown interest in low salinity (less than 1,000 ppm TDS) waterflooding. Core flooding experiments and early results from pilot tests are encouraging and demonstrated possibilities to recover more oil from both carbonates and clastics. The exact mechanism of this method is not yet fully understood, however it appears that water-wetness is increased which promotes higher oil recovery (Morrow, 2011; Nasralla, 2011).

3.3.3. Miscible gas injection

For light type oil reservoirs injection of miscible gas injection is the most appropriate EOR method. The three most common injection gases are CO₂, N₂ and CH₄. Although all of them have specific pressures at which miscibility is achieved, ranges overlap and at greater depths the choice of the gas usually depends on local availability. CO₂ is the most common and most effective injection gas. N₂ is most inert, but requires the highest pressures. (Taber et al., 1997)

3.3.4. Outlook

Currently the industry average (estimated) ultimate recovery of oil fields is about 34% (Schulte, 2005), meaning that two third of the oil-in-place is left behind in the reservoir after abandonment. With enhanced oil recovery methods, the recovery factor could be raised to 60% or more (Falcone et al., 2007). Raising the recovery factor of existing fields has an advantage that there are no exploration costs and existing infrastructure can be used. Meanwhile, high oil prices and the continuously growing energy demand help to make EOR economically feasible.

However, offshore EOR applications are limited as these locations remain economically and environmentally challenging. Furthermore, the large well spacing offshore introduces more uncertainty in reservoir quality and a longer time lag between injection and results. On the other hand, a number of miscible gas projects in the North Sea have proved to be successful (Lake, 2008).

4. Methods

The statistical part of this thesis involves a series of data mining techniques in order to extract meaningful information from the data. The process steps of data mining and the hereby involved methods are the subject of this chapter. Firstly, this chapter will discuss the input data followed by the adopted workflow. Additionally, a more detailed description of the used data mining techniques of the preprocessing and analyzing phase will be presented.

4.1. Input data

The data analyzed in this project is sourced from a static reservoir engineering database stored in the TQ EUR TOOL. The TOOL is designed to compare the complexity and recovery factor of reservoirs worldwide. The TQ EUR TOOL is used as a tool to search for analogues and to benchmark reservoir performance. The data set consists of 1964 data objects (oil- and gas reservoirs) and more than hundred features (geological-, reservoir engineering- and economic parameters) describing each object.

4.1.1. *Economic bias*

It is of vital importance to realize that there is a strong economic filter over the data set. Only the reservoirs that are of economic significance could be brought to production. This is reflected as a bias to reservoirs with good qualities. An indication is the amount of oil reservoirs, 54%, that were deposited in a coastal environment. This depositional setting is known for its good reservoir properties: good sorting and rounding of the grains homogeneity, and lateral extent of the sand bodies. On the other hand, fluvial reservoirs that are generally more complex only account for 16% of the reservoirs. We can also observe that the reservoir rocks of younger age are relatively more common. Assuming that the younger the reservoir rock, the better the reservoir properties this is again an example of the economic bias. See Figure 4.1.

In economically complex locations such as deepwater only the best reservoirs could be developed. Consequently, we find some positive correlations between reservoir quality and depth. This should not be mistakenly interpreted as a causal relationship between the both of them. In fact, some properties such as porosity actually deteriorate with depth. See Figure 4.2. Besides the trend we can observe an economic cut-off porosity that increases with depth. As a consequence the sample group of deep reservoirs is smaller than that of shallow reservoirs. A bias is also present when we look at the reservoirs subjected to EOR methods. Due to screening criteria, the bandwidth of property ranges is much smaller than that of the primary recovery data set.

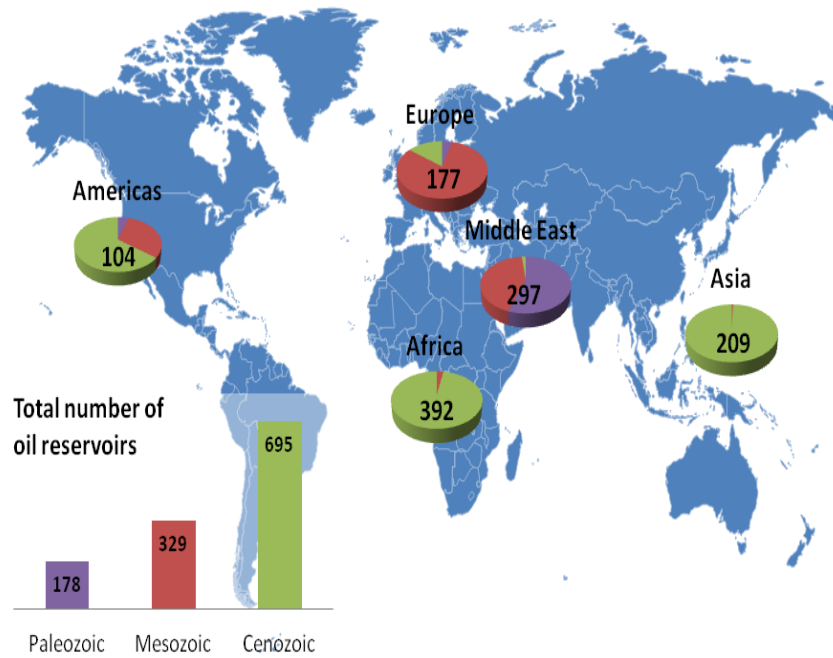


Figure 4.1: Number of oil reservoirs per era. Note that the younger reservoirs are relatively more common. However, the distribution of oil reservoirs over geologic time is not the same for all the regions.

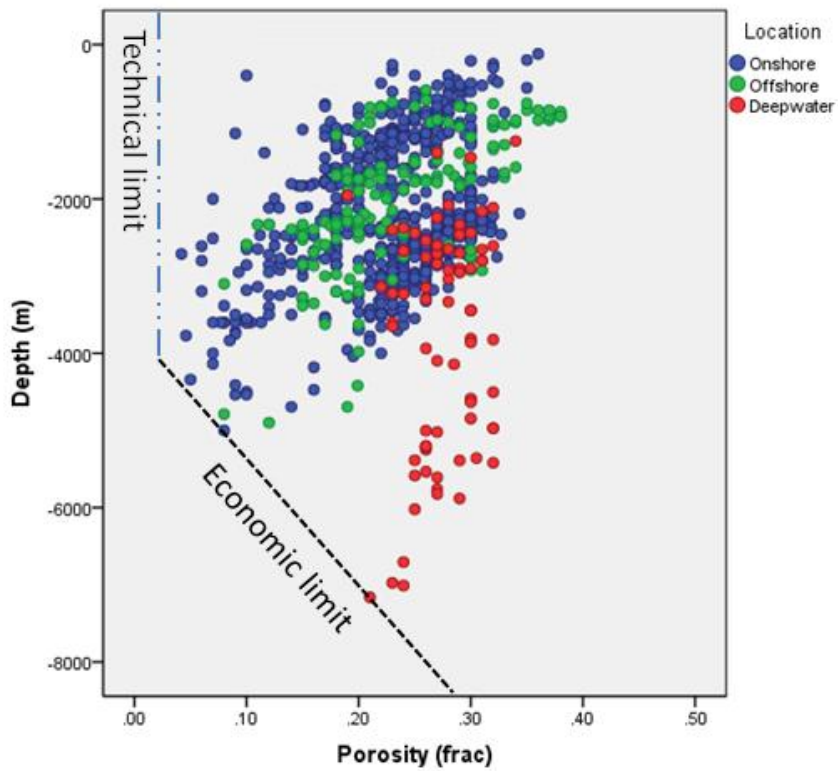


Figure 4.2: Porosity versus depth trend per location.

4.1.2. Data uncertainty

When considering the input data, the uncertainty should be acknowledged. There are a variety of sources for uncertainty in reservoir engineering data sets. According to Rijnders, 1974 these can be classified as: random error, systematic error, interpretational uncertainties and limited areal influence of data points. Rijnders, 1974 stretches that limited areal influence of the data points is possibly the greatest source of uncertainty in reservoir engineering. This might be even more so in our case as there is only one dimension for each parameter. For example, the permeability of the entire reservoir is represented by only one value. In most cases this is the average matrix permeability measured through core analysis, which in heterogeneous reservoirs may strongly differ from the field scale permeability.

It is also important to realize that the data that is stored in the TQ EUR Tool results from a combination of many different measuring methods. See Table 4.1. All of these methods introduce different boundary conditions, flow geometries, scales of accuracy and potential biases. The data sources are not always clear and can differ for a certain variable from one reservoir to another. Hence, the introduction of a parameter that represents the uncertainties for a reservoir is suggested. See also chapter 6 paragraph 6.3 ‘Statistical Analysis’.

Table 4.1: Different measurement methods encountered in the tool.

Method	Parameter	Scale
Seismic	Reservoir area Gross thickness	Reservoir scale
Well logs	Porosity Initial water saturation Net to gross ratio	Vertically: >0.5 m. Limited areal influence
Fluid samples	API gravity Oil viscosity Initial gas oil ratio Bubble point pressure Oil formation volume factor	Matrix scale: core diameter between 6 cm-10cm, Direct measurement
Core and rock samples	Porosity Permeability Irreducible oil saturation	Matrix scale: core diameter between 6 cm-10cm, Direct measurement
Well tests	Permeability	Reservoir scale

4.1.3. Calculated variables

In addition to the recorded data, there are a number of variables that were calculated within the dataset. The data points for the calculated are internally consistent with the contributing components. As a consequence, one wrong entry in any of the components will lead to an erroneous calculated variable.

- *Ultimate Recovery Factor*

In the TQ EUR TOOL the ultimate recovery factor is calculated automatically using Equation 9.

$$\text{Ultimate Recovery Factor (\%)} = \frac{PROD+EXP+SFR}{EXP.STOIIP} * 100\% \quad \text{Equation 9}$$

In which: *PROD* is cumulative production in mln bbl, *EXP* is expected oil in mln bbl (proven and probable), *SFR* is scope for recovery resources in mln bbl (possible) and *EXP.STOIIP* is expected stock tank oil initially in place in mln bbl (P50). It should be noted that the expected oil and scope for recovery can include recovery from secondary and enhanced methods. This is the case if such methods are part of the field development plan even for reservoirs that are currently in the primary recovery phase. Unfortunately, there is no variable that measures the expected oil recovery from primary recovery only.

- *Current Recovery Factor*

In addition to ultimate recovery factor, we calculate current recovery.

$$\text{Current Recovery Factor (\%)} = \frac{PROD}{EXP.STOIIP} * 100\% \quad \text{Equation 10}$$

In which: *PROD* is cumulative production in mln bbl and *EXP.STOIIP* is expected stock tank oil initially in place in mln bbl (P50). Since the current recovery factor does not include planned improved recovery methods, this is a better measure to compare the effectiveness of the natural drive mechanisms. On the other hand, this variable is strongly affected by the maturity of the reservoir.

- *Sweep efficiency*

Recovery efficiency can be split into microscopic- (equation 11) and macroscopic sweep efficiency (equation 12).

$$\textit{Microscopic Sweep Efficiency} = \frac{S_{oi} - S_{or}}{S_{oi}} \quad \text{Equation 11}$$

In which: S_{oi} is initial oil saturation and S_{or} is irreducible oil saturation. Initial oil saturation is mostly measured through well logging and represents the in situ conditions in the vicinity of the wellbore.

$$\textit{Macroscopic Sweep Efficiency} = \frac{\textit{Ultimate Recovery Factor}}{\textit{Microscopic Sweep Efficiency}} \quad \text{Equation 12}$$

In the paper of Tyler and Finley, 1991 unrecovered mobile oil was introduced to express the opposite of sweep efficiency on the reservoir scale.

$$\textit{Unrecovered Mobile Oil} = \textit{Microscopic Sweep Efficiency} - \textit{Recovery Factor} \quad \text{Equation 13}$$

4.2. Workflow

The first step is inspection and understanding of the data through analysis of the statistical properties of the raw data set. As part of the raw data exploration phase, basic statistics were calculated using the ‘descriptive statistics’ tool in EXCEL. Secondly, preprocessing of the raw data is necessary to ensure the data quality which has a large impact on the quality of the statistical analysis. After, the data is ready to be analyzed. The outputs of the data mining process are the patterns that after interpretation reveal information present in the data. See Figure 4.3.

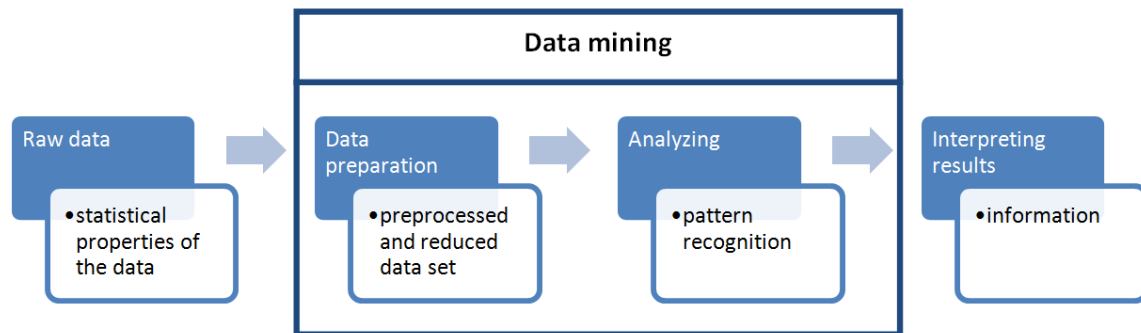


Figure 4.3: Workflow showing the sequential steps of the data mining process and the resulting output for each step. Modified after (Sharma, 2010).

The workflow presented in Figure 4.3 should be considered as an iterative process in which each data mining cycle provides new information and improves quality of the data set. The output of the first cycle is the preliminary result of visual inspection of univariate and bivariate plots. Based on which the features are selected that are subjected to pattern recognition techniques. Observed patterns can be either interpreted directly if meaningful information is obtained. Alternatively, when further refinement is needed they are again fed into the data preparation.

Reservoir Engineering principles are applied during the entire process. More specifically in the following steps:

- Data preparation: Parameter selection
- Data preparation: Identification of extreme values
- Analyzing: Forming of rational subgroups
- Interpretation of results: Conclusions and recommendations

4.3. Preprocessing

The preprocessing phase of the data mining process has three main objectives: reduction of dimensionality of the data set, improving of data quality through removal of erroneous data points and identifying and correcting missing features.

4.3.1. Feature selection

The TQ EUR TOOL is a data set which contains a large amount of features and is therefore high dimensional. In most of the data mining methods, especially those involving matrix operations, dealing with high dimensional data can become extremely difficult. Additionally, if a high number of dimensions is involved the amount of data objects required to guarantee accuracy will increase significantly. In order to overcome the issues arising from a large number of features the dimensionality can be reduced. The goal of reducing the data dimensions is to end up with about ten to fifteen features or less without losing significant information. A traditional technique to reduce data dimensionality is through principal component analysis (PCA). However, analysis of principal components is also a powerful tool in analyzing correlation and variation patterns and was used for more purposes than just reducing the number of features. Therefore, the first selection of suitable parameters relevant to the problem is based on theoretical background and literature study. Further quantitative selection of the relevant features through PCA will be done in the analysis phase.

4.3.2. Data quality

The data set contains such a large amount of information that it is almost inevitable that the data also contains errors. Data errors can have known or unknown causes and they can be detected or remain undetected. Examples of data errors are typing errors (e.g. Jurrassic instead of Jurassic) or use of a different scale (e.g. porosity in % instead of fractions). Additionally data errors can produce outliers, which are extreme values compared to the entire variable distribution. Especially in regression analysis, outliers can have a large influence on the parameters and hence removal of these extreme values is required. Fortunately, outliers are relatively easy to identify using visual aids such as a frequency distributions, box and whiskers-plots and cross plots. Model-based methods to remove outliers such as Grubbs' test assume normal distribution. Not all variables are normally distributed and even the variables that are approximately normally distributed can deviate significantly around the tails. Therefore, it was opted not to use an outliers removing routine but instead to remove outliers based on visual inspection only.

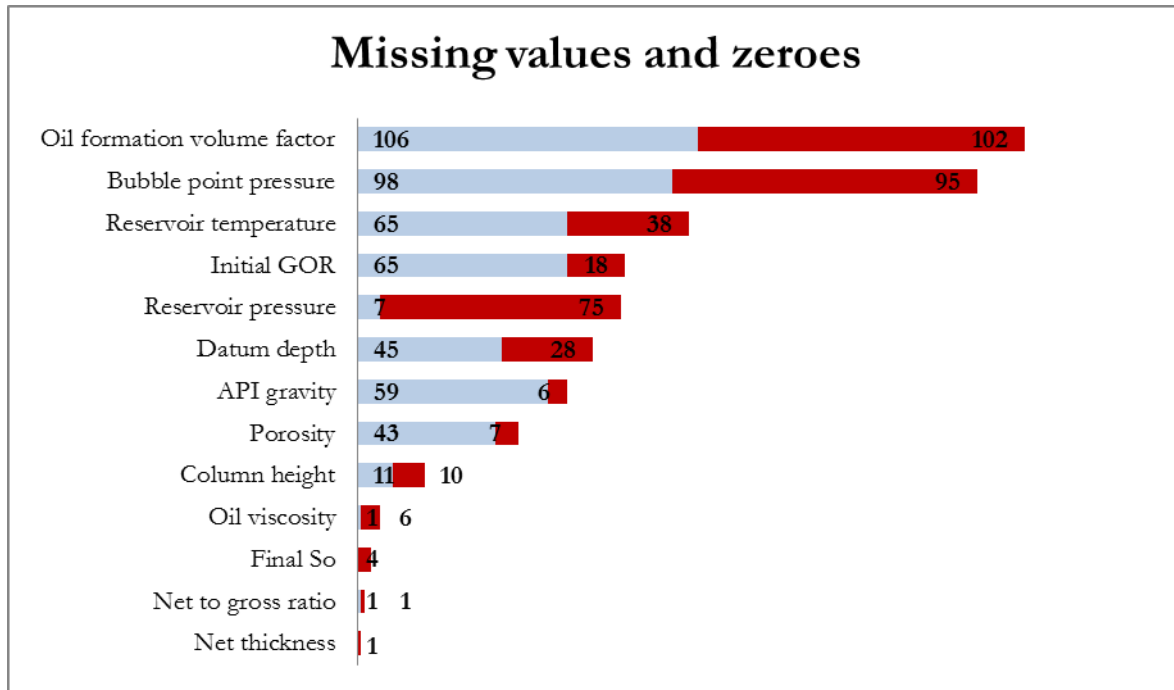


Figure 4.4: Number of missing values (blue) and zeroes (red) per variable. Exp. STOIP, Ultimate Recovery Factor, Permeability and Initial oil saturation have values for all 1203 reservoirs.

Figure 4.4 shows that there are a considerable number of blank entries. Like data errors, blank data entries can cause problems in the analysis of the data. A potential solution is to assign a value, for example the mean value, to the blank entries. The disadvantage of this solution is that it will reduce the variability and therefore gives a false impression of accuracy. Therefore it was opted to remove the blank data entries rather than assign a value to them. Removing all reservoirs that contain one or more blank entries would drastically reduce the sample number and is in most cases unnecessary as not all features are required for analysis. For example, the low and high estimates for STOIP and GIIP are poorly reported but they are not critical as they exert no physical influence on other parameters. For each individual test and analysis it was ascertained that no blank values were taken into account.

In addition to missing values it sometimes happens that the desired data features itself are missing. In those cases it was tried to capture the missing dimensions using a combination of features. An example of a data feature that is not recorded in the TQ EUR TOOL is the year of production or production time, which was why the time could only be expressed in dimensionless production time (cumulative production over expected hydrocarbons in place). More missing and desired data features will be dealt with in chapter 6 'Additional data'.

The analysis of the data is directed towards an evaluation of oil reservoirs. Including gas reservoirs would have required a similar but completely separate analysis. The main reason to focus on oil was that gas recovery factors are usually higher and less variable. Additionally, the success of gas assets is assumed to depend more on economic factors rather than physical parameters.

The filter criteria that were applied to the TQ EUR database are:

- All oil reservoirs with zero or no value for Expected STOIP
- All oil reservoirs with zero or no value for Oil Recovery Factor

The resulting data set consists of 1203 oil reservoirs. The univariate calculations are based on this screened data set unless noted otherwise. Principal component analysis and multivariate linear regression were applied to a subset of 916 reservoirs without any zeroes for all relevant variables. It should be noted that the reservoirs with low current recoveries (<5%) have a larger uncertainty due to scarcity of production data.

4.4. Analyzing

In the analyzing phase the pre processed data set is subjected to data mining techniques in order to extract patterns and obtain potentially useful information from the database. Pattern recognition is followed by identifying the nature of variation patterns in order to gain insight in the phenomena causing the variation.

4.4.1. Univariate analysis

Descriptive statistics are calculated values that describe the population such as mean, median, standard deviation, variance, skewness, and kurtosis. Most of the statistical tests used in the analysis are based on the concept of hypothesis testing. In hypothesis testing there are always two hypotheses, the null hypothesis H_0 and an alternative hypothesis H_a . The null hypothesis is a contradiction to what you want to prove. The aim of the test is to accept one of the hypotheses and reject the other one.

Since most of the statistical methods require that the variables are normally distributed it was verified whether and to what extent the distribution of selected features could be approximated by a normal distribution. The SPSS software package offers two of the most common tests for normality: the Shapiro-Wilk test and the Kolmogorov-Smirnov test.

4.4.2. Subgroup analysis

The data set is characterized by a large variance in recovery factors. This makes it difficult to identify trends that are statistically significant. Especially in the univariate analysis there is a risk that fundamental trends are obscured by other effects. Causes for the variance in the data set are:

- Large amount of independent variables affecting recovery factor
- Wide ranges for independent variables
- Different data sources for same variable
(See chapter 4 paragraph 4.1.2 'Data uncertainty')
- Uncertainties in the measurements of independent variables and recovery factor

In order to reduce the impact of variance on the analysis reasonable subgroups were formed. The aim was to create groups with lower variance within the groups and maximum differences between the groups. The lower variance within the groups makes it possible to obtain better results. Such groups could also be found through cluster analysis, but with reasonable subgroups we have a better understanding of the differences between the groups that can explain differences in performance.

A statistical method to compare groups based on the mean value is one-way ANOVA. In one-way ANOVA similarity of variance between the groups that are compared is assumed. To test this

assumption Levene's test of Homogeneity of Variances is used. When the significance value for this test is higher than 0.05 the assumption is met. Even when the assumption of homogeneity of variances is violated, we can still analyse the difference between groups based on their mean values. In this case robust tests of equality of means are applied like Welch and Brown-Forsythe. See Figure 4.5.

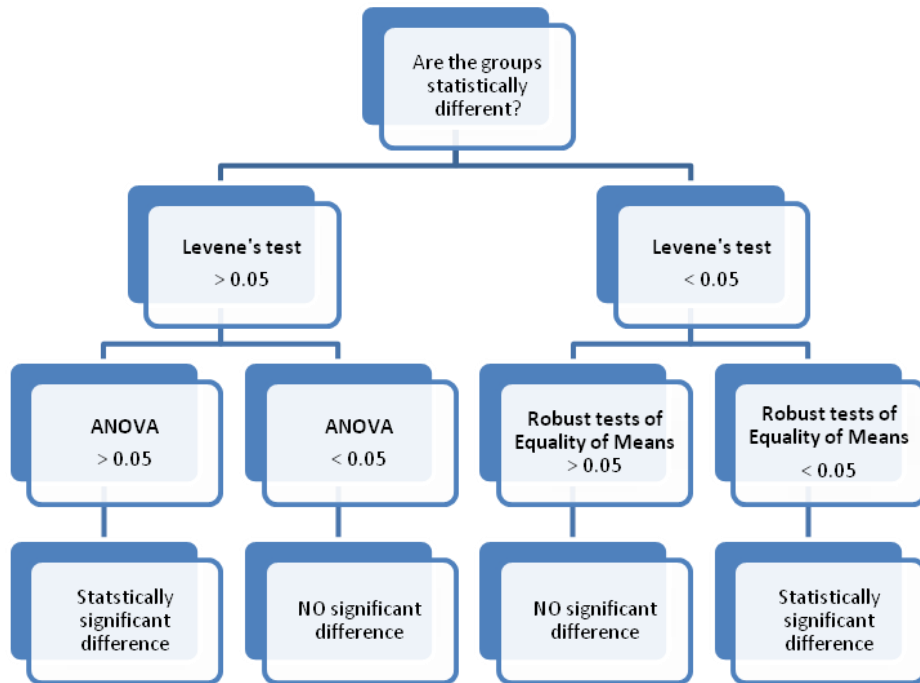


Figure 4.5: Analysis of Variance (ANOVA) flowchart.

When more than two groups are compared, the above discussed tests only test whether there is a significant difference between the groups but does not give information on which groups deviate. Therefore post-hoc tests have to be done.

4.4.3. Regression

The purpose of regression techniques is to extract a correlation that can be used for prediction. Regression techniques used in this research are linear regression and multi linear regression.

Linear regression aims to find linear equations of one or more dependent variable(s) to predict or estimate the dependent interest response. Linear regression is therefore different from correlation which investigates the association between variables but does not make an a priori assumption of the causation between predictor and response. In simple linear regression only one predictor variable is used and the resulting regression line is mathematically described by Equation 14.

$$Y = c + a x \tag{Equation 14}$$

In which, y is the dependent variable or response, c is a constant, a describes the slope of the line and x is the independent variable or predictor. Constant c is also called the intercept as this is the point where the regression line intercepts with the y-axis.

Aiming to find an equation to predict recovery efficiency and learn relationships between features in order to formulate rules, simple linear regression was applied to all independent variables that potentially exert influence on the recovery factor. However, previous studies already showed that recovery factor does not show a strong correlation with any of the single predictors but rather that the response is a result of the combined influence of various variables. To allow for a regression based on the combined effect of independent variables on the response, Multiple Linear Regression (MLR) analysis was used. Equation 14 is expanded to Equation 15 that describes the result of MLR.

$$Y = c + a_1 x_1 + a_2 x_2 + \dots \tag{Equation 15}$$

The constants a_i are called regression parameters and their absolute value shows the degree of dependence. Estimation of regression parameters is based on mean square error minimization.

4.4.4. Principal component analysis

A prediction model resulting from regression analysis neglects the interaction between parameters. Since we know that some of the parameters influencing oil production, for example porosity and permeability, are interrelated this may lead to erroneous results. For this reason we will now consider a technique called principal component analysis, which does take the interrelationships into account. The first step is to subtract the mean from the data points in each dimension. The result of this operation is a dataset with an average of zero. This adjusted dataset can be used to construct a covariance matrix. Formula to calculate covariance:

$$cov(X, Y) = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)} \tag{Equation 16}$$

Covariance matrix of 2 dimensional data:

$$C = \begin{pmatrix} cov(X, X) & cov(Y, X) \\ cov(X, Y) & cov(Y, Y) \end{pmatrix} \tag{Equation 17}$$

We can calculate the eigenvectors and eigenvalues of the resulting symmetric square matrix. The eigenvalues is the amount of variance captured by the eigenvectors. The eigenvectors with the

highest eigenvalue are the principal components of the dataset. Now the data can be reduced, by removing the eigenvectors with lower eigenvalues.

Principal components analysis is a technique used for pattern recognition in high dimensional datasets. It also reduces the number of features and is therefore also considered a data reduction technique. Firstly, it expresses the in components. Secondly, the process gives information on the relative importance of the input variables. To identify parameters that only share variance with another variable and not with the rests of the variables Kaiser's Measure of sampling adequacy (MSA) can be used. Those parameters should be removed or new parameters that are believed to correlate with them should be included. According to Kaiser (Kaiser, 1970) MSA values below .5 are unacceptable.

The number of components generated is equal to the amount of variables, but after inspection only components that account for a significant amount of the variance are kept. There are several ways to select the number of components. One can opt to keep the components with eigenvalues higher than one, which means the variance they describe is greater than that of a single variable. A visual method to extract the amount of components by inspecting is the so called 'screeplot'. The screeplot shows how much of the variance in the data is captured by the principal component. The screeplot is named after scree, which is the geologic term for rubble at the bottom of a sloping cliff.



Figure 4.6: 'Scree', derived from the Old Norwegian word: 'skriða', is the term for the rock debris covering a sloping cliff. Source: http://en.wikipedia.org/wiki/File:Yamnuska_bottom_cliff.jpg

5. Results and Discussion

The previous chapters provided an introduction to the dataset and described the data mining approach. This chapter will now presents and discusses the results of the analysis. Firstly, we will shortly discuss the statistics of the dataset that were obtained in the exploratory phase of the data mining. Secondly, we will deal with univariate trends that were revealed mainly through subgroup analysis. The final part will explain the structure of the dataset that was revealed through multivariate linear regression and principal component analysis.

5.1. Statistical overview of the data

The statistics of the selected reservoir engineering parameters are listed in Table 5.1. Frequency distribution diagrams are included in Appendix I. The Shapiro-Wilk and Kolmogorov-Smirnov numerical tests showed that none of the independent variables are normally distributed. Average well density, initial oil saturation, bubble point pressure, API gravity, reservoir- temperature and pressure are positively skewed. Initial oil saturation, porosity and net to gross ratio exhibit a negative skew. This can be explained by the economic bias, which introduces a cut-off for unfavorable values for most of the reservoir properties. Through visual inspection of the frequency distributions and Q-Q plots we could confirm that porosity and net to gross ratio are approximately normally distributed. Permeability, net thickness and expected STOIP are approximately log normally distributed.

5.1.1. Recovery efficiency

Both visual inspection and numerical tests show that estimated ultimate recovery (Oil RF%) is non-normally distributed. See Figure 5.1. As many statistical tests assume normally distributed variables, the non normality can become problematic later on. Fortunately, from the visual inspection we can conclude that although the oil recovery is not normally distributed deviation from normal behavior is not large. We observe that the distribution of oil recovery has too many low values and is positively skewed. Compared to the TORIS data (Sharma, 2010) with an median Oil RF% of 38%, our Oil RF% of 34.4% is significantly lower. Differences in maturity of the reservoirs in the dataset was identified as the main cause for this observation. The reservoirs in early stages of development have relatively lower estimated recoveries. We find that for our data the median Oil RF% of the producing fields (cum. Prod > 0) is significantly higher than that of non producing fields (cum. Prod = 0) with respectively 37% versus 28%. This phenomena is referred to as 'reserve appreciation' in literature. Some papers (Laherrère, 2007; Watkins, 2002) claim that this is the result of pessimistic in-place estimates. Unfortunately, there was not enough data (in-place versus time or maturity) available to test this hypothesis on our dataset.

Table 5.1: Ranges of independent variables. The first three columns give the minimum, median and maximum for the all reservoirs. The ranges are also presented for clastic and carbonate reservoirs separately. Please note that there are (current and ultimate) recovery factors higher than 1. This is the effect of pessimistic estimates of in-place volumes.

Variables	Unit	TOTAL			Clastics			Carbonates		
		minimum	median	maximum	minimum	median	maximum	minimum	median	maximum
<i>Rock properties</i>										
Permeability	mD	0.02	502	10000	0.02	643	10000	0.1	10	535
Porosity	-	0.042	0.24	0.38	0.07	0.24	0.38	0.042	0.18	0.35
Initial oil saturation	-	0.25	0.72	0.95	0.25	0.72	0.94	0.26	0.73	0.95
Net to gross ratio	-	0.025	0.77	1	0.03	0.76	1	0.025	0.945	1
Net thickness	m	1.2	29.7	560	1.2	26.8	475	1.3	50.0	560
Column height	m	2.1	52.7	3000	2.1	52.0	3000	5	59	800
EXP. STOIIP	mln. bbl	0.67	53.8	25000	0.67	48	25000	0.69	222.26	21013
<i>Fluid properties</i>										
API gravity	Deg.	8.0	33.2	115.4	8.0	33.0	115.4	13.0	36.9	51.0
Oil viscosity	cP	0.11	0.96	150000	0.11	0.96	150000	0.17	0.875	3000
Oil formation volume factor	rb/stb	1	1.253	131	1	1.246	131	1	1.3285	2.9
Bubble point pressure	psi	10	2760	10400	10	2900	10400	72	2001	8172
Initial GOR	scf/b	0.5	477	5500	0.5	477	5500	8.0	466.5	4300
<i>Reservoir environment</i>										
Reservoir temperature	degC	10	78	305	10	77	305	22	80	236
Reservoir pressure	psi	145	3578	16910	145	3578	16883	377	3376	16910
Datum depth	m	120	2380	7163	120	2387	7163	200	1999	5000
Water depth	m	0	0	2134	0	0	2134	0	0	83.8
Average well density	w/km	0	1	1500	0	1	1500	0	0.55	500
<i>Recovery</i>										
Current Recovery Factor	-	0	0.147	1.02	0	.1676	1.02	0	0.0228	0.57
Ultimate Recovery Factor	-	0.006	0.344	1.297	0.006	0.351	1.297	0.006	0.2865	0.68
Final oil saturation	-	0.01	0.20	0.59	0.05	0.20	0.59	0.01	0.19	0.40

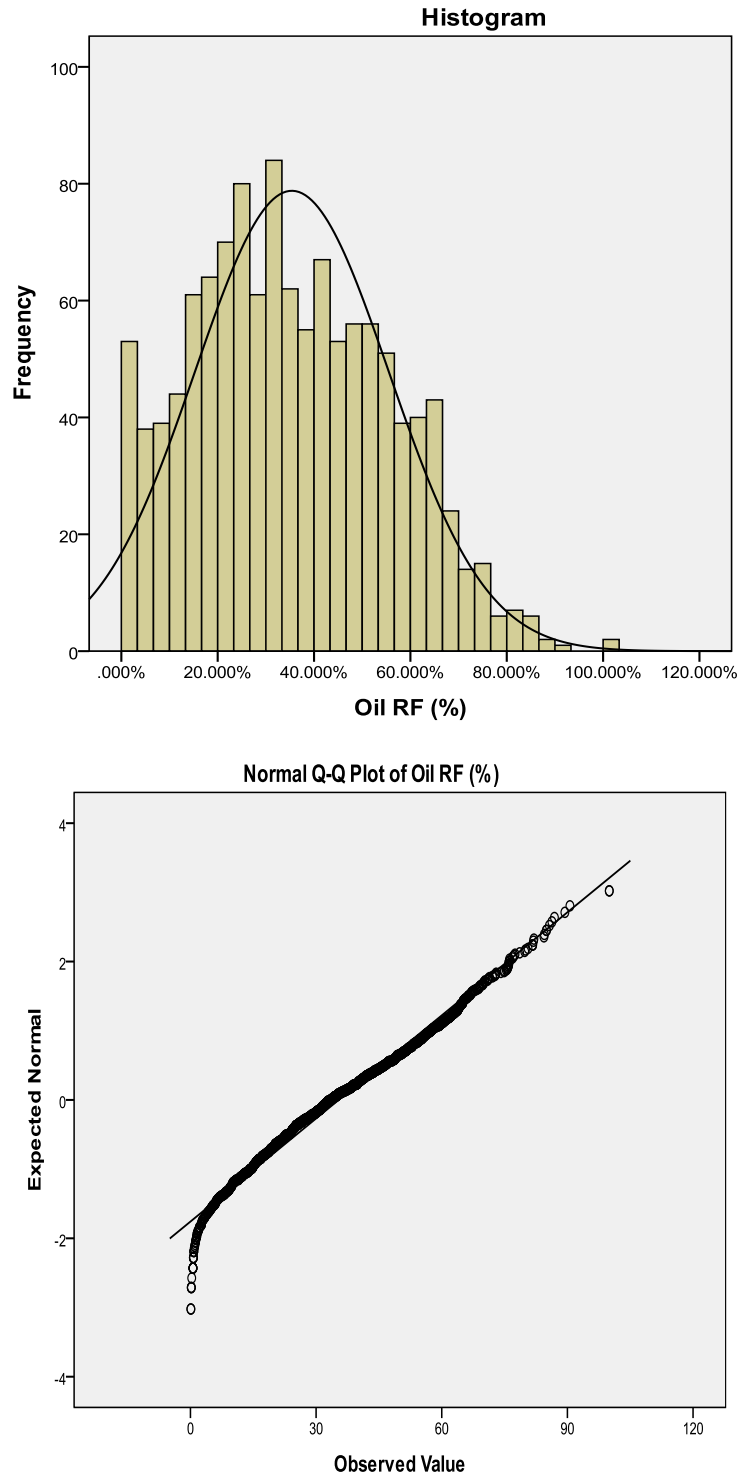


Figure 5.1: Top: Frequency distribution of oil recovery factor. Normal distribution is fitted over the distribution. Bottom: Normal Q-Q plot of the oil recovery factor. The straight line indicates a normal distribution and the dots represent the data. Both plots indicate that there are more reservoirs with low recovery factors than if it would be a normal distribution.

5.2. Subgroup analysis

This chapter presents and discusses the results of the subgroup analysis. Subgroups were chosen based on reservoir engineering principles and geologic concepts. The statistical significance of the subgroups was tested using ANOVA.

5.2.1. Drive mechanisms

Figure 5.2 shows the current primary recovery factor for different drive mechanisms. All distributions are skewed to the right. Again, this is effect of the reservoirs that are in early stages of the development. The spread of the gas cap reservoirs is somewhat smaller than that of the others because the sample size of this group is smaller. The average recovery factor ranges from 40% for edge water drive reservoirs to below 25% for gas cap reservoirs. Generally, the water drive reservoirs are performing better than the gas drive reservoirs. However, if we look at the ultimate recovery factor which includes expected oil production and scope for recovery we find that only gas cap reservoirs show significantly lower recovery whereas the solution gas reservoirs score as well as the water drives. See Figure 5.3.

Recall from chapter 4 'Methods' paragraph 4.1.3 'Calculated variables' that ultimate recovery factor can include secondary and enhanced recovery. Most likely, a relatively large percentage of the solution gas reservoirs will be subjected to secondary recovery applications. This is reflected in a distinct difference between ultimate recovery factor and current recovery factor for solution gas drive reservoirs. In support of this claim, Table 5.2 shows that most reservoirs that are currently on water flooding have solution gas drive as primary recovery mechanism. This can explain why gas reservoirs seem to perform as good as the other natural drive mechanisms. However, it does not explain why the gas cap reservoirs show such low recoveries. A possible explanation was found through multivariate analysis. See chapter 5, paragraph 5.4.5 'Performance of gas cap reservoirs'

There is no significant difference in lithology distribution over the different drive mechanisms. Although the solution gas drive has a slightly larger percentage of carbonate reservoirs than water drive reservoirs the proportion is similar to that of gas cap reservoirs.

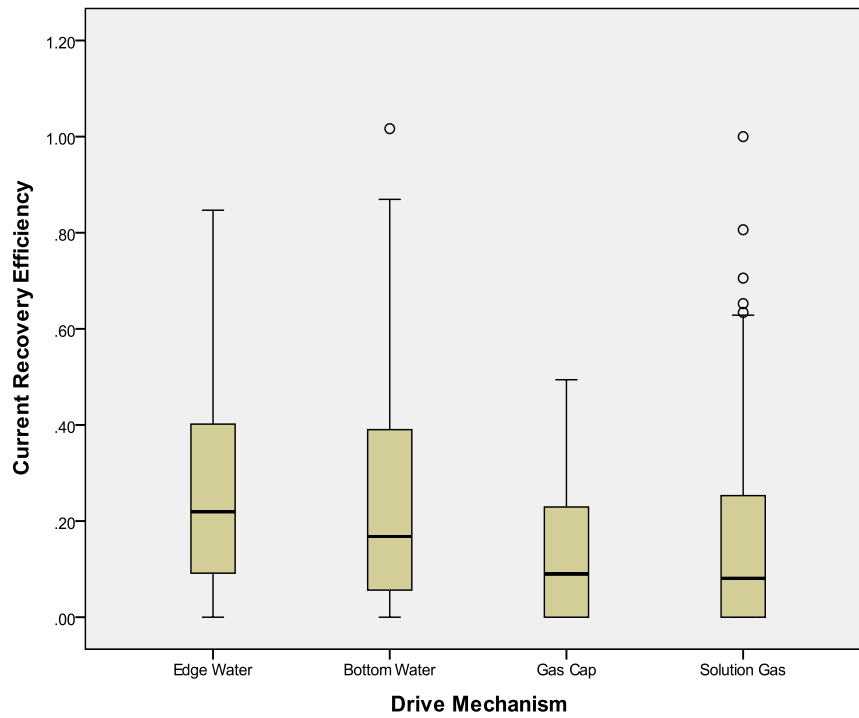


Figure 5.2: Box plot of Current Recovery Efficiency versus drive mechanisms. Five basic statistics can be obtained from the box plot: the median, represented by the bar inside the box, the upper quartile and the lower quartile, represented by respectively the upper and lower outline of the box, the minimum and the maximum value. Between 1.5 and 3 times the box length we find the outliers that are represented by the circles. Not present in this plot are values further than three box lengths away which would be flagged as extreme values and represented by asterisks.

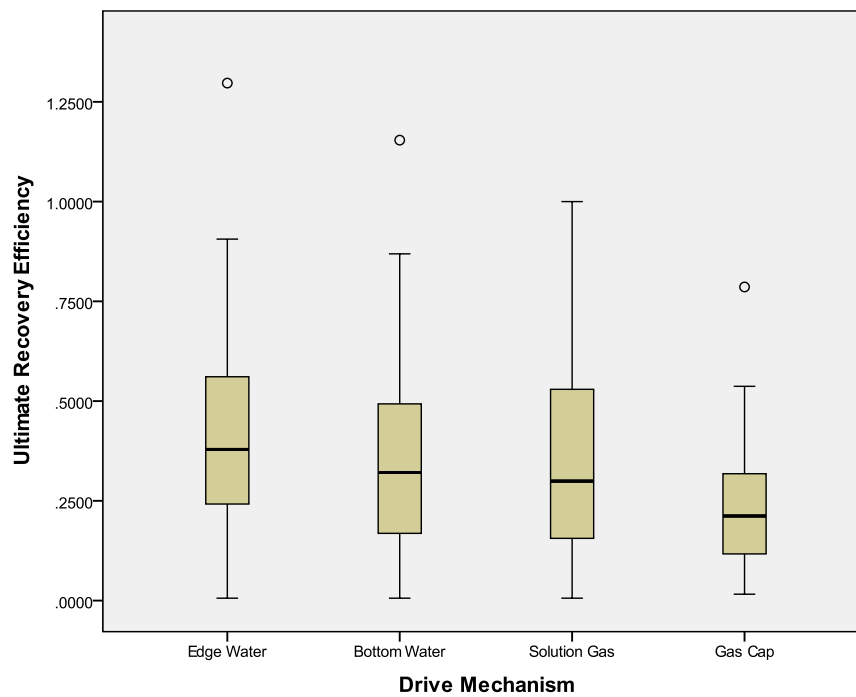


Figure 5.3: Ultimate Recovery Efficiency versus drive mechanisms.

Table 5.2: Natural drive mechanism versus secondary recovery method.

	None	Water Injection	Gas Injection	TOTAL
Bottom Water	204	22	1	227
Edge Water	284	91	1	376
Gas Cap	41	15	7	63
Solution Gas	255	124	12	391
Compaction	3	1	0	4
Gravity Drainage	2	4	1	7
Combination	3	5	4	12
Unknown	77	45	1	123
TOTAL	896	307	27	1203

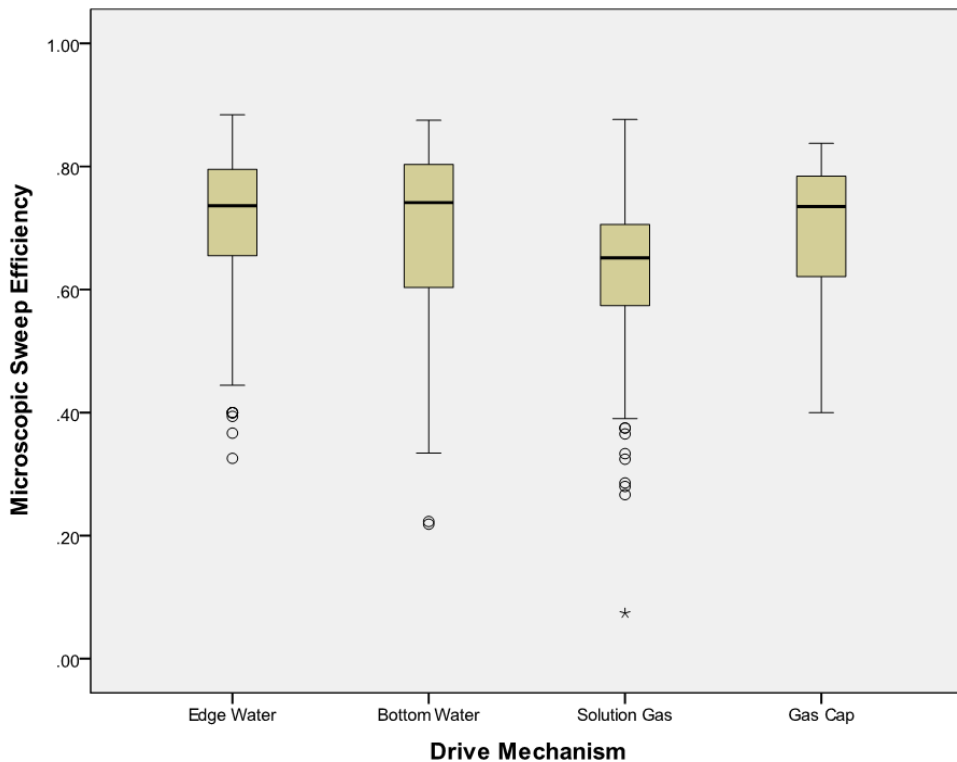


Figure 5.4: Microscopic Sweep Efficiency versus Natural Drive mechanism. The microscopic sweep efficiency is significantly lower for solution gas drive reservoirs. This is caused by the fact that this group is characterized by higher residual oil saturations.

Drive mechanisms affect the recovery efficiency both on the microscopic as well as on the macroscopic scale. See Figure 5.4. shows the average microscopic sweep efficiency versus drive mechanism under primary recovery. Edge water-, Bottom water- and Gas Cap drive show similar microscopic sweep efficiencies of 70 percent of the oil in place, however that of solution gas drive is with an average of 63 percent significantly lower due to higher residual oil saturations. Theoretically, the microscopic sweep should be better for gas drive reservoirs than for water drive reservoirs due to lower irreducible oil saturation. However, this could not be retrieved from the data. Most likely due to the fact that the recorded 'final oil' does not in all cases represent irreducible oil saturation obtained from special core analysis (SCAL). Instead final oil seems to be the remaining oil that is left after the energy of the drive mechanism is exhausted. This explains high 'final oil' saturations for solution gas drive reservoirs as those types of reservoirs often run out of energy before the irreducible saturation is reached. On the other hand, it could be that the low microscopic sweep efficiency for solution gas reservoirs would be an artifact of the SCAL measurements that is not representative for the actual field conditions. In that case that would mean that the macroscopic sweep efficiency for solution gas reservoirs will be overestimated in most cases. This is important as we stated earlier a large part of the reserves from solution gas drive reservoirs comes from the expected oil obtained through water flooding.

Figure 5.5 shows the calculated macroscopic sweep for the different drive mechanisms. Maximizing the ultimate recovery from gas cap reservoirs seems most challenging. While the edge water drives have both high microscopic- and macroscopic sweep efficiency, the performance of the gas cap reservoirs is dominated by the poor macroscopic sweep. According to (Ahmed, 2010), conservation of energy stored in the compression of the gas is a key determinant for the performance of gas cap reservoirs.

Less intuitive might be the result that bottom water drive reservoirs have an ultimate recovery efficiency that is significantly lower than that of edge water drive reservoirs. Although the displacement process in bottom water drive is assumed to be more stable because of the high gravity number (L.P. Dake, 1994. p.343), often the gravity component is small compared to the viscous component and water coning may become a problem. Additionally, in our dataset the permeability and oil column height are higher in edge water reservoirs. This has a positive influence on the reservoir performance.

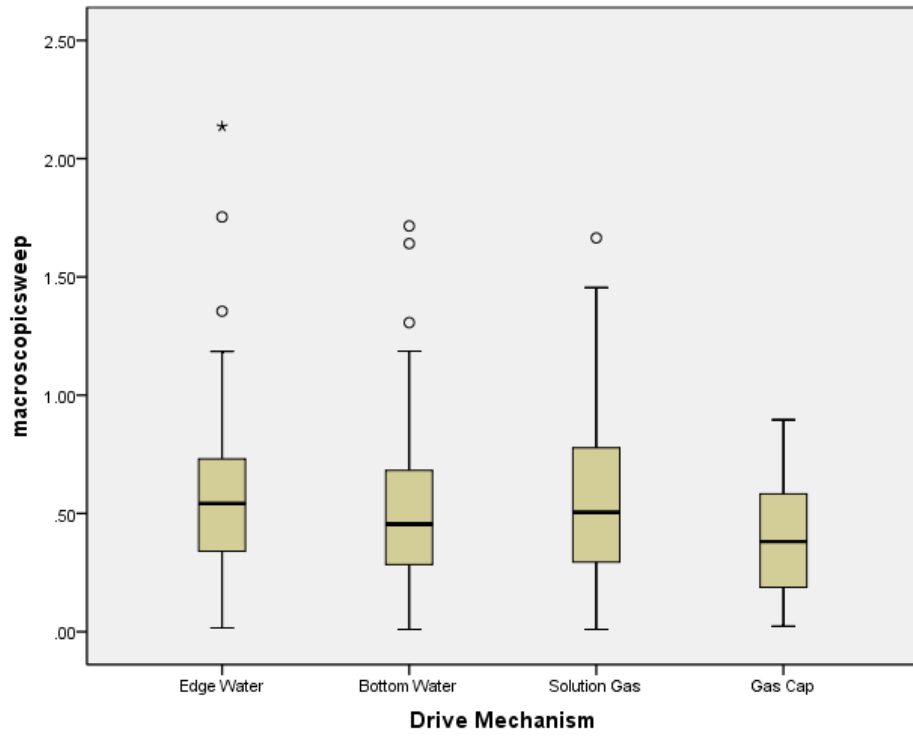


Figure 5.5: Macroscopic Sweep Efficiency versus Natural Drive Mechanism.

5.2.2. *Lithology*

Lithology is a key determinant for recovery efficiency. Carbonates are notorious for highly variable and on average lower recovery factors. Differences between carbonates and clastics in macroscopic sweep are mainly contributed to fractures and reservoir heterogeneity. On the microscopic scale carbonates are often more oil wet than sandstones. Shales differentiate from the sandstones by having low permeabilities and some clay minerals tend to be oil wet.

Figure 5.6 shows the initial oil saturation for different lithology. Initial oil saturation is not significantly different between sandstone and carbonates. However, a significant difference between carbonates and sandstones on the one hand and shaly sandstones and shales on the other hand can be observed. In the latter high connate water saturations are the result of low permeability and porosity. As pores and pore throats are smaller it is harder to push the fossil water out and replace it with oil.

Secondly, even though for carbonates matrix- permeability and porosity are low, we find high initial oil saturations. This can be explained by the fact that most carbonates are oil-wet and hence the preferential displacement of water by oil has led to higher oil saturations. However, this is in contrast with the idea that wettability does not play a considerable role in the initial water saturation. (Zinszner, 2007)

In contrast, irreducible oil saturation is believed to depend on wettability which explains the findings that carbonates have significantly higher irreducible oil saturations. No statistic difference was observed for the clastic rocks. See Figure 5.7.

Another major difference between clastics and carbonates in the dataset is the percentage of reservoirs that is subjected to secondary and enhanced recovery methods. Currently there are 46% percent of the carbonate and only 23% of the sandstone reservoirs on water flooding and respectively 10% versus 1% on gas injection. Also the application of enhanced methods is larger in carbonates than in clastics. So the effect of secondary recovery or EOR programs on the ultimate recovery factor is stronger for carbonates than sandstones. This is reflected in the ratio of current recovery factor over ultimate recovery factor, where the former includes only primary recovery whereas the latter includes secondary and enhanced recovery. The resulting ratios are $36.9/21.3=1.732$ for clastics and $30.2/8.0=3.775$ for carbonates.

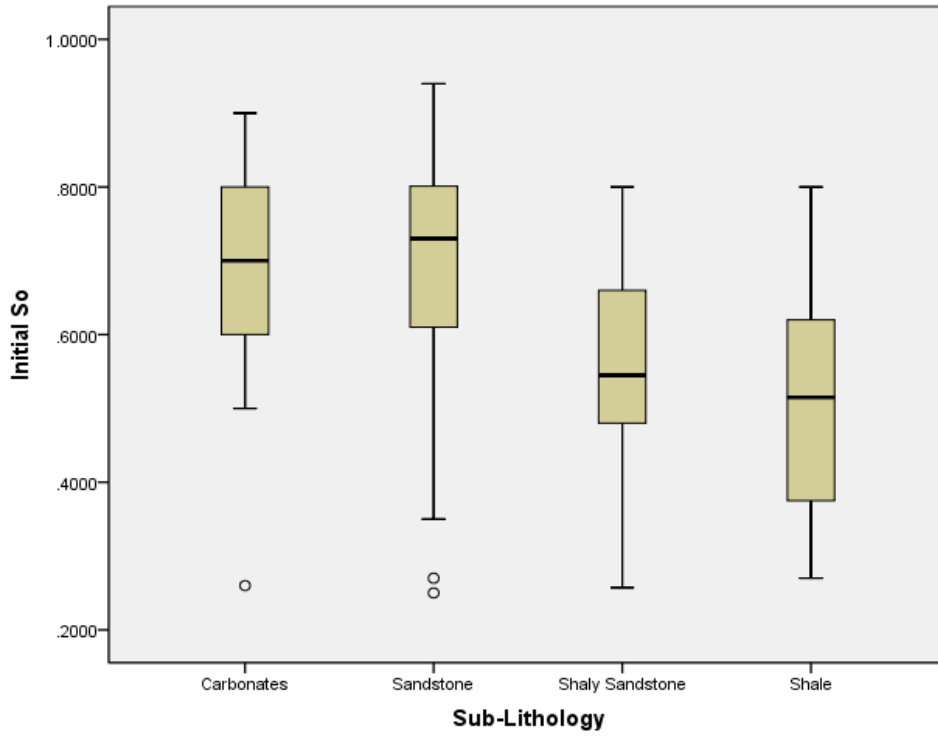


Figure 5.6: Initial oil saturation versus lithology.

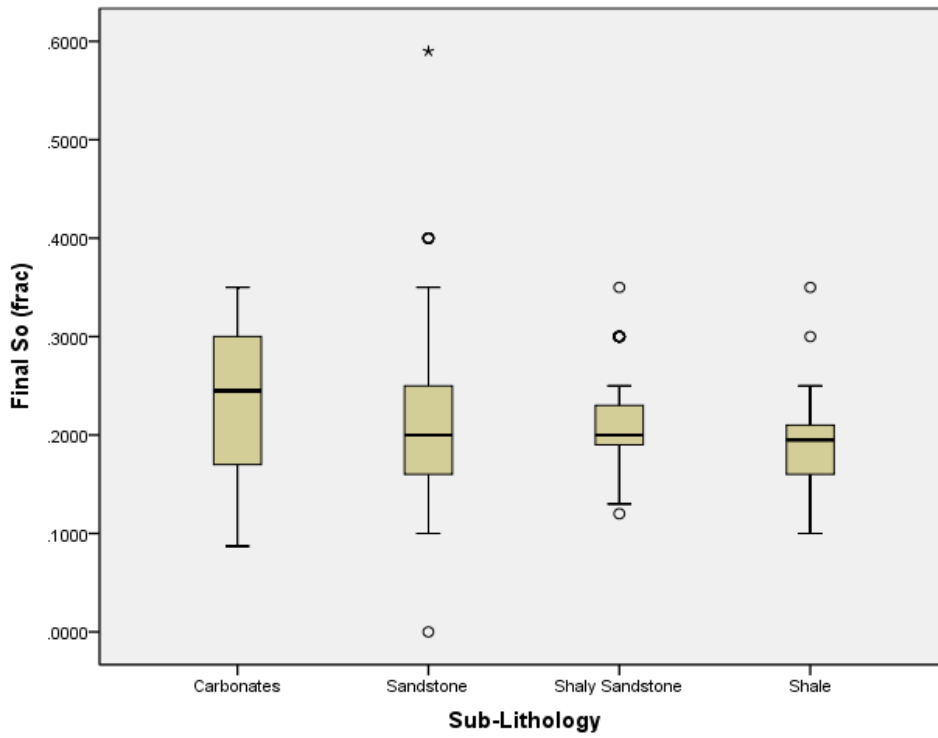


Figure 5.7: Final oil saturation versus lithology.

5.2.3. Aquifer strength, lithology and fracture intensity

Figure 5.8 and Figure 5.9 show the effect of the aquifer strength on the ultimate recovery efficiency. For sandstone reservoirs we find a positive correlation between aquifer strength and both current primary recovery and ultimate recovery efficiency. The explanation for this observation is that an active aquifer accommodates good pressure maintenance.

A reason that a strong aquifer drive can also have a negative effect on recovery efficiency is that it can cause early breakthrough. This risk becomes larger when the reservoir heterogeneity increases. Besides their heterogeneity carbonate reservoirs are generally more fractured than clastics. Qing Sun, 2003, differentiated brittle rocks in which fractures extent into the aquifer from the ductile rocks such as chalks and chalky limestones. In the cases where the fracture networks connect to the aquifer, we can expect that the fracture intensity plays a large role in the way aquifer strength affects recovery efficiency.

Figure 5.10 and Figure 5.11 show the effect of aquifer strength on average recovery efficiency for different fracture intensities. The measure for fracture density used in the TQ EUR TOOL is a nominal scale ranging from 1 to 5. See table 2. We find that only the highly fractured carbonates have a decreasing reservoir performance with increasing aquifer strength. Highly fractured carbonate reservoirs therefore require extra careful reservoir management to prevent early water breakthrough. Additional data to monitor and optimize reservoir performance would be production and injection rate and water cut.

Table 5.3: Relative measure of fracture intensity in the TQ EUR TOOL.

1	No fractures/High perm streaks or fracture conductivity and connectivity and conductivity is very favorable to reservoir recovery
2	Low fracture density or degree of high perm streaks and/or degree of connectivity and conductivity favorable to reservoir recovery
3	Moderate fracture density or degree of high perm streaks and/or degree of connectivity and conductivity has little effect on reservoir recovery
4	High fracture density or degree of high perm streaks and/or degree of connectivity and conductivity is unfavorable to reservoir recovery
5	Intensely fractured or degree of high perm streaks and/or degree of connectivity and conductivity is very unfavorable to reservoir recovery

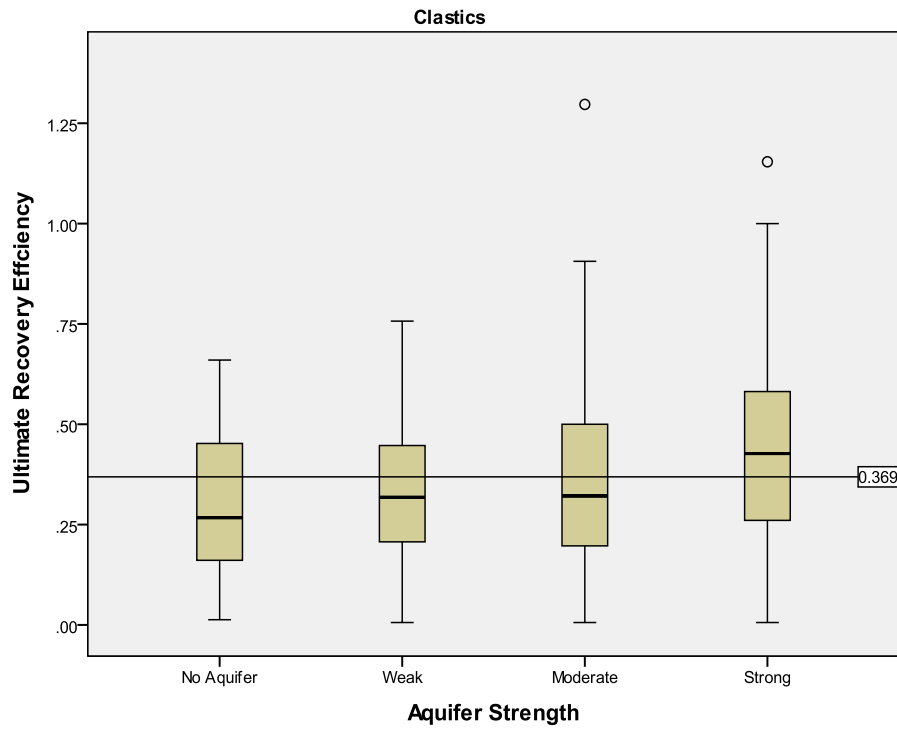


Figure 5.8: Ultimate recovery factor for clastic reservoirs as a function of aquifer strength.

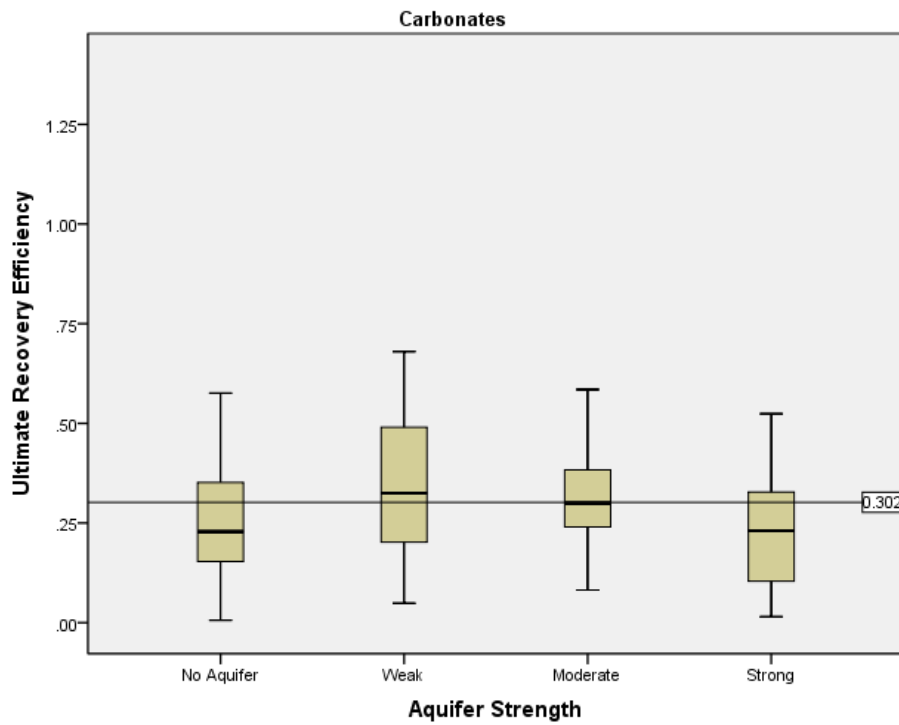


Figure 5.9: Ultimate recovery factor for carbonate reservoirs as a function of aquifer strength.

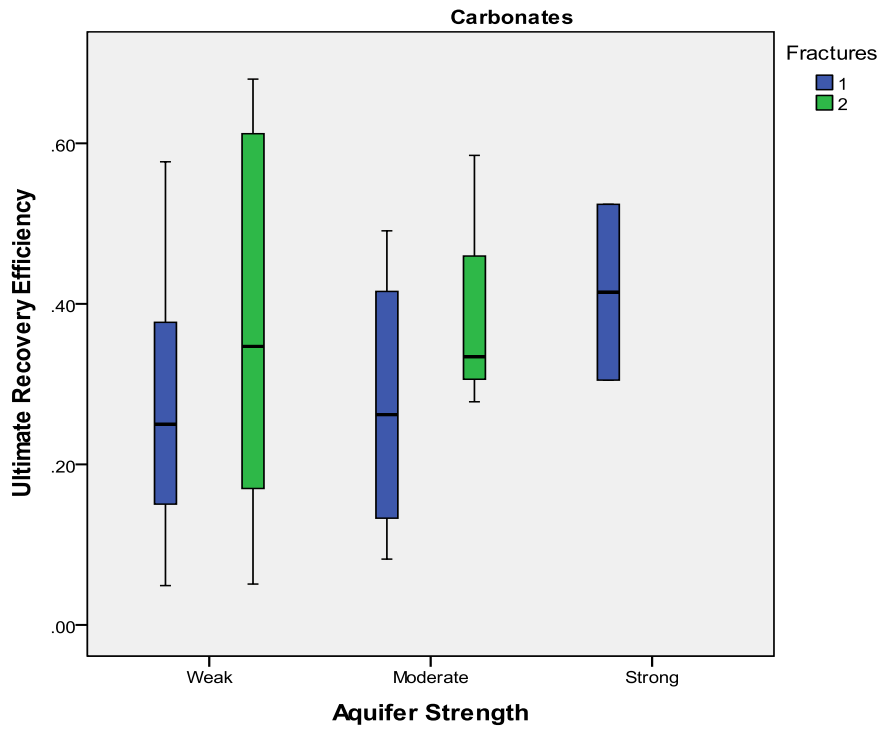


Figure 5.10: Ultimate recovery efficiency for carbonate reservoirs with no fractures and low fracture intensity. The plot shows that for these reservoirs the aquifer strength has a positive influence on recovery efficiency.

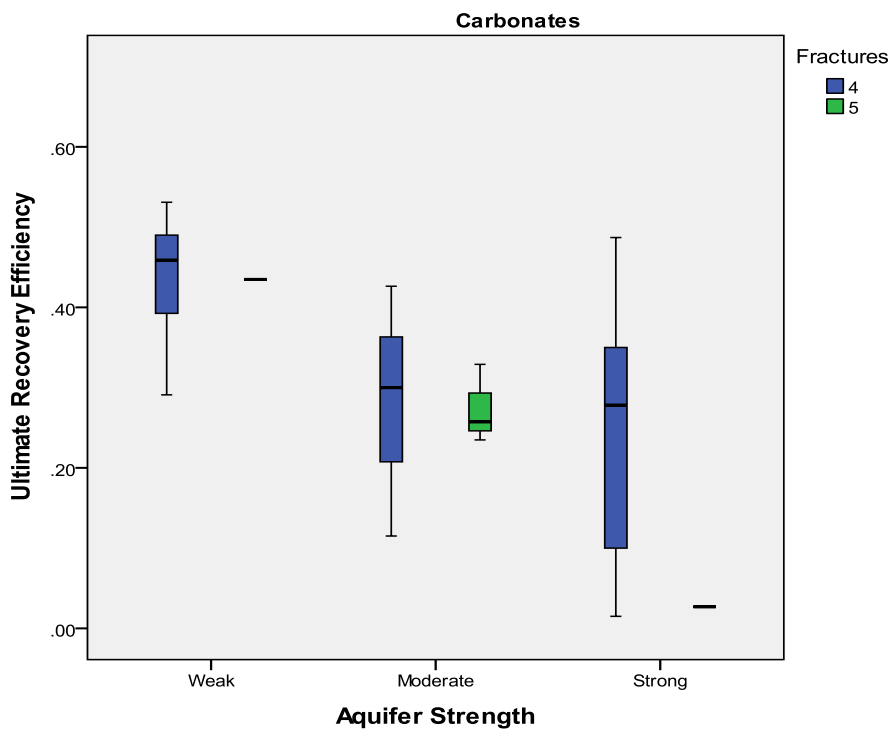


Figure 5.11: Ultimate recovery efficiency for carbonate reservoirs with high fracture intensity. For highly fracture reservoirs the aquifer strength has a negative impact on the recovery efficiency.

5.2.4. Depositional environment

Tyler and Finley, 1991, proposed a classification of reservoirs in terms of heterogeneity based on geological environment of deposition.

In our data we find a trend between oil recovery efficiency that is consistent with the net to gross ratio per depositional environment. See Figure 5.12. To verify whether this observation is indeed related to the impact of reservoir heterogeneity on macroscopic sweep efficiency the unrecovered mobile oil was plotted. See Figure 5.13. The groups of reservoirs with low heterogeneity and high recovery efficiency also have the highest EUR per well. See Figure 5.14.

Exceptions are the reservoir formations deposited in a deep marine setting. This group is dominated by Miocene sandstones which allows for a high microscopic sweep and high recovery efficiencies compared to other reservoirs in the dataset. However, most of the deep marine reservoirs are found in present day offshore and deepwater locations. The number of wells in such locations is constraint by economic limits which results in lower recovery factors. Most likely, reduction of the well costs would lead to additional recovery from those types of reservoirs.

Drilling more wells will increase the recovery of mobile oil by reducing the amount of unconnected oil. When we compare the well density we find that deltaic, shallow marine and fluvial reservoirs have similar ranges. We can conclude that this might identify infill drilling opportunities for the fluvial reservoirs which have significantly lower recoveries. Generally, the most successful infill wells are the ones that tap into undrained areas. Identifying such targets in complex geological settings as fluvial environments requires a very good understanding of the geology and high level reservoir modeling. The technology of 4D seismic has proven to be successful in providing this kind of improved understanding of the reservoir drainage performance.

According to Tyler and Finley, 1991, the unrecovered mobile oil can be divided into oil that is bypassed and oil that is unconnected to the well. The amount of bypassed oil relates to vertical heterogeneity whereas the oil that is not connected is a function of lateral heterogeneity. An additional parameter that would be interesting to differentiate unconnected oil from bypassed oil is water cut. High water cut with low recovery would indicate early breakthrough and mean bypassed rather than unconnected. In that case EOR methods such as polymer flooding would be more appropriate than infill drilling.

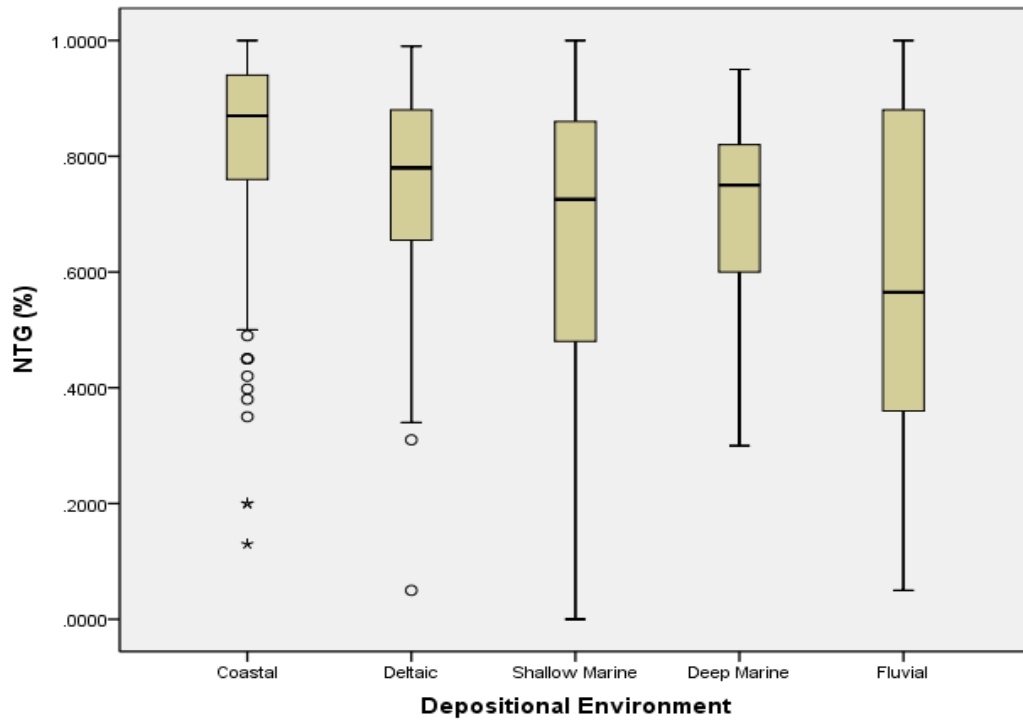


Figure 5.12: Net to gross ratio versus depositional environment.

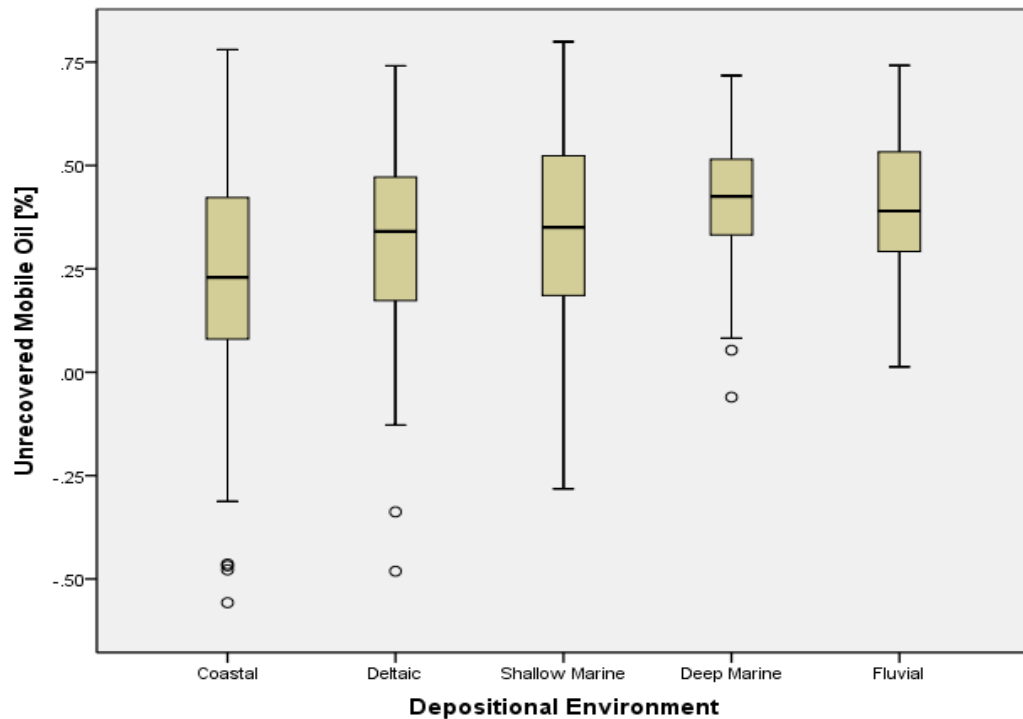


Figure 5.13: Unrecovered mobile oil versus depositional environment. Note that the unrecovered mobile becomes negative in some cases. This is unphysical and is caused by either too low estimates of microscopic sweep efficiency, which is a common problem in the dataset due to uncertainties around the origin of the 'Final So' value. See chapter 5, paragraph 5.2.1 'Drive mechanisms. It could also be the result of too high estimates of ultimate recovery, caused by pessimistic in-place estimates.

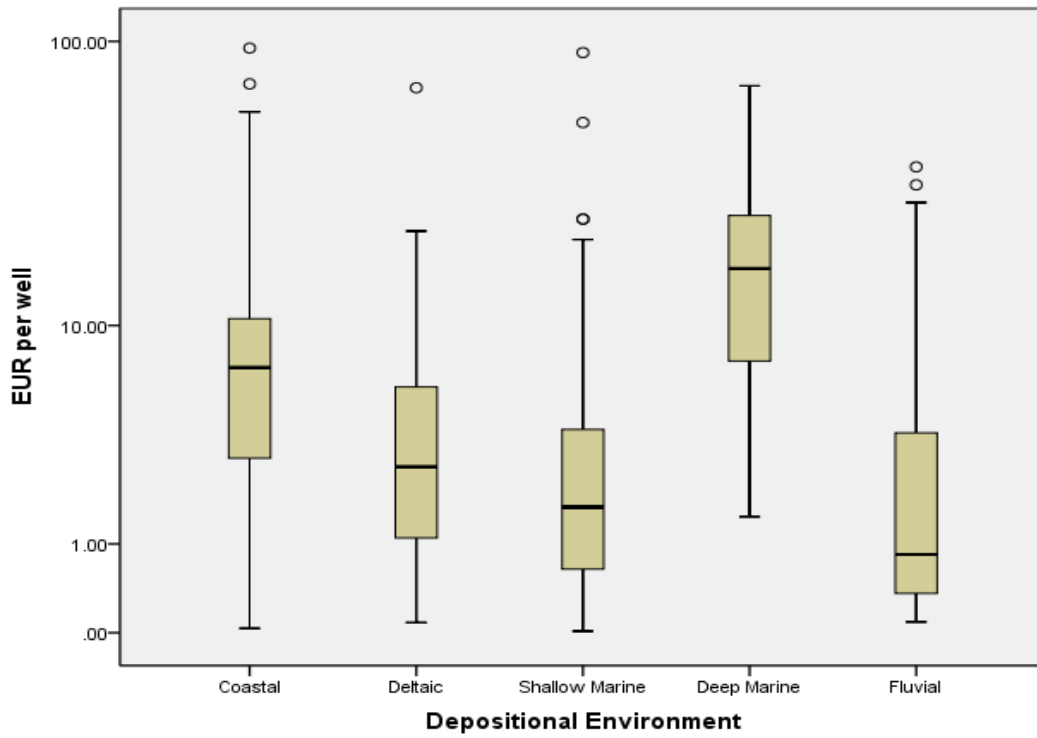


Figure 5.14: EUR per well versus depositional environment. There appears to be a relation between reservoir heterogeneity and EUR per well. Only Deep Marine reservoirs have a significantly higher EUR per well. This is due to the fact that all Deep Marine reservoirs are located in current day deepwater. The economic bias allows only for reservoirs with high EUR in this economically difficult location.

5.2.5. Sedimentary basins, geological age and depth

The stratigraphic record of sedimentary basins reflects both local conditions as well as global scale events such as tectonic activity, climate change and eustatic sea-level change. (Hancock, 2000). Due to variable local conditions the correspondence between reservoirs is strongest within individual basins.

Additionally, we may expect that geological units from a certain time period show some degree of similarity as they were deposited in the same worldwide climate. Moreover, in clastic rocks porosity reduces upon burial due to rearrangement of the grains under increased stress of overburden and cementation. The older the rocks, the longer they could have been exposed to diagenesis. In reality the result may not be as simple as burial histories are basin specific and could vary greatly. Another (imperfect) proxy to quantify diagenesis is depth of the reservoir. Processes that can cause differences in diagenesis for similar age or depths are basin inversion, regional volcanic activity and early migration of hydrocarbons.

The above suggests a possible trend between geological age of rocks and rock properties that influence the reservoir performance such as mineral composition, porosity and permeability. A link between fluid properties and age of the reservoir is less obvious. Oil properties are mainly determined by the type of source rock, maturation and migration path. It is not necessary that the age of the source rock is related to that of the reservoir. Even though younger sediments are originally deposited on top of older (law of superposition) certain structural settings allow reservoirs to be charged by younger source rocks. Unfortunately, source rock data itself is scarce and for many hydrocarbon accumulations the source is unknown. During deposition pore spaces are filled with water which varies from low salinity in fluvial and lacustrine settings to seawater for marine environments. Over time the formation water salinity is altered due to chemical interactions with the rocks upon diagenesis. As water chemistry influences fluid-rock interactions (wettability) it has direct influence on the microscopic sweep efficiency. Additionally water viscosity is a function of water salinity, but the ranges of water viscosity are narrow compared to that of oil.

Figure 5.15 shows porosity categorized by geological age. Carbonates were excluded from the analysis because they are more abundant in certain time periods. Besides, the porosity alteration of carbonates upon diagenesis deviates strongly from that of clastics. Insufficient data was available to make a similar analysis for carbonates only.

A potential global trend of decreasing porosity with increasing age is broken by the reservoirs from the Paleozoic. However, closer investigation of those particular reservoirs revealed that all of them are in the same country and the same basin. The same holds for the Jurassic reservoirs of which all

but one field are located in the same basin. Reservoir rocks from the other time periods are more equally spread around the world. Figure 5.16 shows a plot of porosities for sandstone reservoirs in the North Sea Basin. Caution must be taken in the interpretation of this plot as the amount of data points is limited. A similar but weaker global trend could be obtained for ultimate recovery efficiency. It is to be expected that better correlations could be obtained when the analysis is confined to a specific basin.

No correlation between age of reservoir rocks and fluid properties or macroscopic sweep was found. The relationship between geological age or depth on one hand and rock properties and reservoir performance on the other hand is weak. The dataset contains too few samples to test the statistical significance of trends per basin. Furthermore, classification of the data based on geologic age is not ideal as reservoirs are not equally distributed throughout time. The dataset is dominated by reservoirs that were deposited during the Neogene period (Miocene sandstones). Table 5.4 and Table 5.5 show the number of reservoirs in a cross table of geological age versus depositional environment.

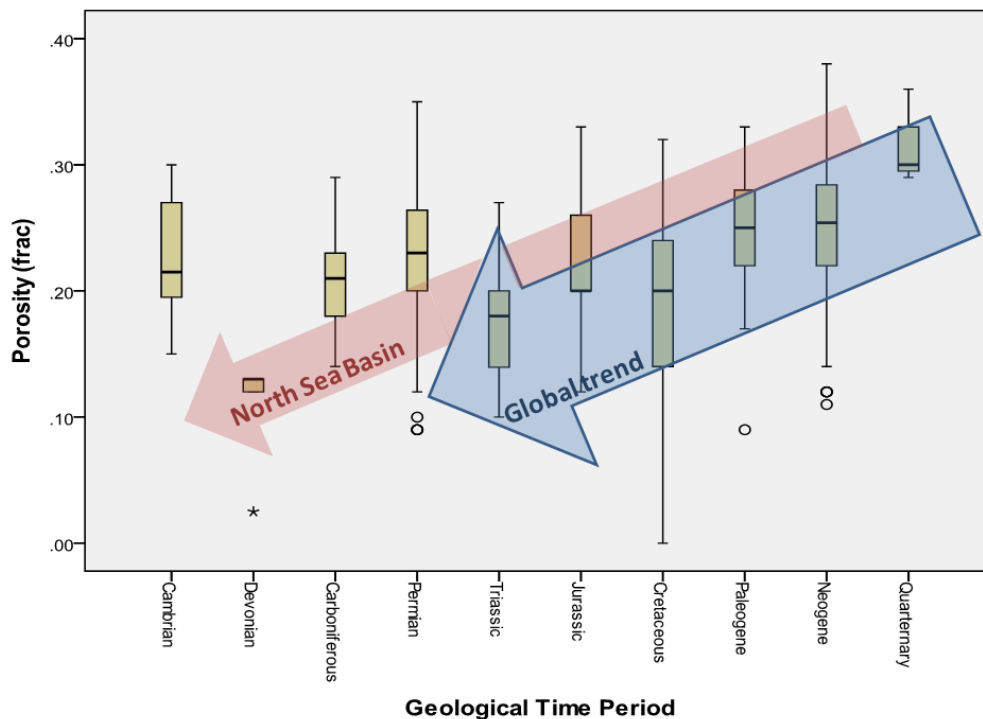


Figure 5.15: Porosity versus geological age. A potential trend of decreasing porosity with age seems to be violated by Paleozoic reservoirs. More detailed inspection of the data revealed that most of those reservoirs are located in the same basin. Also, the Jurassic and Devonian reservoirs are restricted to the North Sea basin.

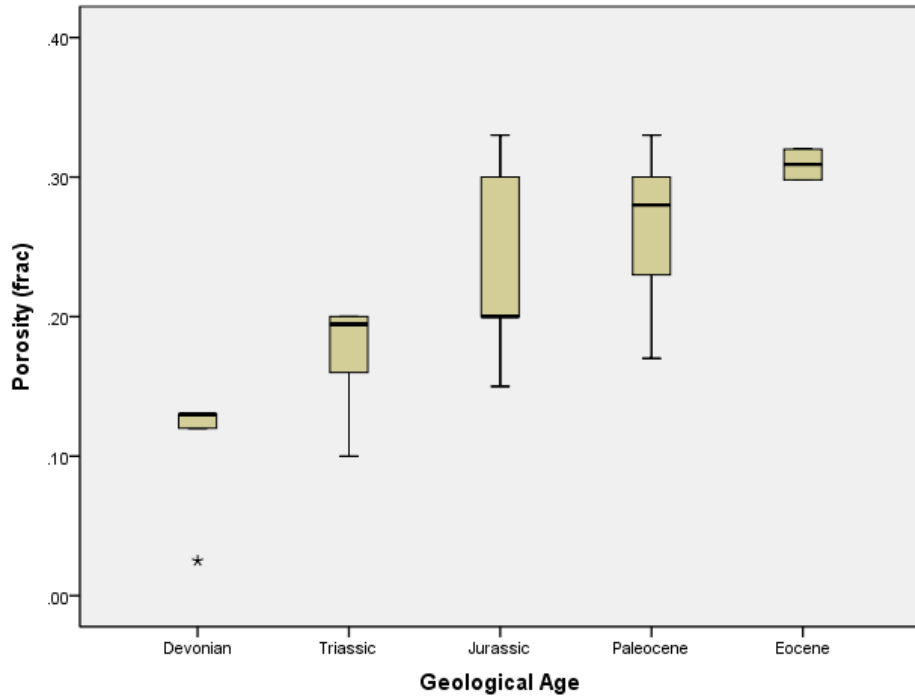


Figure 5.16: Porosity versus Geographical Age for reservoirs in the North Sea Basin.

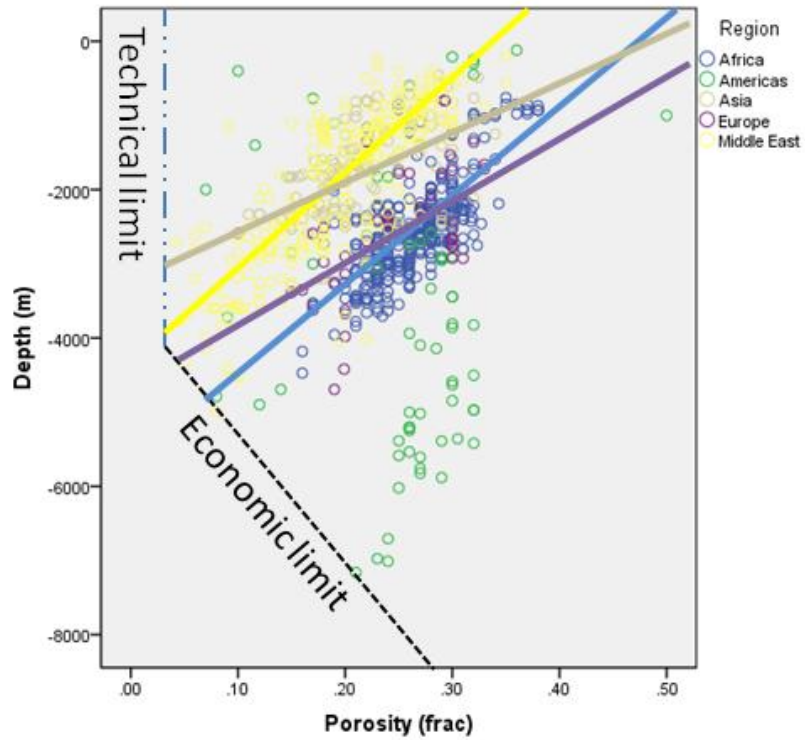


Figure 5.17: Porosity versus depth trend per region. Note that no trend for the Americas region could be obtained. This is because the reservoirs in the Gulf of Mexico follow a different trend than the other reservoirs.

Table 5.4: Clastic reservoirs in the dataset per geological age versus depositional environment.

	Aeolian	Coastal	Deep Marine	Delta	Fluvial	Glacial	Lacustrine	Total
Cambrian	3	0	2	0	14	0	3	22
Devonian	0	0	0	0	6	0	0	6
Carboniferous	0	0	0	0	3	6	0	9
Permian	0	8	0	2	77	24	1	112
Triassic	1	0	0	2	17	0	0	20
Jurassic	1	33	4	31	21	0	0	90
Cretaceous	1	83	5	3	24	0	2	118
Paleogene	0	25	33	38	3	0	0	99
Neogene	0	424	82	75	3	0	3	587
Quarternary	0	0	2	0	1	0	0	3
Total	6	573	128	151	169	30	9	1066

Table 5.5: Carbonate reservoirs in the dataset per geological age versus depositional environment.

	Coastal	Fore slope/Basin	High-Energy Carbonat Sand	Low-Energy Carbonat Mud	Organic Buildup	Total
Cambrian	0	0	0	21	0	21
Devonian	0	0	0	1	1	2
Carboniferous	0	0	0	0	1	1
Permian	0	0	2	2	1	5
Triassic	0	0	0	1	0	1
Jurassic	1	0	1	4	0	6
Cretaceous	6	1	37	47	3	94
Paleogene	0	1	0	1	1	3
Neogene	0	0	0	2	1	3
Quarternary	0	0	0	0	0	0
Total	7	2	40	79	8	136

5.3. Multivariate analysis results

Hydrocarbon production is by nature a multivariate process. Therefore reservoir performance can only be adequately described by considering not solely the individual variables but also and possibly more importantly their correlation. The main technique used to discover and extract the patterns of correlations is principal component analysis (PCA). To reveal the most important predictors, multivariate linear regression was used.

5.4. Principal component analysis

5.4.1. Correlations

The first result we obtain from the principal component analysis is a correlation matrix of all the independent variables that influence the recovery factor. Removed from the analysis are the parameters rock compressibility and reservoir area because they have a low correlation with RF and the other variables. Rock compressibility only plays a key role in reservoirs with high rock compressibility where compaction is a major contribution to the drive such as chalk reservoirs. However, such reservoirs only make up a small amount of the entire dataset. Water depth was included as a proxy for installation and operating costs.

In describing the strength of relationships we will distinguish between strong (0.6 and higher), moderate (0.3-0.6) and low (0.3 or less) correlations. Initially, using all reservoirs except those with missing values or zeroes resulted in poor correlations. Therefore, variables which are (approximately) log normally distributed were transformed before the analysis. This resulted in better differentiate strong and weak correlations. See Table 5.6.

Table 5.6: Correlation matrix. The darker the cell, the stronger the correlation.

	logP	logD	Pb	logGOR	logT	logOFVF	Soi	logWell	Porosity	Final So	logPerm	NTG%	logNET	logHc	API	logSTOIIP	Water D	logmu
logP	1.000	0.830	0.668	0.703	0.974	0.625	0.035	-0.582	-0.848	0.105	-0.410	-0.451	0.033	0.539	0.608	-0.090	-0.175	-0.668
logD	0.830	1.000	0.728	0.600	0.692	0.529	0.053	-0.297	-0.536	0.171	-0.063	-0.232	-0.163	0.712	0.274	0.227	0.365	-0.524
Pb	0.668	0.728	1.000	0.874	0.641	0.801	0.132	-0.459	-0.591	0.132	-0.027	-0.287	0.232	0.407	0.391	-0.288	-0.021	-0.677
logGOR	0.703	0.600	0.874	1.000	0.720	0.927	-0.108	-0.660	-0.656	0.398	-0.322	-0.359	0.188	0.213	0.759	-0.453	-0.176	-0.941
logT	0.974	0.692	0.641	0.720	1.000	0.645	0.034	-0.654	-0.913	0.077	-0.505	-0.522	0.133	0.420	0.675	-0.257	-0.390	-0.680
logOFVF	0.625	0.529	0.801	0.927	0.645	1.000	-0.203	-0.586	-0.596	0.217	-0.305	-0.208	-0.025	-0.063	0.765	-0.426	-0.185	-0.886
Soi	0.035	0.053	0.132	-0.108	0.034	-0.203	1.000	-0.148	-0.008	-0.397	0.631	0.241	0.501	0.268	-0.289	0.384	-0.073	0.260
logWell	-0.582	-0.297	-0.459	-0.660	-0.654	-0.586	-0.148	1.000	0.777	-0.230	0.309	0.073	-0.506	-0.144	-0.699	0.159	0.439	0.674
Porosity	-0.848	-0.536	-0.591	-0.656	-0.913	-0.596	-0.008	0.777	1.000	-0.053	0.577	0.475	-0.299	-0.347	-0.586	0.336	0.524	0.580
Final So	0.105	0.171	0.132	0.398	0.077	0.217	-0.397	-0.230	-0.053	1.000	-0.418	-0.422	-0.134	0.252	0.319	-0.210	0.261	-0.517
logPerm	-0.410	-0.063	-0.027	-0.322	-0.505	-0.305	0.631	0.309	0.577	-0.418	1.000	0.713	0.182	0.034	-0.533	0.622	0.472	0.423
NTG%	-0.451	-0.232	-0.287	-0.359	-0.522	-0.208	0.241	0.073	0.475	-0.422	0.713	1.000	0.086	-0.313	-0.215	0.635	0.361	0.293
logNET	0.033	-0.163	0.232	0.188	0.133	-0.025	0.501	-0.506	-0.299	-0.134	0.182	0.086	1.000	0.206	0.088	-0.187	-0.469	-0.058
logHc	0.539	0.712	0.407	0.213	0.420	-0.063	0.268	-0.144	-0.347	0.252	0.034	-0.313	0.206	1.000	-0.144	0.309	0.352	-0.108
API	0.608	0.274	0.391	0.759	0.675	0.765	-0.289	-0.699	-0.586	0.319	-0.533	-0.215	0.068	-0.144	1.000	-0.409	-0.397	-0.898
logSTOIIP	-0.090	0.227	-0.288	-0.453	-0.257	-0.426	0.384	0.159	0.336	-0.210	0.622	0.635	-0.187	0.309	-0.409	1.000	0.662	0.385
Water D	-0.175	0.365	-0.021	-0.176	-0.390	-0.185	-0.073	0.439	0.524	0.261	0.472	0.361	-0.469	0.352	-0.397	0.662	1.000	0.133
logmu	-0.668	-0.524	-0.677	-0.941	-0.680	-0.886	0.260	0.674	0.580	-0.517	0.423	0.293	-0.058	-0.108	-0.898	0.385	0.133	1.000

Expected relationships with a physical explanation such as correlations between API-viscosity, STOIP- Net thickness and porosity-permeability could be confirmed. The relation between porosity and permeability is weaker than one might expect, but bear in mind that there are many different rock types in the dataset. Strong are the positive correlations in the upper left corner of the matrix between pressure (logP), depth (logD), bubble point pressure (Pb), Gas-Oil-Ratio (logGOR), temperature (logT), oil formation volume factor (logOFVF) and API gravity. These correlations can be explained as the change of fluid properties as a function of depth controlled independent variables pressure and temperature. Negative correlations are present between the previous variables and viscosity (logmu), well density (logWell) and porosity. The two latter don't represent a causal relationship but a bias effect on the selection of reservoirs. It seems that for lighter reservoir fluids lower porosities and lower well density are acceptable. Overall, the reservoir fluid properties seem to correlate better with each other than the reservoir rock properties. This indicates that there is more variance in the reservoir rock properties which might not be adequately described by the variables included in the analysis.

5.4.2. Components

From inspection of the correlation matrix we can already conclude that there are some high correlations between the independent variables. Now, we wish to express our dataset in principal components to describe the variance in recovery efficiency. The scree plot helps in identifying how many components should be considered. In our case we analyzed three principal components. However, further analysis of plots with principal component 1 against principal component 3 did not result in interpretable results. This is due to the fact that the amount of variance explained by component 3 is already very low (<15%).

5.4.3. Loadings and principal scores

The loadings represents the relative importance of each variable on a component. The loadings were rotated to increase orthogonality of the solution. The rotation method VARIMAX was used, which maximizes the high loadings and minimizes the low loadings. The original and rotated solution is shown in Table 5.7. The first component has heavy loadings on variables that describe PVT data and fluid properties. The second component has heavy loadings on rock properties such as permeability but also on water depth.

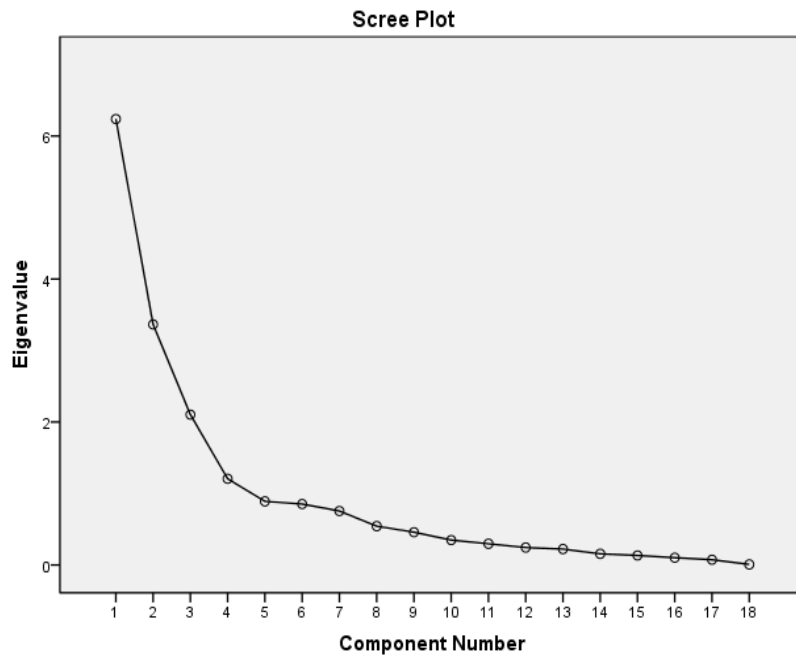


Figure 5.18: Scree plot. This plot can be used to determine the required amount of components by determining the ‘elbow’ of the graph, here at the elbow sits at 5 components. Alternatively, one can select the number of components by selecting ones with the eigenvalues of each component. In this case 4 have eigenvalues greater than 1. If the eigenvalue is less than 1, a component does not explain more variance than a single variable itself would.

Table 5.7: Left: Loadings of input variables on the components. Right: Loadings on the rotated solution. Rotation VARIMAX maximizes high loadings and minimizes low loadings per component.

	Component Matrix ^a					Rotated Component Matrix ^a				
	1	2	3	4	5	1	2	3	4	5
logP	.868	.323	.039	-.218	-.273	.624	-.261	.692	-.039	-.211
logD	.631	.660	.352	-.090	-.162	.486	.163	.831	-.195	.022
Bubble Point (psi)	.762	.381	-.023	.154	.129	.718	.034	.445	.195	.116
logGOR	.921	.068	.044	.289	.171	.910	-.164	.224	.098	.229
logT	.908	.156	-.105	-.241	-.256	.644	-.415	.570	.058	-.260
logOFVF	.839	-.028	.067	.458	-.113	.953	-.092	.050	-.113	.019
Initial So	-.147	.602	-.631	-.076	.116	-.129	.391	.253	.693	-.294
logWell	-.739	-.060	.406	-.185	-.119	-.724	.182	-.109	-.434	.069
Porosity (frac)	-.868	-.044	.264	.273	.172	-.592	.505	-.454	-.211	.269
Final So (frac)	.345	-.173	.574	.024	.640	.224	-.220	.097	-.141	.872
logPerm	-.566	.634	-.233	.379	.100	-.239	.880	-.047	.298	-.082
NTG (%)	-.529	.327	-.222	.654	-.166	-.045	.819	-.387	.062	-.252
logNET	.165	.124	-.794	-.041	.497	.116	-.061	-.039	.951	-.005
logHc	.299	.695	.180	-.527	.285	-.054	.052	.922	.213	.253
API Gravity (deg)	.802	-.277	-.025	.355	-.112	.862	-.314	-.099	-.080	-.020
logSTOIIIP	-.459	.729	.126	.091	-.209	-.301	.757	.322	-.106	-.179
Water Depth (m)	-.349	.571	.695	.212	.124	-.195	.704	.300	-.427	.429
logmu	-.889	.049	-.166	-.375	-.120	-.927	.173	-.112	.062	-.265

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Extraction Method: Principal Component Analysis.

Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 7 iterations.

We can plot the first two components on a so called loadings plot. Additionally, we can calculate and plot the principal scores of all reservoirs on the rotated component space. See Figure 5.19. None of the reservoirs are located in the lower left corner which is the effect of the economic bias of our dataset. We can also observe that there are two clusters of data points and most of the data is accumulated in the left cluster. The group of reservoirs on the right could be identified as the deepwater reservoirs. See Figure 5.20.

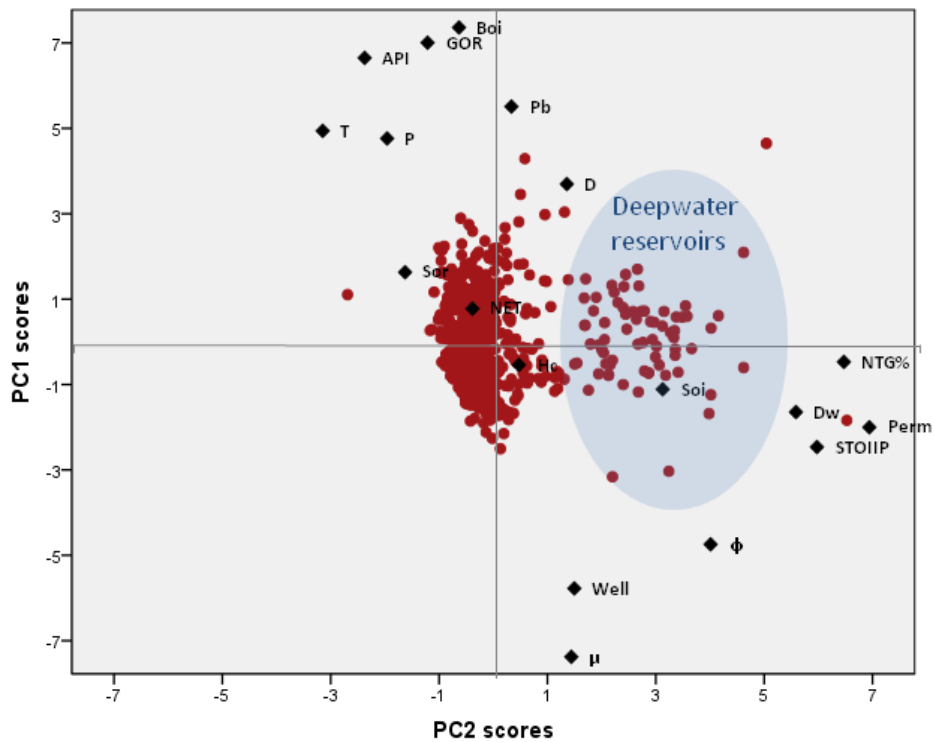


Figure 5.19: Combined loadings (black) and principal scores (red) plot for all reservoirs. The first principal component has heavy loadings on fluid properties. Reservoirs scoring high on this component are light type oils. The second component has heavy loadings on rock properties but also on water depth. Reservoirs scoring high on this component form a distinct cluster and could be identified as deepwater reservoirs.

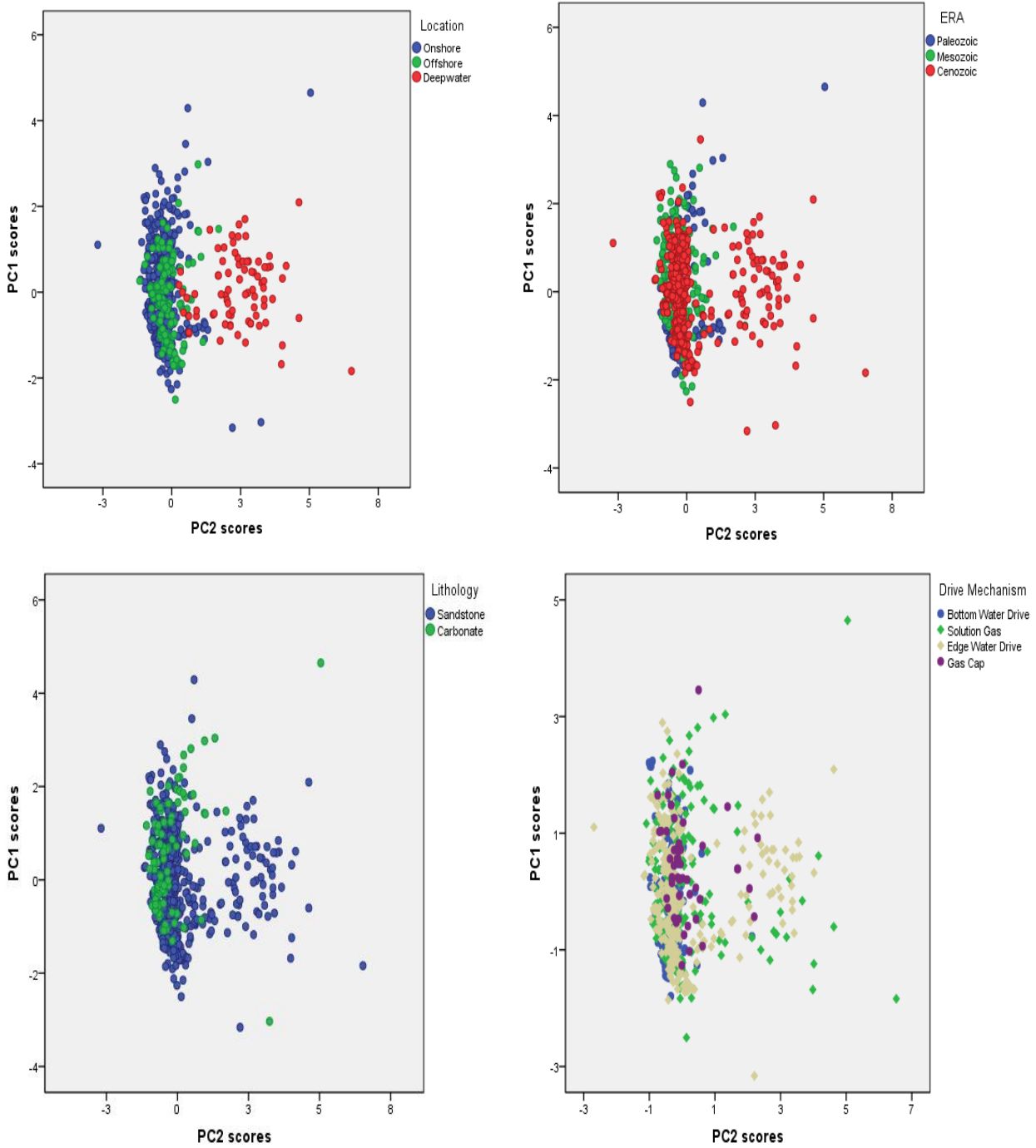


Figure 5.20: Principal scores of reservoirs with markers indicating: location, geological age, lithology and drive mechanism. The principal components represent the same variables as shown in Figure 5.19.

5.4.4. Deepwater reservoirs

In our dataset ‘deepwater’ refers to reservoirs located in areas with water depths exceeding 200 metres. Reservoirs consisting of sediments deposited in a deepwater environment are referred to as ‘deep marine’. Through both the subgroup- and the principal component analysis we find that deepwater reservoirs are significantly different from the other reservoirs. This is again the effect of an economic bias, which seems to apply to reservoir rock properties and not so much to reservoir fluid properties. Because deepwater drilling is costly and risky, only the reservoirs with the best reservoir properties could be economically developed. Hence the resulting average well density is much lower for deepwater reservoirs while the porosity and permeability are much higher. See Table 5.8. Also, it is the explanation for why no statistically different estimated ultimate recovery was found by dividing per location: the most difficult location deepwater compensates by having better reservoir rock properties.

Table 5.8: Comparison of mean values of key parameters for deepwater, offshore and onshore fields in our dataset. The deepwater reservoirs are characterized by low average well densities but high porosity and permeability.

	Well density [wells/km ²]	Exp.STOIP [mln. bbl]	Net to Gross [frac]	Porosity [frac]	Perm [mDarcy]
Deepwater	2.68	1067	68	0.27	956
Offshore	3.56	1024	68	0.23	687
Onshore	5.08	860	73	0.23	736

In our dataset, most of the deepwater reservoirs are submarine fans deposited during the Miocene in a deep marine environment. Deep marine reservoirs have potentially large lateral extents, good connectivity and excellent sand quality. Due to their young age and rapid deposition by turbidity currents, the reservoirs have maintained high porosities and permeabilities. A key risks for such reservoir types is reservoir compaction (Ostermeier, 2001). According to Ostermeier, 2001, compaction can cause three major issues, namely: subsidence affecting the integrity of the structure, stresses damaging and affecting the integrety of the casing and reduction of porosity and permeability affecting the production performance.

Analysis of the reservoirs excluding deepwater reservoirs resulted in two new components. The first component represents the porosity-depth relationship that was also found through univariate analysis. This relationship is over shadowed by the economic bias when the entire dataset is considered. The second component is again a measure of the fluid properties, which was found as component one in analysis of all the reservoirs. Further analysis of the results did not reveal any distinct subgroups such as the deepwater reservoirs.

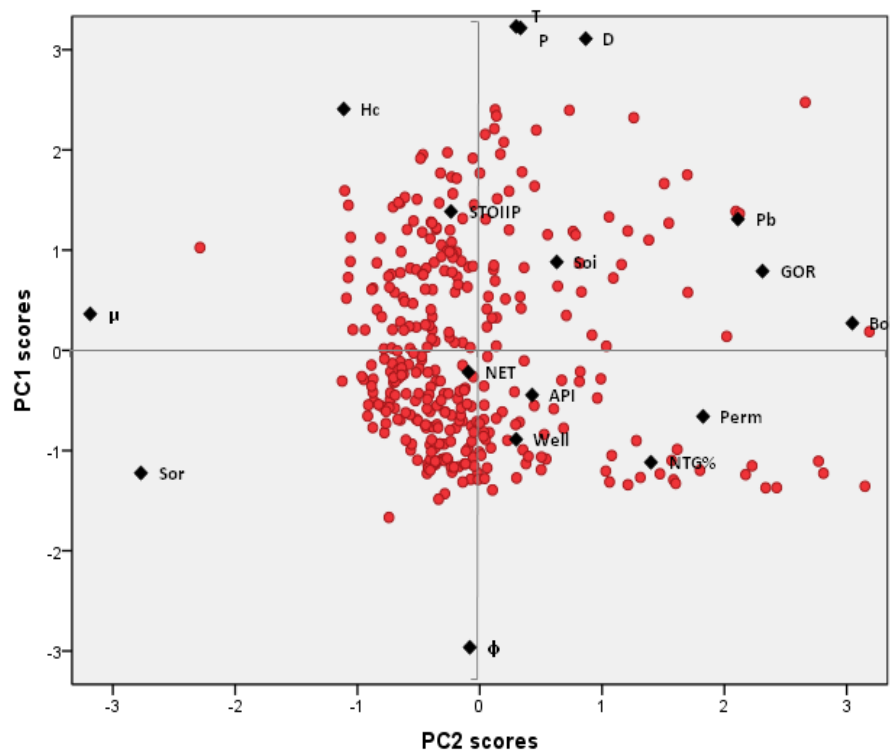


Figure 5.21: Combined loadings (black) and principal scores (red) plot for all Non-Deepwater reservoirs.

5.4.5. Performance of gas cap reservoirs

Recall from Chapter 4 ‘Methods’, paragraph 4.4.2 ‘Subgroup analysis’ that bottom water drive reservoirs and gas cap drive reservoirs in the dataset are performing less than respectively edge water- and solution gas reservoirs. From Figure 5.22 it can be seen that the edge and solution gas have a considerable amount of reservoirs which score high on the second component. In other words, they seem to have slightly better rock properties than which can explain the difference in performance. This corresponds with the fact that permeability is higher in edge water drive reservoirs. For gas cap reservoirs such results could not be obtained from univariate analysis, as for the distributions of independent parameters did not show statistically different results. This results highlights the importance of multivariate statistical analysis in reservoir engineering. As hydrocarbon production is in essence a multivariate process, interaction between predictor variables has to be taken into account.

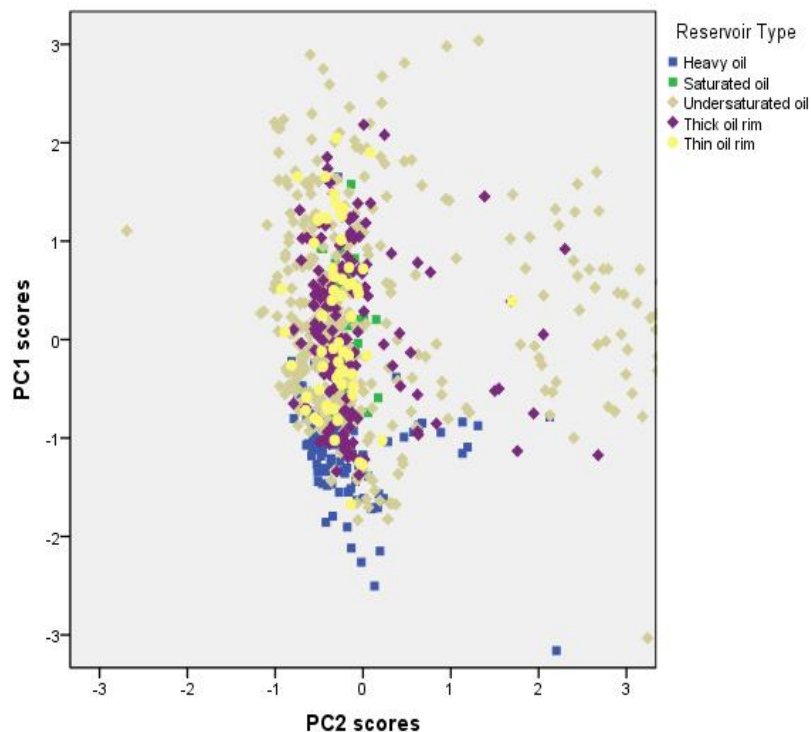


Figure 5.22: Principal scores plot with markers indicating reservoir type. The principal components represent the same variables as shown in Figure 5.19. To increase visibility, the axes are cut at PC scores of 3. Note that the heavy oils form a coherent group at the bottom left of the cloud. Under saturated reservoirs plot over the entire range, even outside the plotted are.

5.4.6. EOR methods

Figure 5.23 shows that the first principal component, representing fluid properties, is most critical in determining which method is applied. This is not a surprising result as depth and the corresponding API gravity are the main screening criteria for EOR methods. Generally, steam injection is applied to heavy oil reservoirs at shallow depths to minimize wellbore heat losses. Miscible gas injection on the other hand is often applied to light type oils at great depths and high pressures to achieve miscibility. (Taber et al., 1997). We also find that the second principal component, representing rock properties, does not seem to affect the choice of EOR method. For reservoirs on tertiary recovery principal component analysis does not give a lot more information. Simple crossplots of for example depth versus API gravity yield similar result that are easier to interpret.

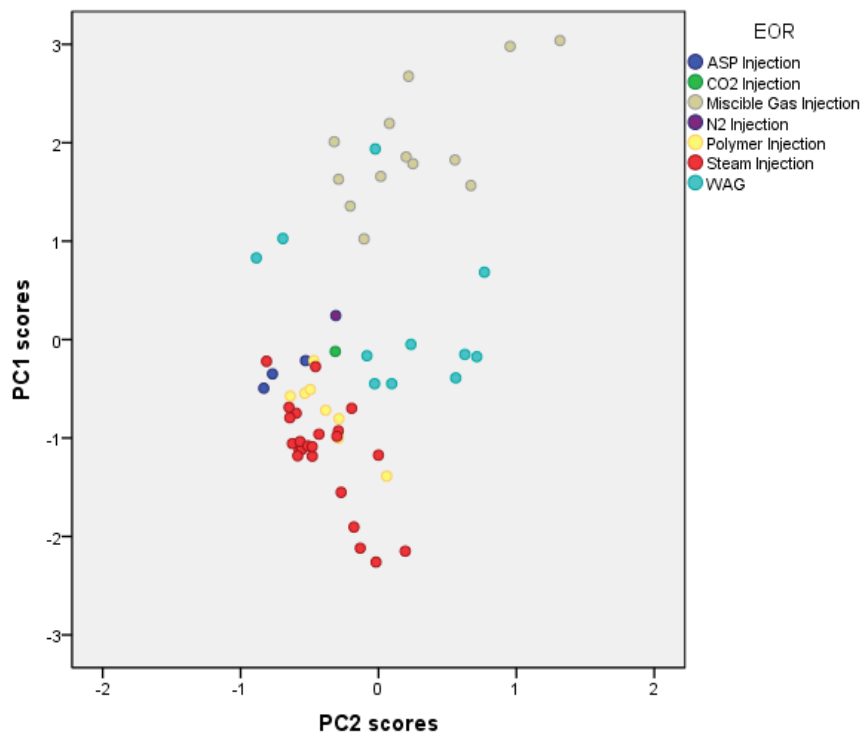


Figure 5.23: Principal scores with markers indicating EOR method. The principal components represent the same variables as shown in Figure 5.19. The plot shows that the first principal component, representing fluid properties, is most important in determining which method is applied. Also note that in this plot we can recognize the heavy oils from Figure 5.22 as reservoirs on steam flooding.

5.5. Multi linear regression

The structure of the dataset has been identified in terms of two principal components (fluid- and rock properties). Now we would like to know which of the independent are the most important for the prediction of recovery efficiency. Therefore multivariate linear regression technique has been applied.

5.5.1. Subgroups and sample size

Univariate and subgroup analysis showed that a lot of variance in the reservoir performance can be explained by discrete variables such as basin, depositional environment and drive mechanism. When we analyze the entire dataset as a whole this variance will be unaccounted for. This results in only low to moderate fits for multi linear equations. However, if we apply the same analysis to subgroups of the data the correlations we can extract are become more apparent. For this reason the analysis was applied to a subgroup with the largest amount of cases: primary RF coastal sandstone reservoirs of Miocene age. The advantage of analyzing this subgroup is that while we have reduced the variance significantly we are still left with a large amount of reservoirs. On the other hand, we are not using the full range of our dataset and findings are hard to expand to other subgroups. Therefore, we also consider a random training sample of approximately 5% of the entire dataset. This random sample was used to check whether the findings from the defined subgroup could be expanded to the entire dataset. Initially, the results which were obtained using the random sample are presented. Later, paragraph 5.5.3 'Expansion to the dataset' will discuss how well these results relate to the entire dataset.

5.5.2. Most important variables

Multivariate linear regression was used to extract the most important predictor variables from the first two principal components that were obtained through principal component analysis. The variables with the highest correlation coefficients are also the most significant according to t-test. This is because the spread of the distributions is in the same order of magnitude. The most important variable for is reservoir temperature which has loadings on the first principal component. Secondly, we find permeability which has loadings on the second principal component. The third important parameter is API gravity. It reflects the fluid composition on which other fluid properties depend. The other extracted parameters are not statistically significant in the prediction equations for recovery factor using t-test criteria of 0.1 (10%). However, it should be noted that the R^2 values of the resulting equations were low. Beware that unlike the principal component analysis the multi linear regression does not account for correlation between the independent variables.

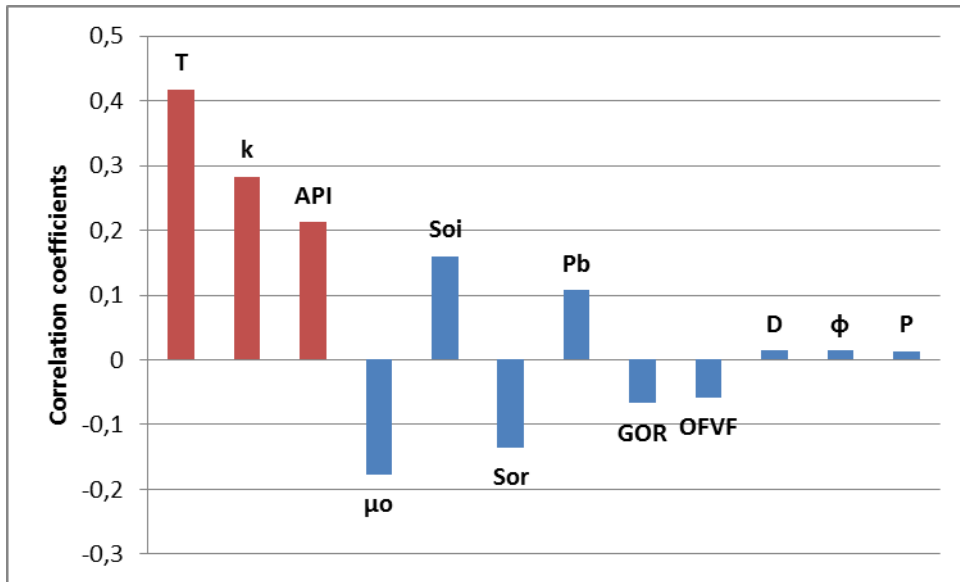


Figure 5.24: Standardized weights for predictors of recovery efficiency. In red, the significant variables according to t-test with tolerance level of 0.10.

5.5.3. Expansion to entire dataset

The R^2 values of the obtained multivariate equations for recovery factor are low and the results are therefore not appropriate for prediction purposes. We may conclude that it is not possible to obtain useful correlation from a reservoir engineering dataset with so many different reservoirs. This is in contrast to previous studies that successfully obtained correlations for much smaller sample groups. Hence if the objective of a study is to formulate such equations, one should focus on a subgroup to reduce the variance. Considering the observations of the univariate analysis, it is suggested that limiting such research to a specific basin would be a good place to start. Nevertheless, the multivariate regression highlighted the most important predictor variables. Moreover, they seem to be consistent if we expand the analysis to the entire dataset. See Table 5.9.

Table 5.9: Percentage of variance in oil recovery factor accounted for by the two most important variables of component 1. Note the change of order for all the reservoirs.

	Reservoir temperature	API gravity
Random sample	17.5%	6%
Defined subgroup	6%	4%
All oil reservoirs	3.6%	6.7%

5.5.4. Predict missing variables

Because some variables share a lot of variance with other variables it is in theory not necessary to know all properties. This is especially useful in a large but inhomogeneous (many missing values and zeroes) dataset such as the TQ EUR tool. Also it reduces the dimensionality of the problem. We have derived the following linear regression models to predict the most important variables: temperature, permeability and API gravity.

Reservoir temperature

$$\log T = -1.113 + 0.106 \log \mu_{oil} + 0.004 API + 0.811 \log P \quad \text{Equation 18}$$

$$R^2=0.825$$

The positive correlation between temperature and oil viscosity is unphysical. This term seems to appear as a correction on the term of API. Removal leads to a much lower value of R^2 (0.310) of the model.

Permeability

$$\log perm = -0.111 + 6.716\phi + 1.471 S_{oi} \quad \text{Equation 19}$$

$$R^2=0.317$$

It appears to be hard to use other variables for the prediction of permeability. The reason is that the dataset contains a wide variety of rock types. This provides also an explanation for why permeability is a dominant predictor for recovery efficiency. It describes variance in the dataset that is unique and not shared with other variables. Fortunately, permeability is a well recorded variable, all cases in the dataset report a non-zero value for permeability. Recall Figure 4.4 from Chapter 4 ‘Methods’, paragraph 4.3.2 ‘Data quality’.

API gravity

$$API = -60.913 + 26.585 \log GOR - 0.004 Pb + 13.532 \log T + 24.435 \log OFVF \quad \text{Equation 20}$$

$$R^2=0.867$$

Additionally, we would like to know if we can use parameters that we can measure in the field such as reservoir pressure, reservoir temperature and API gravity to adequately estimate fluid properties

required in reservoir studies (bubble point pressure, solution gas oil ratio, and oil formation volume factor and oil viscosity).

Solution gas oil ratio

$$\log GOR = -2.087 + 1.638 \log T + 0.016 API + 0.241 \log P \quad \text{Equation 21}$$

$R^2=0.357$

Bubble point pressure

$$Pb = -9702.988 + 2157.412 \log P + 24.567 API + 1974.214 \log T \quad \text{Equation 22}$$

$R^2=0.389$

Oil formation volume factor

$$\log OilFVF = -0.491 + 0.061 \log P + 0.002 API + 0.162 \log T \quad \text{Equation 23}$$

$R^2=0.288$

Oil viscosity

$$\log \mu_{oil} = 14.7 - 7.352 \log P + 6.739 \log T - 0.015 API \quad \text{Equation 24}$$

$R^2=0.824$

The equation for prediction of oil viscosity is the only one with high R^2 values. Again we find a positive correlation between oil viscosity and reservoir temperature. Removal of temperature did strongly reduce the R^2 value of the model. It appears that is a correction for the way that pressure influences recovery factor.

The obtained models are used to demonstrate that in some cases missing values can be substituted by using other variables to predict them. However, the insights in the correlations are more valuable than the obtained mathematical expressions themselves. The models were generated using the full range of data. Moreover, the models are the result of linear regression. Non linear regression could lead to better results.

6. Additional data

One of the objectives of this study was to investigate what key parameters are currently not reported in the tool. Based on both the literature study, dimensional analysis and the results of the statistical analysis additional data that should be included was identified.

6.1. Literature study

Based on the literature study a summary of the key determinants for recovery efficiency was made. See chapter 2 'Literature Review', Table 2.1. The parameters used in different literature studies were then compared to the parameters stored in the TQ EUR Tool. This comparison results in the recommendation to include the following variables: dip angle, flow rate, fractional water cut, and pressure drop. These parameters will provide more information on the operating strategy and the control we have of the reservoir. This information is currently missing but is expected to have a large influence on recovery efficiency, especially for difficult reservoir such as gas cap reservoirs or highly fractured carbonates. The parameters are also necessary when secondary and tertiary recovery methods are applied.

6.2. Dimensional analysis

Besides statistical analysis another approach to determine which variables are fundamental predictors for recovery factor is via dimensional analysis. The dimensionless groups become increasingly more important when secondary and tertiary recovery methods.

The advantages of using dimensionless groups are:

- the reduction of the number of variables
- making the system independent of units
- deduction important drive mechanisms
- prediction of effect changing one parameter by testing effect of this parameter on dimensionless group.

Dimensional analysis and statistical analysis are similar in the sense that all relevant parameters should be included in order to obtain meaningful results.

Geerstma, 1956 derived dimensionless groups from relevant basic reservoir engineering equations for cold water-, hot-water- and solvent flooding reservoirs. According to Geertsma, 1956, dimensionless groups are very useful for the appropriate scaling of model experiments.

However, because of the complexity of reservoir engineering problems, dimensional analysis often results in a large number of dimensionless parameters which makes it almost impossible to take all of them into account as scaling parameters. Bai, 2005, therefore states the need to focus on the

essential dimensionless groups only. His study comprised of a numerical approach to analyze the sensitivity of the dimensionless parameters in order to rank them in order of importance. The results of this study show that the following dimensionless groups are the key (scaling) parameters for water-flooding reservoirs.

$$\frac{K_{cwo}}{K_{row}}, \frac{K_o}{K_{cwo}}, \frac{K_w}{K_{row}}, \frac{s_{wi} - s_{cw}}{\Delta s}, \frac{\mu_o}{\mu_w}, \frac{\rho_{o0}}{\rho_{w0}}$$

The ratio of the oil permeability under the condition of the irreducible water saturation to the water permeability under the condition of the irreducible oil saturation, the dimensionless permeability of oil and water, the density and viscosity ratios between water and oil and the reduced initial water saturation. Relative permeability ratios are important determinants for recovery efficiency, especially for secondary and tertiary recovery methods. Currently, relative permeability data is not recorded in the tool. However, there are separate datasets from which this data can be obtained. Despite the importance of relative permeabilities it is not recommended to include them in our dataset due to the structure of the TQ EUR Tool.

6.3. Statistical analysis

In chapter 5, paragraph 5.2.3 ‘Aquifer strength, lithology and fracture intensity’ the influence of fracture intensity on recovery efficiency was highlighted. However, fracture intensity is currently reported with a subjective ordinal scale. See Table 5.3. Therefore the use of an objective continuous variable for fracture intensity is proposed. The same holds for vertical and lateral heterogeneity.

Chapter 4, paragraph 4.1.3 mentions that Estimated Ultimate Recovery (OilRF%) in some cases includes production from secondary and tertiary methods if those are part of the field development plan. As we do not know exactly when those methods are included in this value it becomes hard to compare reservoirs. Therefore it is suggested to split the OilRF% in: primary-, secondary- and tertiary recovery factor. This will be of special interest when further research would be directed towards an assessment of the effectiveness of IOR and EOR methods.

In the analysis all the reservoirs were treated equally. Alternatively, we could have assigned weights to each reservoir or even to each individual parameter describing every reservoir. Reasons for doing so are for example the fact that maturity has a distinct effect on estimated ultimate recovery factor. However, a good measure for maturity that could be used to assign such weights is currently not available. Including the origin of measurements for all data would be a massive if not impossible exercise, increases the dimensionality of the problem and probably not worth the effort. As an alternative a variable to specify the level of confidence of all values attached to one reservoir could be included. Additionally, it could be considered to include the measurement method only for

parameters that are essential. Especially when large differences depending on the measure method are to be expected. In our dataset that are: 'Final So' and 'Permeability'.

Chapter 5, paragraph 5.1.1 'Recovery Efficiency' discusses the effect of maturity on the distribution of Estimated Ultimate Recovery. As there is a direct relationship between the stage of development and the reliability of the ultimate recovery estimate, this should be accounted for in the analysis. Therefore, the use of a categorical variable with the following categories 'Exploration', 'Appraisal', 'Development', 'Production' and 'Abandonment' is proposed.

6.4. Formation water properties

When we inspect the dimensionless groups we find that most of the dimensionless number include a ratio of oil and water properties. Historically reservoir engineers have not focussed on the properties of formation water. Formation water- salinity, density and viscosity are industry wide poorly reported parameters. Also the TQ EUR Tool does not include any formation water properties. The C&C dataset on the other hand records water salinity from which the viscosity and density could be estimated. In this estimation the brine composition is oversimplified and represented as an NaCl solution. Error in the estimations of viscosity and density from salinity are 5% (Craft & Hawkins, 1981; McCain, 1990). The average brine salinity of the reservoirs in the C&C data is 4.2%. This average was used to estimate the water viscosity and density using reservoir temperature and pressure for the reservoirs in the TQ EUR TOOL. The estimation is based on the assumption that the average water salinity for both datasets would be the same. The same procedure was applied to the C&C data and cross checked with viscosity approximations using individual salinity values. The average error introduced by using average salinity instead of individual values was 14%.

Water viscosity varies from 0.3 to 1.4 cP, depending on the composition and concentration of salts. Sensitivity analysis showed that it is not necessary to include the missing water properties. Like oil properties strongly correlated with reservoir- temperature and pressure. Moreover, their variance is even less than that of oil.

However, salinity is of importance for secondary and enhanced oil recovery methods as a high salinity contrast between injection fluid and formation water affects the injectivity. Furthermore, in high saline water the viscosifying effect of polymers is strongly reduced. So in the case of screening for EOR opportunities water salinity is a required variable.

7. Conclusions and Recommendations

7.1. Conclusions

The global and industry wide dataset stored in the so called TQ EUR TOOL was subjected to data mining. The following insights in the origin of variance in reservoir performance could be extracted from this analysis:

- There is a strong economic filter resulting in a bias towards reservoir with good rock and fluid properties. As a result of the economic bias we find positive correlations between reservoir properties and economic complexity of the reservoirs. Available proxies for economic complexity and development costs are location, water depth and reservoir depth. Especially in the assessment of the relationship between well spacing and recovery factor economic overprint plays a key role.
- The spread of recovery factors is rather wide. Despite the variance some global trends could be distilled. We find that the oil recovery factor decreases with increasing heterogeneity of the depositional environment. Exceptions are the fluvial reservoirs which have low well densities with respect to their heterogeneity and recovery factors indicating potential infill drilling opportunities.
- Basin specific analysis yielded good correlation between rock- and fluid properties versus depth. Recovery factor versus depth correlations were not found, probably due to the economic bias. If the objective is to produce correlations that can be used for accurate predictions, they should be restricted to basins.
- Using ANOVA the following statistically different subgroups were identified: drive mechanism, reservoir type, depositional environment, lithology and region. Division by fiscal regime and location (deepwater/offshore/onshore) did not lead to statistically different subgroups with respect to recovery factor. Use of reasonable subgroups proved to be successful in reducing variance of ultimate recovery factor. Now further statistical analysis of the dataset is suggested to focus on cluster analysis to discover unknown groups and hidden structures in the dataset.
- Principal component analysis was used to reveal the structure of the dataset. The analysis showed that most of the variance in recovery factor can be contributed to fluid properties and rock properties. The volumetric parameters reservoir area, gross thickness and expected STOIP don't show significant influence on Recovery Factor. However, the variance in recovery factor is much larger for smaller fields. The variance seems to originate from the possibility of success per well and decreases with an increasing number of wells.

- In addition, automatic multivariate linear regression was applied to extract the most important variables: API gravity, permeability and reservoir temperature. These findings suggest that in using analogues for benchmarking special attention should be given to those parameters.

7.2. Recommendations

Future research is suggested to be directed towards effectiveness of IOR/EOR projects and reservoir control. These topics could not be adequately addressed in this analysis due to the absence of key parameters. To improve the overall structure of the TQ EUR tool and the following are suggested:

- In chapter 6 'Additional data' the necessity to include additional parameters in the database is discussed. Especially in the performance prediction of secondary and tertiary recovery methods more variables are required. Based on dimensional analysis and literature review it is recommended to include the following variables: dip angle, flow rate, fractional water cut, and pressure drop. The suggested parameters are also of use when making an assessment of the operating strategy and control of the reservoir.
- In addition to data gathering, some of the current variables should be improved. Firstly, it is suggested to separate the Estimated Ultimate Recovery in: primary recovery, secondary and tertiary recovery factor. This would also allow for a better to benchmark secondary and tertiary methods against conventional field development.
- Furthermore, heterogeneity and fracture intensity are represented by ordinal variables. As a consequence these parameters could not be included in the multivariate analysis. From the analysis it follows that both have a strong impact on oil recovery factor. Therefore it is justified to record both factors as continuous variables.
- The dataset does not record any formation water properties such as water salinity, viscosity or density. For enhanced recovery methods the water salinity is critical as it affects both the injectivity as well as the effect of polymers on the mobility ratio.
- To improve the quality of further analysis weighing of reservoirs based on the uncertainties of the recorded variables is recommended. A categorical variable to express the data quality and confidence level for each reservoir is suggested.
- Acknowledging the effect of maturity on the distribution of estimated ultimate recovery factor it is recommended to record the stage of development.

References

1. API Bulletin D14: A Statistical Study of Recovery Efficiency. API 1967
2. API Bulletin D14: Statistical Analysis of Crude Oil Recovery and Recovery Efficiency. API 1984
3. Ahmed, T. 2010. Reservoir Engineering Handbook. 4th edition. Elsevier Science.
4. Allan, J. 2003. Controls on Recovery Factor in Fractured Reservoirs: Lessons learned from 100 Fractured Fields. SPE 84590.
5. Arps, J.J. 1967. A statistical study of recovery efficiency.
6. Arps, J.J. and Roberts, T.G. "The Effect of Relative Permeability Ratio, the Oil Gravity, and the Solution Gas-Oil Ratio on the Primary Recovery from a Depletion Type Reservoir." AIME (Production Development and Technology) 1955, 204, 120.
7. Arps, J.J. "Reasons for Differences in Recovery Efficiency." SPE 2068.
8. Bai, Y., Li, J. & Zhou, J. 2005. "Sensitivity analysis of dimensionless parameters for physical simulation of water-flooding reservoir." Science in China, Vol.48. No. 4 441.
9. Craft, B.C. & Hawkins, M. Revised by Terry, R.E. 1981. Applied Petroleum Reservoir Engineering. Prentice Hall.
10. Craig Smalley, P. 2009 Reservoir Technical Limits: A framework for Maximizing Recovery from Oil Fields. SPE 109555
11. Craze, R.C. & Buckley, S.E. 1945. "A Factual Analysis of the Effect of Well Spacing on Oil Recovery." API 1945-144
12. Dake, L.P. 1994. The Practice of Reservoir Engineering. Elsevier Science B.V. Amsterdam
13. Van Everdingen, A.F & Kriss, H.S. 1980. "A Proposal to Improve Recovery Efficiency." SPE 9088-PA
14. Falcone, G., Harrison, B. and Teodoriu, C. 2007 Can We Be More Efficient in Oil and Gas Exploitation? A Review of the Shortcomings of Recovery Factor and the Need for an Open Worldwide Production Database. Journal of Physical and Natural Sciences. Volume 1, Issue 2, 2007.
15. Geertsma, J., Croes, G.A., & Schwarz, N. 1956. "Theory of Dimensionally Scaled Models of Petroleum Reservoirs."
16. Gulstad, R.L. 1995 The Determination of Reservoir Recovery Factors by Using Modern Multiple Linear Regression Techniques. Texas Tech university
17. Guthrie, R.K. & Greenberger, M.H. 1955. "The Use of Multiple-correlation of Analyses for Interpreting Petroleum-engineering Data." API 1955-130.
18. Hagoort, J. 1974. "Displacement Stability of Water Drives in Water Wet Connate Water Bearing Reservoirs", SPE pp 63-74, Trans. AIME
19. Kaiser, H. F. 1970. A second generation Little Jiffy. Psychometrika, 35, 401-415.
20. Kumar, M. 2008 High-Mobility-Ratio-Waterflood Performance Prediction: Challenges and New Insights. SPE Reservoir Evaluation & Engineering, February 2008.
21. Laherrère, J. 1997. Distribution and evolution of "recovery factor" via <http://www.hubbertypeak.com/laherrere> on august, 2011.
22. Lake, L.W. 2008 Enhanced Oil Recovery (EOR) Field Data Literature Search.
23. McCain, W.D. 1990. The Properties of Petroleum Fluids. PennWell books
24. Morrow, N. and Buckley, J. 2011 Improved Oil recovery by Low-Salinity Waterflooding. JPT may, 2011.
25. Mullane, J.J. 1944. Reservoir Performance. API Drilling and Production Practice, 1944
26. Nasralla, R. 2011 Coreflood Study of Low Salinity Water Injection in Sandstone Reservoirs.
27. Ostermeier, R.M. 2001. "Compaction Effects on Porosity and Permeability: Deepwater Gulf of Mexico Turbidites". SPE Distinguished Author Series.
28. Qing Sun, S. and Sloan, R. 2003. Quantification of Uncertainty in Recovery Efficiency Predictions: Lessons Learned from 250 Mature Carbonate Fields. SPE 84459.

29. Rijnders, J.P. 1974. Incorporation of Uncertainty in Reservoir Engineering: A Data Management Problem. SPE 4845.
30. Schulte, W.M. 2005. Challenges and Strategy for Increased Oil Recovery. IPTC 10146.
31. Sharma, A. 2010 Classification of oil and gas reservoirs based on recovery factor: a data-mining approach. SPE 130257.
32. Steffensen, R. J.(Editor) 2007 Petroleum Engineering Handbook, Chapter 37: "Solution-Gas-Drive Reservoirs."
33. Surguchev, L. 2011. "Fast Screening and Evaluation of IOR/EOR Methods." Poster presented 73th EAGE conference Vienna 2011.
34. Taber, J.J., Martin, F.D. and Seright, R.S. 1997. "EOR Screening Criteria Revisited-Part 1: Introduction to Screening Criteria and Enhanced Recovery Field Projects. SPE Reservoir Engineering.
35. Thomas, C.E. 2007 Petroleum Engineering Handbook, Chapter 44: "Water-Injection Pressure Maintenance and Waterflood Processes."
36. Tyler, N. and Finley, R.J. 1991 Architectural controls on the recovery of hydrocarbons from sandstone reservoirs
37. Vietti, W.V., Mullane, J.J., Thornton, O.F., and van Everdingen, A.F. "The Relation between Well Spacing and Recovery." API 1945-160.
38. Watkins, G.C. 2002. "Characteristics of North Sea oil reserve appreciation." Massachusetts Institute of Technology. Reprint Series Number 180.
39. Wyckoff, R. 1940. "Factors Affecting Reservoir Performance." API 1940-105.
40. Zinszner, B. and Pellerin, F.M., 2007. A geoscientist's guide to petrophysics. IFP Publications

Appendix 1.

