



University of
Stavanger

FACULTY OF SCIENCE AND TECHNOLOGY

MASTER'S THESIS

Study programme/specialisation:

Geoscience engineering

Spring/ Autumn semester: 2021

Open / Confidential

Author: Miranda Ebsworth Skjeldal

Programme coordinator:

Supervisor(s): Pål Østebø Andersen and Ingrid Carita Augustsson

Title of master's thesis:

Machine Learning techniques for Prediction of Rock Properties from Reservoir Well Logs

Credits:

30

Keywords:

Machine learning, well logging, porosity prediction, oil saturation prediction, water saturation prediction, Artificial Neural Networks (ANN), Levenberg-Marquardt optimisation

Number of pages: 105

+ supplemental material/other: 16

Stavanger, 15. June 2021.

Title of thesis

Machine Learning techniques for
Prediction of Rock Properties from
Reservoir Well Logs

Faculty of Science and Technology

Acknowledgements

I would like to express my sincerest appreciation to my supervisors for the guidance and help they have provided in this study.

I feel grateful to complete my master program.

Abstract

Estimation of reservoir parameters is important in reservoir evaluation and estimation of petroleum volume. Reservoir parameters such as oil saturation, water saturation and porosity are derived from petrophysical logs or time-consuming, expensive core analyses. Not all wells are cored in a field, and the number of fully cored wells and recovery is limited. In this study, a time-efficient and economical method to estimate oil saturation, water saturation and porosity is employed. An artificial neural network (ANN) model, a multilayer feedforward network (MLP) is developed to predict the reservoir parameters. The model is based on 1042 petrophysical log data, oil saturation and water saturation data, and 1697 petrophysical log data and porosity data from five wells in the Varg field, Central North Sea. The MLP model optimised using a backpropagating algorithm, the Levenberg-Marquardt algorithm. In the study, six wells are used in total. The sixth well is excluded from the data set of the calibrated model and employed to validate the performance of the calibrated models. Feature selection is conducted on the petrophysical logs in the study: Gamma-ray, Self-potential, Acoustic, Neutron porosity, bulk density, deep resistivity, and medium resistivity. Feature selection aims to identify the most relevant petrophysical logs and remove those that are considered less relevant. The feature selection is conducted using correlation coefficients and a combination of trial-and-error and a stepwise regression approach. The estimated oil saturation shows a very weak linear (Pearson) and non-linear (Spearman and Distance) correlation (R^2 (Pearson) = 0.1, R^2 (Spearman)=0.17 and R^2 (Distance)=0.11). The estimated water saturation shows weak linear (Pearson) and non-linear (Spearman and Distance) correlation (R^2 (Pearson) = 0.29, R^2 (Spearman)=0.41 and R^2 (Distance)=0.40). The estimated porosity shows moderately strong linear (Pearson) and non-linear (Spearman and Distance) correlation (R^2 (Pearson) = 0.67, R^2 (Spearman)=0.45 and R^2 (Distance)=0.74). The estimated parameters are not fully reliable on data outside the calibrated models data sets range, and the most reliable estimation is the porosity. Increasing the number of wells may increase the data set range and improve the accuracy of the models. For future works, by increasing the number of wells, the models can be employed and tested in other fields.

Table of contents

Acknowledgements	ii
Abstract	iii
Table of contents	iv
List of tables	vi
List of figures	viii
1 Introduction	1
2 Well logs and key parameter	4
2.1 Porosity	4
2.2 Permeability	4
2.3 Fluid saturation	5
2.4 Gamma-ray (GR)	7
2.5 Spontaneous potential log (SP)	8
2.6 Caliper log (CAL)	10
2.7 Sonic log (AC)	11
2.8 Neutron porosity log (NPHI)	12
2.9 Density log (RHOB)	14
2.10 Resistivity logs (RD and RM)	15
3 Machine learning techniques	18
3.1 Application of machine learning techniques	22
3.2 Artificial Neural Network (ANN)	23
3.2.1 Network Architecture	26
3.2.2 Feedforward networks	28
3.2.3 Other neural networks	31
4 Optimisation techniques	32
4.1 Ant colony optimisation (ACO)	32
4.2 Particle swarm optimisation (PSO)	32
4.3 Back-propagation algorithms	33
4.3.1 Gradient Descent (GD) and Stochastic Gradient Descent (SGD)	34
4.3.2 Newton and Gauss-Newton algorithm	35
4.3.3 Levenberg-Marquardt Algorithm	35
4.4 Advantages and disadvantages of the Levenberg-Marquardt algorithm	36
4.4.1 Overfitting	37
5 Methodology	38

5.1	Study area	38
5.2	Stratigraphy	39
5.3	Data set (Petrophysical logs and core data).....	41
5.4	Data preparation	51
5.5	Feature selection	61
5.5.1	Statistical evaluation of performance	63
5.6	MLP-LM model and design	63
5.6.1	Network architecture	67
6	Results	69
6.1	Model validation and calibration	74
6.2	Model performance comparison	78
6.3	Best results.....	85
6.4	Sensitivity analysis of input variables	87
6.4.1	Partial derivatives method (PaD)	87
6.4.2	Variation of input variables.....	91
7	Discussion	95
8	Conclusion.....	97
9	References	99
	APPENDIX A	106
	APPENDIX B	109
	APPENDIX C	110
	APPENDIX D	111
	APPENDIX E.....	113
	APPENDIX F.....	119

List of tables

Table 2.1 Overview of common gamma-ray reference values from Baker et al.,2015	8
Table 2.2 Common sonic travel times and velocities after Tixier et al., 1959.....	12
Table 2.3 Common matrix densities and different lithologies from (Donaldson Erle C & Tiab Djebbar, 2012).....	15
Table 3.1 Simplified illustrations and a brief description of relevant network architectures in this study	28
Table 4.1 The parameters used in the Particle Swarm Optimisation (PSO) algorithm with descriptions of each parameter.	33
Table 5.1 Well information of the selected wells in the study	38
Table 5.2 Stratigraphic column of the units in used in the study.	39
Table 5.3 The depth of the core samples and the corresponding formation in well 15/12-1 ..	41
Table 5.4 The depth of the core samples and the corresponding formation in well 15/12-4 ...	43
Table 5.5 Lithological description of the cores from well 15/12-5.....	45
Table 5.6 Lithological description of the cores from well 15/12-6S	46
Table 5.7 Lithological description of the cores from well 15/12-9S	48
Table 5.8 Lithological description of the cores from well 15/12-A8.....	50
Table 5.9 The errors between the cored depth and the well logging depths	53
Table 5.10 Statistical indexes of the original data used in the model without well 15/12-4....	57
Table 5.11 The statistical indexes of the data used in the training set of the model.....	58
Table 5.12 The statistical indexes of the validation set of the model	58
Table 5.13 The statistical indexes of the testing set of the model.....	59
Table 5.14 The statistical indexes of the well 15/12-4.....	59
Table 5.15 A summarised overview of the parameter selections for the Levenberg-Marquardt	66
Table 5.16 The number of layers in the network and a the vectors of the number of neurons in each layer.....	68
Table 5.17 A summary of the network architecture	68
Table 6.1 The feature selections obtained from the correlation analysis of the oil saturation, water saturation and porosity.	72
Table 6.2 Summary of feature selection from trial-and-error.	73
Table 6.3 Results from stepwise regression	73
Table 6.4 The parameter selection from the stepwise regression analysis.....	74
Table 6.5 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using all petrophysical logs.	75
Table 6.6 The coefficient of determinant R^2 of the training set, validation set, testing set and well 15/12-4, based on the calibrated model attained from the Pearson, Spearman and Distance selection.....	77
Table 6.7 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using the petrophysical logs in Table 6.1 .	79
Table 6.8 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using the petrophysical logs in Table 6.2.	80
Table 6.9 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using the petrophysical logs in Table 6.4 .	83

Table 6.10 summary of the relative importance of each input variable using partial derivatives of from the oil saturation, water saturation and porosity predictions from the training and validation set.	88
Table 6.11 Reference values to evaluate how the model is affected by variations of the input variables.	92
Table 6.12 Summary of the calibrated models performance of porosity on the data set from well 15/12-4.....	92

List of figures

Figure 2.1 A conventional core analysis (RCAL) of well 15/12-4, Varg field (Statoil, 1984). nmp = no measurement possible, nvpp = no vertical plug possible.....	6
Figure 2.2 A special core analysis (SCAL) report from Well 15/12-9S, Varg field (Statoil, 1991).....	7
Figure 2.3 A synthetic log with log readings representing lithologies (Baker et al., 2015).....	9
Figure 2.4 A synthetic log showing common characteristics of the Caliper log and bit size modified from (Baker et al., 2015).....	11
Figure 2.5 A synthetic log of gamma ray readings and neutron porosity readings with common lithologies.	13
Figure 2.6 A neutron density cross plot, used to determine the porosity and lithology of simple lithologies (Schlumberger Educational Services., 1989).....	14
Figure 2.7 Self-potential and resistivity log in a permeable formation and the different measured resistivities (Selley & Sonnenberg, 2015).....	16
Figure 3.1 Illustration of a simple linear model function with one variable: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ (error term).	19
Figure 3.2 An overview of some of the most common machine learning techniques modified from (Sharma & Wang, 2018; Duc et al., 2019)	20
Figure 3.3 Two examples with one prediction (above) and two predictions, classification (below) showing illustrations of underfitting(left), good fitting (middle) and overfitting (right) (Boulesteix et al., 2020).	21
Figure 3.4 (a) Learning curve of the accuracy of the model as a function of samples. (b) Learning curve of an artificial neural network modified from (Sammut & Webb, 2011)	22
Figure 3.5 Model of a simple network with one input and one hidden layer where information is forwarded from left to the right.	24
Figure 3.6 A simplified network structure with multiple inputs, with connected weights (w_{ij} and w_{jk}) from the input layer and hidden layer, respectively.	25
Figure 3.7 A simple illustration of a Sigmoid function.....	26
Figure 3.8 A simple perceptron where the information is forwarded from left to right.	29
Figure 3.9 A schematic illustration of a multilayer perceptron with one input layer, one hidden layer and one output layer. The hidden layer consists of 4 neurons in this illustration.	30
Figure 5.1 Location of the Varg field in the Central North Sea. The black circles are the wells used to calibrate the model, and the red circle is the well the calibrated model is tested on. ..	38
Figure 5.2 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from core samples of Hugin FM and Sleipner FM of well 15/12-1.....	42
Figure 5.3 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-1- The petrophysical logs are taken from core sample of Hugin FM and Sleipner FM of well 15/12-1	42
Figure 5.4 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from core sample of the Heather FM and Hugin FM of well 15/12-1.	44
Figure 5.5 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-4. The petrophysical logs are taken from core sample of Heather FM and Hugin FM of well 15/12-1.....	44

Figure 5.6 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the cored depth of well 15/12-5.	45
Figure 5.7 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-5. The petrophysical logs are taken from the cored depth of well 15/12-5.	46
Figure 5.8 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the cored depth of well 15/12-6S.	47
Figure 5.9 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-6S.	48
Figure 5.10 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the core depth of well 15/12-9S.	49
Figure 5.11 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-9S.	49
Figure 5.12 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the core depth of well 15/12-A 8.	50
Figure 5.13 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-A 8.	51
Figure 5.14 The amount of data points of oil saturation in each wells.	52
Figure 5.15 The amount of data points of water saturation in each wells.	52
Figure 5.16 The amount of data points of porosity in each wells.	53
Figure 5.17 A statistic summary of the outliers in the Gamma-ray (GR), Self-Potential (SP), Caliper (CAL) and Sonic (AC) log in each wells.	55
Figure 5.18 A statistic summary of the outliers in the Neutron porosity (NPHI), Bulk density (RHOB), deep resistivity (RD) and medium resistivity (RM) log in each wells.	56
Figure 5.19 A statistic summary of the outliers in the desired parameters: Oil saturation, water saturation and porosity in each wells.	57
Figure 5.20 The data distribution of the oil saturation, water saturation and porosity from core data.	61
Figure 5.21 Flow chart of the Levenberg-Marquardt parameter optimisation.	65
Figure 5.22 A Pseudo-code of the workflow used to construct the calibrated models.	67
Figure 6.1 Pearson correlation results of the training set and validation set.	70
Figure 6.2 Spearman correlation results of the training set and validation set.	71
Figure 6.3 Distance correlation results of the training set and validation set.	72
Figure 6.4 The results of the calibrated models selecting the models with highest R^2 using Pearson correlation coefficient (Table 6.5).	76
Figure 6.5 The results of the calibrated models selecting the models with highest R^2 using Spearman correlation coefficient (Table 6.5).	76
Figure 6.6 The results of the calibrated models selecting the models with highest R^2 using Distance correlation coefficient (Table 6.5).	77
Figure 6.7 The results of the calibrated models from the feature selection with highest R^2 using Pearson correlation.	79
Figure 6.8 The results of the calibrated models from the feature selection, the models with highest R^2 using Spearman and Distance selection.	80

Figure 6.9 The results of the calibrated models from the second feature selection, using models with highest R^2 based on Pearson	81
Figure 6.10 The results of the calibrated models from the second feature selection, using models with highest R^2 based on Spearman.....	82
Figure 6.11 The results of the calibrated models from the second feature selection, using models with highest R^2 based on Distance.....	82
Figure 6.12 The results of the calibrated models from the feature selection with highest R^2 using Pearson selection.	83
Figure 6.13 The results of the calibrated models from the feature selection, the models with highest R^2 using Spearman selection.	84
Figure 6.14 The results of the calibrated models from the feature selection, the models with highest R^2 using Distance selection.	84
Figure 6.15 Regression plot of the predicted oil saturation and oil saturation from core data.	85
Figure 6.16 Regression plot of the predicted water saturation and water saturation from core data.	86
Figure 6.17 Regression plot of the predicted porosity and porosity from core data.	87
Figure 6.18 Partial derivatives of each input features vs sensitivity of the predicted oil saturation	90
Figure 6.19 Partial derivatives of each input features vs sensitivity of the predicted water saturation	91
Figure 6.20 Partial derivatives of each input features vs sensitivity of the predicted porosity	91

1 Introduction

Understanding the shape and spatial distribution in reservoirs is vital in estimating the petroleum volume in the petroleum industry. The key petrophysical parameters such as porosity, permeability and fluid saturation are crucial for characterising a petroleum reservoir. Porosity, permeability and fluid saturation can be used to estimate the volume of hydrocarbons in the reservoir. Accurate estimation of these properties can be crucial in improving the petroleum recovery, CO₂ sequestration, reducing costs, and optimising the production of a field (Ahmadi & Chen, 2019). Furthermore, evaluating hydrocarbon accumulation, fluid migration and identifying potential pressure seals to reduce drilling hazards also relies on accurate estimations of reservoir parameters (e.g. fluid saturations and porosity) (Helle et al., 2001). In addition, it can also be necessary for the improvement of developing geothermal energy schemes, management of water supplies and radioactive waste storage (Ahmadi & Chen, 2019).

A relationship exists between petrophysical well logging data and approximation of fluid saturations and porosity measurements. Petrophysical logs such as sonic, density and resistivity logs can be used to determine approximate values of fluid saturation (e.g. oil and water) and porosity. Several empirical formulas exist for predicting reservoir parameters such as porosity, permeability and fluid saturation (Wyllie et al., 1958). The permeability is often dependent on the porosity, e.g., through the Carman-Kozeny equation (Helle et al., 2001). However, most petrophysical formulas are only applicable for specific reservoir types, such as sandstones, unconsolidated sands, and homogeneous porous media (Ahmadi & Chen, 2019). The empirical formulas often contain terms and factors that depend on specific lithologies or specific locations (Helle et al., 2001). In addition, the vertical resolution of petrophysical logs is larger than for core data (Ahmadi & Chen, 2019).

Well log data cannot determine porosity directly; this is determined from cores in laboratories. Not all wells are cored, and fully cored recovery is often not achieved. The process can be expensive, time-consuming and it is not a routine operation. Core data from wells are important in constructing representative simulation models, evaluating well locations, and perforation strategies (Stiles & Hutfilz, 1992). Determining porosity is dependent on the number of drilled cores. If there is not enough data, this will affect the results (Newman & Martin, 1977). Having a limited number of cores may increase the

uncertainty related to the prediction of properties. Thus, mathematical models that simulate potential possibilities are used. Machine learning approaches create models that can predict or estimate values using mathematical algorithms. Machine learning techniques can be applied as a quick, cost-effective solution for reservoir evaluation (Ahmadi & Chen, 2019). Many machine learning approaches such as artificial neural networks have been applied to predict reservoir parameters (Huang et al., 1996; Huang & Williamson, 1997; Helle et al., 2001; Helle & Bhatt, 2002; Rwechungura et al., 2011; Mahmoudi & Mahmoudi, 2014; Saputro et al., 2016; Ahmadi & Chen, 2019; Hamada et al., 2020).

Neural networks have been applied to various fields, such as biology, chemistry, and the petroleum industry. Artificial neural networks are superior to other methods in the following after Masters (1993):

1. Data which a conclusion is unclear, irregular, or subject to large errors. In this case, the robustness of the neural network is essential.
2. The patterns in the data are important, and one of the advantages of neural networks is their ability to recognise patterns in the data.
3. The data exhibits significant, unpredictable non-linearity, which is not a problem for neural networks since they are adaptable.
4. The data is noisy, and neural networks are robust with inputs of this type.

These conditions overlap with the features of data commonly used in geoscience and petroleum engineering. Artificial neural networks can be used to predict future values of noisy multivariate data from previous data values (Adamowski et al., 2012).

The study aims to use machine learning techniques to solve a complex non-linear problem and find the best models using statistical approaches. The model aims to predict oil saturation, water saturation and porosity using a series of petrophysical well logging data. Three different models are calibrated with each of the petrophysical parameters from the core data to predict oil saturation, water saturation and porosity. The data set consists of petrophysical logs and core data, and it is divided into a training set, test set and validation set. The model is calibrated with the data set from the training set and tuned with hyperparameters, and lastly tested on the test set. Feature selection of the most relevant petrophysical logs for each estimation is employed to increase the accuracy of the calibrated models.

In addition, the accuracy of the calibrated models is verified using another well outside of the data set used for the calibrated models.

The Varg field, Central North Sea, contains the relevant data (petrophysical logs and core data) for the study. The study tries to match the petrophysical log responses with the information from the core data and predict oil saturation, water saturation and porosity using machine learning techniques. The type of machine learning model selected in this study is an Artificial Neural Network (ANN), specifically a feedforward network or multilayer perceptron network (MLP). All the models are optimised using Levenberg-Marquardt optimisation.

The main objectives are summarised by the following:

- Train a model using the Levenberg-Marquardt algorithms to predict reservoir parameters (oil saturation, water saturation and porosity) from petrophysical data.
- Find the models that give the best predictions on the validation set using statistical evaluation methods such as R^2 (coefficient of determinant).
- Evaluate the influence of the petrophysical logs on the calibrated models through variations of the petrophysical logs and Partial derivatives method (PaD method).
- Compare the predictions from the calibrated model with the data from core analysis.
- Verify the model using another well excluded from the training set and validation used to calibrate the model.

2 Well logs and key parameter

The calibrated models are trained by the petrophysical logs and core data from the Varg field. The petrophysical logs are Gamma-ray, self-potential, Caliper, sonic, neutron porosity, bulk density, and resistivity logs. This study presents a detailed description of these eight petrophysical logs available from the Varg field. In addition, a description of the reservoir parameters from the core data is presented. The petrophysical log in the study were selected on the basis that there exists a relation between the petrophysical logs and reservoir parameters.

2.1 Porosity

Porosity (φ) is defined as the fraction of pore volume (Vp) divided by the total volume, grain and void space, of the rock (Vt) (Eq. 2.1). It is the void space in a porous rock that may be able to store fluids such as hydrocarbons.

$$\varphi = \frac{Vp}{Vt} \quad \text{Eq. 2.1}$$

Porosity can be divided into the total porosity (φt) and the effective porosity (φe). The total porosity is the total pore volume (Vtp) divided by the total rock volume (Vt) while the effective porosity is the interconnected pore volume (Vip) divided by the total rock volume (Vt):

$$\varphi t = \frac{Vtp}{Vt} \quad \text{Eq. 2.2}$$

$$\varphi e = \frac{Vip}{Vt} \quad \text{Eq. 2.3}$$

2.2 Permeability

Permeability is defined as rocks ability for fluids to flow through. It is measured in millidarcies (mD) or darcies (D). Permeable sandstones typically have large pores that are interconnected, while impermeable formations such as shales tend to have smaller pores and less interconnected pores. Darcy's law for fluid flow in permeable rocks can only be used when there is a single fluid or phase present in the rock. The steady-state flow (qi) is defined as the permeability (k), flow area (A) and pressure drop (Δp_i) divided by the fluid viscosity (μ_i) and flow distance (L) of a particular fluid I (i =oil, gas, water) (Eq. 2.4).

$$q_i = \frac{kA\Delta p_i}{\mu_i L} \quad \text{Eq. 2.4}$$

The permeability from the steady-state flow equation is called the absolute permeability. By rearranging the steady-state equation the absolute permeability of a fluid k_i (i =oil, gas or water) is:

$$k_i = \frac{A\Delta p_i}{q_i \mu_i L} \quad \text{Eq. 2.5}$$

If there are more than one single fluid another equation for the permeability applies, relative permeability (k_{ri}). The relative permeability is the ability for another fluid to enter the rock when there are multiple fluids present. The relative permeability (k_{ri}) is the fraction of effective permeability (k_i) of a particular fluid (i =oil, gas or water) divided by the fully saturated permeability (k):

$$k_{ri} = \frac{k_i}{k} \quad \text{Eq. 2.6}$$

2.3 Fluid saturation

An inverse relationship between permeability and water saturation exists, such that when the water saturation increases the permeability decreases.

The hydrocarbon pore volume is the total volume in the reservoir that is filled with hydrocarbons:

$$HCPV = V_t \phi (1 - S_{w_{irr}}) \quad \text{Eq. 2.7}$$

Where S_w is the irreducible (connate) water saturation expressed as a fraction of the pore space in the rock. The saturation in the rock volume is the fraction of the interconnected void space occupied by a given phase such as oil, gas and water.

$$S_w + S_o + S_g = 1 \quad \text{Eq. 2.8}$$

Porosity can be obtained either directly from cores or indirectly from well logging tools, while the permeability is obtained from either well testing or core analysis in laboratories. Core data defines the basis for determining permeability distribution, a good understanding of the permeability distribution is critical for planning a successful secondary oil recovery by water-flooding of a stratified interval (Pyle & Sherborne, 1939; Stiles & Hutfilz, 1992). It is important


to keep in mind that usually not all wells that are drilled in a field are cored, and even in cored wells a full recovery of the core is not achieved. The porosity, permeability and saturations of the formations can be obtained from cores through Special core analyses (SCAL) or routine core analyses (RCAL) in laboratories.

A summary of the measured petrophysical properties obtained from the analyses:

i. Routine core analysis (RCAL)

The routine core analysis or conventional core analysis measures basic petrophysical properties of the core sample (reservoir formations or intervals of interest). The routine core analysis is widely used, and it is cheaper than the special core analysis. The basic properties obtained in this analysis is the: grain density, rock dimensions, porosity, gas permeability (absolute permeability), Klinkenberg permeability and water saturation (Figure 2.1). The limitations of the RCAL arise from the laboratory procedure to obtain porosity and permeability. The porosity and permeability are obtained using gases on preserved clean, dried core samples at room conditions, not at reservoir conditions (Stiles & Hutfilz, 1992).

The routine special core analysis procedure is used to establish a relationship between the porosity and permeability. The routine core data and the porosity data calculated from well logs can be used to estimate the permeabilities in the wells without cores. If there are large variations between the porosity and permeability data for the same formations in different wells suggests that a routine core is not appropriated for the selected reservoir (Stiles & Hutfilz, 1992)

COMPANY :		STATOIL		FINAL REPORT		PAGE: 1			
WELL :		15/12-4		CORE NO.: 1		DATE: NOVEMBER 1984			
FIELD :		15/12							
STATE :		NORWAY							

Plug No.	Depth (meter)	Permeability (mD),				Porosity (%)		Pore saturation		Grain dens. g/cc	Formation Description
		horizontal Ka	horizontal K1	vertical Ka	vertical K1	He	Sum.	So	Sw		
	2439.00	KAH	KLH	KAV	KLK	FOR				GREENS	
12	2439.00	0.091	0.067	nvpp		26.0				2.65	Mrl-st.Lt-gry.Consol
13	2439.90	0.075	0.055	nmp		27.8				2.64	A.A.
14	2440.10	0.090	0.067	nmp		28.3	24.5	0	95.1	2.66	A.A.
15	2440.40	npp									
16	2440.70	0.20	0.15	0.28	0.21	15.8				3.02	Calc.sltst.Lt-brn.Consol.gry.mott.w/Sld
17	2441.05	nmp		nvpp		29.8	24.8	0	93.7	2.65	Mrl-st.Lt-gry.Consol.fis.
18	2441.40	nmp		nvpp		25.1				2.68	A.A.Br-n-mott.
19	2441.70	0.11	0.081	nvpp		22.7				2.67	A.A.w/o Brn-mott.fis.
20	2442.05	0.044	0.032	0.004	<0.01	25.5	24.6	0	92.1	2.64	A.A.
21	2442.50	0.053	0.039	0.006	<0.01	27.9				2.65	A.A.
22	2442.85	nmp		0.012	<0.01	21.9				2.66	A.A.fis.
23	2443.20	nmp		0.006	<0.01	nmp	24.7	0	94.8		A.A.

Figure 2.1 A conventional core analysis (RCAL) of well 15/12-4, Varg field (Statoil, 1984). nmp = no measurement possible, nvpp = no vertical plug possible.

ii. Special core analysis (SCAL)

The special core analysis is a more advanced test on the core samples compared to RCAL. The procedure includes flow experiments on core samples of two fluids or two-phases, obtaining capillary pressure, relative permeability, Archie factors and wettability. A potential problem in the SCAL studies is the use of oil based drilling mud in deviated wells from the North Sea, which may alter the wettability of the reservoir rock (Stiles & Hutfilz, 1992). The cores used for the SCAL is preserved using either wax coat, samples are kept under simulated brine or in deoxygenated formation brine. A special core analysis on 16 samples from well 15/12-6S is shown in Figure 2.2. The brine permeability was measured with simulated formation water.

Sample no.:	Depth (m)	Φ (frac.)	GD (g/cm^3)	Ka (mD)	KI (mD)	Corr. coeff.	Kw (mD)	Kw/KI
148.11	2875.10	0.111	2.72	5.69	5.00	0.966	3.26	0.65
153.11	2876.33	0.256	2.68	741	727	0.917	529	0.73
166.11	2879.55	0.101	2.76	1.22	0.94	0.752	-	-
170.11	2880.55	0.290	2.65	1001	967	0.990	702	0.73
189.21	2885.37	0.314	2.65	526	503	0.997	490	0.97
243.11	2899.11	0.140	2.69	0.37	0.19	0.647	< 0.1	-
255.11	2902.12	0.240	2.67	5.21	4.08	0.997	2.98	0.73
260.11	2903.33	0.260	2.64	14.1	11.4	1.000	10.2	0.89
266.11	2904.86	0.203	2.66	1.30	0.90	0.978	0.48	0.53
269.11	2905.59	0.232	2.62	4.41	3.42	0.997	2.54	0.74
273.21	2906.68	0.207	2.64	0.99	0.67	0.992	0.29	0.43
278.11	2910.35	0.251	2.66	6.40	4.97	0.999	3.57	0.72
289.21	2913.19	0.170	2.65	0.55	0.34	0.996	0.12	0.35
339.11	2926.10	0.257	2.66	39.4	34.2	1.000	30.3	0.89
352.21	2929.41	0.213	2.65	168	157	1.000	150	0.96
367.21	2933.17	0.209	2.64	2.42	1.76	0.992	1.04	0.59

Figure 2.2 A special core analysis (SCAL) report from Well 15/12-9S, Varg field (Statoil, 1991).

2.4 Gamma-ray (GR)

The gamma-ray log is one of the most common well logging tools (REF). The well logging tool measures the natural radiation in the penetrated formations. In shales there are three naturally occurring elements that can be easily detected by the gamma ray log, such as potassium, thorium, and uranium (Donaldson Erle C & Tiab Djebbar, 2012).

A clean sandstone has low shale content and will produce a lower gamma-ray reading compared to a shale layer. However, if there are heavy mineral present in the sandstone the gamma-ray readings may be unreliable. The presence of potassium rich feldspars and heavy minerals such as micas, glauconite and uranium rich waters may produce a higher gamma-ray log reading (Asquith & Krygowski, 2004). If there are heavy minerals present in the formation a spectral gamma-ray log is the preferred option.

The gamma-ray readings can be useful to identify lithologies, shale volume calculations and geological assessments (continuity between wells). Typical GR log readings and the corresponding lithology are listed in Table 2.1.

Table 2.1 Overview of common gamma-ray reference values from Baker et al.,2015

Lithology	API°
Shale	80-140
Sandstone	15-30
Dolomite	8-15
Limestone	10-20
Gypsum	5-10
Coal	5-10
Salt	5-10
Anhydrite	~ 15

2.5 Spontaneous potential log (SP)

The spontaneous potential log or self-potential log (SP) is a useful tool to identify permeable rocks, the shaliness of the interval, the salinity of the formation. The SP log can be used to determine the formation-water resistivity (R_w).

The SP log measures the electrical potential between an electrode in the borehole and a reference electrode at surface. The electrical potential is the sum of the contacts between formations, contact between different fluids and fluid flow in the formation. The electrical potential between the formations is called the membrane potential. Secondly, the electrical potential between the contacts with different fluids is called the fluid junction potential. Lastly, the electrical potential of the fluid flow is the electrokinetic potential (Baker et al., 2015).

i. Membrane potential

Shales mainly consists of clays that are permeable to positive ions (Na^+) and impermeable to negative ions (Cl^-). The positive ions move from the saline fluids (e.g saline formation water) towards less saline fluids (e.g less saline mud) (Baker et al., 2015). In the presence of nonconductive mud, the SP log readings are not recorded, the tool requires conductive borehole fluids e.g water-based mud (Glover, 2000).

ii. Fluid junction potential

If the mud filtrate is in contact with the formation water it creates an electrical potential. The negative ions (Cl-) are more mobile than the positive ions and generated a negatively charged current.

iii. Electrokinetic potential

In a permeable rock the flow of an electrolyte can generate an electrokinetic potential and current. The electrokinetic potential occurs in the mud-cake and in the formation.

The SP log is shown with negative and positive log readings between -10 and 10 millivolts [mV] (Figure 2.3). When a shale baseline is defined using the GR log as a guide, the deflections to positive or negative current can be identified. Information about the salinity of resistivity of the formation water and the mud filtrate can be used to indicate the salinity of the formation water. (Glover, 2000). Factors that affects the SP log readings is the baseline shift, high-resistivity formations, and invasion effects (Baker et al., 2015). In many offshore wells the use of saltwater-based drilling muds results in ineffective estimation of the resistivity of the formation water (R_w) and the SP log can not be used in cased holes (Selley & Sonnenberg, 2015)

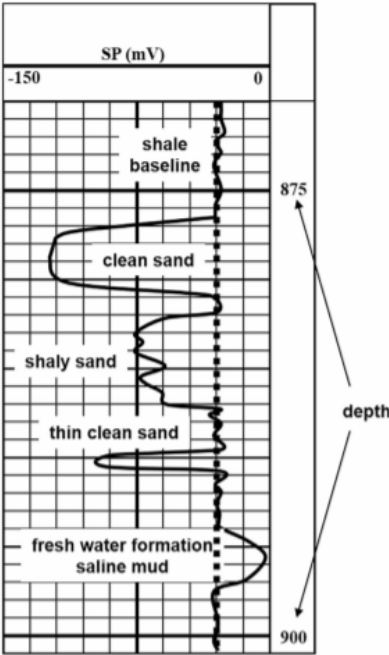


Figure 2.3 A synthetic log with log readings representing lithologies (Baker et al., 2015)

In addition, to use the SP log readings quantitatively a value of the total potential drop must be derived. The value can be derived directly from the SP log deflections or indirectly from correction charts (Glover, 2000). It is called the static spontaneous potential (SSP) and it is

obtained from clean sandstones if no current is flowing, however in shaly formations after correction of layer thickness it is called the pseudo-static spontaneous potential (PSP).

2.6 Caliper log (CAL)

The Caliper logging tool measures the changes in the wellbore diameter and shape. The tool has several spring-loaded arms that can detect electrical signals when they are released or withdrawn. The well bore diameter and the well bore shape may change during drilling when different lithologies are penetrated. In addition, the occurrence of mud cake and caving along the wellbore can be identified from the Caliper log. A simple illustration of an openhole logging in a vertical well with some key elements that can affect the quality of the petrophysical data is shown in Figure 2.4. The accuracy of the petrophysical data is affected by the presence of mudcake, invasion or caving (sloughing) (Baker et al., 2015). The Caliper log reading and bit size shows different scenarios that can be summarized into the following:

- i. Caliper reading (wellbore diameter) smaller than bit size

If the wellbore diameter is smaller than the bit size this may indicate the presence of mudcake in the formation. The most probable lithology at this depth is a permeable rock layer e.g. permeable sandstone, carbonate, or swelling shales (Glover, 2000).

- ii. Caliper reading (wellbore diameter) equal to the bit size

If the wellbore diameter is equal to the bit size it may indicate that the tool is going through a well consolidate rock layer. Possible lithologies are well consolidated sandstones or tight lithologies e.g thick sandstone layers, calcareous shales, igneous and metamorphic rocks (Glover, 2000). When there is a match between the bit size and the Caliper log, the other log readings are considered reliable.

- iii. Caliper reading (wellbore diameter) larger than the bit size

If the well bore diameter is larger than the bit size the tool is penetrating a formation that is possible soft or unconsolidated. During this drilling interval caving or sloughing may occur. Suggested lithologies when this occurs e.g. unconsolidated sands, brittle shales or salt formations drilled with freshwater drilling mud (Glover, 2000).

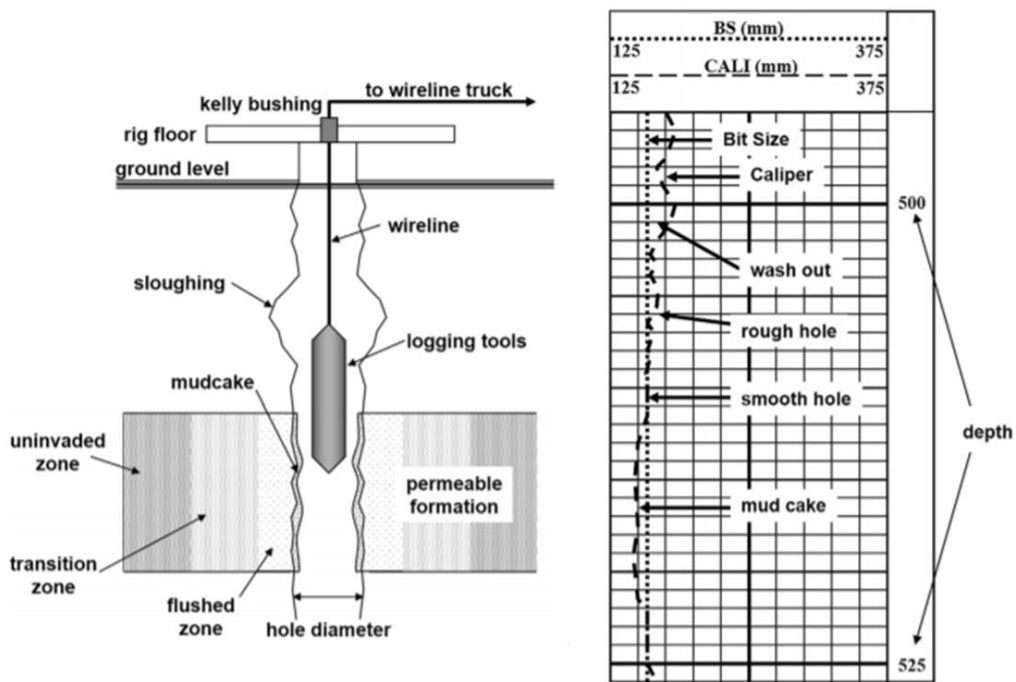


Figure 2.4 A synthetic log showing common characteristics of the Caliper log and bit size modified from (Baker et al., 2015).

Using the Caliper log in combination with the bit size measured in diameter can be used to calculate the mud-cake thickness (h_m). The mud-cake is used as an indicator of permeability, which is typically only present in permeable rocks. The size of the mud-cake can be used to delimit the potential reservoir.

$$h_m = \frac{\text{Bit size (diam)} - \text{Caliper log readings (diam)}}{2} \quad \text{Eq. 2.9}$$

2.7 Sonic log (AC)

The sonic log, also called acoustic log measures the travel time or transit time of sound waves through the formations. The tool has two transmitters and two or four receivers, where the averaged travel time for the transmitters are used to compensate for borehole quality (size of the borehole) (Baker et al., 2015). The sonic log is a useful tool to link the petrophysical data and the seismic data (calibrate), it can also be used to derive the porosity of the formation. Some common lithologies and the corresponding sonic travel times and velocities is presented in Table 2.2.

Table 2.2 Common sonic travel times and velocities after Tixier et al., 1959

Lithology	T_{ma} μ /ft.	V ft. /s
Oil	232	4.3
Water	189-200	5.3-5
Sandstones	55.6	≤ 18
Dolomite	42	24
Anhydrite	50	20
Carbonates	43.5-47.6	23-21
Shales	62.5-167	16-6
Salt	15	66.7

The sonic log can be used to calculate the porosity of the formations, however porosity derived from density and neutron porosity logs are superior. The velocity of the sound waves through a given lithology can be described as a function of porosity. Wyllie's equation or the time average equation is only applicable in homogeneous rocks (Wyllie et al., 1958).

$$\varphi_{sonic} = \frac{\Delta t_{log} - \Delta t_{matrix}}{\Delta t_{fluid} - \Delta t_{matrix}} \quad \text{Eq. 2.10}$$

where Δt_{log} is the transit travel time observed in the log readings, and Δt_{matrix} is the corresponding travel time through the rock matrix. The Δt_{fluid} is the transit travel time through fluids in the formation if it is present.

2.8 Neutron porosity log (NPHI)

The neutron logging tool have a neutron source that emits neutrons, which are absorbed by nuclei of atoms in the rock formation and the wellbore. The neutrons are affected by the presence of hydrogen atoms in the formation such as hydrogen atoms in e.g. water or hydrocarbons. The measured neutrons will vary with the pore space (porosity) in the rock formation. High porosity and low porosity results in low and high measurements of neutrons, respectively (Donaldson Erle C & Tiab Djebbar, 2012). The resolution of the neutron porosity log is less than the density log (approximately 3 feet or 0.94 m) (Baker et al., 2015). The neutron log is affected by rare earth elements (REE), chlorine and boron, which are often present in shales. The presences of these elements lead to overestimation of the porosity in formations interbedded with shales (Figure 2.5).

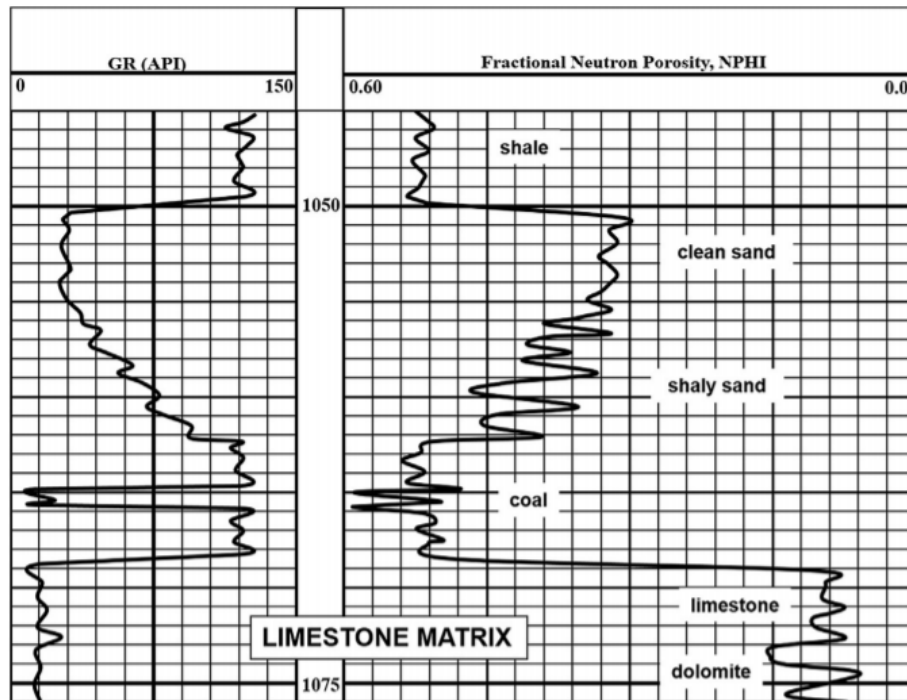


Figure 2.5 A synthetic log of gamma ray readings and neutron porosity readings with common lithologies.

The neutron porosity logs can be used to identify gas zones and fluid contacts. If the neutron porosity log and density log is combined, it provides a clear indication of fluid contacts (Donaldson Erle C & Tiab Djebbar, 2012). The neutron density log is affected by the hydrogen content, in oil and water the hydrogen content is similar while the hydrogen content in gas is much lower. The effect of the presence of gas in the formation results in lower amounts of neutron and underestimated porosity (Baker et al., 2015).

The neutron log is usually calibrated to a limestone matrix or sandstone matrix, such as the density log and also requires correction for all other rock types (Baker et al., 2015) (Figure 2.6).

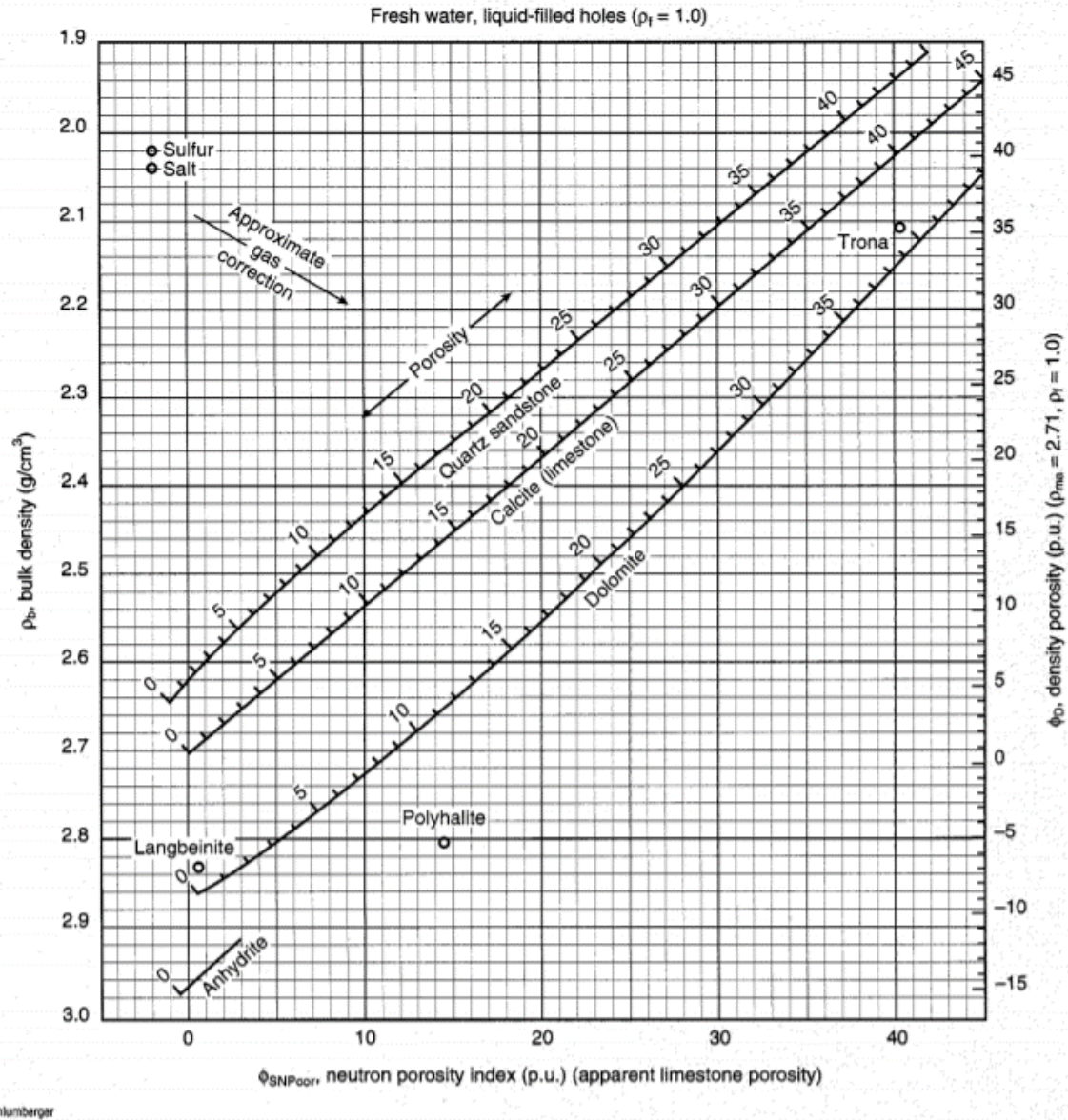


Figure 2.6 A neutron density cross plot, used to determine the porosity and lithology of simple lithologies (Schlumberger Educational Services, 1989)

2.9 Density log (RHOB)

The density log measures the bulk density of the formations using a radioactive source and detectors. The gamma rays enter the formation and the logging tool measured the reduced gamma ray count due to Compton scattering and photoelectric absorption (Glover, 2000). The density logging tool measures the matrix bulk density with corrections for mud cake thickness and irregularities in the wellbore (Donaldson Erle C & Tiab Djebbar, 2012). The total density or bulk density (ρ_{bulk}) of a formation is the average densities of matrix (ρ_{matrix}) and fluid (ρ_{fluid}) in the pores. The porosity in the formation is affected by the presences of shale, which

have a density range that varies depending on the clay minerals present in the shale (Donaldson Erle C & Tiab Djebbar, 2012). In addition, it can be used to identify evaporites and gas-bearing formations.

The density log is used to estimate the effective porosity, the effective porosity derived from the density log can be obtained by the following:

$$\varphi_{density} = \frac{\rho_{matrix} - \rho_{bulk}}{\rho_{matrix} - \rho_{fluid}} \quad \text{Eq. 2.11}$$

The fluid density is normally assumed to be 1.0 gm/cc in wells with water-based mud and 1.1 gm/cc for saline muds (Alger & Raymer, 1963; Glover, 2000). This may lead to incorrect porosities if there are gas, oil or no mud invasion. The fluid density in an uninvaded zone of oil and gas is typically 0.9 gm/cc and 0.7 gm/cc, respectively. If there is salt-water in the formation the density readings increase (Alger & Raymer, 1963). The typical matrix density is limestone density (2.71 g/cm³) or sandstone matrix density (2.65 g/cm³) (Table 2.3). If a log is based on a specific matrix density (e.g limestone or sandstone) it gives the correct measured porosity value of the specified matrix filled with fresh water. The porosity require corrections for all other lithologies. (Baker et al., 2015)

Table 2.3 Common matrix densities and different lithologies from (Donaldson Erle C & Tiab Djebbar, 2012)

Lithologies	ρ_{matrix} g/cm³
Carbonate(limestone)	2.71
Calcareous sand	2.69
Consolidated sand	2.65
Unconsolidated sand	2.60
Shaly sand	2.6
Sand	2.2-2.85

2.10 Resistivity logs (RD and RM)

The three main ways of measuring the electrical resistivity of formations penetrated by the wellbore: conventional log, laterolog and induction log. The conventional resistivity log measures the electric potential and flow of the current between a transmitter and a receiver at the surface. For low resistivity or salty muds laterologs are generally used, while the induction log is generally used for freshwater or oil-based muds with low resistivity (Selley &

Sonnenberg, 2015). The electrical resistivity measured from the formations varies greatly. Rocks filled with hydrocarbons and solid rocks are highly resistive, while shales and permeable rocks have low resistivities (Figure 2.7).

The electrical resistivity along a wellbore measures the resistivity of the invaded (R_{xo}) and uninvaded zone (R_t) (Figure 2.7). The invaded or flushed zone is where the mud cake squeezes into the formation and original pore fluid is displaced. The responses of R_t is measured in deep resistivity logs, while the resistivity of the flushed zone (R_{xo}) and resistivity of the mud filtrate (R_{mf}) is measured by medium and shallow resistivity logs, respectively. The shallow and medium resistivity logs is used in the invaded zones (Baker et al., 2015). The shallow resistivity log measures the resistivity of the flushed zones, and the medium resistivity logs measures the invaded zones (flushed zone and transition zone). Using the resistivity log and SP log combined allows for a qualitative interpretation of lithology and fluids in the formation. The resistivity log measures the resistivity of the formation which depends on the porosity, water saturation and the salinity (Baker et al., 2015).

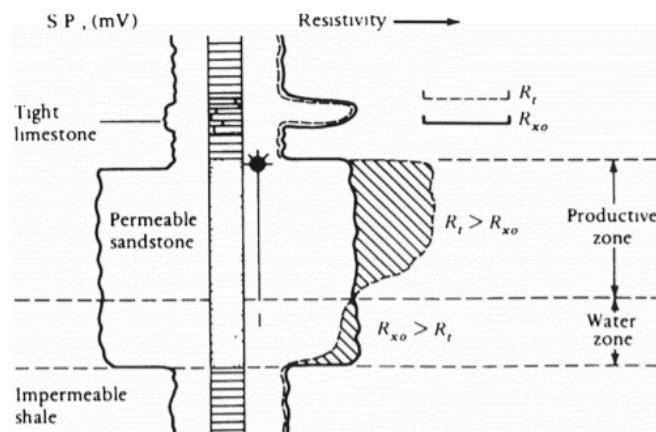


Figure 2.7 Self-potential and resistivity log in a permeable formation and the different measured resistivities (Selley & Sonnenberg, 2015).

The resistivity logs are used to determine the water saturations in the formation using Archie's law (Archie, 1942). The general relationship is given by the formation factor resistivity (F) (Eq. 2.12).

$$\left(\frac{R_o}{R_w}\right) = F = \left(\frac{a}{\phi^m}\right) \quad \text{Eq. 2.12}$$

Where R_o is the resistivity of the rock at 100% saturated water resistivity, R_w is the water resistivity, a is a constant and m is the cementation factor. The water resistivity can be

determined from the SP log, a geographical constant R_w , from formation water samples or measure resistivity and porosity of the water zone assuming 100% water saturation. There exist several values for the cementation factor and the constant. In a flushed zone the equation for the formation resistivity is:

$$F = \left(\frac{R_{xo}}{R_{mf}} \right) \quad \text{Eq. 2.13}$$

In general the Humble formula for soft formations (e.g. sandstone) is used (Donaldson Erle C & Tiab Djebbar, 2012; Selley & Sonnenberg, 2015). The constant a is 0.62 and the cementation factor is 2.15 for the Humble formula, given by the following:

$$F = \left(\frac{0.62}{\varphi^{2.15}} \right) \quad \text{Eq. 2.14}$$

Then the saturation in the uninvaded zone (Eq. 2.15) and in the flushed zone (Eq. 2.16) yields:

$$S_w = \left(\frac{FR_w}{R_t} \right)^{\frac{1}{2}} \quad \text{Eq. 2.15}$$

$$S_{xo} = \left(\frac{FR_{mf}}{R_{xo}} \right)^{\frac{1}{2}} \quad \text{Eq. 2.16}$$

If the mud invasion is too deep the water saturation may give an incorrect water saturation (Baker et al., 2015).

3 Machine learning techniques

Machine learning can be defined as a computational method that uses experience to improve performance or make accurate predictions (Mohri et al., 2018). The experience that is referred to is past information available to the learner. The success of the predictions relies on the quality and size of the data in the training sets (Mohri et al., 2018).

In the learning stage of machine learning algorithms, there are examples used as a guide, for instance, the data used for learning and evaluation. The data set is divided into the training set, validation set and test set. The training set is used to train the learning algorithm, while the validation set is used to tune the parameters of the learning algorithm, and lastly, the test set is used to evaluate the performance of the learning algorithm (Mohri et al., 2018). The data set used in the training stage consists of input data (X) and the desired outputs, target data (T), where the network aims to describe the relationship between the input data and the target data.

The main aim of machine learning is to construct a model that learns a function to describe a certain pattern of a dataset where another unknown pattern exists. The model learns prediction rules to predict data where the desired output (T) is unknown (Boulesteix et al., 2020). The models in Artificial Neural Networks (ANN) consists of computational stages and an algorithm that minimizes the error value of the model. In addition, there are a series of free parameters such as hyper-parameters that can be obtained to make the machine learning model the best description of the selected data set. A hyperparameter is a free parameter that is user-defined and needs to be manually adjusted rather than from the learning stage, such as input weighting or bias value etc.

In machine learning, there are two main types of learning: active learning and passive learning. Active learning is used to describe a learning problem where the data set is somewhat controlled by the user. On the other hand, passive learning has a training set that is not controlled by the user. Active learning is used when obtaining target data for the data set is expensive or time-consuming.

The standard machine learning model is either classification (prediction) from data or regression, which is prediction of a continuous variable (Nichols et al., 2019). Linear regression is the simplest form of machine learning, assuming a linear function for the model the slope (β_1) and intercept (β_0) is the training of the model (Figure 3.1).

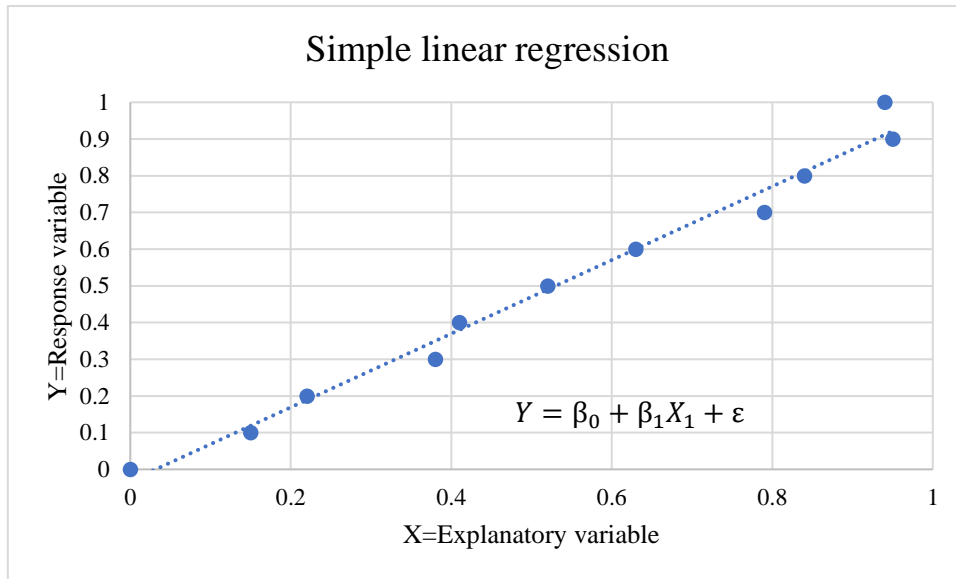


Figure 3.1 Illustration of a simple linear model function with one variable: $Y = \beta_0 + \beta_1 X_1 + \varepsilon$ (error term).

The machine learning process can be divided into supervised, unsupervised, and reinforced or semi-supervised learning. These machine learning techniques are used to solve different problems (Figure 3.2). The supervised learning method uses a method that trains the network by using a vast amount of target data. The supervised machine learning process mainly used for classification and regression since the output can be divided into categories or classes. The contrast to supervised learning is unsupervised learning that aims to recognise patterns in the data set. Unsupervised learning processes are typically self-organising maps, association rules and clustering. The last machine learning process is the reinforced or semi-supervised learning, which is like supervised learning. It uses sequential decision making and does not have examples such as target data.

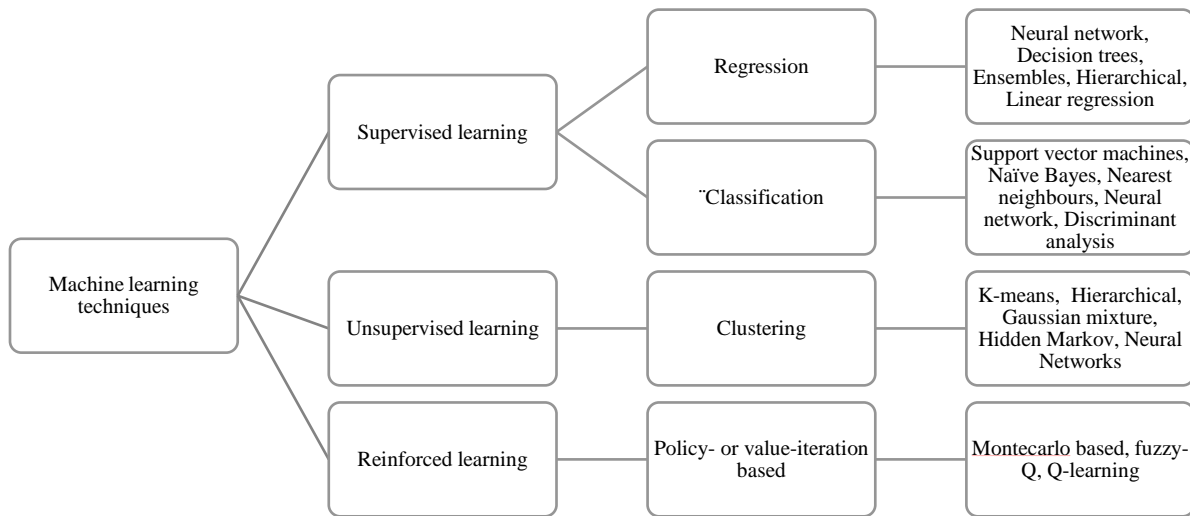


Figure 3.2 An overview of some of the most common machine learning techniques modified from (Sharma & Wang, 2018; Duc et al., 2019)

Prediction of rock properties from petrophysical logs such as porosity or permeability is a supervised process. The model aims to predict the rock properties using the petrophysical logs and desired output values or target values, such as core data.

The main problems for the supervised learning method are overfitting and underfitting of the data (Müller & Guido, 2016). Underfitting occurs when the supervised model fails to capture the patterns of most of the data in the training set. The constructed model is too simple to capture the variation in the data. On the other hand, overfitting is when the supervised model is particularly fitting to a set of data rather than capturing the pattern of the remaining training set (Figure 3.3). The overfitted model is unable to be used for new data, and the fit is too good to be realistic. An appropriate fit is a fit that lies between overfitting and underfitting, and a good fit of the supervised model is the main aim.

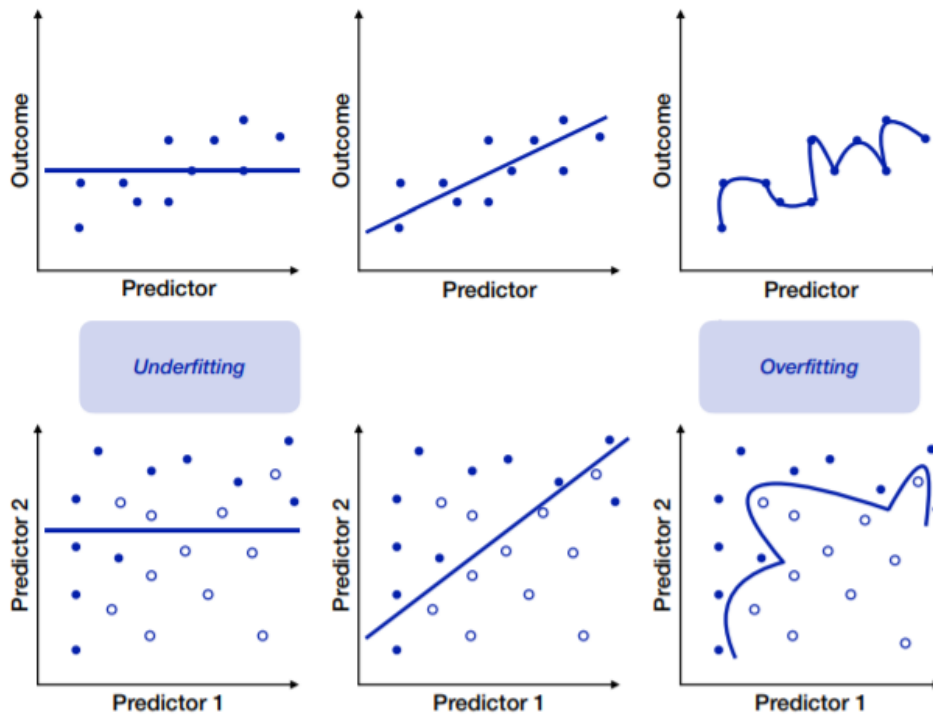


Figure 3.3 Two examples with one prediction (above) and two predictions, classification (below) showing illustrations of underfitting(left), good fitting (middle) and overfitting (right) (Boulesteix et al., 2020).

The performance of the machine learning model over time when it is learning is important for the fitting of the data. The machine learning model learns over a certain period and tries to reduce the error of the model on the training set. The learning curves in machine learning is used in two different contexts: the accuracy of the predictions for a certain number of training examples and the inaccuracy of the predictions over a certain number of training iterations (Sammut & Webb, 2011). In the learning curve, the error in the prediction decreases proportionally with the errors in the training set. If the model using the training set is overly trained, the performance on the training set may continuously decrease, and result in overfitting. If the model is overfitting the model's ability to generalise decreases and the error in the training set increases. The iteration where the generalisation and training error initially have decreased and reached a minimum prior to the generalisation error increasing is the optimal iteration for a good fit of the model (Figure 3.4).

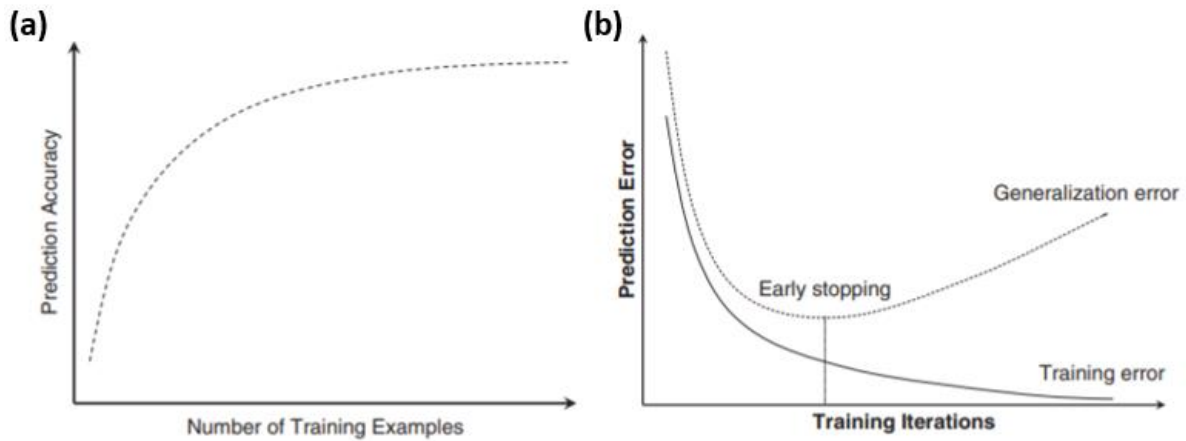


Figure 3.4 (a) Learning curve of the accuracy of the model as a function of samples. (b) Learning curve of an artificial neural network modified from (Sammut & Webb, 2011)

There are two types of machine learning, shallow and deep learning. Shallow learning is based on user-defined information requiring additional information to aid in the learning process. Deep learning has flexible architectures that can learn directly from raw data, such that the accuracy of the prediction increases with the number of data samples (Motamedie, 2020). Some examples of shallow learning are support vector machines (SVMs) or decision trees. The examples of deep learning are multilayer feed-forward neural networks or recurrent neural networks.

3.1 Application of machine learning techniques

Machine learning techniques have been utilised to model progression and the treatment of cancerous patients (Kourou et al., 2015). Machine learning techniques have also been proven successful in clustering, classification or regression (Schmidt et al., 2019). Self-driving cars are also a result of machine learning techniques (Bojarski et al., 2016). Image classification using machine learning techniques have been applied and as a result have resulted in image and speech recognition, web-searches and email or spam filtering (Kourou et al., 2015). Using machine learning to optimise existing water supply systems, such as forecasting urban water demand (Adamowski et al., 2012)

In the petroleum industry various machine learning techniques have been applied to predict porosity, permeability, and fluid saturations. Neural networks have been applied to predict porosity and/or permeability (e.g. Huang et al., 1996; Huang & Williamson, 1997; Helle et al., 2001; Rwechungura et al., 2011; Saputro et al., 2016; Ahmadi & Chen, 2019). In addition, there

are studies where neural networks are applied to predict porosity and/or fluid saturation (Helle & Bhatt, 2002; Mahmoudi & Mahmoudi, 2014; Hamada et al., 2020).

3.2 Artificial Neural Network (ANN)

The first proposed computational model using neurons as architecture was in 1943 (McCulloch & Pitts, 1943). The findings compared the model parameter values to determine the output, thus the model was not able to learn since some of the model parameters (weights) were fixed (YEGNANARAYANA, 2009). The introduction of computers allowed to develop and test artificial neurological networks (Vandeginste et al., 1998). The first artificial neural networks on computers were the perceptron (Rosenblatt, 1957, 1958) and the Adaptive linear element (Adaline) (Widrow & Hoff, 1960) (Vandeginste et al., 1998). In a paper by Minsky & Papert, 1969 it was proven that a network with two layers is incapable of representing or approximate functions outside a narrow range. Eventually, the first published article on neural network application was published in 1989 using a neural network where the input data propagates through the network in a forward direction from input to output (feedforward neural network) (Hornik et al., 1989). The main difference between the networks is mainly the learning rules or functions. Learning strategies such as back propagation was proposed originally in 1974 and rediscovered in 1986 with a new technique and a clear framework (Werbos & John, 1974; Rumelhart et al., 1986).

Artificial Neural Networks (ANN) are computational models that are constructed based on the structure of biological neural networks, like a human brain (Sammut & Webb, 2011). The computational model can be defined as a network of simple processing neurons that can perform simple numerical manipulations (Adamowski et al., 2012). The data-driven process with ANNs using mathematical algorithms can solve complex problems that are non-linear using the relationship between the input data and output data (Adamowski et al., 2012). In most cases ANNs are adaptive systems that changes structured based on external or internal information used in the training phase of the network (Sammut & Webb, 2011).

The structure of an artificial network consists of three key elements: Input layer, hidden layers, and output (Wang, 2003). A simple illustration of a single neural network with one input(x), hidden layer and output (y) (Figure 3.5). In addition, the connections between the input layer and hidden layer are represented by input weights and the connections between hidden layer and the output layer are layer weights. If there are multiple hidden layers the connections have

several layer weights. Other elements are bias (b) and the activation or transfer function of the hidden layer ($f(y)$) (Anastassiou, 2011).

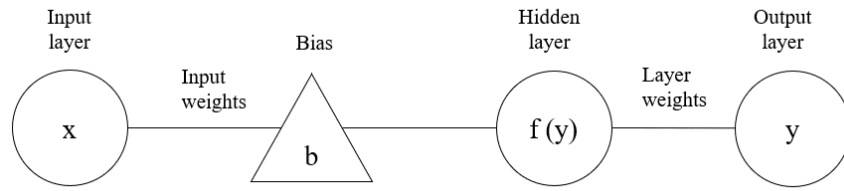


Figure 3.5 Model of a simple network with one input and one hidden layer where information is forwarded from left to the right.

The artificial neural network can be expressed mathematically by the following equations in this chapter.

The function of a single network with one input neuron can mathematically be expressed by:

$$y = f(y) = f(wx + b) \quad \text{Eq. 3.1}$$

Where the predicted output (y) is obtained from the activation or training function ($f(y)$) calculated from the input (x), weight (w) and bias (b). A network can consist of several neurons that connect one to several hidden layers between the input neurons and output neurons. When the number of nodes in the layers increases or varies the network architecture changes. In a network with multiple inputs the output (y_k) is the weighted sum of all inputs, and biases. A simple example with multiple inputs (x_i) and one hidden layer j is shown in Figure 3.6. The illustration is a special case with simple connections between the neurons in the input layer and hidden layer, usually, the connection from input neurons are connected to every neuron in the hidden layer (see section 3.2.1, Table 3.1). The multiple inputs ($x_1, x_2 \dots x_i$), are connected to weights ($w_{1,1}, w_{1,2} \dots w_{i,j}$) from the input layer i to a hidden layer j and from the hidden layer to the output layer ($w_{2,1}, w_{2,2} \dots w_{j,k}$). The weights in a network can be divided into the input weights ($w_{i,j}$) and the layer weights ($w_{j,k}$). The number of biases in the network are the sum of the number of nodes in the input (b_i), hidden layers (b_j) and the output layer (b_k)

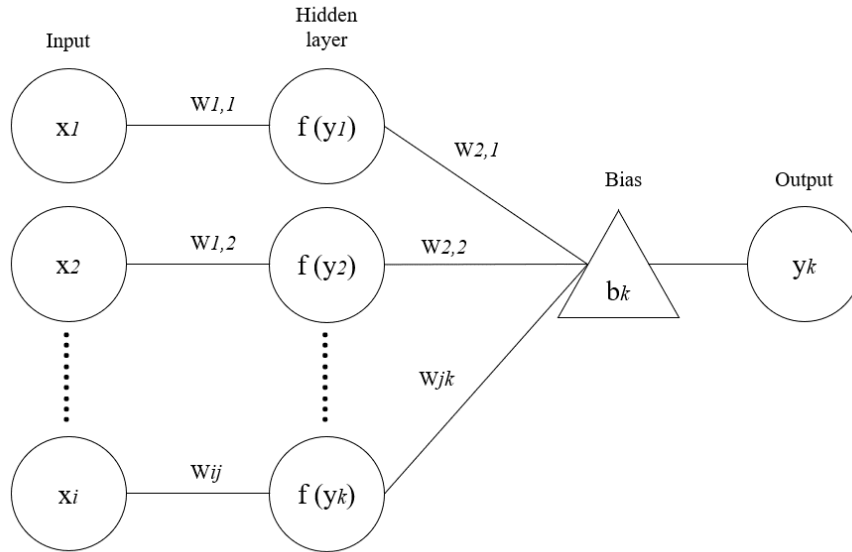


Figure 3.6 A simplified network structure with multiple inputs, with connected weights (w_{ij} and w_{jk}) from the input layer and hidden layer, respectively.

Then the net output of the network can be expressed as the sum of weights, inputs and biases from all the layers (Eq. 3.2). The predicted output, y_k , in the output layer k in a hidden layer is generated using an activation or transfer function (f). The predicted output depends on the transfer function that is chosen.

$$\sum y_k = f \sum (w_{i,j}x_i + b_{i,j}) + (w_{j,k}x_i + b_{j,k}) \quad \text{Eq. 3.2}$$

There are different types of transfer function, such as Sigmoid function, arc tangent and hyperbolic tangent, hard limit function and linear function. sigmoid and hyperbolic tangent functions are usually selected (Özbek & Fidan, 2009).

The three most used activation functions are: linear function, Hard Limit function and log-sigmoid function. Other activation or transfer functions are arc tangent and hyperbolic tangent.

I. Linear functions

The linear function gives a result where the input is equal to the output:

$$f(y) = y \quad \text{Eq. 3.3}$$

II. Hard Limit function

The Hard Limit function is defined by the following function:

$$f(y) = \begin{cases} 1 & \text{if } y \geq 0 \\ 0 & \text{if } y < 0 \end{cases} \quad \text{Eq. 3.4}$$

If the hard Limit function is symmetrical the function yields:

$$f(y) = \begin{cases} 1 & \text{if } y \geq 0 \\ -1 & \text{if } y < 0 \end{cases} \quad \text{Eq. 3.5}$$

III. Log-sigmoid function

The most common example of transfer function is the sigmoid (or logistic) function (Wang, 2003). It is one of the most used transfer functions for back-propagation. The transfer function distributes the outputs into a range between 1 and 0 (Figure 3.7). The log-sigmoid function is expressed by the following equation.

$$f(y) = \frac{1}{1 + e^{-y}} \quad \text{Eq. 3.6}$$

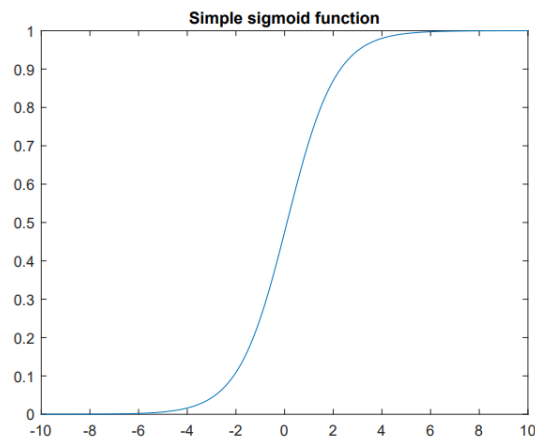


Figure 3.7 A simple illustration of a Sigmoid function

3.2.1 Network Architecture

Selecting a network type that is robust for approximation and the connectivity between the nodes are important. The connection between the nodes determines how the information is forwarded through the network (Vandeginste et al., 1998). Many researchers agree that the quality of the result is dependent on the network size used to solve the problem (Bebis & Georgiopoulos, 1994).

A crucial step for the construction of a neural network consists of selecting an appropriate architecture or network size (Bebis & Georgiopoulos, 1994). The size of the network is related

to the number of hidden layers, neurons in the hidden layers, connectivity between each of the layers. In general, the network size affects the following:

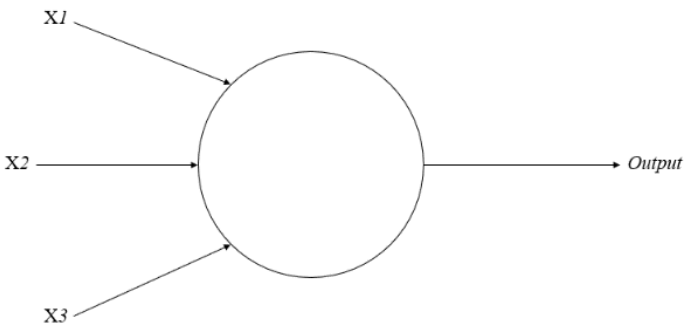
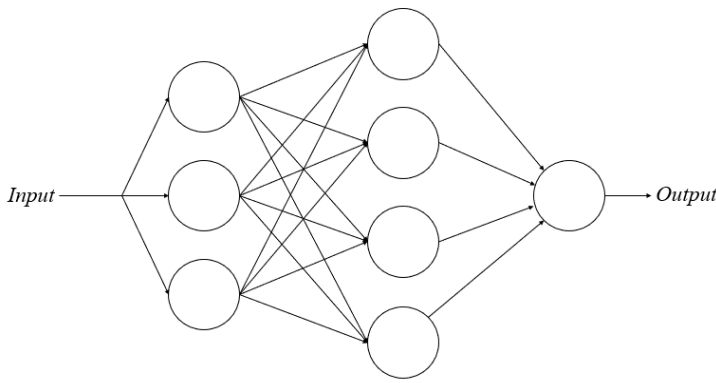
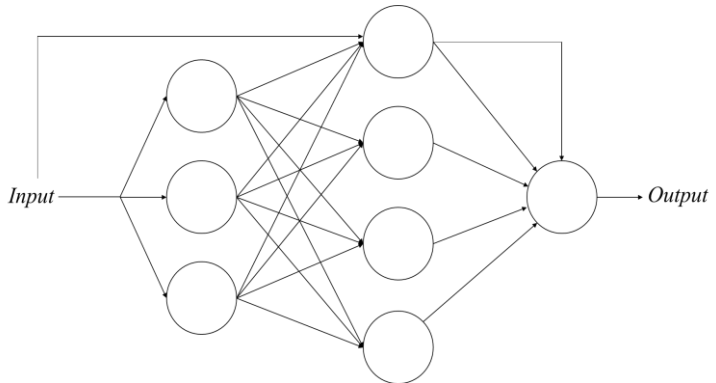
- I. The complexity of the network
- II. learning time of the network
- III. Generalisation capabilities

The complexity of the network is related to the number of unknowns (e.g. weights and biases) generated from the network. The generalisation capabilities are the ability of the neural network to produce accurate outputs that are outside the range of the training set. The accuracy of the model is dependent on the weights and biases in the network since they are used to fine-tune the model fitting. The generalisation of the network can be influenced by the number of training samples in the training set. With a large number of data points in the training set, the influence on the generalisation abilities are preserved, even with a larger number of neurons (Hunter et al., 2012). The number of training samples required depends on the size and distribution of the data (Kavzoglu, 2009). The maximum number of training samples in a neural network can be specified by the Vapnik-Chervonenkis (VC) dimension to obtain a perfect fitting for almost any data set (Friedland & Krell, 2018).

The number of hidden layers in a neural network is user-defined. According to research a network with two hidden layers is considered as acceptable for approximating any non-linear functions for classification problems (Hecht-Nielsen, 1989). It is also proven that one hidden layer is enough for a close approximation of any continuous function by a continuous neural network using a sigmoidal function (Cybenkot, 1989). In many cases, the neural networks are large and trained to very small errors since it is easier to train with a larger number of neurons. The problem with training the large networks to small errors is that the networks respond poorly to patterns outside the training set (Hunter et al., 2012).

The selected method in this study is neural networks, precisely feedforward network types. The neural networks can vary significantly in the structure or architecture depending on the problem. The feedforward network type in the thesis is multilayer perceptron or multilayer feedforward network architecture (MLP). Networks that have similar architecture to the MLP is Cascade network and Single-Layer Perceptron (SLP) (Table 3.1). The most widely used neural network type is the Multiple-Layer Perceptron (MLP) (Pun et al., 2019). In a comparison study the MLP was proven as one of the most robust methods with only one or two hidden layers (Yu et al., 2019).

Table 3.1 Simplified illustrations and a brief description of relevant network architectures in this study

Network types	Network description	Architecture
Feed forward network	<p>Single-layer perceptron:</p> <p>Single-layer perceptron, with one hidden layer and multiple nodes. The information is sent in a forward direction.</p>	
	<p>Multilayer feedforward network/Multilayer perceptron:</p> <p>Consists of a single hidden layer or multiple hidden layers, with connections from input to hidden layer(s) and to output. The information is sent in a forward direction.</p>	
Others	<p>Cascade network:</p> <p>There are some similarities between the MLP feedforward net and cascade network. The network has additional connections from the input to every layer, and from each layer to all following layers.</p>	

3.2.2 Feedforward networks

There are different types of feed forward networks, however they are similar in the way the information is transported through the network. The feedforward neural network has a hidden

layer with sigmoid activation function and in the output layer a linear activation function that would be able to estimate each complex function.

3.2.2.1 Perceptron network (SLP and MLP)

The earliest artificial network is the perceptron after Rosenblatt 1957. A simple perceptron network architecture where the input is connected to one output layer (Figure 3.8). A single perceptron is mainly used for classification of the input vectors into two categories of output (output is either 0 or 1).

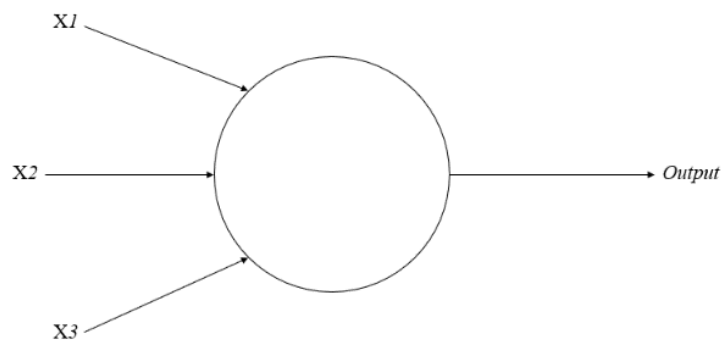


Figure 3.8 A simple perceptron where the information is forwarded from left to right.

The perceptron network uses the hard limit transfer function (see section 3.2) and the output of the network is given by:

$$y = f(wx + b) = \text{hardlimit}(wx + b) \quad \text{Eq. 3.7}$$

Where x , the input vector is connected to the weight vector (w) and the bias vector (b). The perceptron network is limited to give a combination of weights and bias that yields a correct classification, which relies on the number of hidden layers in the network (Vandeginste et al., 1998).

The single perceptron architecture was limited to two outputs and only able to solve simple linear problems. An new version of the perceptron was proposed in 1958, the multiple layer perceptron (MLP) (Rosenblatt, 1958). A multilayer perceptron usually has three layers one input layer, one hidden layer and one output layer, with log-sigmoid transfer function in the other layers and a linear transfer function in the output layer (Figure 3.9). It has been proven that a perceptron with only two hidden layers is incapable of predicting values outside a narrow and special range (Hornik et al., 1989).

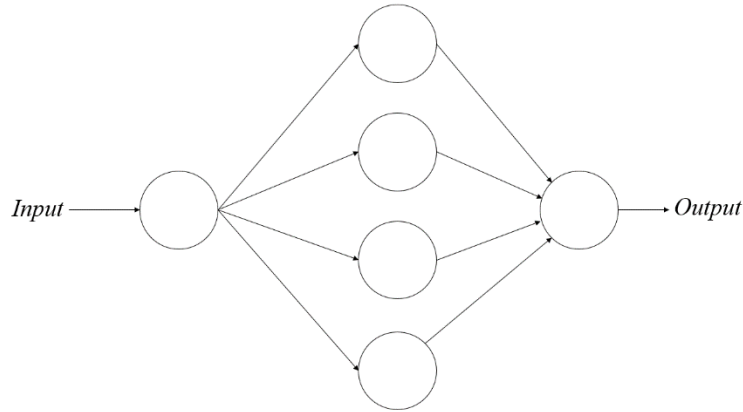


Figure 3.9 A schematic illustration of a multilayer perceptron with one input layer, one hidden layer and one output layer. The hidden layer consists of 4 neurons in this illustration.

The MLP is the most widely used neural network (Pun et al., 2019). The MLP network consists of multiple hidden layers that can classify inputs into several categories, compared to the perceptron. The multi-layer perceptron is also called multilayer feedforward network with layers that are fully connected to each other from the input to the output. The input is sent in a forward direction from the input layer to the output layer, where the output of each layer is forwarded to the following layer. At the output layer, the error between the predicted outputs and target values is compared. The error (ε) from the perceptron network and the MLP is the difference between the for the targeted output (T) and the predicted output (Y) for the k th iteration.

$$\varepsilon_k = T_k - Y_k \quad \text{Eq. 3.8}$$

After the error is calculated the network weights are adjusted using a supervised learning algorithm. The updated weights for the Perceptron are expressed by:

$$w_{k+1} = w_k + \varepsilon_k x_k \quad \text{Eq. 3.9}$$

$$b_{k+1} = b_k + \varepsilon_k \quad \text{Eq. 3.10}$$

One of the most popular neural network models is multilayer networks, trained by a learning algorithm called Back-propagation (BP) algorithm (Li et al., 2012).

3.2.2.2 Adaline

The Adaptive Linear Neuron network was proposed by Widrow and Hoff in 1960. The Adaline network have the same architecture as the perceptron (Figure 3.8), however the transfer function

applied is linear. The Adaline network can solve linear problems and classifications, the predicted output is described by:

$$y = \text{purelin}(wx + b) \quad \text{Eq. 3.11}$$

The Adaline network uses a delta rule or least mean square (LMS) algorithm instead of hard-limit. The LMS algorithm minimises mean squared error making it more powerful than the single perceptron (M. Hagan T. et al., 2002). The LMS learning algorithm is the squared errors between the predicted and desired outputs (Eq. 3.12Eq. 2.1).

$$\varepsilon(k) = [T_k - Y_k]^2 \quad \text{Eq. 3.12}$$

The weights and bias are adjusted by the following equations:

$$w_{k+1} = w_k + k x_k \quad \text{Eq. 3.13}$$

$$b_{k+1} = b_k + k b(k + 1) \quad \text{Eq. 3.14}$$

3.2.3 Other neural networks

An improved version of the MLP is the Bridged Multilayer perceptron (BMLP), with connections across the hidden layers (Hunter et al., 2012). In a Bridged Multilayer Perceptron, the connections between the layers is from the input layer to all the hidden layers.

The cascade-forward neural network architecture is similar to the MLP, however the connections are different. The cascade network has connections from the input and every previous layer to the next layers. requires at least 12 to 19 neurons in the hidden layers and if some of the connections are removed the network will be reduced to a Bridged multilayer perceptron (Hunter et al., 2012). The cascade networks use backpropagation algorithms like the MLP.

4 Optimisation techniques

Many algorithms have been developed for training a neural network aiming to update weights and biases and reduce the errors. The main aim of the learning algorithm is to reduce the learning error of the network to very small values. When the errors are very small the generalisation abilities of the network may be poor.

The most popular training algorithm is the Error Back Propagation (EBP). The simple EBP algorithm is inefficient and improved versions of the EBP algorithm have been proposed such as e.g momentum, stochastic learning rate, flat-spot elimination, RMSPROP and QUICKPROP (Hunter et al., 2012). Better results can be obtained using second-order methods since the search process in the EBP algorithm follows the gradient and can be trapped in a local minimum (Hunter et al., 2012). The Levenberg Marquardt (LM) algorithm is the most efficient second-order learning algorithm, and it is written such that it is only applicable for training multilayer networks. (Hunter et al., 2012).

Swarm intelligence is another discipline that studies the collective behaviour of systems with many individual points that interacts locally with each other, like e.g colonies of ants, flock of birds. There are two types of optimisation algorithms using swarm intelligence: ant colony optimisation and particle swarm optimisation.

4.1 Ant colony optimisation (ACO)

As the name suggest the ant colony optimisation was inspired by the behaviour of ant species. In the paper of Beckers et al., 1992, the results show that ants tend to find and select the shortest path between two points.

The ant colony optimisation is a population-based algorithm that is used to solve complex mathematical problems, such as estimation of reservoir parameters. In the ACO each individual data point in the population is a computational agent or ant that builds a solution to the considered problem (Maniezzo et al., 2004). The algorithm is based on moving stepwise and each step define which solution components that will be added in the final solution. In addition, a probabilistic model is associated with the movements and used to bias the choices of the computational agents in the population, to increase the accuracy (Kakas et al., 2011).

4.2 Particle swarm optimisation (PSO)

Particle Swarm Optimisation (PSO) was inspired by the behaviour of animals (such as a flock of birds) (Kennedy & Eberhart, 1995). The PSO algorithm is used to optimise continuous non-

linear functions (Rwechungura et al., 2011). The main aim of the algorithm is to optimise the model parameters and increase the model performance according to a minimising function e.g Mean Squared Error (MSE).

The algorithm describes the position of a particle (P_i) in a population of 20 to 50 with several random decision values. Each of the particles consists of three parameters, the new particle position (P_{new}), velocity (v_i) of the particle and the best particle position (P_{best}). Each of the particles moves through the problem space with a random velocity aiming to find the best solution for the particle and the best solution on the whole population (P_{best} and P_{global}). The parameters in the PSO algorithm are listed in Table 4.1.

Table 4.1 The parameters used in the Particle Swarm Optimisation (PSO) algorithm with descriptions of each parameter.

Variable	Description
Pi	Particle position
vi	Particle velocity
Pbest	Best solution of individual particle
PGlobal	Best solution of the whole population
C1, C2	Cognitive social parameters
nVar	Number of variables
minVar	Lower bound range
maxVar	Upper bound range
w	Inertia weight

The new velocity is obtained by the following equation:

$$v_{i,new} = v_{i,old} + C1 \times random[minVar, maxVar](P_{best} - P_i) + C2 \times random[minVar, maxVar](P_{global} - P_i) \quad \text{Eq. 4.1}$$

Then the new particle position is updated by adding the velocity to the old particle position (Eq. 4.2).

$$P_{new} = P_{old} + v_{i,new} \quad \text{Eq. 4.2}$$

4.3 Back-propagation algorithms

The Backpropagation learning algorithm is a steepest descent algorithm, using the gradient of the error function (E). In addition, the error function can also be expressed as a function of the

weights or bias. The flow of the network input is sent in a forward direction and backward direction. The term backpropagation network is often used to describe a multilayer perceptron architecture (Bishop, 2006). The simplest optimization technique uses gradient descent (Rumelhart et al., 1986). In the Error backpropagation algorithm, the error function of a specific input pattern (E_n) set can be defined as:

$$E_n = \frac{1}{2} \sum_k (T_{nk} - Y_{nk})^2 \quad \text{Eq. 4.3}$$

The gradient of the error backpropagation with respect to a given weight ($w_{j,i}$).

$$\frac{\partial E_k}{\partial w_{j,i}} = (T_{nk} - Y_{nk})x_{ni} \quad \text{Eq. 4.4}$$

The backpropagation optimisation has some limitations such as slow convergence, incapable of handling multiple objectives and a high probability to be trapped in local minima during training (Santos et al., 2012).

4.3.1 Gradient Descent (GD) and Stochastic Gradient Descent (SGD)

The Gradient Descent and Stochastic Gradient Descent algorithm are similar optimisation algorithms. Both optimisation algorithms update the weights in an iterative process to minimize the errors in the network.

In the Gradient Descent algorithm, the weights are updated every time the training set runs through the algorithm, after each epoch. In the Gradient descent the algorithm updates the weights and biases by moving small steps for each epoch until the algorithm reaches a cost gradient. The Gradient Descent uses all the training samples, while the Stochastic Gradient Descent uses one sample or a subset of the training samples to update the weights. Both optimisation algorithms use the same equation to update the weights (Eq. 4.5).

$$w_{k+1} = w_k - \alpha \frac{\partial E_k}{\partial w_k} \quad \text{Eq. 4.5}$$

The updated parameter vector at i th iteration is equal to the parameter vector at i th iteration with the learning rate and the gradient of the transfer function. In addition, if a momentum is added to reduce oscillation (Murphy 2012) in the updated parameter vector is called the Stochastic Gradient Descent with Momentum (SGDM). The equation is expanded to the

following equation. The momentum value (γ) is the contribution of the previous iteration to the current iteration.

$$w_{k+1} = w_k - \alpha \frac{\partial E_k}{\partial w_k} + \gamma(w_k - w_{k-1}) \quad \text{Eq. 4.6}$$

4.3.2 Newton and Gauss-Newton algorithm

The Newton algorithm calculates the second order derivatives of the sum of errors in a matrix called the Hessian matrix. Thus, the updated weights by the Newton's algorithm is:

$$w_{k+1} = w_k - H_k \frac{\partial E_k}{\partial w_k} \quad \text{Eq. 4.7}$$

The Gauss-Newton algorithm minimises the sum of squared errors, and converges much faster than the Gradient-descent method (Gavin, 2020). The Gauss-Newton algorithm uses the Jacobian matrix with the first order derivatives of the sum of errors. This results in an approximation of the Hessian matrix (Eq. 4.9). The updated weights with the Gauss-Newton algorithm is then defined as:

$$w_{k+1} = w_k - [J^T J]^{-1} J^T e \quad \text{Eq. 4.8}$$

The convergence in the Newton algorithm and the Gauss-Newton algorithms are fast, however the convergence is unstable (Mayyahi et al., 2015).

4.3.3 Levenberg-Marquardt Algorithm

The Levenberg-Marquardt algorithm is a nonlinear least square method for neural networks (M. T. Hagan & Menhaj, 1994). The main aim of the algorithm is fitting the model to the data samples and minimising the sum of squared errors between the model and data samples (Gavin, 2020). The Levenberg-Marquardt algorithm combines the gradient descent method and the Gauss-Newton method. In the Gauss-Newton method the sum of squared errors is reduced by assuming that the function is quadratic and finding the minimum of this. If the parameters in the Levenberg-Marquardt method is further from the optimal value it acts like the Gradient-descent method and like Gauss-Newton method when parameters are close to the optimal value (Gavin, 2020).

The Levenberg-Marquardt algorithm does not compute the Hessian matrix with second order derivatives, it is approximated by the Jacobian matrix with first order derivatives of the errors

using a minimised sum of errors of the form from Eq. 4.3. The errors are computed using a standard backpropagation technique, which is less complex than computing a Hessian matrix. The Hessian matrix is approximated by the following:

$$H = J^T J \quad \text{Eq. 4.9}$$

And the gradient of the approximated Hessian matrix can be computed by:

$$\frac{\partial H}{\partial} = J^T e \quad \text{Eq. 4.10}$$

Where the Jacobian matrix J^T which contains the first derivatives of the network errors (e) with respect to the weights and biases. Then the updated weights in the Levenberg-Marquardt algorithm is:

$$w_{k+1} = w_k - [J^T J + \mu I]^{-1} J^T e \quad \text{Eq. 4.11}$$

Where μ is the controlling parameter or damping factor of the algorithm and I is the identity matrix. If μ is equal to zero the algorithm follows the Newton method, if the parameter is higher it becomes gradient descent with a small step size (Zayani et al., 2008; Gavin, 2020). The convergence in the Levenberg-Marquardt algorithm is stable and fast (Mayyahi et al., 2015).

4.4 Advantages and disadvantages of the Levenberg-Marquardt algorithm

The network type in this study is a feedforward network, using the Levenberg-Marquardt algorithm for optimisation. Some of the advantages and disadvantages of the selected method are presented in this section.

The Levenberg-Marquardt algorithm finds the local minima of a function and not necessarily the global minimum, compared to the Particle Swarm Optimisation algorithm. The Levenberg-Marquardt may get stuck at a local minimum (Jabri & Jerbi, 2009)

In comparison study between Gradient Decent algorithms and Newton methods the Levenberg-Marquardt was deemed as the fastest and ensures the best convergence towards a minimum of error (Zayani et al., 2008). The Levenberg-Marquardt algorithm performs better than steepest descent and other conjugate methods (Arif et al., 2009)

The Levenberg-Marquardt algorithm is sensitive to the input weights and does not consider outliers in the data set. This is countered by quality checking and pre-processing the data prior to it being loaded to the model. The outliers may lead to overfitting of data samples that are outside the range of the general distribution in the data set. In general, regularisation techniques can also be applied to avoid overfitting of these data samples. In the following section some regularisation techniques will be briefly explained.

4.4.1 Overfitting

Regularisation is the most common way to reduce overfitting/generalization of the data set. In a neural network the regularisation term modifies the performance function, which is normally chosen as the sum of squared errors in the neural network.

A regularisation term for the weights and the transfer function can be added to reduce overfitting of the network (Bishop, 2006; Murphy, 2012). L1 regularisation (sum of absolute values of model parameters) and L2 regularisation (sum of squared absolute values of model parameters), are two commonly used regularisation methods (Demir-Kavuk et al., 2011). The regularisation is added to the loss function that minimises the errors. It is important to keep in mind that the biases are not regularised (Murphy, 2012). Selecting correct regularisation values for the function is the main problem with implementing regularisation (Kayri, 2016).

Another regularisation technique is early stopping to avoid overfitting, however early stopping can also reduce the variance and increase bias (Kayri, 2016). The Levenberg-Marquardt algorithm is an algorithm that mainly relies on the training process to stop when certain conditions are met, early stopping, to reduce overfitting (section 5.6).

Another way to reduce overfitting is using a Bayesian regularisation backpropagation network (BR), which has an added regularisation term. The BR network updates the weights according to the Levenberg-Marquardt algorithm with using Baye's rule and the regularisation parameters α and β (Eq. 4.12).

$$P(w \mid data\ set, \alpha, \beta, M) = \frac{P(data\ set|w, \beta, M)P(w|\alpha, M)}{P(data\ set|\alpha, \beta, M)} \quad \text{Eq. 4.12}$$

In studies the Bayesian regularisation have been proven to give better performance based on the mean absolute error, however the required time for training is much longer than the Levenberg-Marquardt algorithm (Jazayeri et al., 2016; Kayri, 2016)

5 Methodology

5.1 Study area

The Varg field is in the central part of the North Sea, located to the South of the Sleipner Øst field (Figure 5.1). The field was discovered in 1984 and production started in 1998. A decommissioning plan was approved in 2001, and a new plan was submitted in 2015, eventually it was shut down in 2016. The field mainly produced oil from Upper Jurassic sandstones at approximately 2700 metres. The available and comparable petrophysical logs and core data in Well 15/12-1, 15/12-4, 15/12-5, 15/12-6S, 15/12-9S and 15/12-A8 is used in this study. A short summary of the well information is presented in Table 5.1.

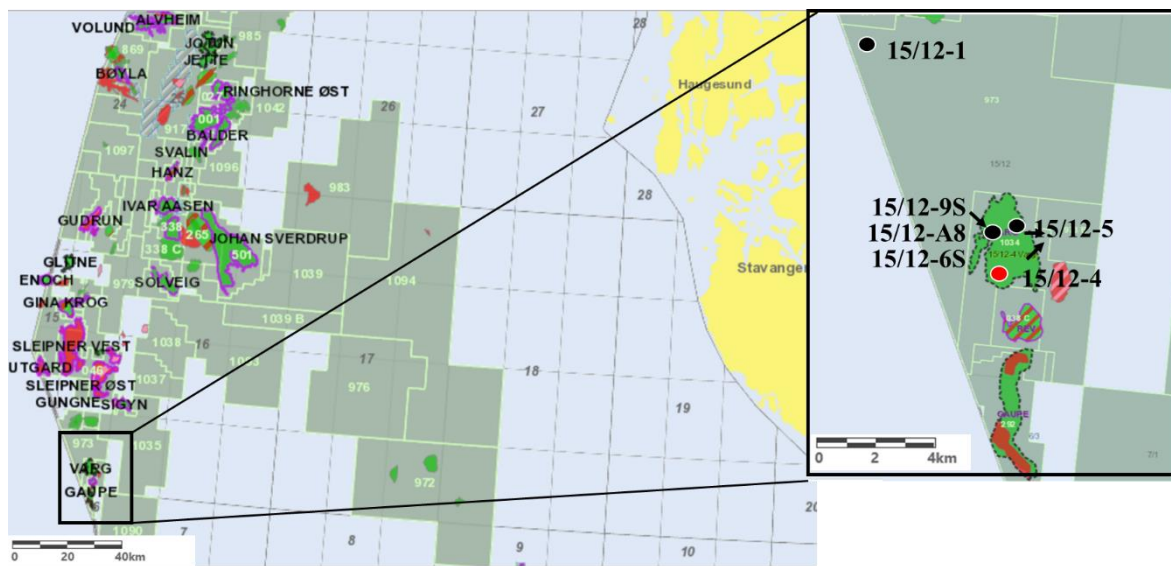


Figure 5.1 Location of the Varg field in the Central North Sea. The black circles are the wells used to calibrate the model, and the red circle is the well the calibrated model is tested on.

Table 5.1 Well information of the selected wells in the study

Wells	Type and Purpose	Total depth [m RKB]	Oldest penetrated age	Top depth core [m]	bottom depth core [m]
15/12-1	Exploration, Wildcat	3269	Late Triassic	2635.8	3150.1
15/12-4	Exploration, Wildcat	3157	Middle Jurassic	2439	2919.7
15/12-5	Exploration, Appraisal	3150	Late Triassic	2892	2967

15/12-6S	Exploration, Appraisal	3050	Triassic	2838	2988.3
15/12-9S	Exploration, Appraisal	3848	Triassic	3389	3670.5
15/12-A8	Development, Injection	3345	-	3785	3959.7

The petrophysical data used in the study are limited to the depth of the core samples presented in Table 5.1. The depth of the cores defines the sections of interest, thus and the lithological description from each well will focus on the section at these depths (see section 5.3). The lithological description of the cores samples is divided into seven main units. The lithological description in the completion log and well report might differ. Thus, the general lithology of the formations may differ from the lithology described in the core analysis.

5.2 Stratigraphy

The study is focused on these units: Lista formation, Maureen formation, Tor formation, Heather formation, Hugin formation and Sleipner formation (Table 5.2).

Table 5.2 Stratigraphic column of the units in used in the study.

Chronostratigraphy		Formations (FM)	Group (GP)
Paleocene	Late	Lista FM	Rogaland GP
	Early	Maureen FM	
Cretaceous	Late	Tor FM	Shetland GP
Jurassic	Late	Heather FM	Viking GP
	Middle	Hugin FM	Vestland GP
		Sleipner FM	
Triassic	Late	Skagerrak FM	Hegre GP
	Middle		

This section summarises the lithological description of the main units in this study.

Lista Formation

The Lista Formation is a part of the Rogaland Group. It consists of poorly laminated shales, stringers of limestone, dolomite, pyrite, and locally thin sections of sandstones closer to the formation base. The depositional environment of the formation is deep marine with low-energy conditions (Isaksen & Tonstad, 1989).

Maureen Formation

The Rogaland Group's Maureen Formation consists mainly of slightly to very calcareous sandstones, reworked limestones and interbedded shales. The depositional environment of the formation is open marine, dominated by sand influx from the northwest and local erosion (Isaksen & Tonstad, 1989).

Tor Formation

The Tor formation of the Shetland Group consists of hard, chalky limestones, alternating mudstones or wackestones chalky limestones or limestones. The formation rarely consists of packstones and calcareous shale. The depositional environment in the formation is open marine with calcareous debris (Isaksen & Tonstad, 1989).

Heather Formation

The Heather Formation of Viking Group consists of two divisions: the upper division consists of silty claystone, often micromicaceous and calcareous. The lower division consists of silty claystone with some carbonaceous sections (limestone streaks). The depositional environment of the formation is open marine (Vollset & Doré, 1984).

Hugin Formation

The Hugin Formation consists of sandstones (often calcareous and glauconitic) with shale, siltstone partings, carbonaceous material, and coal fragments are abundant in the formation and occasionally thin coal beds. The depositional environment is shallow marine with some influence of continental fluviodeltaic conditions (Vollset & Doré, 1984).

Sleipner Formation

The Sleipner formation of the Vestland Group consists mainly of non-calcareous sandstone and micromicaceous, hard and slightly fissile silty claystone with coal measures. The

depositional environment of the formation is a continental fluviodeltaic coaly sequence (Vollset & Doré, 1984).

Skagerrak Formation

The Skagerrak Formation is characterised by interbedded conglomerates, sandstone (orthoquartzitic or highly lithic), siltstones and shales. The depositional environment of the formation is a prograding system of alluvial fans (Deegan & Scull, 1977).

5.3 Data set (Petrophysical logs and core data)

In this section key information about the petrophysical logs and core data is presented. The models are calibrated with the measurements of from routine core analysis, the evaluated lithologies are sandstones and carbonates (e.g. chalk). The information about the findings from each well is either from the Norwegian Petroleum Directorate (NPD), well reports or core analysis (Statoil-Esso, 1975; Statoil, 1984; Statoil-Esso, 1986; Statoil, 1993; Saga Petroleum, 1999; Statoil, 2004; NPD, 2020).

The well measurements for well 15/12-1 was taken in feet units, in this study it is converted to meters, since the other wells have measurements in meters. 51.5 m of core were taken for well 15/12-1 in intervals at 2612-3181 m depth (Table 5.3). They include shale and sandstone from Palaeocene sand of the Lista and Maureen formations, the siltstone and shale of the Jurassic Heather Formation and sandstone of the Jurassic Hugin Formation, as well as coal, shale, siltstone and sandstone of the Triassic Sleipner formation. The Palaeocene sandstones at 2633 m to 2643 m were water-wet with good porosity (26%) (Statoil-Esso, 1975). The petrophysical logs from well 15/12-1 shows low gamma-ray readings (30-35 API°), possibly tight lithology, with no significant hydrocarbon indicators (Figure 5.2 and Figure 5.3).

Table 5.3 The depth of the core samples and the corresponding formation in well 15/12-1

Core depth [m MD RKB]	Formation
2612.1-2635	Lista FM
2635-2651	Maureen FM
3067.2-3073.3	Heather FM
3125.7-3143.7	Hugin FM
3143.7-3181.5	Sleipner FM

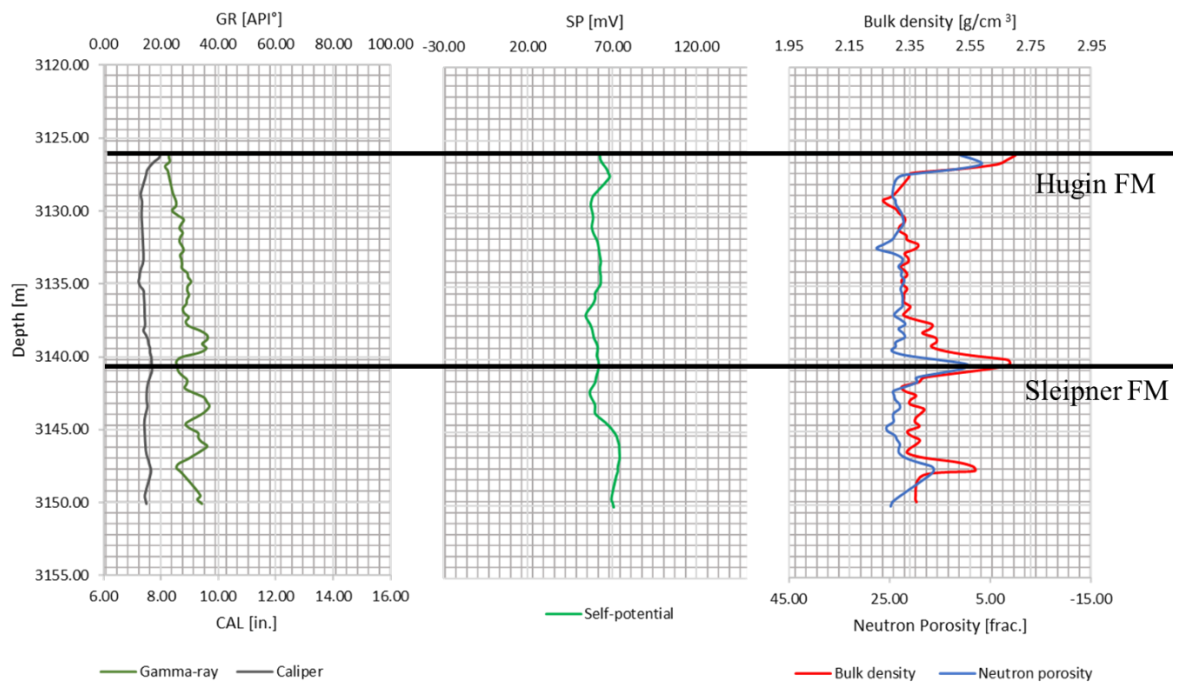


Figure 5.2 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from core samples of Hugin FM and Sleipner FM of well 15/12-1.

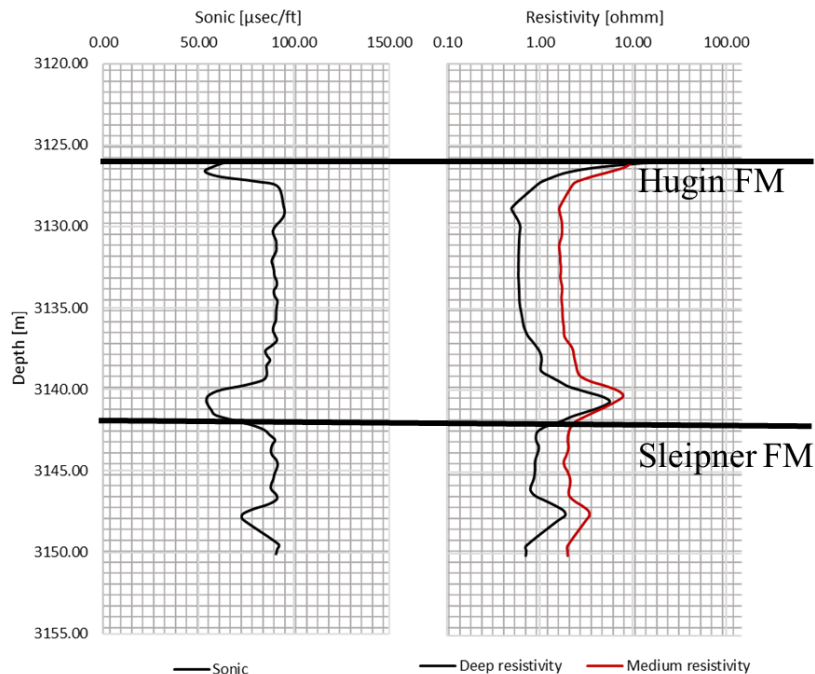


Figure 5.3 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-1- The petrophysical logs are taken from core sample of Hugin FM and Sleipner FM of well 15/12-1.

46 m of core were taken for well 15/12-4 in intervals 2439-2919.7 m depth (Table 5.4). They include shale and sandstones from the Paleocene Lista and Maureen formations, siltstone and shale from the Jurassic Heather Formation, and sandstones from the Jurassic Hugin Formation. In addition, Late Cretaceous chalk of the Tor Formation is cored in this well.

The petrophysical logs shows that the cored interval have positive separation in the crossplot of the neutron and bulk density plot in the section with lower gamma-ray readings (Figure 5.4). In the upper part of the Cretaceous chalk the core analysis indicates poor reservoir properties with low permeabilities (0.01-0.5 mD) and high water saturation (60-80%). In the Jurassic and Triassic age an oil column of 1.5 m was proven at 2911.5 to 2913 m (transition zone down to 2915.5) (Figure 5.5). The thickness of the oil zone and the interbedded shale may affect the resistivity log.

Table 5.4 The depth of the core samples and the corresponding formation in well 15/12-4

Depth [m MD RKB]	Colour
2439-2450	Lista FM
2494.5-2513.5	Tor FM
2902-2920	Heather FM
2912-2919.7	Hugin FM

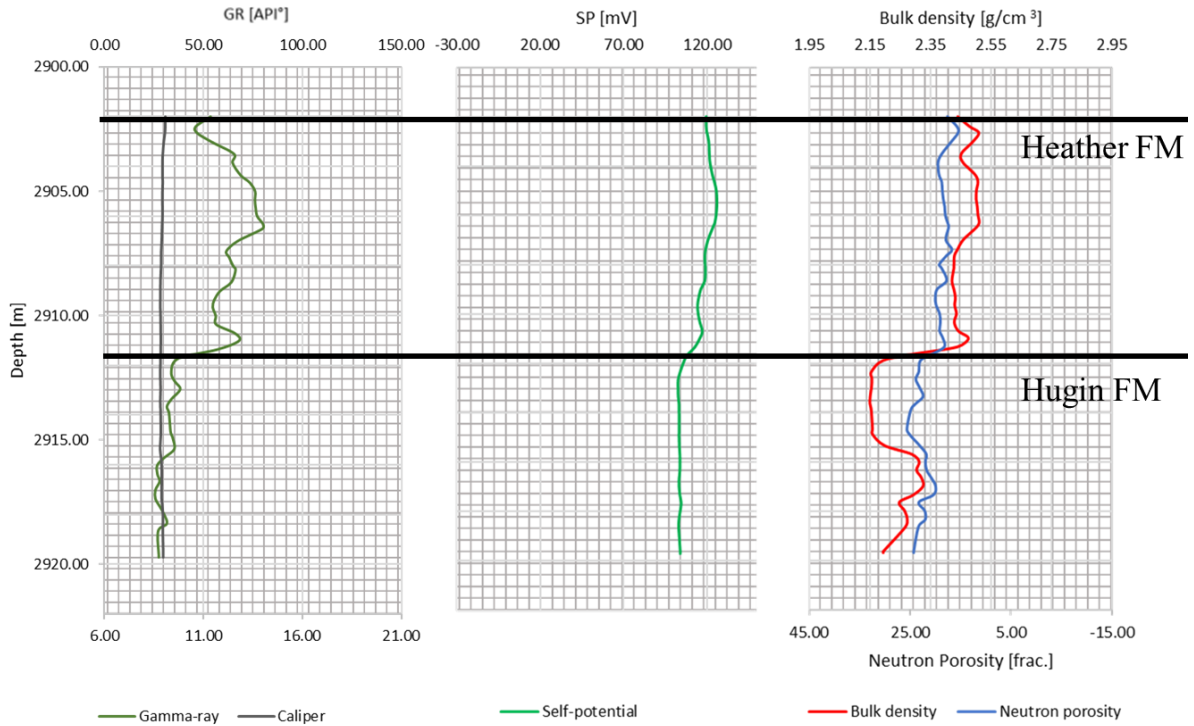


Figure 5.4 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from core sample of the Heather FM and Hugin FM of well 15/12-1.

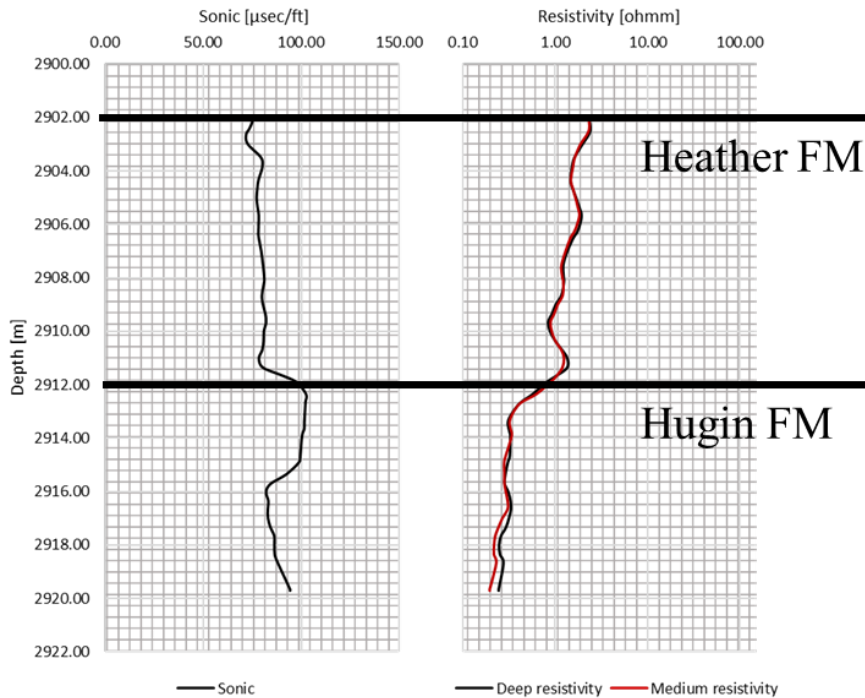


Figure 5.5 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-4. The petrophysical logs are taken from core sample of Heather FM and Hugin FM of well 15/12-1.

68 m of core were taken for well 15/12-5 in intervals 2892-2967 m depth (Table 5.5). They include siltstone, claystone and shale from the base of the Jurassic Heather Formation and sandstones from the Jurassic Hugin Formation. The top of the reservoir was encountered at 2918 m and an oil-water contact (OWC) was observed at 2942 m. An oil column of 25 m was proven in the Hugin Formation. The gamma-ray and self-potential logs shows higher readings until around 2918 metres. The cross plot with the neutron density and bulk density log shows positive separation indicating hydrocarbons as well as the separation and decrease in the resistivity logs (Figure 5.6 and Figure 5.7).

Table 5.5 Lithological description of the cores from well 15/12-5

Depth [m MD RKB]	Colour
2892-2918	Heather FM
2918-2967	Hugin FM

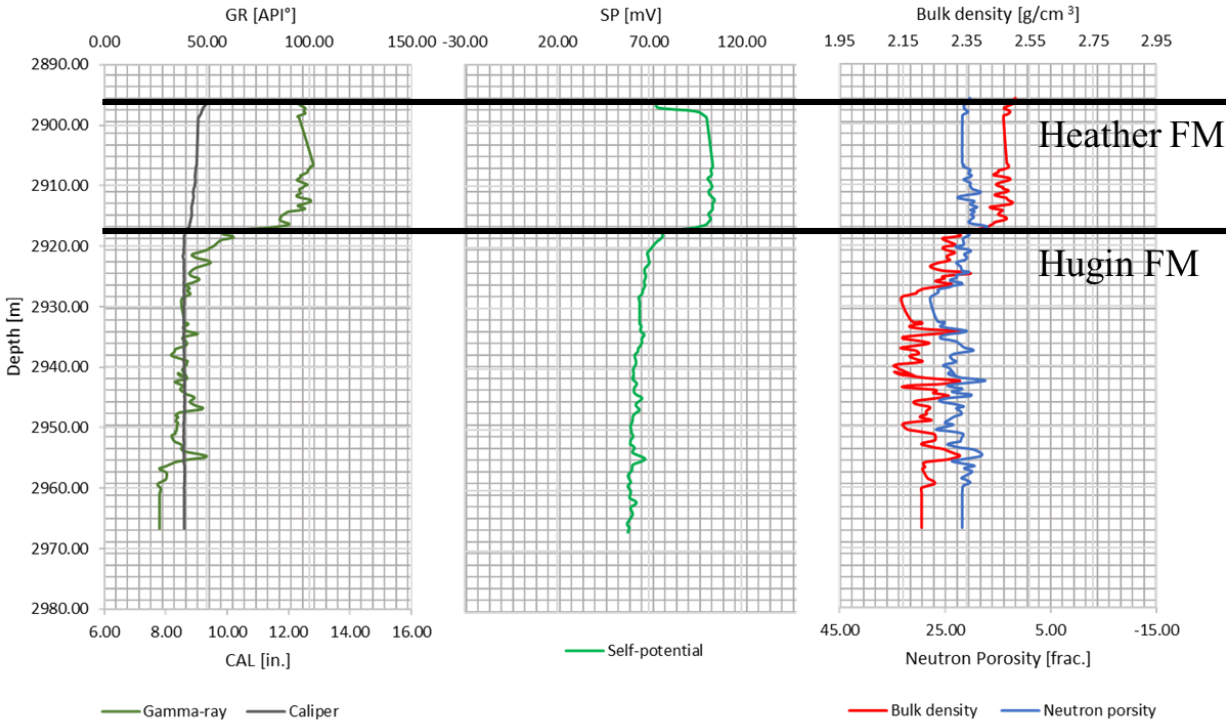


Figure 5.6 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the cored depth of well 15/12-5.

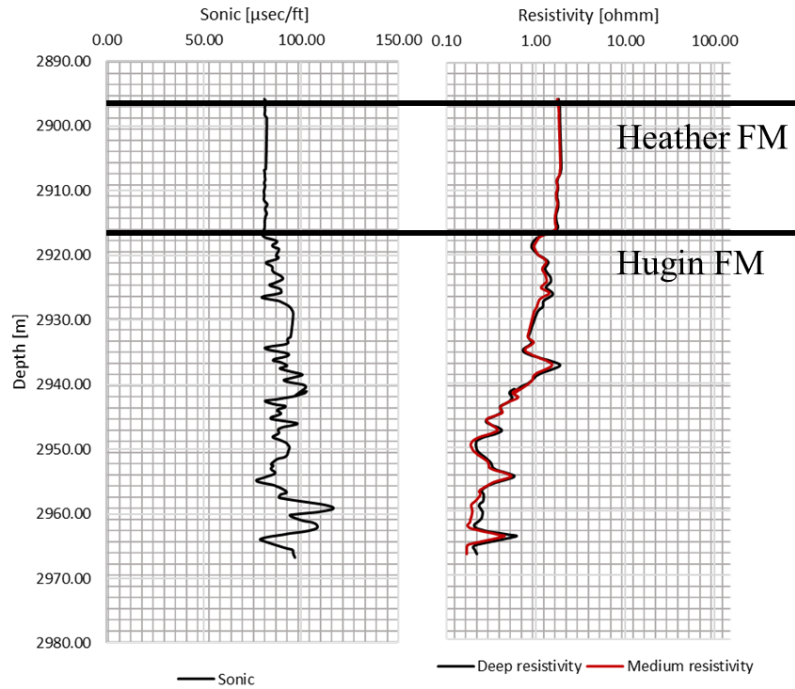


Figure 5.7 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-5. The petrophysical logs are taken from the cored depth of well 15/12-5.

131.7 m of core were taken for well 15/12-6S in intervals 2838-2988 m depth (Table 5.6). one of the samples contained a mix of mudfiltrate and formation water with traces of hydrocarbons, at 2935.5 m. The cores include claystone, siltstone and interbedded sandstone from the Jurassic Heather Formation, sandstones from the Jurassic Hugin Formation. The cored interval of the Middle Jurassic Sleipner Formation consists of claystone, interbedded sandstone, coal and minor siltstone. The last cored formation is the Late to Middle Triassic Skagerrak formation consisting of claystone, interbedded sandstones and some stringers of siltstone.

The cross plot of the bulk density and neutron porosity, and the resistivity logs indicate hydrocarbons around 2870 m (Figure 5.8 and Figure 5.9). In well 15/12-6S Late Jurassic Oxfordian sandstones (Hugin Formation) contained oil and an oil-water contact (OWC) was found at 2943. There were no hydrocarbons below the depth of this OWC. The well changed to a development well (15/12-A-2) after being suspended as an oil appraisal well.

Table 5.6 Lithological description of the cores from well 15/12-6S

Depth [m MD RKB]	Colour
2838-2870.5	Heather FM
2870.5-2949	Hugin FM

2949-2978.5

Sleipner FM

2978.5-2988.5

Skagerrak FM

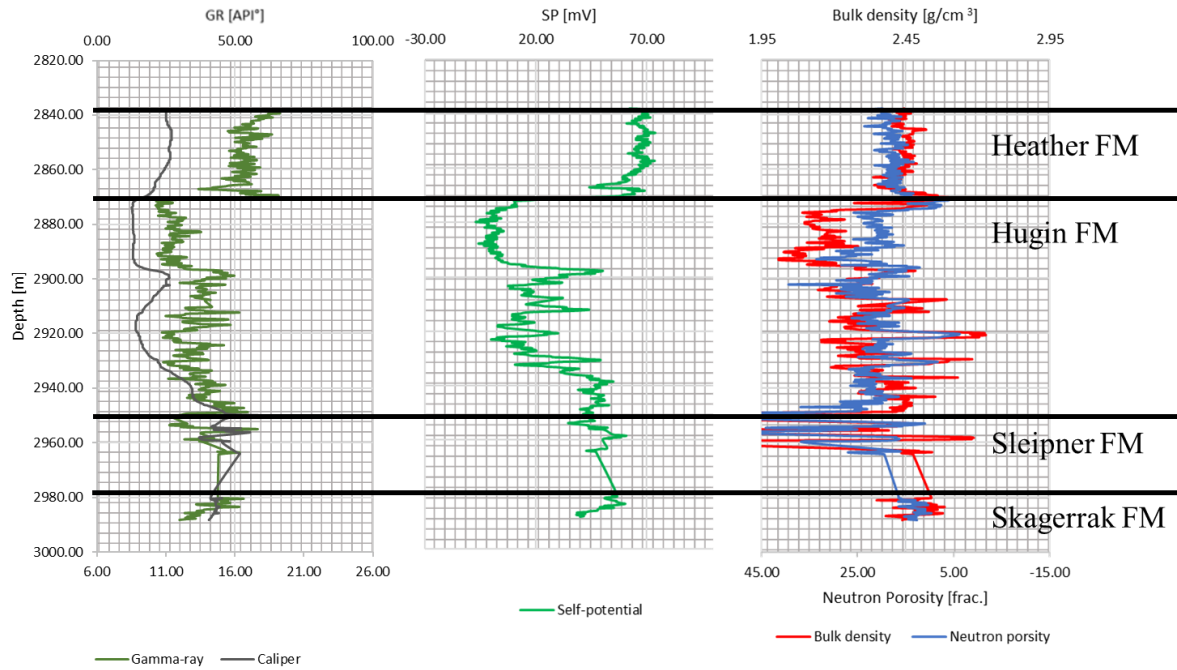


Figure 5.8 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the cored depth of well 15/12-6S.

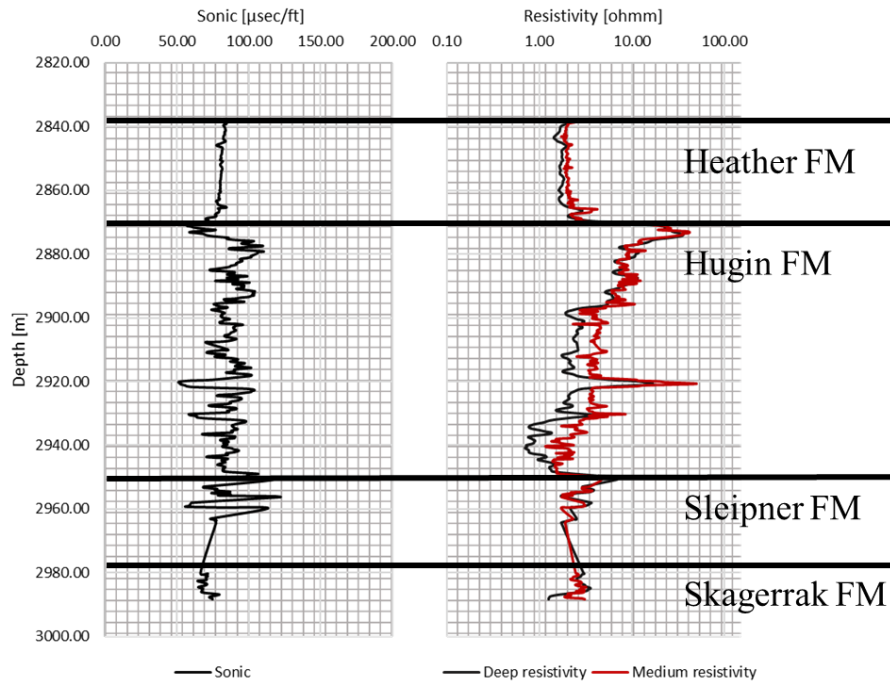


Figure 5.9 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-6S.

184 m of core were taken for well 15/12-6S in intervals 3389-3670 m depth (Table 5.7). The cores include sandstone, siltstone, argillaceous sandstone and stringers of limestone from Middle Jurassic Oxfordian sandstone unit, and sandstone, coal and shale from the Middle Jurassic Sleipner Formation.

The petrophysical logs indicate hydrocarbons to 3550 m, due to the positive separation in the neutron porosity and bulk density cross plot, and the resistivity log (Figure 5.10 and Figure 5.11). The reservoir was oil-bearing down to an oil-water contact (OWC) at around 3501 m.

Table 5.7 Lithological description of the cores from well 15/12-9S

Depth [m MD RKB]	Colour
3389-3620	Hugin FM
3620-3670	Sleipner FM

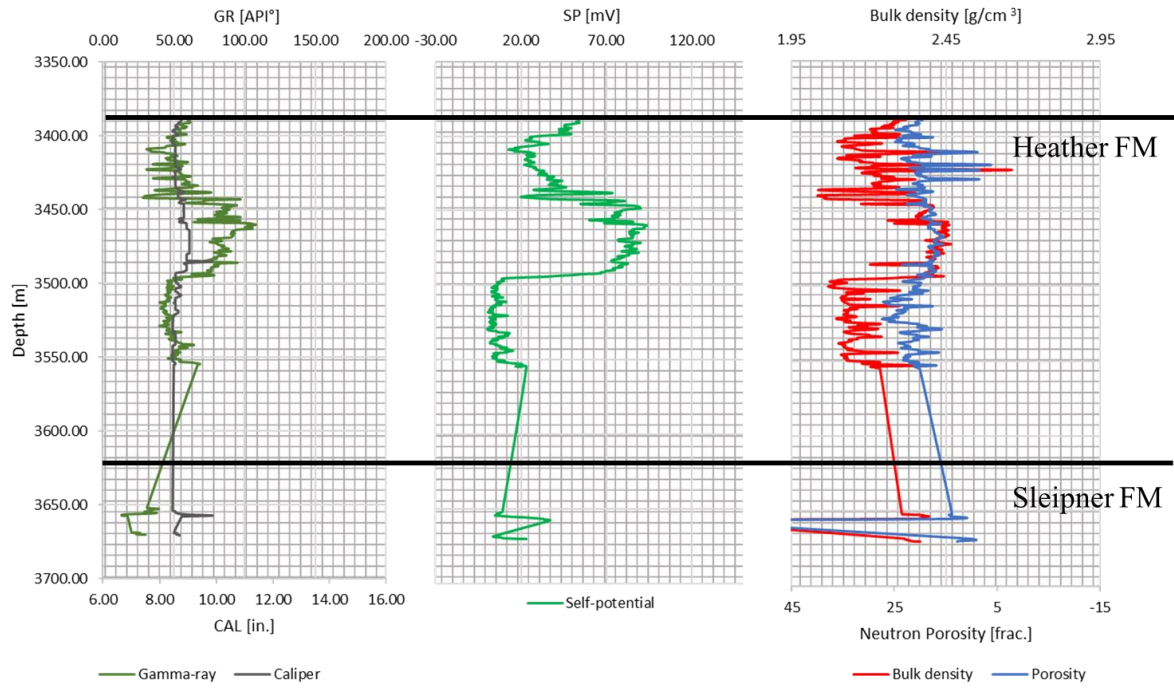


Figure 5.10 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the core depth of well 15/12-9S.

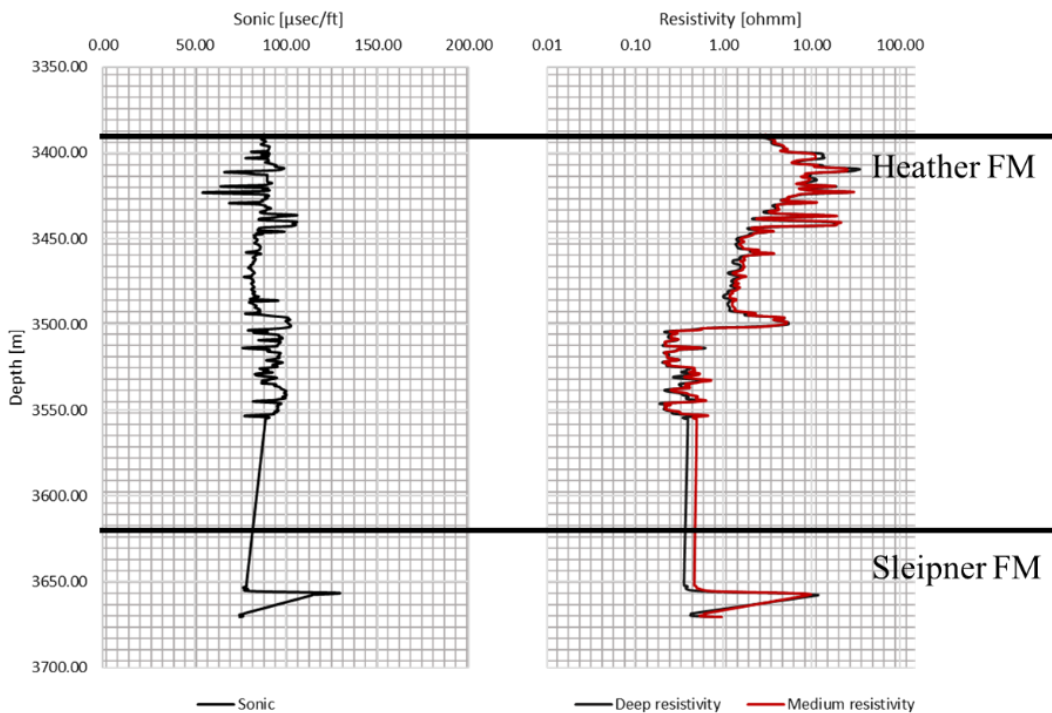


Figure 5.11 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-9S.

116.2 m of core were taken for well 15/12-A8 in intervals 3125-3254.2 m depth (Table 5.8). The cores include sandstones from the Jurassic Heather Formation.

Around 3215 m there is a positive separation in the neutron porosity and bulk density crossplot, and the separation in the resistivity logs indicates hydrocarbons at this depth (Figure 5.12 and Figure 5.13). There is no oil-water contact in this well, it is considered oil-down to (ODT) (Saga Petroleum, 1999).

Table 5.8 Lithological description of the cores from well 15/12-A8

Depth [m MD RKB]	Colour
3135-3251	Heather FM
3251-3254.2	Sleipner FM

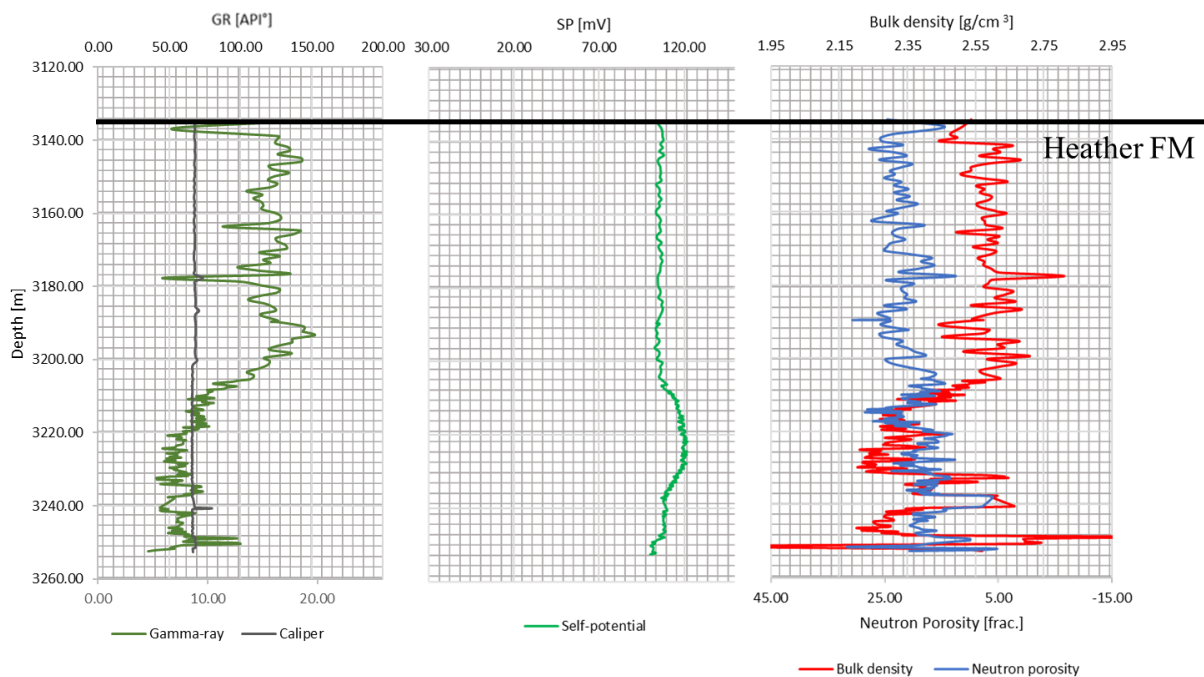


Figure 5.12 From left to right the gamma-ray, self-potential and a cross plot of the neutron porosity and bulk density is displayed. The petrophysical logs are taken from the core depth of well 15/12-A 8.

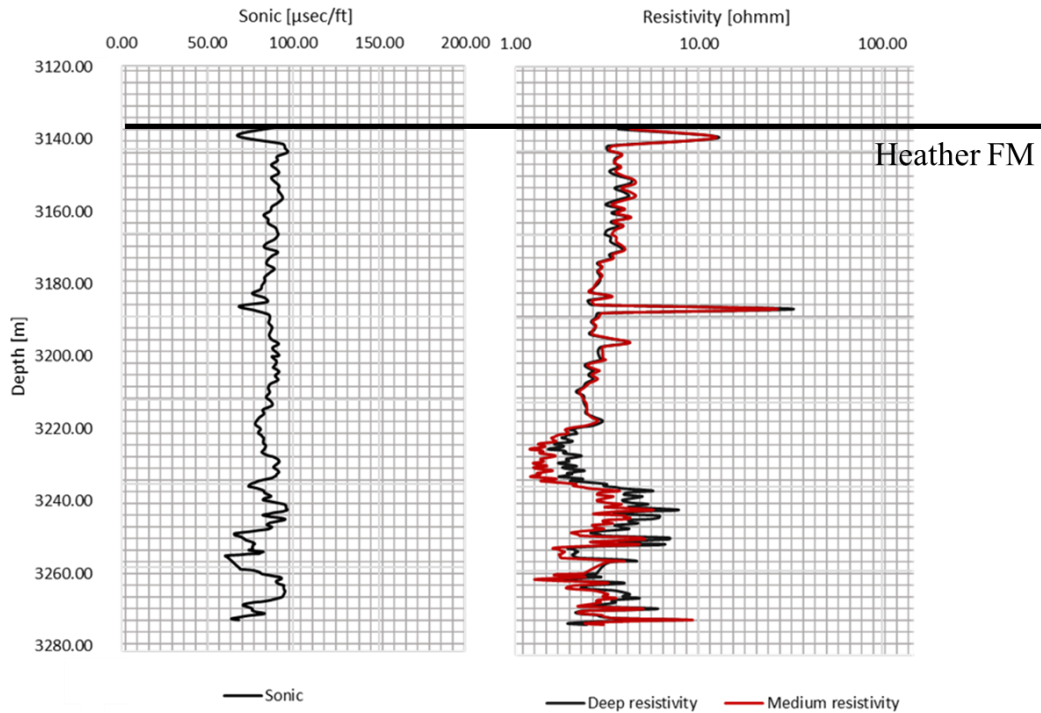


Figure 5.13 From left to right the Sonic log and the resistivity logs (deep and medium resistivity) from well 15/12-A 8.

5.4 Data preparation

The original data set and the selected model parameters can have a high impact on the accuracy, performance, and quality of the proposed model in the thesis. Consequently, the data pre-processing and removing outliers are essential. The quality check of the data is required on the original data set, such that that the data set is representative. The data set used in the model consists of wells from the Varg field. The total number of wells in this study is of 6 wells, five wells for the training, validation and test set, and one well used to test the calibrated model. The total number of data points from the 6 wells are: 1094 data points of oil saturation and water saturation, and 1697 data points of porosity. Each datapoints have log readings of oil saturation, water saturation and porosity.

An outline of the statistics in the data set is important to identify outliers and the distribution of the data. The data points that are fed to the models are the pre-processed data points, the amount of data from each of the wells are shown in Figure 5.14, Figure 5.15 and Figure 5.16. The data set for the models are randomly divided into training set, validation set and test set with a user-defined percentage. The number of data points for the training process of the calibrated models are 1052 for oil saturation and water saturation. The data set is divided into

796 data points (75%), 156 data points (15%), and lastly 107 data points (10%), respectively. The number of data points for the porosity prediction from the wells are 1697. The data set is divided into 1270 data points (75%), 254 data points (15%), and lastly 173 data points (10%), respectively.

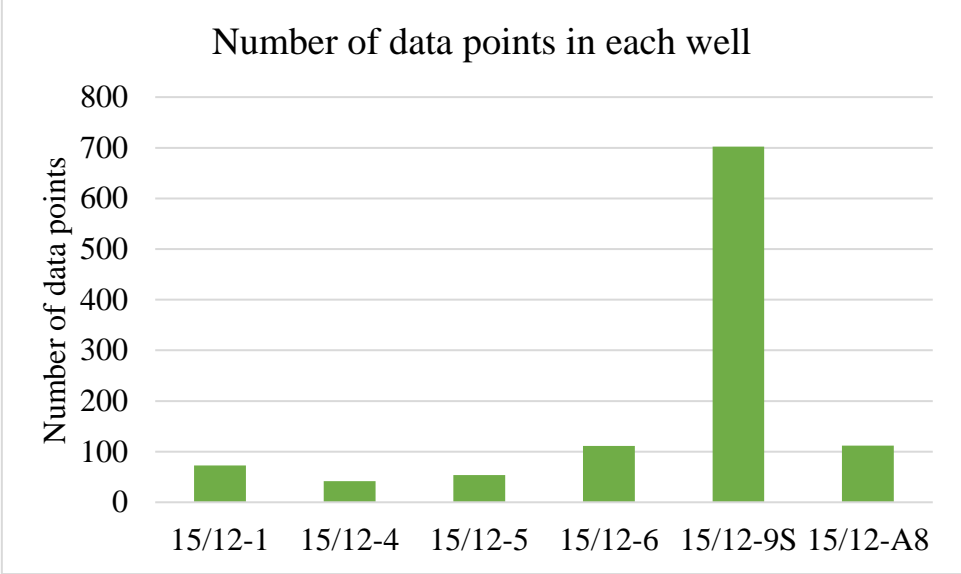


Figure 5.14 The amount of data points of oil saturation in each wells.

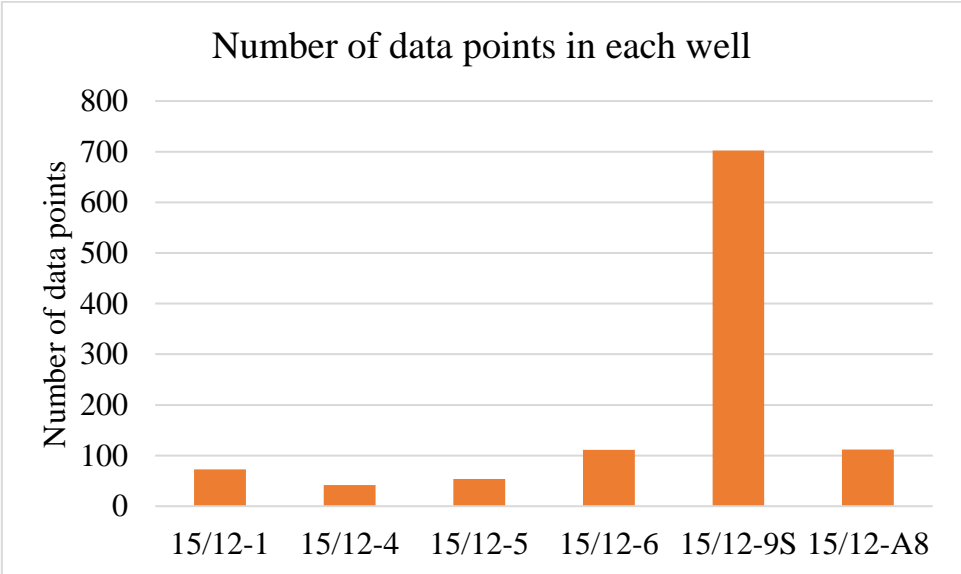


Figure 5.15 The amount of data points of water saturation in each wells.

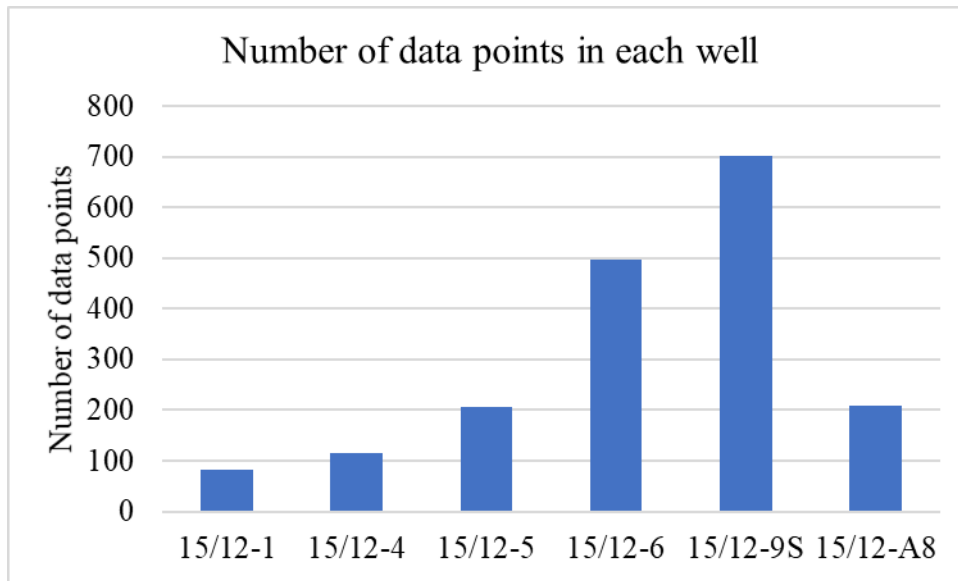


Figure 5.16 The amount of data points of porosity in each wells.

The uncertainty in the match between the well logging data and core data was calculated using the difference between the well logging depth and core depth. The resulting depth error was calculated as less than 0.05 m. The minimum and maximum difference between the well logging depth and core depth is shown in Table 5.9. In addition, some values in the data set are set as lower than 0.01, 0.02 or 0.04. To avoid selecting values, a random distribution range between 0 and 0.01, 0.02 or 0.04 is used.

Table 5.9 The errors between the cored depth and the well logging depths

	Difference [m]
Min	0.0002
Max	0.09
Average	0.05

The data set for the model have well logging data for Caliper (CALI), Gamma-ray (GR), Self-Potential (SP), Sonic Transit time (AC or DT), Bulk density (RHOB), Density correction (DRHO), Neutron porosity (NPHI), Resistivity log (RD) and Resistivity log (RM). In addition, the data set have core data for oil saturation (So), water saturation (Sw) and porosity (POR). The number of datapoints in the oil saturation and water saturation is less than the data points in the porosity, due to the frequency of the measurements.

For large data set the difficulty in validating the results increases, since the data distribution of the data samples may vary largely. For data sets with many variations in the data distribution

the constructed model could become overfitted. The constructed model may mistake the variations as important variance which leads to larger errors (Hira & Gillies, 2015). Other variations are from measurement errors or the natural variation in the data set, such as variations in the data distribution of wells. Identifying noise and which features to include is necessary to improve the efficiency of the model. There are two techniques: feature selection (selecting a smaller subset) and feature extraction (e.g. filtering) (Hira & Gillies, 2015). Thus, the data set is pre-processed to avoid or reduce unnecessary complexity of the problem the model tries to address (Kothari & Oh, 1993).

Scaling, normalising, or rescaling the data set are common methods to handle data. In this study the data is normalised using Eq. 5.1. The technique is called Min-Max normalization where the data is fitted to the boundary [-1, 1] in this study. The shape of the original data distribution will remain the same since the z-score is preserved (Patro & sahu, 2015).

$$data_{rescaled} = -1 + \frac{data - data_{min}}{data_{max} - data_{min}}(1 - 1) \quad \text{Eq. 5.1}$$

Prior to the data being loaded to the model outliers in the data set is identified. The outliers can either be removed or replaced (e.g. with mean value). Boxplot displays a statistical summary of variables such as mean, median, upper quartile (P75), lower quartile (P25), upper and lower extremes (P10 and P90). The upper and lower extremes are also called whisker length, the values extend from the inter quartile range (IQR) from the upper and lower quartiles. The values that are outside the upper and lower extremes are considered as outliers Eq. 5.2.

$$Outlier = if \begin{cases} value > P75 + 1.5(P75 - P25) \\ value < P25 - 1.5(P75 - P25) \end{cases} \quad \text{Eq. 5.2}$$

The statistical distribution of each of the well logs and core data is conducted (Figure 5.17 and Figure 5.18). The outliers are indicated by red crosses in the figures. The boxplots indicate that some outliers appear in the well logs and the core data. The number of outliers for well 15/12-1, 15/12-6S and 15/12-9S are significant in the NPHI, RHOB, RD, RM. The presence of hydrocarbons and fluids in these logs affect the well log readings. Thus, outliers in these wells can indicate the presence of hydrocarbons or fluids. The data points may not have an impact on the model performance since the values are within specific ranges. There are no extreme, impossible values identified in the boxplots such as porosity values above 100 or other extremely high unnatural values. The outliers in the data set contains valuable information and will be kept, no modification of the data set is needed.

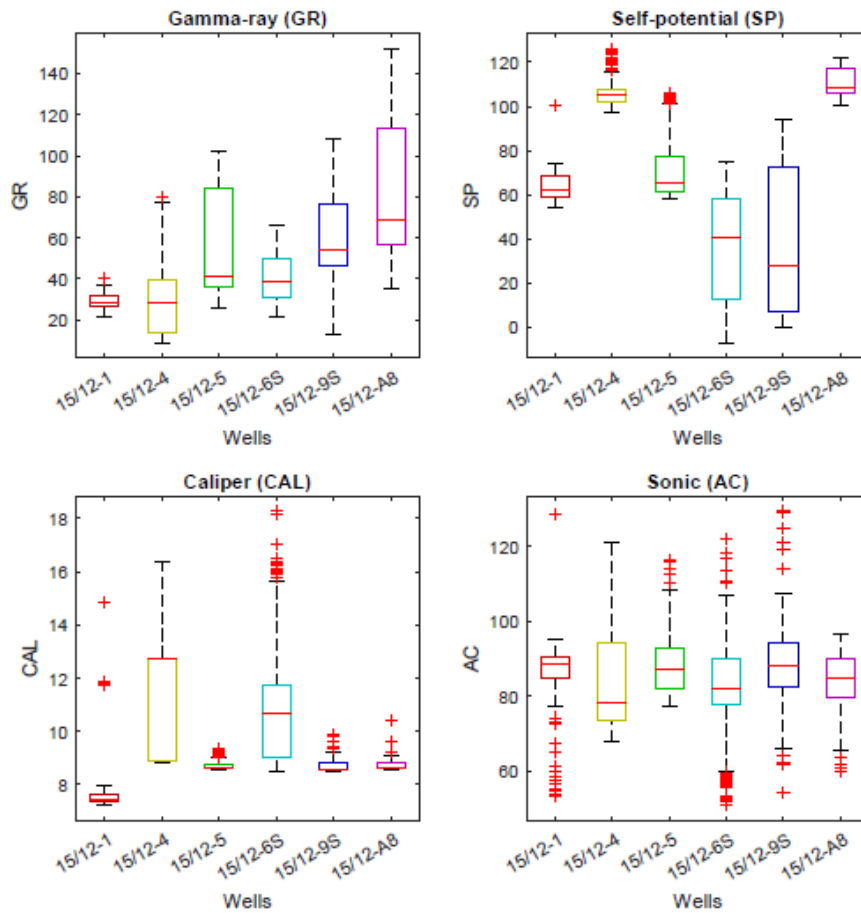


Figure 5.17 A statistic summary of the outliers in the Gamma-ray (GR), Self-Potential (SP), Caliper (CAL) and Sonic (AC) log in each wells.

From observation there are few outliers in the gamma-ray log and self-potential log and the sonic log shows a few outliers in all the wells except well 15/12-4 (Figure 5.17).

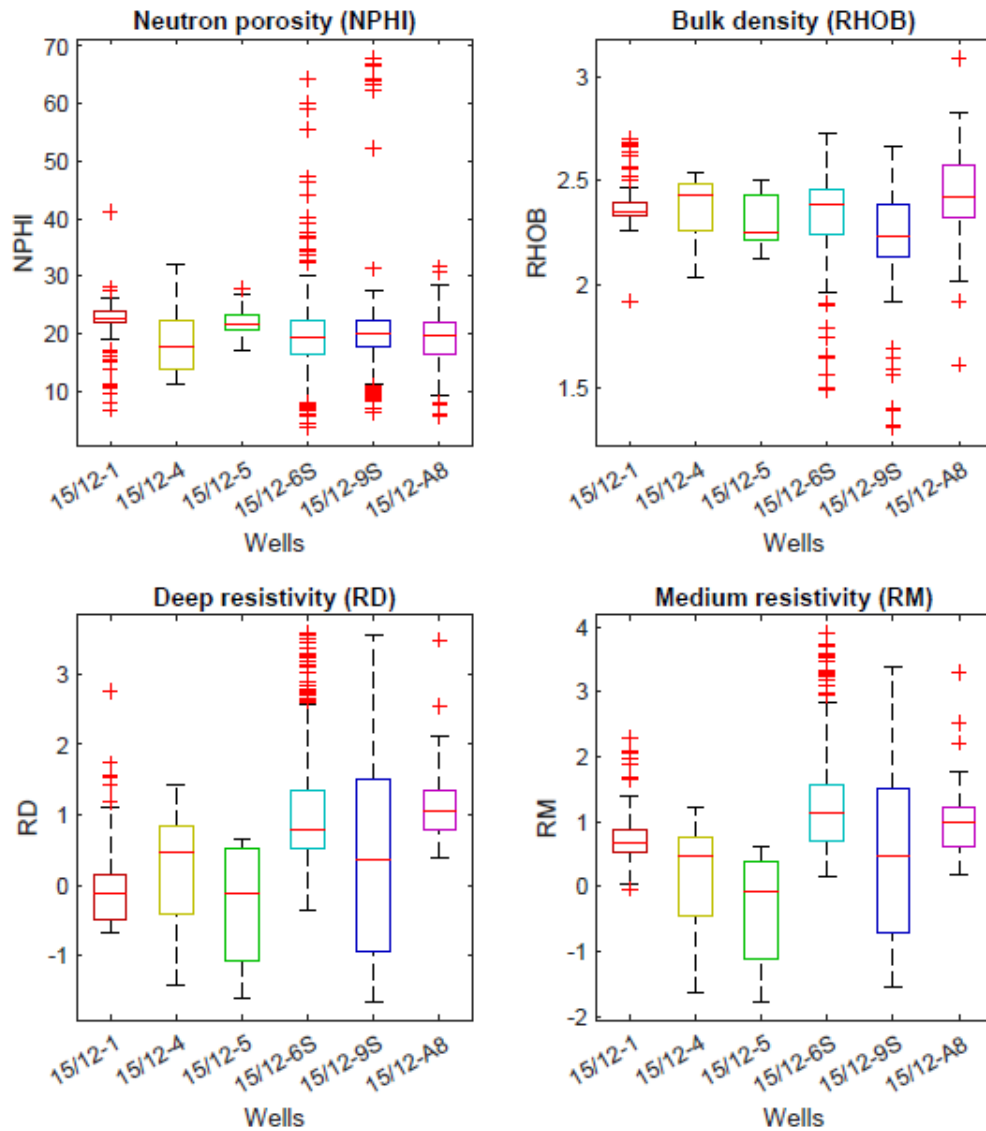


Figure 5.18 A statistic summary of the outliers in the Neutron porosity (NPHI), Bulk density (RHOB), deep resistivity (RD) and medium resistivity (RM) log in each wells.

The boxplot of the NPHI, RHOB, RD and RM shows many outliers in well 15/12-6S and 15/12-9S, and some in well 15/12-1.

The statistical distribution of the desired parameters, oil saturation, water saturation and porosity is shown in Figure 5.19.

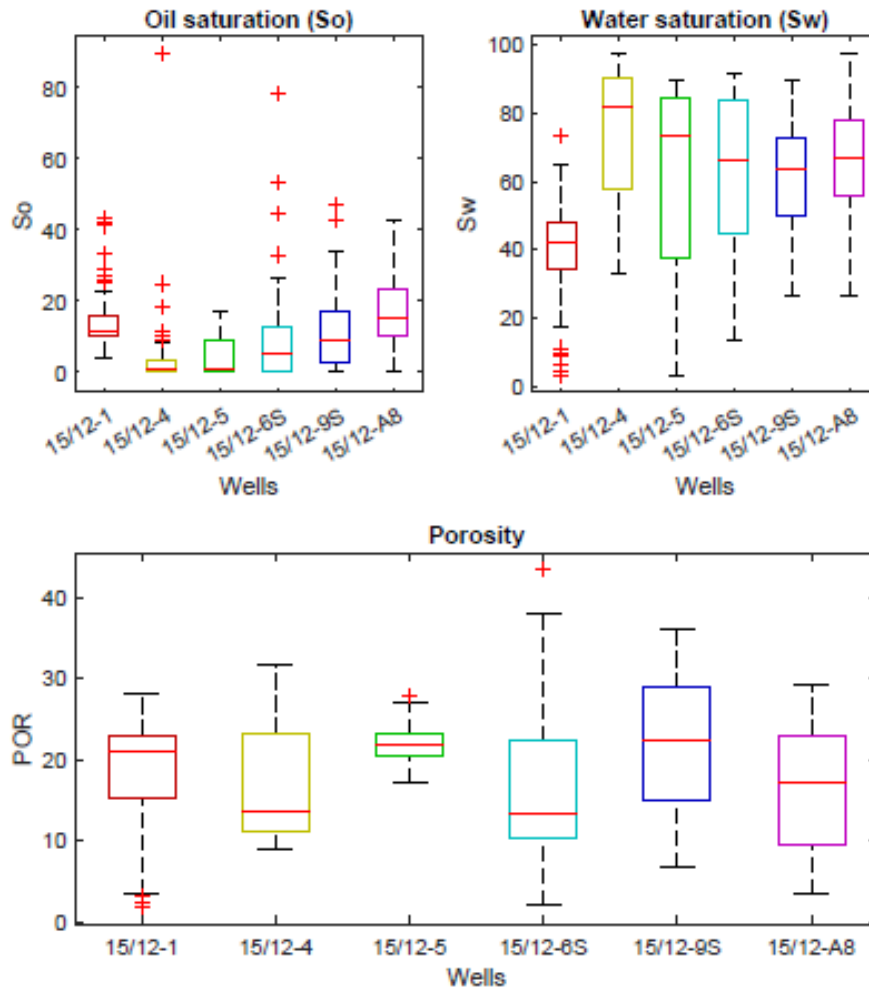


Figure 5.19 A statistic summary of the outliers in the desired parameters: Oil saturation, water saturation and porosity in each wells.

The boxplot of the desired values shows high water saturation and oil saturation, the highest water saturation is above 95% in well 15/12-4.

The boxplots are used as a visual representation of the data distribution, and a quantitative measure is shown using numeric values (Table 5.10, Table 5.11, Table 5.12, Table 5.13 and Table 5.14). In Appendix A, the data distribution of the individual wells and petrophysical logs is summarised. The original petrophysical data of 5 wells excluding well 15/12-4. is an overview of the data used prior to dividing it into training set, validation, and test set.

Table 5.10 Statistical indexes of the original data used in the model without well 15/12-4

	max	min	mean	P10	P25	P50	P75	P90
GR [API°]	152.10	6.22	38.56	17.71	20.74	28.85	47.93	68.07

SP [mV]	121.78	-7.04	45.38	14.00	18.73	25.16	65.96	105.92
CAL [in.]	67.95	6.22	14.00	8.60	8.65	11.22	18.89	22.14
AC [μs/sec]	128.56	6.22	58.40	17.71	20.74	76.36	86.40	92.45
NPHI [frac.]	67.95	3.74	20.26	15.25	17.59	20.23	22.58	24.68
RHOB [g/cm³]	67.95	1.49	9.61	2.21	2.33	2.50	18.89	22.14
Log (RD) [Ωm]	4.22	-1.61	1.62	-0.12	0.57	1.45	2.95	3.11
Log (RM) [Ωm]	4.22	-1.78	1.71	0.30	0.70	1.50	2.96	3.11
So [frac.]	97.23	2.90	33.02	16.88	18.51	21.91	43.43	74.91
Sw [frac.]	78.40	0.00	17.44	5.77	13.90	18.43	21.40	24.14
POR [frac]	67.95	1.70	18.69	9.50	14.77	19.65	22.49	25.30

Table 5.11 The statistical indexes of the data used in the training set of the model

	max	min	mean	P10	P25	P50	P75	P90
GR [API°]	152.10	6.22	38.20	17.75	20.74	28.50	47.50	66.92
SP [mV]	121.40	-7.04	44.86	13.99	18.69	24.90	65.82	105.51
CAL [in.]	67.95	6.22	14.08	8.60	8.67	11.26	19.05	22.33
AC [μs/sec]	128.56	6.22	58.00	17.75	20.74	75.07	86.20	92.51
NPHI [frac.]	67.95	3.74	20.24	14.96	17.61	20.26	22.65	24.65
RHOB [g/cm³]	67.95	1.65	9.73	2.21	2.33	2.52	19.05	22.33
Log (RD) [Ωm]	4.22	-1.61	1.63	-0.11	0.57	1.48	2.96	3.11
Log (RM) [Ωm]	4.22	-1.78	1.72	0.30	0.71	1.53	2.96	3.11
So [frac.]	97.23	2.90	32.41	16.86	18.46	21.79	42.26	73.30
Sw [frac.]	78.40	0.00	17.66	6.01	14.09	18.56	21.45	24.26
POR [frac]	67.95	1.70	18.68	9.50	14.36	19.61	22.53	25.33

Table 5.12 The statistical indexes of the validation set of the model

	max	min	mean	P10	P25	P50	P75	P90
GR [API°]	144.44	9.91	39.64	17.68	20.56	31.17	49.46	67.50
SP [mV]	121.78	-5.58	46.92	15.88	19.11	35.09	66.44	106.23

CAL [in.]	27.15	7.32	13.67	8.59	8.67	11.31	18.27	21.37
AC [μs/sec]	122.16	9.91	59.42	17.68	20.56	79.13	86.57	91.88
NPHI [frac.]	58.93	5.99	20.25	15.78	17.55	19.94	22.15	24.70
RHOB [g/cm ³]	27.15	1.56	9.04	2.21	2.33	2.49	18.27	21.37
Log (RD) [Ωm]	3.37	-1.59	1.53	-0.25	0.55	1.33	2.94	3.09
Log (RM) [Ωm]	3.70	-1.78	1.62	0.29	0.69	1.37	2.94	3.08
So [frac.]	94.80	9.91	35.07	17.18	18.74	22.02	47.58	77.69
Sw [frac.]	53.30	0.00	16.34	1.03	13.05	18.05	20.82	23.37
POR [frac]	35.40	3.51	18.03	9.49	14.50	19.23	21.93	23.60

Table 5.13 The statistical indexes of the testing set of the model

	max	min	mean	P10	P25	P50	P75	P90
GR [API°]	80.17	9.00	31.48	10.79	13.42	28.01	39.79	65.66
SP [mV]	126.14	97.24	107.09	98.49	102.24	104.94	107.76	120.11
CAL [in.]	16.35	8.81	11.59	8.84	8.91	12.70	12.75	15.08
AC [μs/sec]	121.29	67.73	85.11	72.08	73.61	78.19	94.31	114.59
NPHI [frac.]	32.09	11.26	18.44	12.41	13.81	17.87	22.43	25.37
RHOB [g/cm ³]	2.54	2.03	2.37	2.15	2.26	2.43	2.48	2.50
Log (RD) [Ωm]	1.42	-1.42	0.21	-1.17	-0.40	0.48	0.84	1.10
Log (RM) [Ωm]	1.21	-1.63	0.14	-1.23	-0.45	0.47	0.76	0.93
So [frac.]	97.70	32.70	75.22	40.56	57.80	81.70	90.40	94.89
Sw [frac.]	89.70	0.00	5.42	0.00	0.00	0.85	3.60	13.43
POR [frac]	31.60	8.90	17.08	10.01	11.25	13.70	23.30	29.19

Table 5.14 The statistical indexes of the well 15/12-4

	max	min	mean	P10	P25	P50	P75	P90
GR [API°]	40.02	21.29	29.02	24.01	26.39	28.42	31.72	35.20
SP [mV]	100.40	54.05	63.87	57.40	58.92	61.82	68.81	73.15

CAL [in.]	14.86	7.21	7.98	7.32	7.39	7.45	7.62	11.76
AC [μs/sec]	128.56	53.20	84.47	60.94	84.96	88.68	90.53	91.19
NPHI [frac.]	41.07	6.82	21.85	15.28	21.88	22.78	23.98	24.58
RHOB [g/cm ³]	2.70	1.92	2.38	2.29	2.33	2.35	2.40	2.56
Log (RD) [Ωm]	15.90	0.50	1.35	0.59	0.62	0.88	1.17	2.10
Log (RM) [Ωm]	9.84	0.96	2.47	1.34	1.68	1.97	2.40	4.36
So [frac.]	73.30	3.10	39.50	10.82	34.15	42.30	47.68	52.12
Sw [frac.]	43.30	4.30	14.38	8.76	10.48	11.70	15.63	25.56
POR [frac]	28.00	1.70	18.28	7.90	15.20	20.85	23.00	23.83

The statistical summary of the oil saturation, water saturation and porosity can be found in APPENDIX B. The saturation data is usually measured in the reservoir interval, missing data in the saturation data may be due to how often the measurements are done between each sample and the location of the reservoir interval. A visual representation of the saturations and the distribution is shown in Figure 5.20. Well 15/12-4 and 15/12-5 have the highest observed oil saturations, the water saturation in all the wells varies greatly between 0 and around 100. Well 15/12-4 and Well 15/12-A8 have water saturations that are above 95%, with 4 data points and 1 data point, respectively.

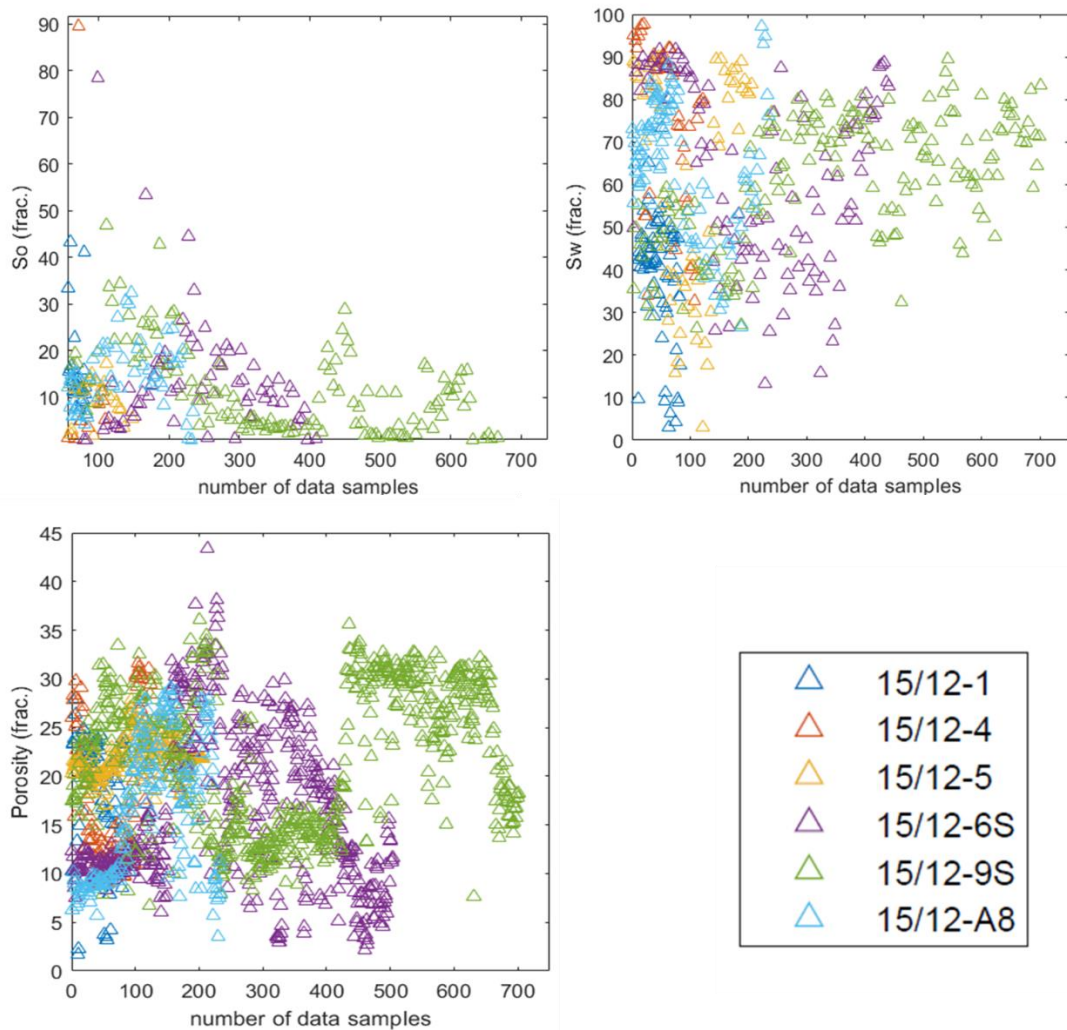


Figure 5.20 The data distribution of the oil saturation, water saturation and porosity from core data.

5.5 Feature selection

Feature selection is a useful process in machine learning, the main aim of feature selection is selecting input features that are relevant for the desired target or problem. Usually feature selection is applied to increase the accuracy and performance of the model by removing irrelevant features. A good feature selection can increase the performance of the model with lower errors and avoid overfitting. Correlation coefficients such as Pearson, Spearman and distance correlation are methods that can be applied to evaluate the relevance between features in the data.

1. Pearson correlation

Pearson correlation coefficient measures the statistical relationship between two variables. The correlation between the input values in the model can be evaluated for the feature selection and

the direction of the relationship (positive or negative). In addition, the statistical relationship between the desired values (T) and the predicted values (Y) can be evaluated. The Pearson correlation coefficient assumes that the variables are normally distributed (parametric), a linear(straight-line) relationship between the variables and that the data is distributed along the regression line.

1. Pearson Coefficient of determination (R):

$$R_{Pearson} = \frac{\sum(T_i - T_{mean})(Y_i - Y_{mean})}{\sqrt{\sum(T_i - T_{mean})^2} \sqrt{\sum(Y_i - Y_{mean})^2}} \quad \text{Eq. 5.3}$$

The Pearson correlation coefficients lies in the range [-1,1], the interpretation of the correlation coefficients values:

- -1, means negative correlation. The values of one variable increases as the other value decreases.
- 0, means no linear correlation between the variables
- 1, means positive correlation, when one variable increase along with the other valuable.
- Other values such as ± 0.8 and ± 0.6 indicate moderately strong and fairly strong correlation, respectively.

2. Spearman linear rank correlation

Spearman's rank correlation coefficient is used to evaluate two variables that can be related in a nonlinear relationship, and the strength of the relationship in the data distribution. It measures the degree of association between the two variables in a range [-1,1]. The Spearman rank correlation coefficient does not have any assumptions about the distribution of the data, such as the Pearson correlation.

$$R_{Spearman} = 1 - \frac{6 \sum d^2}{n_{pairs}(n_{pairs}^2 - 1)} \quad \text{Eq. 5.4}$$

The Spearman linear rank correlation uses the differences between ranks (d) and the number of data pairs (n_{pairs}). The interpretation of the Spearman correlation is the same as the Pearson correlation: ± 1 strong positive or negative correlation, ± 0.8 moderately strong positive or negative correlation, ± 0.6 moderately strong positive or negative correlation. 0 indicates no correlation between the data.

3. Distance correlation.

Measures the dependence between the desired value (T) and predicted value (Y), the distance are based on certain Euclidean distances between the data samples (Székely et al., 2007). The dependence between each variable is given in the range [0,1], where 0 means that the variables are totally independent.

$$Dcorr(T, Y) = \frac{\sqrt{Dcov(T, Y)}}{\sqrt{DCov(T, T)DCov(Y, Y)}} \quad \text{Eq. 5.5}$$

5.5.1 Statistical evaluation of performance

Several statistical evaluation methods are applied in the study to assess the relationship between the reference data (T) and the predicted data. The methods are used to evaluate the performance of the model.

1. Mean squared error (MSE)

$$MSE = \frac{\sum \|Ti - Yi\|^2}{\text{Number of samples}} \quad \text{Eq. 5.6}$$

2. Root-Mean squared error (RMSE)

$$RMSE = \sqrt{\frac{\sum \|Ti - Yi\|^2}{\text{Number of samples}}} \quad \text{Eq. 5.7}$$

3. Sum of squared error (SSE)

$$SSE = \sum \|Ti - Yi\|^2 \quad \text{Eq. 5.8}$$

4. Mean average absolute error (MAE)

$$MAE = \frac{\sum \|Ti - Yi\|}{\text{Number of samples}} \quad \text{Eq. 5.9}$$

5.6 MLP-LM model and design

The study has selected a multilayer network architecture which is trained with a backpropagating algorithm, Levenberg-Marquardt algorithm, to improve the accuracy of the calibrated model. The need for a robust and fast method such as the Levenberg-Marquardt

algorithm, is needed due to the time-constraints of the thesis. The main aim of the Levenberg-Marquardt algorithm is minimising the squared errors in the network. The network is trained with the Levenberg-Marquardt algorithm using log-sigmoid functions in the hidden layers. The mathematical equations applied in the model construction is specified in section 4.3.3. The well data is pre-processed, and quality checked prior to being loaded to the constructed model. A simplified flow chart of the Levenberg-Marquardt parameter optimisation is shown in Figure 5.21. The Jacobian matrix (Eq. 5.10) and the gradient is calculated from Eq. 4.9 and Eq. 4.10, respectively. The sum of squared errors are calculated using Eq. 4.3, The weights and bias are updated according to Eq. 4.11, and the optimisation does not end until a performance goal is met.

$$J(x) = \begin{bmatrix} \frac{\partial E_1(x)}{\partial x_1} & \frac{\partial E_1(x)}{\partial x_2} & \dots & \frac{\partial E_1(x)}{\partial x_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial E_i(x)}{\partial x_1} & \frac{\partial E_i(x)}{\partial x_1} & \dots & \frac{\partial E_i(x)}{\partial x_k} \end{bmatrix} \quad \text{Eq. 5.10}$$

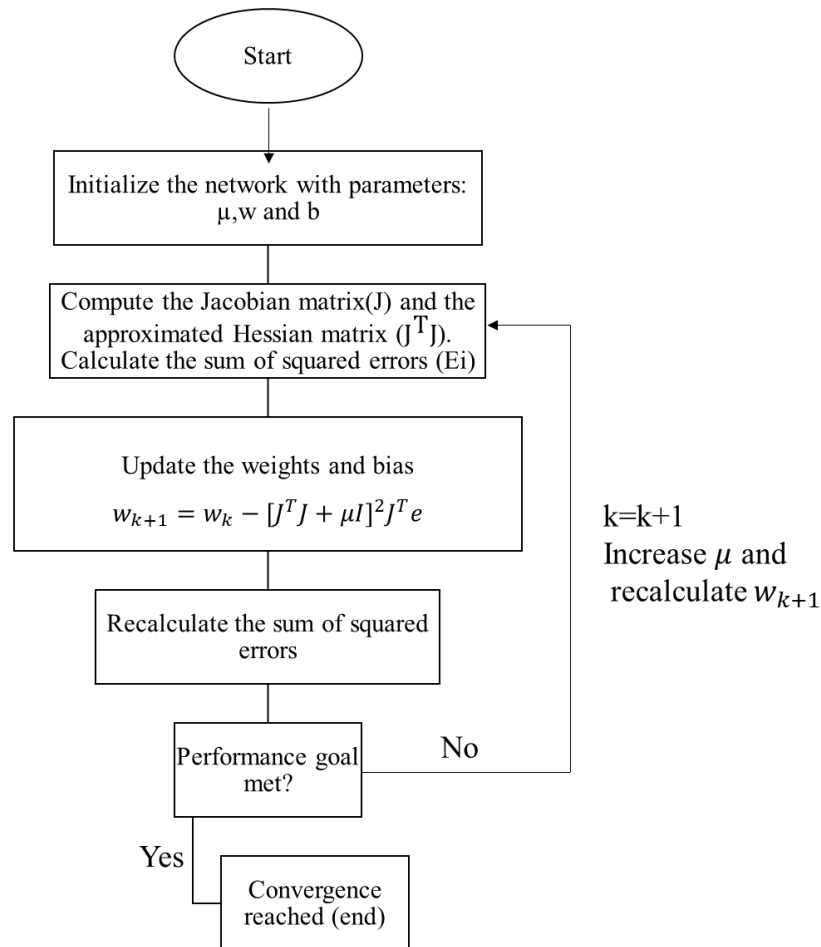


Figure 5.21 Flow chart of the Levenberg-Marquardt parameter optimisation.

The training of the network does not end until one of these performance goals is reached:

- The performance of the model, the sum of squared errors is minimised to 0.
- The control parameter (μ) is adjusted during the training process. If the control parameter exceeds the threshold value of μ , which is 10 in this study the training process stops.
- The performance on the validation set has increased more than the number of maximum failures since the last time it decreased
- The number of epochs for training the model is reached, 1000 in the study.
- A time constraint where the training stops when the maximum time is exceeded. This is set to infinity.
- If the performance gradient is less than the value of the minimum performance gradient (1E-8).

If the conditions are not reached in one epoch the epoch increases by 1, $k=k+1$ epoch (Figure 5.21). The parameter selection in the model is important to obtain convergence of the model (Table 5.15). The controlling (μ), is an important parameter to achieve convergence of the model.

Table 5.15 A summarised overview of the parameter selections for the Levenberg-Marquardt

Parameters	Selected values
μ	0.001
μ_{increase}	0.1
μ_{decrease}	10
μ_{max}	1E10
Maximum failures	6
Minimum gradient	1E8
Performance goal	0

A simplified pseudo-code of the work in the study is presented in Figure 5.22, the first loop creates 10 000 calibrated model and the second loop creates random combinations of hidden layers and neurons. The global seed is set for reproducibility of the calibrated model.

```

Load pre-processed data set
%Define input data and output data
%Divided the data set into training set, validation set and test set.
%Configure for loop to construct the different models
for main_loop=1:10 000 number of models (10 000)
    for sub_loop=1:max(main_loop)
%Configure a minor for loop to select random hidden layer sizes (3 to 5)
with a random number of neurons (1 to 10) in each layer
        Select network architecture: feedforward network
        select Levenberg-Marquardt parameters
        %Parameters are optimised using the Levenberg-Marquardt algorithm
        Errors on the data set are calculated and the weights are updated
according to the algorithm.
        Print ('The best results on the validation set')
    end
end

```

Figure 5.22 A Pseudo-code of the workflow used to construct the calibrated models.

Another method using the Particle Swarm Optimisation algorithm was tested, however the algorithm performed poorly on the data set and the training time was slow with few iterations (50 iterations). Thus, the Particle Swarm optimisation algorithm was not proceeded with any further.

5.6.1 Network architecture

The data set used in the model consists of input variables from the petrophysical logs (X) and target or reference variables, porosity and fluid saturations (oil and water) (T).

$$\begin{bmatrix} X1 \\ X2 \\ \vdots \\ Xi \end{bmatrix} = \text{the values from petrophysical logs, input values (X)}$$

$$\begin{bmatrix} T1 \\ T2 \\ \vdots \\ Ti \end{bmatrix} =$$

the porosity and fluid saturations from core data , targeted output values (T)

The network is calibrated with the target values to predict a porosity value. The training algorithm Levenberg-Marquardt is selected to optimise the model. The network is calculated from the matrix with X input values, bias, and weights (input and layer weights) and calibrated with the desired target values (Porosity, water saturation and oil saturation). Lastly, a training algorithm in each layer of the network to train the network. Eventually the prediction value Y is derived from the network model using input variables X.

Usually the number of hidden layers and nodes are tested through trial and error, in this study a method for randomly selecting number of neurons and layers of a certain range in the network is tested. The range for the number of layers is set to 3 and 5 hidden layers, while the number of nodes varies in a range from 1 to 10. In each layer the number of nodes is randomly generated to check which combinations that are optimal.

The hidden layer size can be represented as a row vector where each layer is represented as number of neurons nL where L is the layer number (Table 5.16). The input data is transported from left to right through the first layer with n nodes (n1) to the fifth layer with n nodes (n5), and lastly resulting in a predicted output data. The number of layers and nodes are selected with the number of unknown variables in mind as it affects the generalisation abilities of the model.

Table 5.16 The number of layers in the network and a the vectors of the number of neurons in each layer.

Layers	Hidden layers (nL)
3	= [n1 n2 n3]
4	= [n1 n2 n3 n4]
5	= [n1 n2 n3 n4 n5]

The number of unknown variables in a network with input variables (I), three hidden layers (n1, n2, n3) and the number of output variables O and the sum of biases from each layer is calculated from the following formula:

$$\theta = I * L1 + [L1 L2 L3][L2 l3 O] + b \quad \text{Eq. 5.11}$$

The results from each generated network is based on the correlation coefficients from the random selector is selected as the best network architecture. The main parameters for the network architecture of the model is summarised below.

Table 5.17 A summary of the network architecture

Parameter	Parameter selection
Network type	Feedforward network
Number of layers	3-5
Layers with bias	1: Yes 2: Yes 3: Yes 4: Yes 5: Yes
Input to Layer connection(s)	Yes (from input layer to first layer)
Layer Connection(s)	Yes (from previous layer to following layer)
Layer to Output connection(s)	Yes (from last hidden layer to output layer)

6 Results

In general, the model accuracy is highly affected by the input features in the training and validation sets since this trains and tunes the model to a specific range. The water saturation and oil saturations in the well selected to be excluded had a water saturation above 95%. The high-water saturation from well 15/12-4 might impact the quality of the model during the blind well test since the model is calibrated against lower water saturation (<95%).

The input features may affect the model's ability to generalise well, so it is necessary to evaluate the input features to the model. In this study, the correlation between the petrophysical data and the core data is measured using correlation coefficients. The feature selection in this study relies on selecting higher correlation coefficient values, petrophysical logs that may be more relevant for the specific predictions. The input features for each desired values, oil saturation, water saturation and porosity, are selected from the correlation relationships. Thus, the petrophysical logs that have higher correlation coefficient values for the individual desired values are selected as input features for the predictions in the model.

The correlation is the only correlation where the correlation is classified as independent only when the value is equal to 0. The distance correlation makes no assumptions about the data, making it a more generalised approach than Pearson correlation or Spearman correlation. The Pearson correlation assumes a normal distribution for the data set (parametric), while Spearman correlation and Distance correlation is non-parametric. The Pearson correlation, Linear Spearman rank correlation and distance correlation are calculated from the complete data set of five wells, and the training and validation set combined since these are the inputs that calibrate and tune the model. The correlations are used to investigate the relationships between the petrophysical logs, oil saturation, water saturation and porosity to determine the inputs in the model. The correlation of the data set used in the training, validation and test set can be seen in APPENDIX D.

The petrophysical logs have different correlations with the oil saturation, water saturation and porosity (Figure 6.1). This section focuses on the correlation between the petrophysical logs and the oil saturation, water saturation and porosity, not the correlation between the reservoir parameters. The Pearson correlation of the oil saturation (S_o) has the strongest positive correlation with AC and NPHI and the strongest negative correlation with RHOB. The correlation with the GR log is weakest. The strongest positive correlation of the oil saturation is ~ 0.49 , and the strongest negative correlation is ~ -0.51 . The water saturation (S_w) correlation

indicates that the strongest correlation is with the SP and CAL logs. The water saturation correlates poorly with the GR, AC, RD and RM logs. The strongest positive correlation of the water saturation is ~0.4, and the strongest negative correlation is ~-0.06. The Pearson correlation of the porosity (POR) shows that the strongest positive correlation is with the AC and NPHI, while the strongest negative correlation is with the RHOB. The negative correlations of the GR and SP are probably due to high readings in both logs, usually indicate less porous rocks such as shale where porosity is lower. The strongest positive correlation of the porosity is ~0.71, and the strongest negative correlation is ~-0.75.

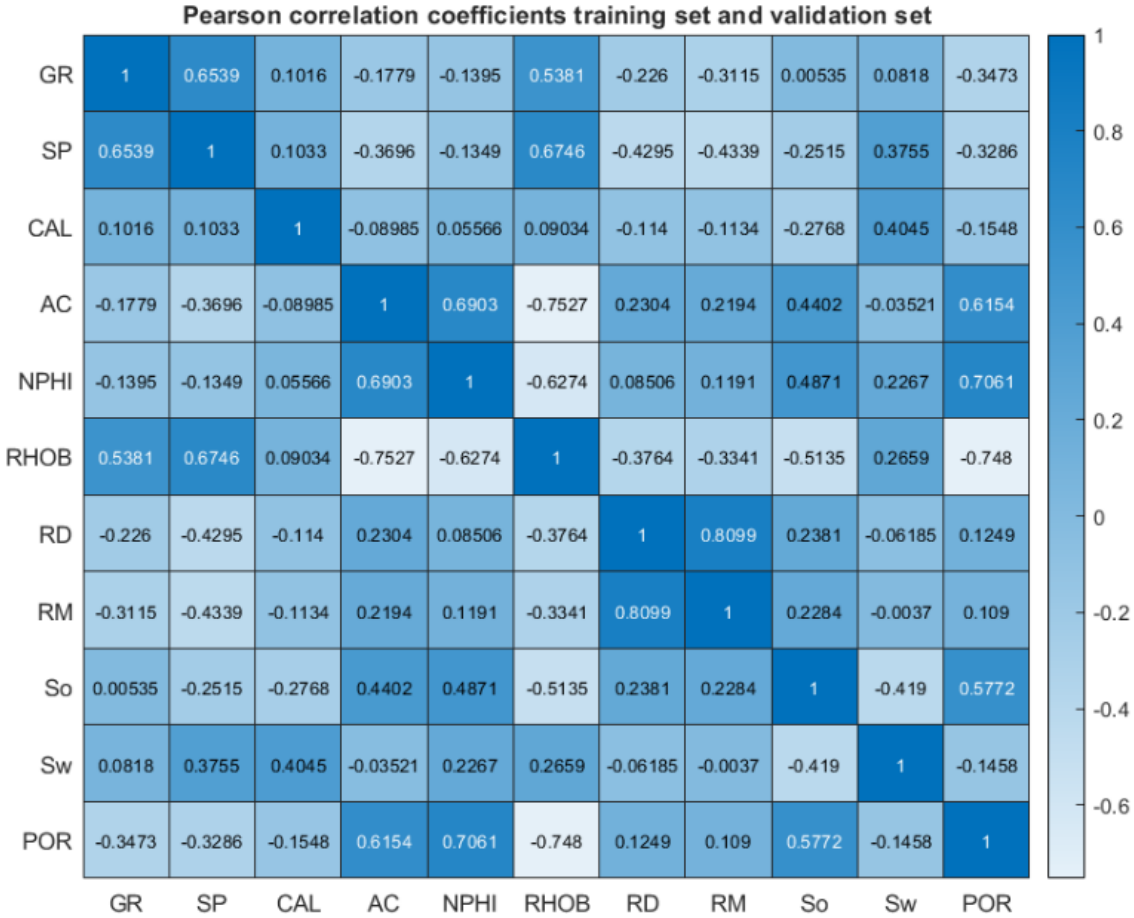


Figure 6.1 Pearson correlation results of the training set and validation set

The Spearman linear rank correlation is presented in Figure 6.2. The Spearman correlation of the oil saturation shows the strongest positive correlations with AC, NPHI, RD and RM, while the strongest negative correlations are observed with SP, CAL and RHOB. The strongest positive correlation is ~0.53 with AC, and the strongest negative correlation is ~-0.58 with RHOB. The water saturation shows that the strongest positive correlation is with the NPHI, while the strongest negative correlations are with the GR, ~-0.55 and ~-0.21, respectively. The

correlation of the porosity shows strong positive correlations with AC and NPHI. The strongest negative correlations are observed with RHOB (~-0.71) and SP (~-0.46).

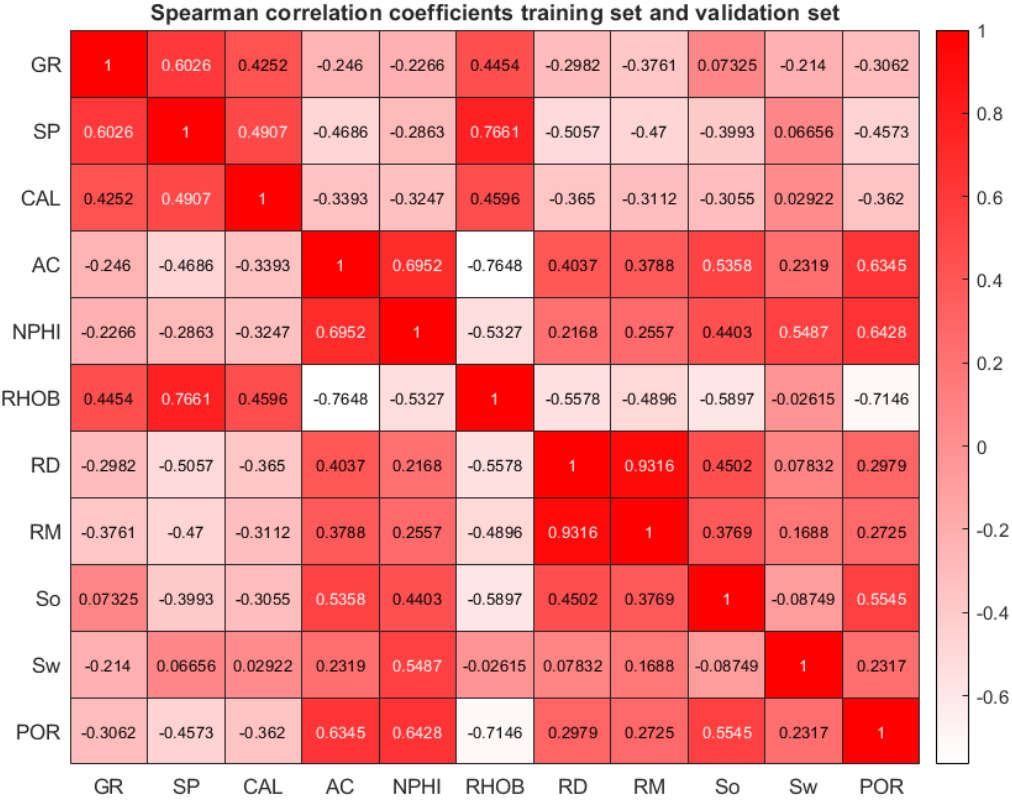


Figure 6.2 Spearman correlation results of the training set and validation set

The results from the distance correlation for the oil saturation shows the strongest positive correlations with CAL, AC, NPHI, and RHOB (Figure 6.3). In general, the RHOB shows a better correlation than the NPHI with the oil saturation, water saturation and porosity. The resistivity logs appear to have the strongest correlation with oil saturation compared to water saturation and porosity in all the correlations (Pearson, Spearman and Distance). The distance correlation of the water saturation shows that the strongest positive correlation is with SP, CAL and RHOB. Lastly, the best distance correlation of the porosity is the CAL, AC, NPHI and RHOB.

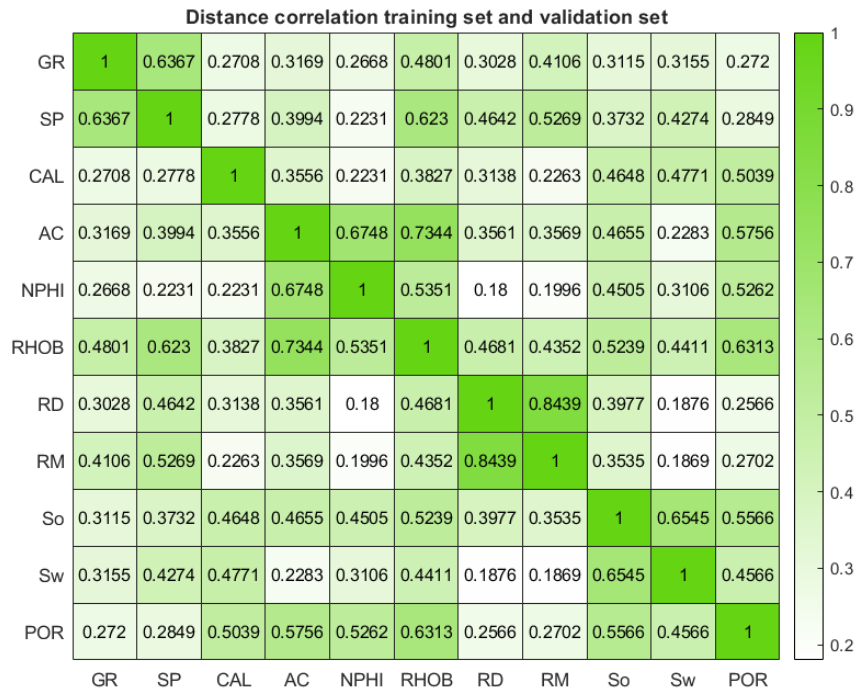


Figure 6.3 Distance correlation results of the training set and validation set

The feature selection results from the correlation analysis from Pearson, Spearman and Distance correlation had different correlations with each of the desired parameters. The four petrophysical logs that displayed the most relevant correlation with the desired parameters were selected as input variable (Table 6.1).

Table 6.1 The feature selections obtained from the correlation analysis of the oil saturation, water saturation and porosity.

First feature selection	
Parameter	Input features petrophysical logs
Oil saturation	CAL AC, NPHI, RHOB
Water saturation	SP, CAL, NPHI, RHOB
Porosity	CAL, AC NPHI, RHOB

The selected input features of the model are selected because it is most likely that a relationship exists between these petrophysical logs and oil saturation, water saturation and porosity. The Caliper log correlates well with the water saturation and the porosity, however no relationship between these parameters are known.

Secondly, a trial-and-error approach by manually adding and removing the petrophysical logs one at the time, that had the lowest correlation values with the desired parameters. By observing

the average changes in R^2 on the validation set of 10 different calibrated models for each desired parameter. The method starts with all the petrophysical logs and removing them one at the time. If the R^2 increased the variable was removed, if the R^2 decreased it was included. This process continued until no improvement was possible, and variables were added again one at the time to check how other combinations of the petrophysical logs affected R^2 . The results from this feature selection is summarised in (Table 6.2).

Table 6.2 Summary of feature selection from trial-and-error.

Parameter	Input features petrophysical logs	Features removed
Oil saturation	GR, AC, NPHI, RHOB, RD	SP, CAL, RM
Water saturation	SP, CAL, AC, NPHI, RHOB, RD	GR, RM
Porosity	GR, CAL, NPHI, RHOB, RD	SP, AC, RM

In addition, a statistical approach using stepwise regression analysis is done. Stepwise regression adds or removes individual variables, using the feature inputs statistical significance to increase the R^2 . Variable selection using stepwise regression is a standard procedure (Chong & Jun, 2005). The petrophysical logs and the core data are evaluated using the stepwise regression to identify variables that increase the R^2 on the validation set. The stepwise regression either adds the most significant variable or removes the least significant variable using a least-squares approach. The approach may be locally optimal, however, not necessarily globally optimal (Thompson, 1989). The significance value (α) is set at 0.05 for the stepwise regression and a variable is added when the p-value is less than 0.05. The variables are removed if the p-value is greater than 0.1 (Table 6.3).

Table 6.3 Results from stepwise regression

Oil saturation				
	Coefficient	Std.error	Status	P
GR	0.1639	0.0117	Include	1.37E-40
SP	-0.031	0.0103	Include	0.0027
CAL	-2.9196	0.2473	Include	4.44E-30
AC	-0.2513	0.0415	Include	1.96E-09
NPHI	0.5252	0.0573	Include	3.08E-19
RHOB	-30.5915	2.8522	Include	2.19E-25
RD	-0.5492	0.3761	Exclude	0.1446

RM	1.3358	0.3082	Include	1.62E-05
Water saturation				
	Coefficient	Std.error	Status	P
GR	-0.4045	0.0274	Include	1.45E-44
SP	0.1892	0.0237	Include	4.43E-15
CAL	9.2328	0.575	Include	2.10E-51
AC	0.7576	0.0961	Include	9.13E-15
NPHI	2.1195	0.1326	Include	5.09E-51
RHOB	125.023	6.5139	Include	1.91E-69
RD	5.3127	0.5361	Include	4.48E-22
RM	0.9665	1.1667	Exclude	0.4077
Porosity				
	Coefficient	Std.error	Status	P
GR	-0.0439	0.0068	Include	1.70E-10
SP	0.0145	0.005	Include	0.0039
CAL	-1.6525	0.0796	Include	2.02E-84
AC	0.0119	0.0211	Exclude	0.5727
NPHI	0.3831	0.0301	Include	2.64E-35
RHOB	-14.4412	1.2641	Include	4.71E-29
RD	-0.1492	0.2129	Exclude	0.4837
RM	-0.3778	0.1706	Include	0.0269

The results from the stepwise regression of which variables to include in the model is summarised in (Table 6.4).

Table 6.4 The parameter selection from the stepwise regression analysis.

Parameter	Input features petrophysical logs
Oil saturation	GR, SP, CAL, AC, NPHI, RHOB, RM
Water saturation	GR, SP, CAL, AC NPHI, RHOB, RD
Porosity	GR, SP, CAL, NPHI, RHOB, RM

6.1 Model validation and calibration

The first models utilise all the petrophysical logs to find the best model in the set boundaries of layers and neurons based on the accuracy performance R2. The accuracy of the predictions on the test set is considered as valid results of the calibrated models performance. However, testing the calibrated models against well 15/12-4 may validate the generalisation abilities and robustness of the model. The model is initialised with the training set and, secondly, run for the validation set to acquire the predicted values of oil saturation, water saturation and porosity on

the testing set. The proposed method of creating 10 000 models to obtain the best-calibrated model based on R^2 on the validation set resulted in three different models prior to feature selection. The selection of the calibrated models from the highest R^2 on the validation set is based on the Pearson, Spearman or Distance coefficient of determinant (R^2)

The calibrated models have a different number of hidden layers and neurons (Table 6.5). The most common network architecture using all the petrophysical logs are either 3 or 4 layers and between 4-10 neurons in each layer. In most of the calibrated models, there are more than 5 neurons in the layers. The smallest number of neurons in the hidden layer is 2 for the water saturation model with 5 layers. The best calibrated model of the porosity remains the same for the Pearson, Spearman and Distance selection with 3 hidden layers and the number of neurons is from 6 to 8.

Table 6.5 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using all petrophysical logs.

		Model #	Architecture
Pearson	Oil saturation	7669	10-8-8-10-7
	Water saturation	2519	8-10-4-7
	Porosity	4730	8-6-7
Spearman	Oil saturation	5617	4-8-4-6
	Water saturation	5636	8-10-4-7
	Porosity	4730	8-6-7
Distance	Oil saturation	3569	6-8-6
	Water saturation	2519	7-5-7-6-2
	Porosity	4730	8-6-7

The detailed performance results of the model on the training, validation and testing set and on new data (Well 15/12-4) is included in APPENDIX E. A visual representation of the results from the calibrated model are presented using histograms of the coefficients of determinant of the predictions and core data. The Pearson selection resulted in the highest value for the prediction of water saturation (Figure 6.4). The results shows that the R^2 of the oil saturation and water saturation predictions on the testing set and well 15/12-4 is in the range 0-20 (Figure 6.4, Figure 6.5 and Figure 6.6).

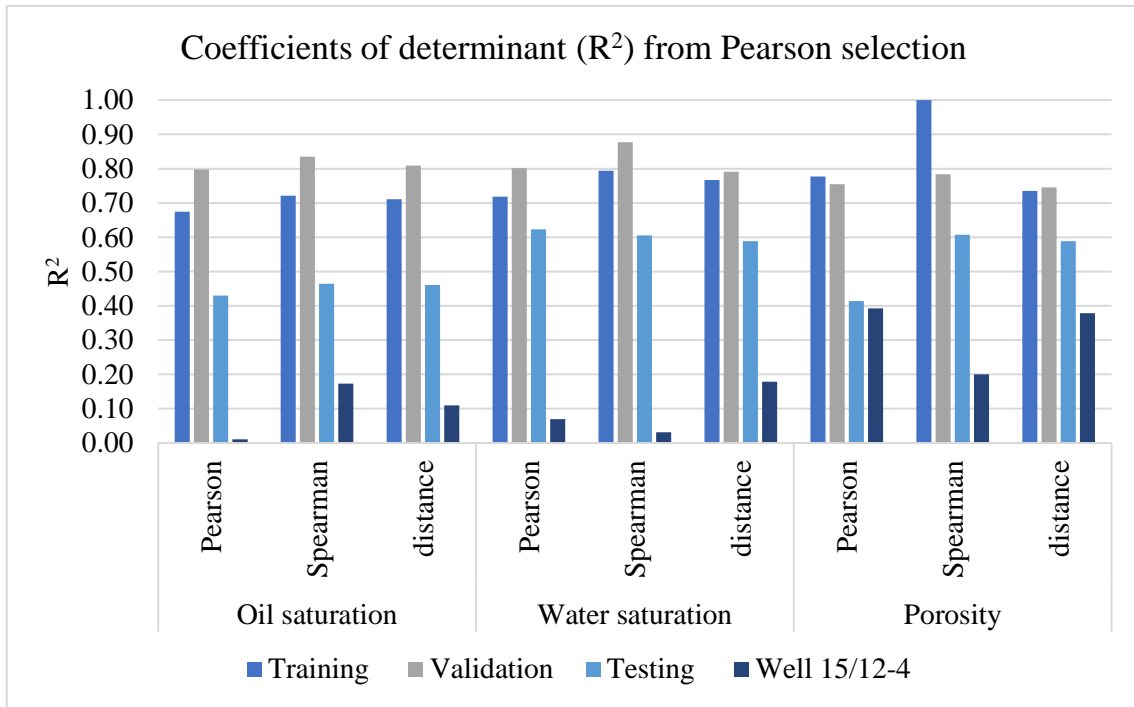


Figure 6.4 The results of the calibrated models selecting the models with highest R^2 using Pearson correlation coefficient (Table 6.5).

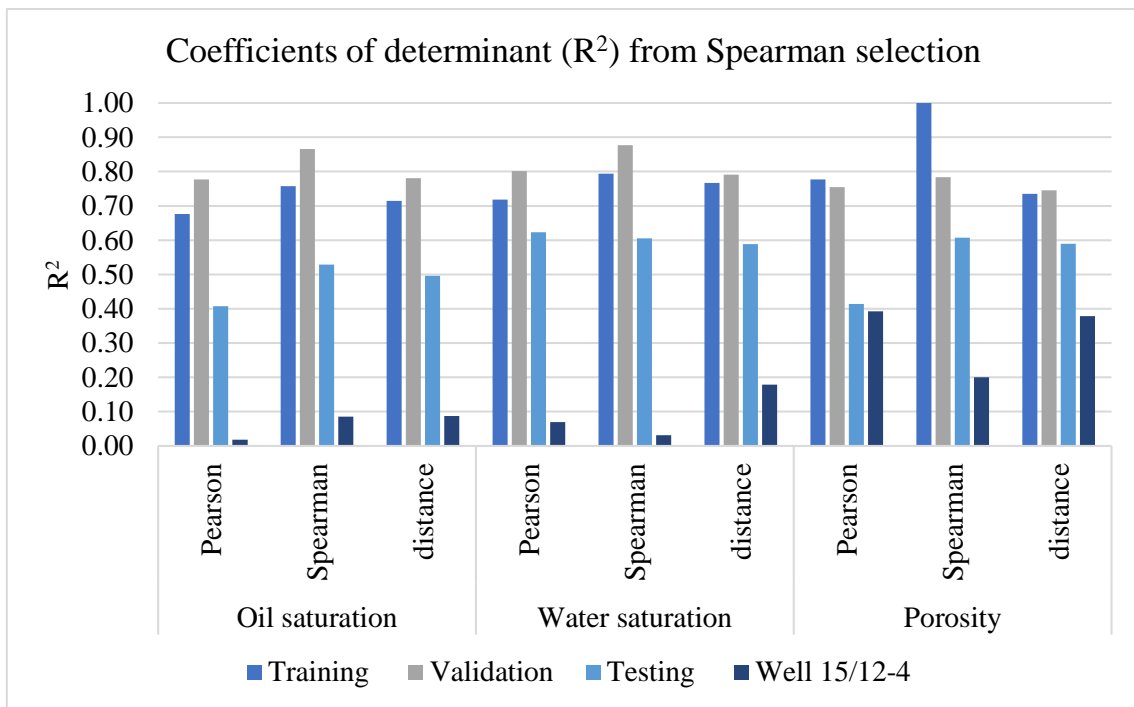


Figure 6.5 The results of the calibrated models selecting the models with highest R^2 using Spearman correlation coefficient (Table 6.5).

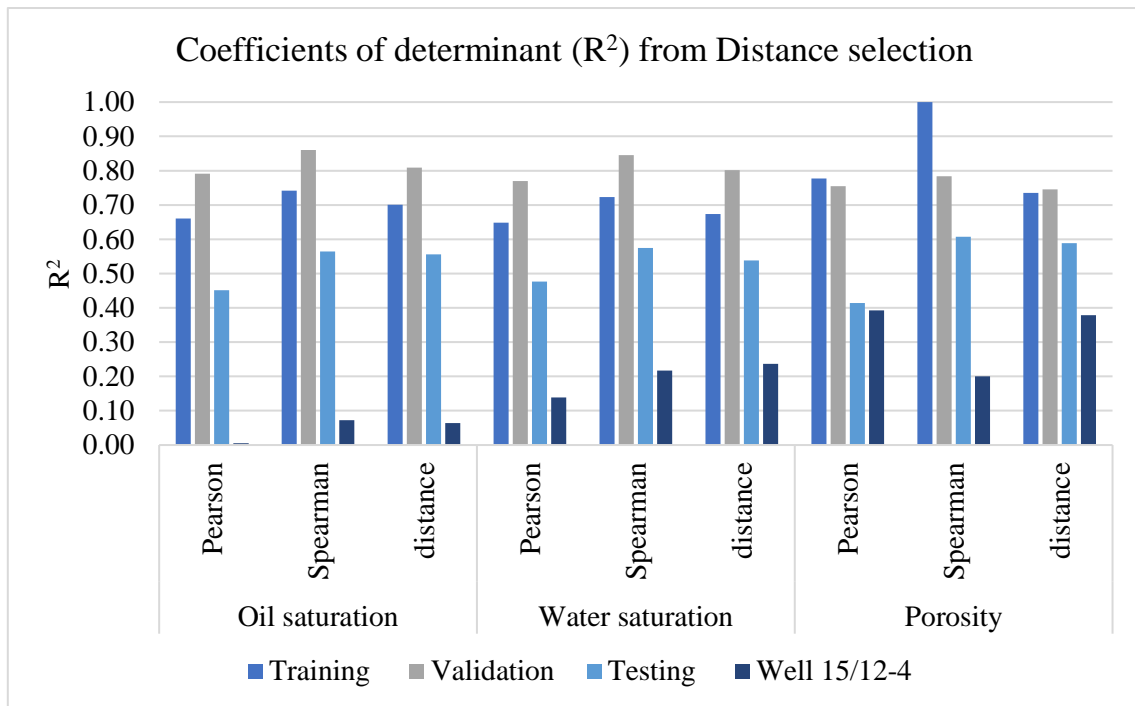


Figure 6.6 The results of the calibrated models selecting the models with highest R^2 using Distance correlation coefficient (Table 6.5).

The numerical results from the selection of the best models are summarised in Table 6.6. The predictions from the calibrated model gave model accuracy, R^2 in the range 0.4-0.6 on the test set for all the predictions, which is considered fairly good. The performance of the calibrated model on well 15/12-4 was reduced significantly, where R^2 is in the range 0-0.4. The calibrated model of the oil saturation has the lowest R^2 result on well 15/12-4.

Table 6.6 The coefficient of determinant R^2 of the training set, validation set, testing set and well 15/12-4, based on the calibrated model attained from the Pearson, Spearman and Distance selection.

Pearson selection		Training	Validation	Testing	Well 15/12-4
Oil saturation	Pearson	0.67	0.80	0.43	0.01
	Spearman	0.72	0.84	0.46	0.17
	distance	0.71	0.81	0.46	0.11
Water saturation	Pearson	0.72	0.80	0.62	0.07
	Spearman	0.79	0.88	0.61	0.03
	distance	0.77	0.79	0.59	0.18
Porosity	Pearson	0.78	0.75	0.41	0.39
	Spearman	1.00	0.78	0.61	0.20
	distance	0.73	0.75	0.59	0.38

Spearman selection		Training	Validation	Testing	Well 15/12-4
Oil saturation	Pearson	0.68	0.78	0.41	0.02
	Spearman	0.76	0.87	0.53	0.09
	distance	0.71	0.78	0.50	0.09
Water saturation	Pearson	0.72	0.80	0.62	0.07
	Spearman	0.79	0.88	0.61	0.03
	distance	0.77	0.79	0.59	0.18
Porosity	Pearson	0.78	0.75	0.41	0.39
	Spearman	1.00	0.78	0.61	0.20
	distance	0.73	0.75	0.59	0.38
Distance selection		Training	Validation	Testing	Well 15/12-4
Oil saturation	Pearson	0.66	0.79	0.45	0.01
	Spearman	0.74	0.86	0.56	0.07
	distance	0.70	0.81	0.56	0.06
Water saturation	Pearson	0.65	0.77	0.48	0.14
	Spearman	0.72	0.85	0.57	0.22
	distance	0.67	0.80	0.54	0.24
Porosity	Pearson	0.78	0.75	0.41	0.39
	Spearman	1.00	0.78	0.61	0.20
	distance	0.73	0.75	0.59	0.38

The predicted results can be compared with the target values from the core analysis. The results from the predictions shows that the accuracy of the porosity prediction is higher on well 15/12-4, than the accuracy of the prediction of oil saturation and water saturation.

6.2 Model performance comparison

In this section the results from the three feature selections are presented. The detailed performance results of the models on the training, validation and testing set and on new data (Well 15/12-4) presented in this section is included in APPENDIX E.

The network architecture from the feature selection varies greatly, and the most common architecture from the feature selection consists of 4 layers with neurons between 1-10 (Table 6.7).

The results from the feature selection resulted in no correlation of the predicted water saturation and the best model performed unsatisfactory (Figure 6.7 and Figure 6.8). The correlation

between the predicted values and desired values may have been reduced with too few petrophysical logs. The best accuracy of the oil saturation predictions improved significantly ($R^2 = 0.6-0.7$), however the performance on the test set was reduced ($R^2 = 0.2-0.3$) (Figure 6.7 and Figure 6.8). The results from the Spearman and Distance selection gave the same results for the best calibrated model and is summarised in one figure (Figure 6.8).

Table 6.7 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using the petrophysical logs in Table 6.1

		Model #	Architecture
Pearson	Oil saturation	8387	2-3-9-1
	Water saturation	8743	8-1-7-8
	Porosity	24	5-10-7-9
Spearman	Oil saturation	6906	2-3-9
	Water saturation	8743	8-10-2-1
	Porosity	1	9-9-10-5-5
Distance	Oil saturation	6906	2-3-9-1
	Water saturation	8743	8-10-2
	Porosity	1	9-9-10-5-5

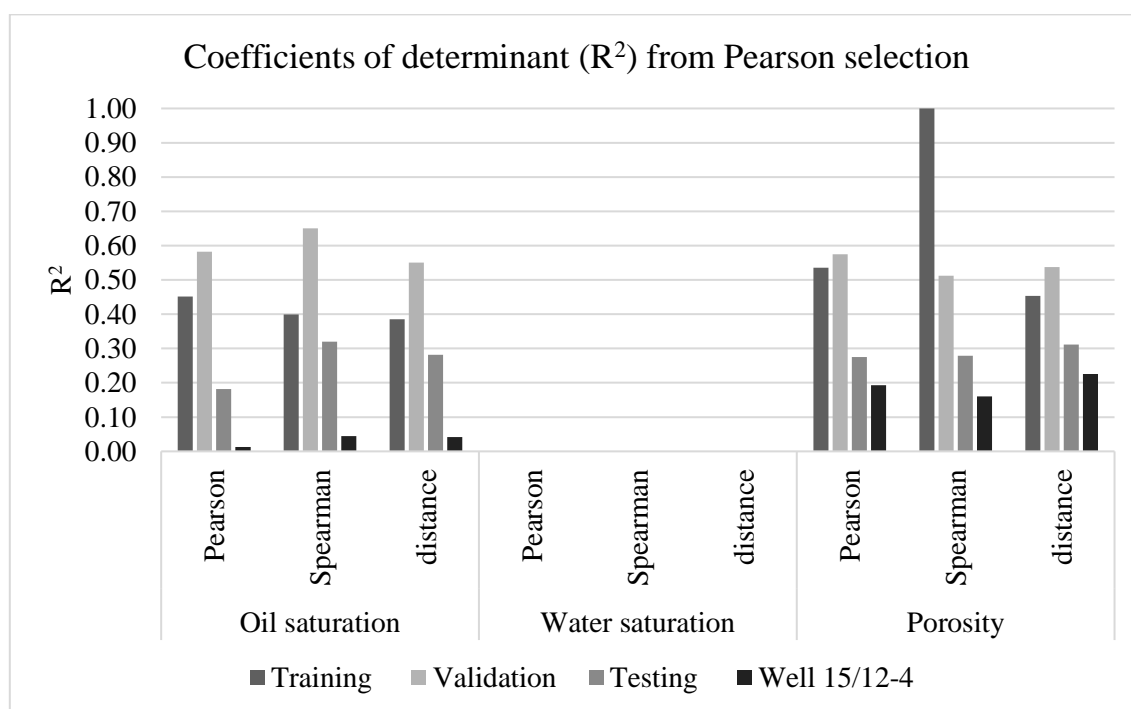


Figure 6.7 The results of the calibrated models from the feature selection with highest R^2 using Pearson correlation.

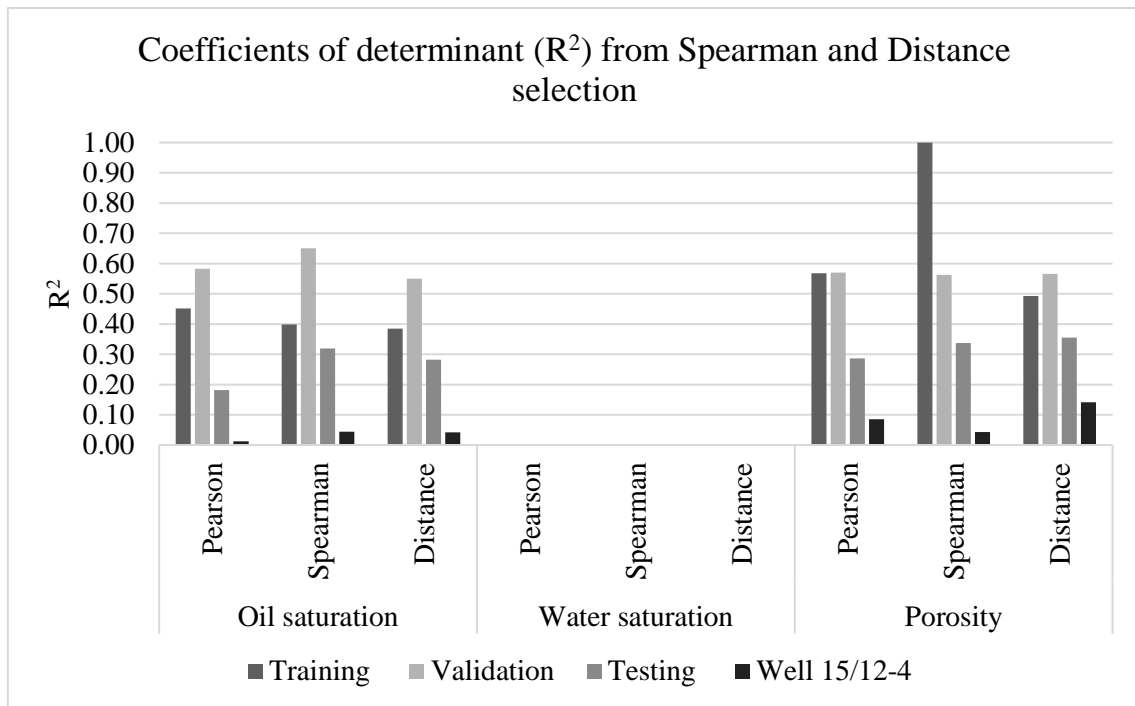


Figure 6.8 The results of the calibrated models from the feature selection, the models with highest R² using Spearman and Distance selection.

The oil saturation prediction improved when irrelevant logs were removed, while the prediction of the water saturation significantly decreased. The petrophysical logs from the feature selection is CAL, AC, NPHI and RHOB for the oil saturation and porosity, and SP, CAL, NPHI and RHOB for the water saturation. The accuracy of porosity prediction was also reduced to R² between 0.2-0.3 using fewer petrophysical logs.

The results from selecting the second feature selection had different network architectures, with layers varying between 3 and five and nodes between 1-10 (Table 6.8).

Table 6.8 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using the petrophysical logs in Table 6.2.

		Model #	Architecture
Pearson	Oil saturation	8387	9-9-10-9
	Water saturation	8743	9-8-1
	Porosity	3005	6-10-6-10
Spearman	Oil saturation	6906	7-2-9-4-3
	Water saturation	8743	6-8-8
	Porosity	7050	9-6-7-6-1
Distance	Oil saturation	6906	10-6-4-5-6
	Water saturation	8743	7-5-4
	Porosity	7050	9-6-7-6-1

The results from Pearson selection improved the accuracy of water saturation predictions on the test set (~0.7), while it accuracy on the predictions of well 15/12-4 where in the range 0-0.3 (Figure 6.9). The porosity prediction in the Spearman selection and Distance selection selected the same model for predictions. The accuracy of the porosity prediction on well 15/12-4 improved to R^2 (Distance)=0.74 (Figure 6.10). The accuracy of the oil saturation improved on the test set, however it was unable to predict reliably on well. 15/12.4 (Figure 6.9, Figure 6.10 and Figure 6.11).

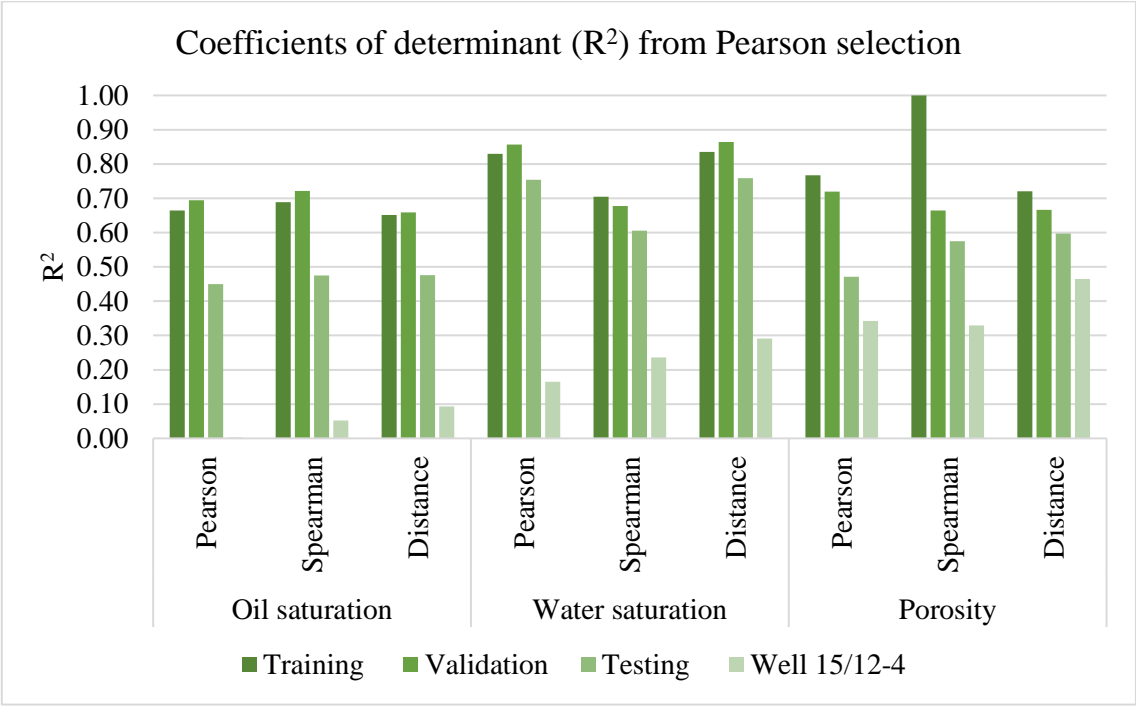


Figure 6.9 The results of the calibrated models from the second feature selection, using models with highest R^2 based on Pearson

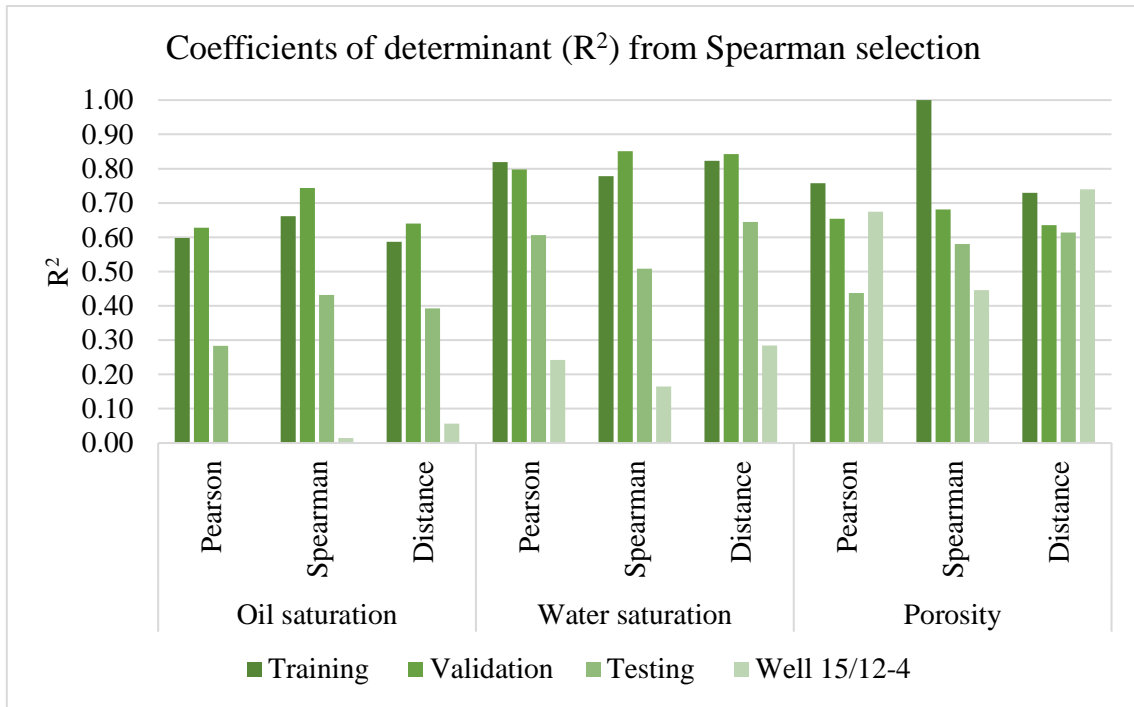


Figure 6.10 The results of the calibrated models from the second feature selection, using models with highest R^2 based on Spearman

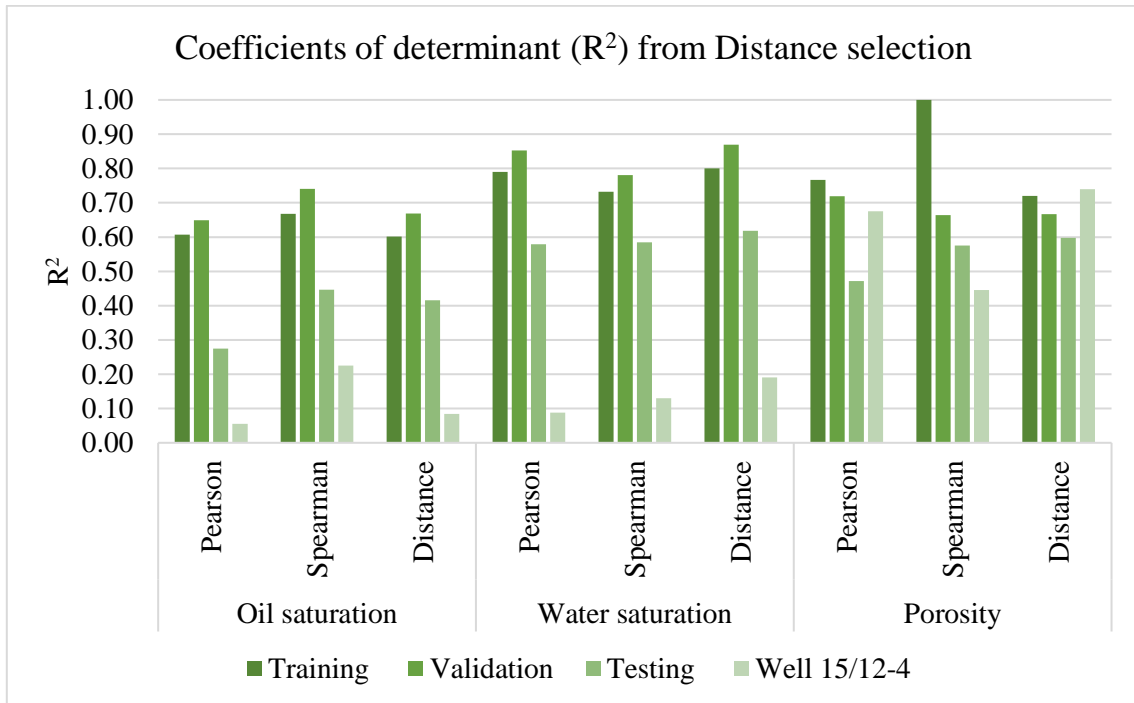


Figure 6.11 The results of the calibrated models from the second feature selection, using models with highest R^2 based on Distance

The architecture of the calibrated models from the stepwise regression approach consists mainly of 3 or 4 layers (Table 6.9). The number of neurons in each calibrated model varies greatly between 1-10.

The stepwise regression selection increased the performance on the predictions of oil saturation ($R^2 = 0.6-0.7$) and water saturation ($R^2 = 0.6-0.7$) on the test set (Figure 6.12, Figure 6.13 and Figure 6.14). The performance of the water saturation prediction increased ($R^2 = 0.3-0.4$) (Figure 6.12). The performance of the porosity predictions was reduced ($R^2 = 0.0-0.1$).

Table 6.9 Summary of the calibrated model architectures and the model number from 10 000 iterations. The results are from calibrated models using the petrophysical logs in Table 6.4

		Model #	Architecture
Pearson	Oil saturation	3005	6-5-9-2
	Water saturation	4738	9-8-1
	Porosity	5791	6-10-6-10
Spearman	Oil saturation	2236	7-2-9-4-3
	Water saturation	4399	6-8-8
	Porosity	1856	9-9-9
Distance	Oil saturation	3005	10-6-4-5-6
	Water saturation	168	5-9-10-9-4-1
	Porosity	260	6-10-6-10

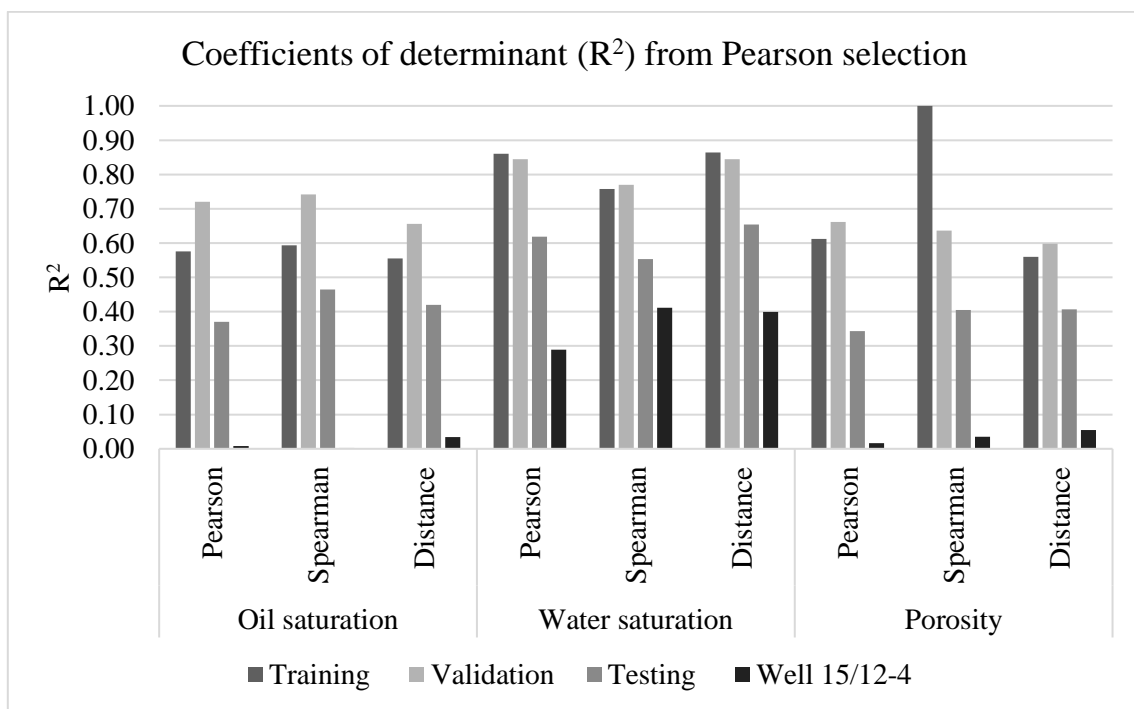


Figure 6.12 The results of the calibrated models from the feature selection with highest R^2 using Pearson selection.

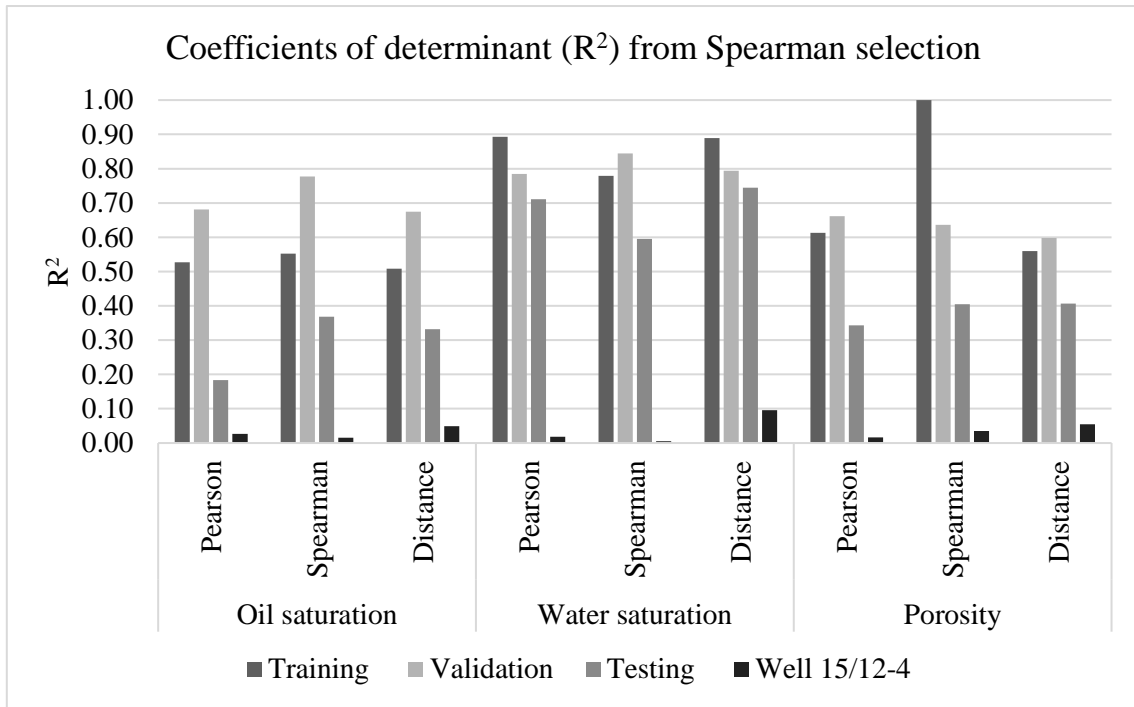


Figure 6.13 The results of the calibrated models from the feature selection, the models with highest R^2 using Spearman selection.

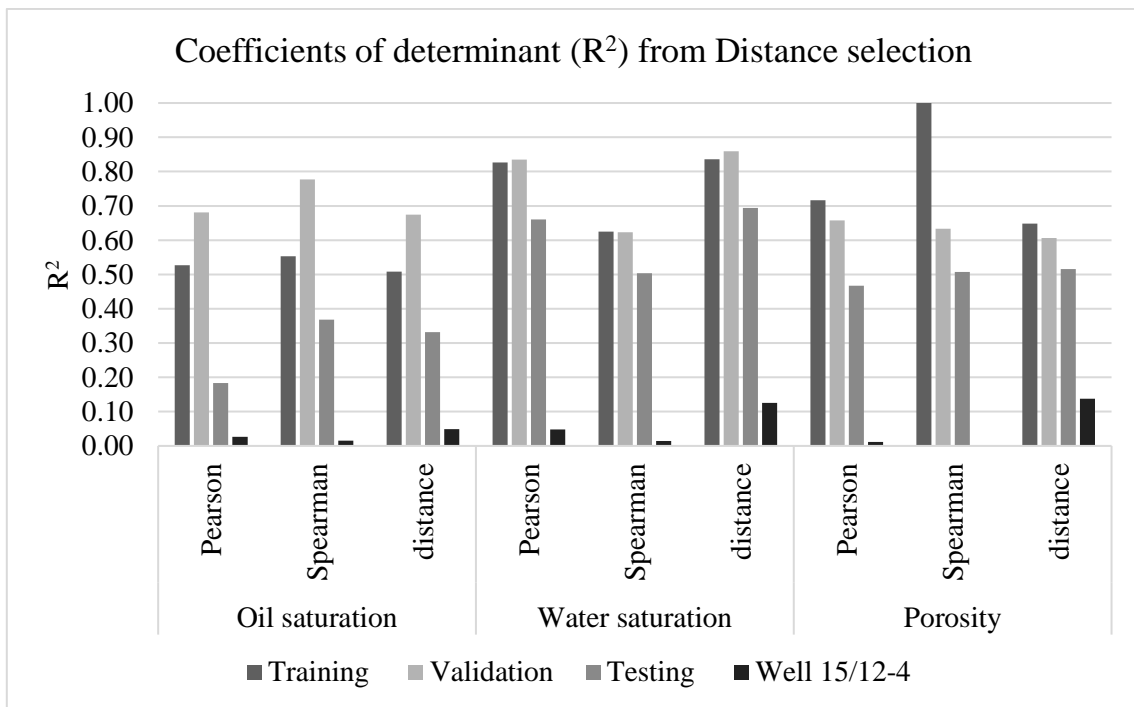


Figure 6.14 The results of the calibrated models from the feature selection, the models with highest R^2 using Distance selection.

6.3 Best results

The best prediction results on well 15/12-4 of the oil saturation, water saturation and porosity are summarised in this section. The results from the predictions are presented using regression plot of the predicted value and desired values.

The best model to predict oil saturation is obtained using the original model using all petrophysical logs. The best model is obtained using Pearson selection, with R^2 (Pearson)=0.01, R^2 (Spearman)=0.17, R^2 (Distance)=0.11. The Pearson coefficient of determinant is low, however the model has the highest Spearman and Distance coefficient of determinant. The model number is 7669 with 5 hidden layers and neurons between 7-10 (hidden layer:10-8-8-10-7). The feature selection did not improve the performance of oil prediction on well 15/12-4. The oil saturation in well 15/12-4 are very low and the model predicts higher oil saturation values of 5-20 (Figure 6.15). There is one sample with very high oil saturation around 90.

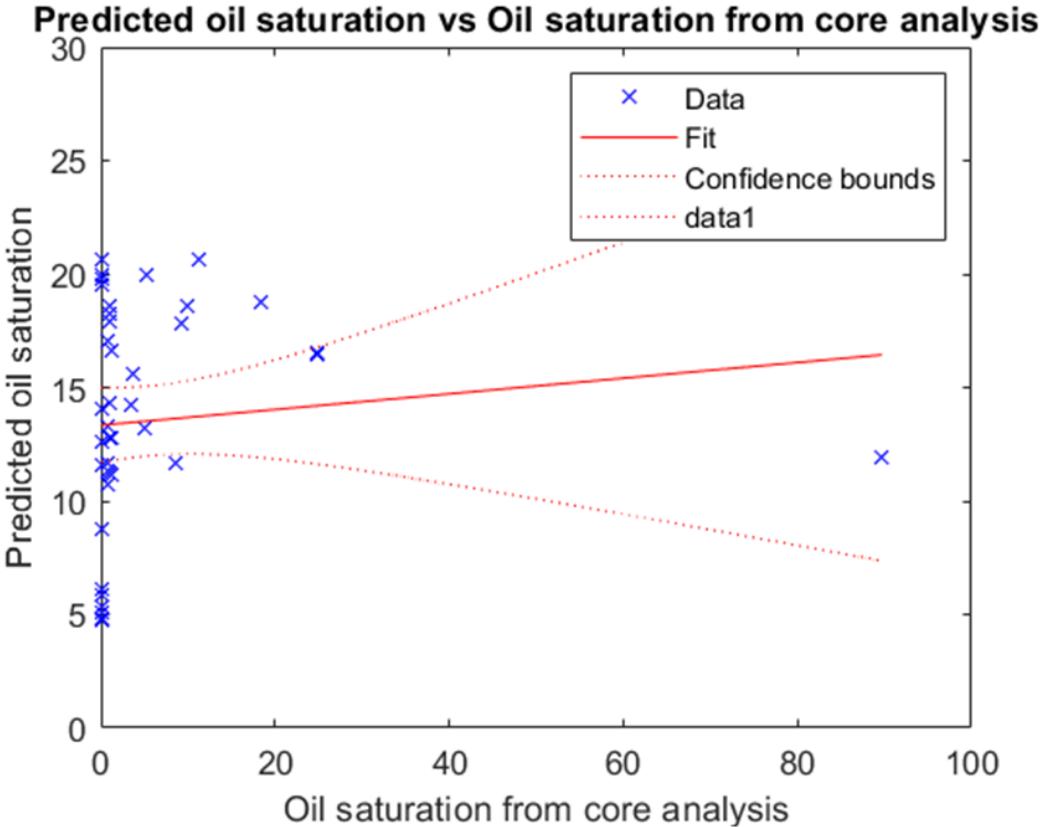


Figure 6.15 Regression plot of the predicted oil saturation and oil saturation from core data.

The best model water saturation is obtained from the stepwise regression method using the Pearson coefficient of determinant. The model best model with R^2 (Pearson)=0.29, R^2 (Spearman)=0.41, R^2 (Distance)=0.40. The model number is 4738 with 3 hidden layers. There

is a higher number of neurons in the first two layers (9 and 8) and only neuron in the last layer. The predictions of the water saturation appears to overestimate in the range 60-80, and underestimate at higher water saturations (Figure 6.16). When the water saturation from core data increases to very high values the model predicts outside the range to negative values.

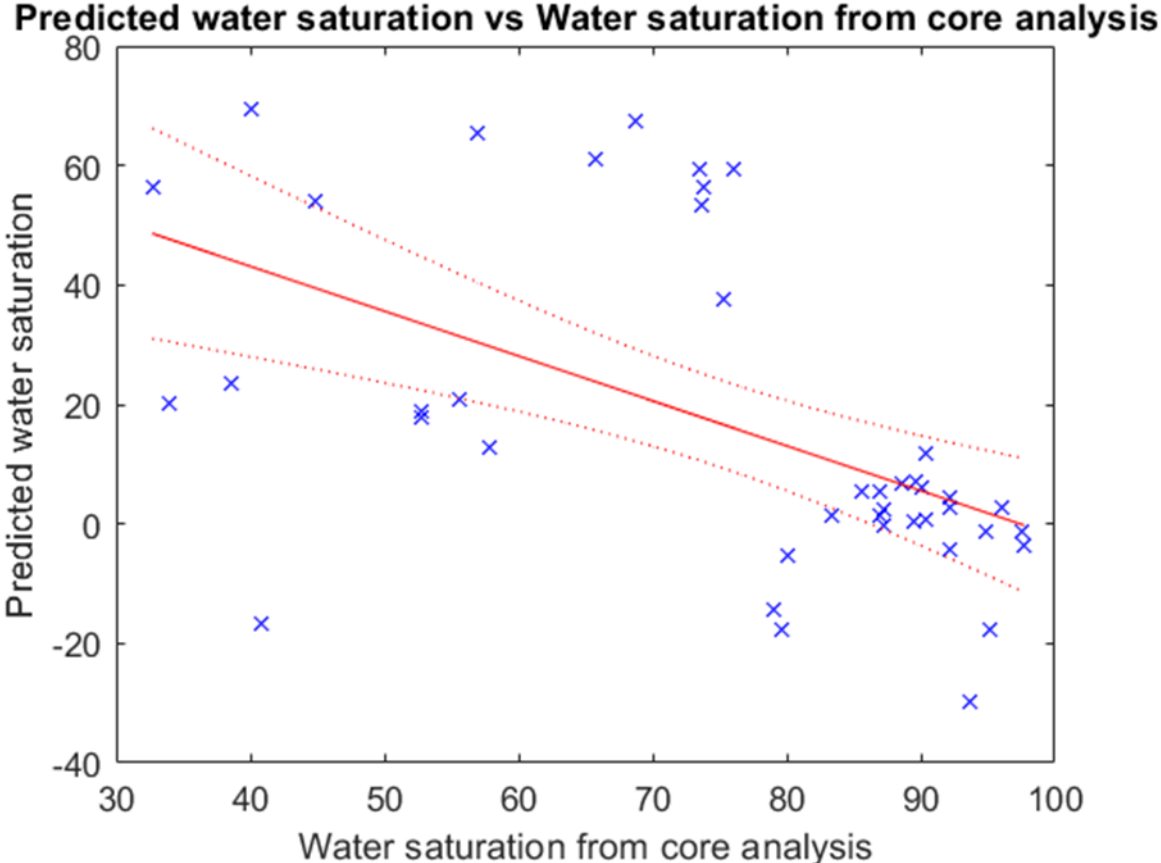


Figure 6.16 Regression plot of the predicted water saturation and water saturation from core data.

The best model of porosity was obtained from the second feature selection in the Spearman and distance selection. The model best model with R^2 (Pearson)=0.67, R^2 (Spearman)=0.45, R^2 (Distance)=0.74. The best model for porosity it is model number 3005 with 4 hidden layers, and 6 and 10 neurons in the hidden layers. The porosity predictions underestimate the porosity around the lower porosity values (15-20) (Figure 6.17). The predictions around the lower values are concentrated in the range 0-10 when it should increase.

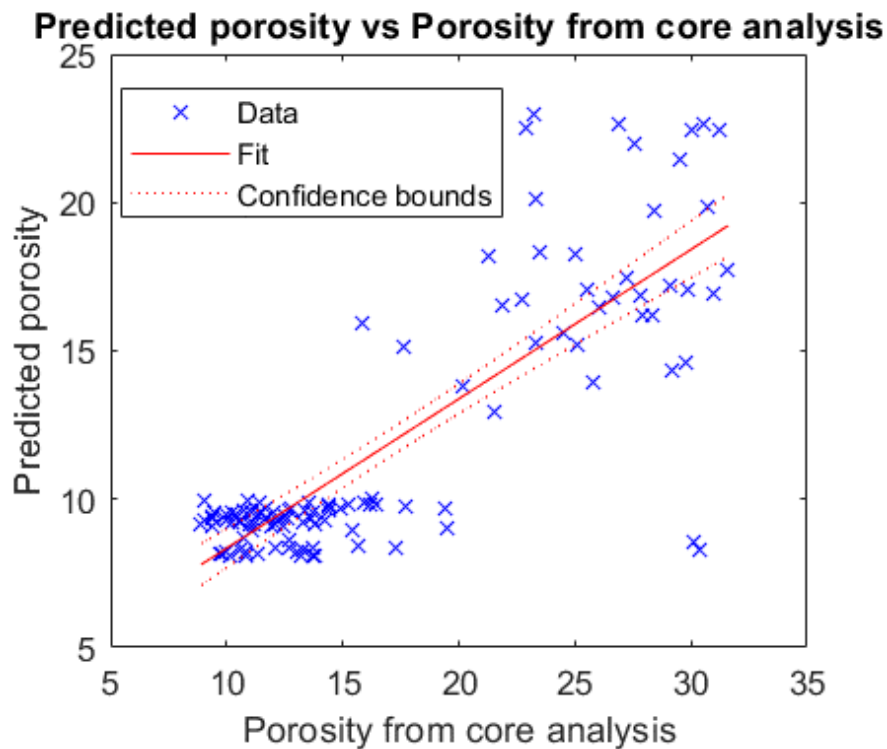


Figure 6.17 Regression plot of the predicted porosity and porosity from core data.

6.4 Sensitivity analysis of input variables

In the following section two sensitivity analysis have been applied in the study. The first one using partial derivatives and a second one by varying the input variables and evaluating how it influences the predictions, if it increases, decreases or remains stable.

6.4.1 Partial derivatives method (PaD)

The partial derivatives of the input variables and the predictions can be used to obtain a plot of prediction variations for small changes in each input variable and the relative contribution of each variable on the predictions (Gevrey et al., 2003). To obtain the plot of variations of the predictions for small changes, the partial derivatives of each input variable and predictions are calculated (Y. Dimopoulos et al., 1995; I. Dimopoulos et al., 1999).

The Partial derivatives (PaD) method calculates the sensitivity of each input feature on the predicted output from the model, using partial derivatives of the predicted output and input values (Eq. 6.1). If the partial derivatives are negative, the output decrease when the input feature increases.

$$S_i = \frac{1}{N} \sum_p \frac{\partial y_k^p}{\partial x_i^p} \quad \text{Eq. 6.1}$$

where N is the total number of data variables and p is the number of patterns. The relative contribution of each input feature is determined from the sum of squared partial derivatives:

$$SSD_i = \sum_p \frac{\partial y_k^p}{\partial x_i^p} \quad \text{Eq. 6.2}$$

Using these equations, the input features can be ranked in order of their influence on the output (Table 6.10). The contribution is calculated from the partial derivatives of the predicted output with respect to the input variables (Y. Dimopoulos et al., 1995; I. Dimopoulos et al., 1999)

$$\text{contribution}_i = \frac{SSD_i}{\sum_i SSD_i} \quad \text{Eq. 6.3}$$

The sensitivity and contribution using partial derivatives of the oil saturation, water saturation and porosity is presented in Table 6.10. This analyses the sensitivity of each individual variable, where positive and negative values demonstrate that an increase in the variable results in increase or decrease in output variable (Gevrey et al., 2003). In addition, the high or low value of the sensitivity implies higher or lower efficiency of the variables on the predicted output.

Table 6.10 summary of the relative importance of each input variable using partial derivatives of from the oil saturation, water saturation and porosity predictions from the training and validation set.

Oil saturation				
Log	Si	SSDi	Relative contribution	Rank
GR	0.51	23.9	0.0000025	8
SP	0.06	113057.1	0.0120322	3
CAL	1.02	209.6	0.0000223	5
AC	0.05	2.4	0.0000003	7
NPHI	0.08	57.7	0.0000061	6
RHOB	1.74	4594.8	0.0004890	4
RD	-37.23	505134263.6	53.7592907	1
RM	121.19	434370054.4	46.2281569	2

Water saturation				
Log	Si	SSD	Relative contribution	Rank
GR	0.44	31.8	0.000409	5
SP	0.06	119664.9	0.000007	8
CAL	1.59	267.3	0.005304	4
AC	0.13	3.0	0.000037	7
NPHI	0.33	70.5	0.000221	6
RHOB	2.07	6091.0	0.008992	3
RD	-128.31	349405842.2	34.481534	2
RM	176.84	369792547.0	65.503495	1
Porosity				
Log	Si	SSD	Relative contribution	Rank
GR	0.95	25.7	0.0000023	7
SP	0.26	194330.1	0.0170713	3
CAL	3.08	159.4	0.0000140	5
AC	0.19	1.5	0.0000001	8
NPHI	0.72	40.6	0.0000036	6
RHOB	4.53	3587.7	0.0003152	4
RD	-21.41	590602998.0	51.8826640	1
RM	155.91	547542482.6	48.0999296	2

The contribution and ranking of each variable show that the two highest ranking variables are the resistivity logs RD and RM. The third highest ranked variable for the Oil saturation and porosity predictions is SP and RHOB for the water saturation predictions. The three lowest ranking variables are the SP, AC and NPHI for water saturation, and GR, AC and NPHI for the oil saturation and porosity predictions. The highest variables, RD and RM, are the variables that influence the output variable most. The dominant direction of sensitivity is positive for all variables, except the RD. That means when the RD increase, the prediction value decreases.

The accuracy of the model is relying on the data in the training and validation set to fit and tune the model. Sensitivity plots can be applied to assess the influence of the variables on the predicted output. The results of the predicted oil saturation, water saturation and porosity from

the original model with all petrophysical logs can be evaluated using the sensitivity of each variable calculated from the partial derivatives method (Figure 6.18).

All variables have positive and negative sensitivity (Figure 6.18, Figure 6.19 and Figure 6.20). The model for the predicted oil saturation is most sensitive to changes in the RD and RM (Figure 6.18). The highest sensitivity of GR, SP CAL, NPHI and RHOB is attained at the lowest value of each variable. The sensitivity of the AC is unclear from the profile, however it is slightly in the positive direction according to the sensitivity of the input variable (Table 6.10). The highest sensitivity of RD and RM is attained at the higher values of each variable.

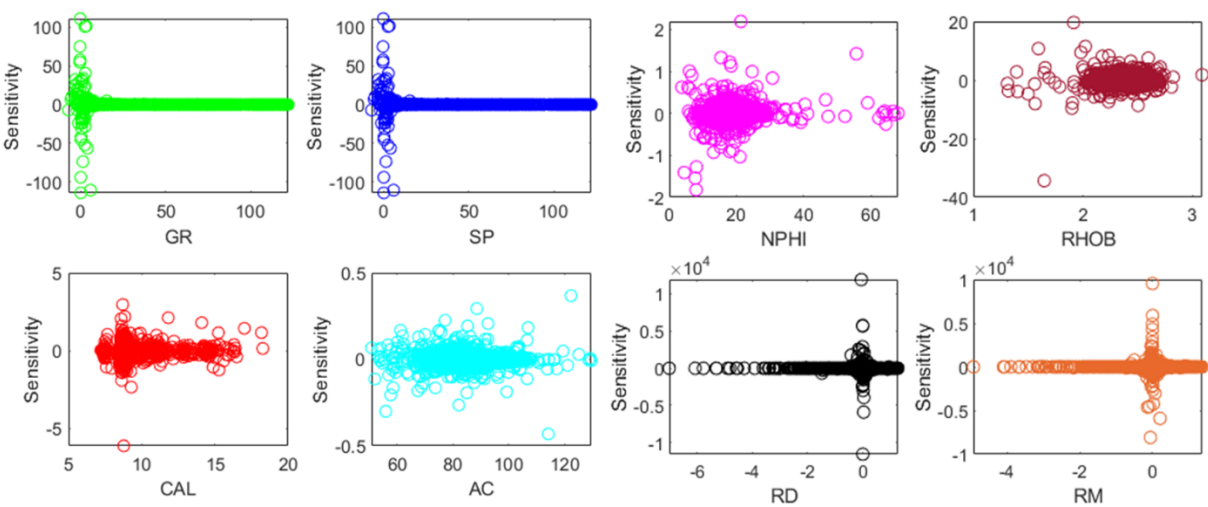


Figure 6.18 Partial derivatives of each input features vs sensitivity of the predicted oil saturation

The model for the predicted water saturation indicates that it is most sensitive to changes in the resistivity logs (RD and RM), and the highest sensitivity is attained at the higher values of these variables (Figure 6.19). The highest sensitivity of GR, SP CAL, NPHI and RHOB is attained at the lowest value of each variable. The sensitivity of AC is the same as the predictions for oil saturation.

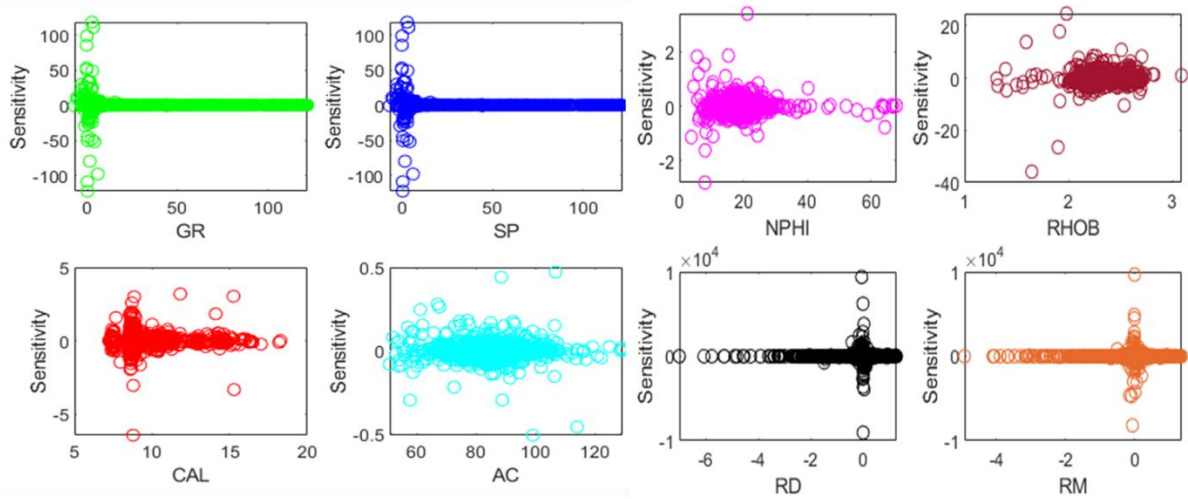


Figure 6.19 Partial derivatives of each input features vs sensitivity of the predicted water saturation

The model for the predicted porosity indicates that it is most sensitive to changes in the resistivity logs (RD and RM), and the highest sensitivity is attained at the higher values of these variables (Figure 6.19). The highest sensitivity of GR, SP, CAL, NPHI and RHOB is attained at the lowest value of each variable. The sensitivity of AC remains the same as oil saturation and water saturation.

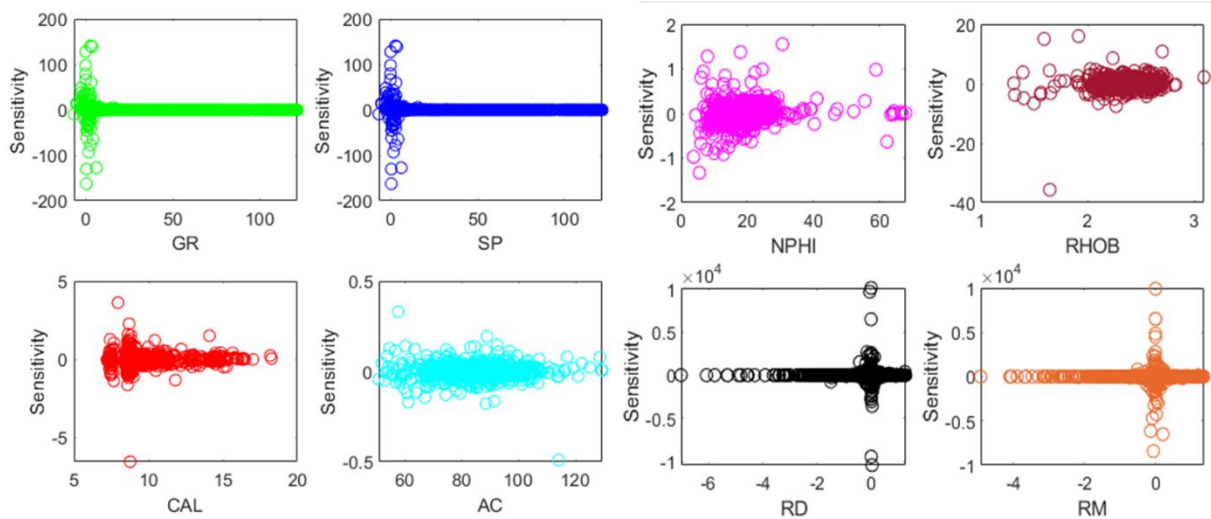


Figure 6.20 Partial derivatives of each input features vs sensitivity of the predicted porosity

6.4.2 Variation of input variables

In the following section, a sensitivity analysis of how variations in each variable affect the prediction accuracy is presented. The sensitivity analysis of the variables is conducted on the data from well 15/12-4. According to the statistical summary of P50 of well 15/12-4, there are

lower gamma-ray readings than most of the wells, high SP value than most wells (except well 15/12-A8), high CAL value, lowest AC value and NPHI value and the highest RHOB value (APPENDIX A). Thus, there are some values outside the data range in the training and validation set. In this case, the model tries to predict values outside the range of the training and validation set, which tests the generalisation ability of the model. The accuracy of the original calibrated models is summarised in Table 6.11.

Table 6.11 Reference values to evaluate how the model is affected by variations of the input variables.

	R pearson	R spearman	R distance
Oil saturation	0.01	0.17	0.11
Water saturation	0.07	0.03	0.18
Porosity	0.39	0.20	0.38

The models that were calibrated with all petrophysical logs are applied to evaluate how variations of each variable affect the prediction accuracy. Each individual variable is increased by 10%, 20% 30%, 40% and 50% by turn, while the other variables are kept at their reference values. A summary of the sensitivity results by measurements of MSE, MAE and R^2 are shown in (Table 6.12). The variations in CAL and AC reduced R^2 significantly to a range of 0.0-0.2, compared to the reference value 0.38. The increase of CAL by 50% improved the accuracy of the model to 0.44.

Table 6.12 Summary of the calibrated models performance of porosity on the data set from well 15/12-4.

	Variation [%]	MSE	MAE	R pearson	R spearman	R distance
GR	10	606.10	21.93	0.42	0.22	0.40
	20	587.64	21.31	0.41	0.22	0.40
	30	565.59	20.55	0.41	0.23	0.40
	40	543.50	19.80	0.41	0.23	0.40
	50	524.16	19.24	0.40	0.23	0.39
SP	10	915.16	23.40	0.29	0.26	0.31
	20	879.89	21.89	0.29	0.29	0.33
	30	561.35	15.81	0.23	0.32	0.28
	40	327.43	10.93	0.15	0.34	0.20
	50	230.85	9.37	0.15	0.39	0.21
CAL	10	80.18	8.11	0.04	0.07	0.16

	20	79.00	8.10	0.03	0.06	0.12
	30	73.34	7.80	0.08	0.10	0.13
	40	61.19	6.95	0.21	0.26	0.29
	50	45.54	5.53	0.37	0.39	0.44
AC	10	548.56	21.66	0.05	0.00	0.14
	20	605.67	22.77	0.03	0.01	0.13
	30	674.94	23.98	0.01	0.01	0.12
	40	746.47	25.09	0.00	0.01	0.12
	50	804.09	25.81	0.00	0.00	0.11
NPHI	10	432.05	17.25	0.41	0.18	0.38
	20	433.02	17.04	0.42	0.21	0.40
	30	436.80	16.95	0.43	0.24	0.41
	40	443.22	16.95	0.44	0.23	0.41
	50	452.20	17.06	0.43	0.22	0.41
RHOB	10	525.91	20.03	0.33	0.23	0.33
	20	519.72	19.74	0.34	0.19	0.34
	30	516.31	19.53	0.34	0.16	0.35
	40	513.14	19.31	0.34	0.15	0.35
	50	509.75	19.09	0.35	0.15	0.36
RD	10	480.72	17.97	0.39	0.21	0.38
	20	478.02	17.89	0.39	0.22	0.38
	30	475.40	17.81	0.39	0.22	0.38
	40	473.60	17.75	0.39	0.23	0.38
	50	472.28	17.69	0.39	0.24	0.38
RM	10	557.45	20.13	0.45	0.27	0.45
	20	542.75	19.75	0.45	0.26	0.44
	30	527.21	19.36	0.44	0.24	0.44
	40	511.42	18.96	0.43	0.23	0.43
	50	496.61	18.54	0.42	0.23	0.41

The results from the oil saturation and water saturation is found in (APPENDIX F).

According to sensitivity analysis results, some observations on how the uncertainties affect the calibrated model's prediction accuracy and the relationship between the petrophysical logs and oil saturation, water saturation and porosity.

- I. The resistivity logs (RD and RM) have the highest contribution and most influence on the predictions of oil saturation, water saturation and porosity according to the PaD method.
- II. The lowest sensitivity contribution on the prediction of oil saturation and porosity are GR and AC according to the PaD method. The lowest sensitivity contribution on water saturation is SP and AC.

- III. The water saturation prediction is less influenced by the changes in RM and more influenced by the RD according to the PaD method.
- IV. According to the PaD method, the calibrated model either predicts high or low values of the predictions when the value of most input variables is low, except for the resistivity logs (RD and RM), which predicts high and low values for higher values of the input variable.
- V. In addition, the porosity prediction of well 15/12-4 is highly affected by uncertainties in the CAL and AC. When the AC is varied slightly (10%), it reduces the accuracy of the predictions to 0.16. Until AC increases (50%), it improves the accuracy to 0.44.
- VI. The porosity prediction of well 15/12-4 is slightly affected by variations in RHOB and SP. When the variables increase, it reduces the accuracy of the predictions from approximately R2 in the range 0.15 to 0.3 of Pearson and Distance, while the R2 of Spearman increases to 0.39.
- VII. The oil saturation prediction of well 15/12-4 is highly affected by variations in SP, AC and NPHI. When the SP has increased, the accuracy of the predictions is initially increased to approximately 0.2 and reduced to 0 as the variable increases. However, when the AC increases up to 50%, the accuracy of the predictions increases to approximately 0.2. The increase in NPHI improves the accuracy of the model.
- VIII. The water saturation prediction of well 15/12-4 is highly affected by variations in SP and RHOB. When the variable increases, the accuracy initially increases to approximately 0.2, and as it reaches 50%, the accuracy is reduced.
- IX. In the sensitivity analysis using variations, the most stable input variables are the RD for porosity and water saturation and RM for oil saturation (See APPENDIX F). The water saturation prediction also shows some stability for RD.

7 Discussion

In this chapter, the results from the models are discussed, and the advantages and limitations of the method are addressed. The focus in this section is the best results from well 15/12-4 to verify the generalisation ability and accuracy of the models. Keep in mind that the accuracy of the predictions on the test set was usually moderately strong for the reservoir parameter ($R^2 = 0.6-0.8$).

The results from the calibrated model showed that the calibrated model of oil saturation and water saturation could predict reservoir parameters, with linear and non-linear very weak positive association ($R^2 = 0-0.2$) and weak negative association ($R^2 = 0.2-0.4$), respectively. There are fewer data points from the water saturation and oil saturation than porosity and it may make it difficult for the model to predict with a limited amount of data. The number of data points and the data range of the training and validation set may be the cause of low accuracy of the models. The results from the calibrated models also showed that the model for porosity prediction can predict porosity values with linear and non-linear moderate to moderately strong association ($R^2 = 0.6-0.8$).

The feature selection aimed to improve the accuracy of the models using targeted petrophysical logs that are more relevant for the reservoir parameters. The oil saturation was predicted using all petrophysical logs since feature selection, unfortunately did not improve the accuracy. The model built on 5 petrophysical logs: GR, AC, NPHI, RHOB and RD gave a similar result as the original model with R^2 (Pearson)=0.05, R^2 (Spearman)=0.23 and R^2 (Distance)=0.31. The R^2 of Spearman is higher in the original model. The model for the water saturation predictions is built on 7 petrophysical logs: GR, SP, CAL, AC NPHI, RHOB, and RD, and the model for porosity predictions is built on 5 logs: GR, CAL, NPHI, RHOB and RD.

The stepwise regression approach was used and obtained the best results on the water saturation. However, using cross-validation of different significance levels may give a different result in the variable selection. The stepwise regression may lead to overfitting in the calibrated model if SSE is underestimated, and the confidence interval is too narrow (Smith, 2018). Stepwise regression is widely used in substantive and validity research (Thompson, 1989). It is also applied as variable selection (Chong & Jun, 2005) or sensitivity analysis (Gevrey et al., 2003).

Feature selection aims to improve the accuracy and generalisability of the model. However, when there are fewer data points from fewer petrophysical logs, it may not recognise patterns. This can be seen in the first feature selection of water saturation, using only four petrophysical

logs. In addition, the accuracy may be improved on the training and validation set, while the generalisation ability is reduced. There is uncertainty in the data distribution when the model tries to predict values outside the range of the training and validation set. In the statistical summary of well 15/12-4 there is very low GR and AC and very high SP, CAL and NPHI (APPENDIX A). In addition, in the training and validation set, there are few wells with very high-water saturation, well 15/12-4 are one of the wells with water saturation above 95%, only one data sample from 15/12-A8. This leads to uncertainty in the predictions, which may result in the underestimated water saturation and predictions outside the data range.

The performance of the oil saturation and water saturation, and porosity models may increase by increasing the number of wells. The model can predict oil saturation, water saturation and porosity. However, the results show some uncertainties between the core data and predictions on well 15/12-4. The best models either overestimates (oil saturation) or underestimate (porosity and water saturation). If the number of wells increases, the models can learn to recognise more patterns and improve the generalisation. If the accuracy is increased by increasing the number of wells, the models can be applied for predictions in other fields in the North Sea.

The performance of the models can be improved by adjusting the control factor in the Levenberg-Marquardt algorithm to improve the accuracy, convergence stability and robustness of the algorithm. This factor can be adjusted with iterative numbers in the Levenberg-Marquardt algorithm, such as the method proposed in Cui et al., (2017).

The advantages of using these models to predict oil saturation, water saturation and porosity are summarised in the following:

- Time-efficient and economical method to obtain reservoir parameters than RCAL.
- Identifying the relevant petrophysical logs that are needed for the model, which can reduce petrophysical logging costs by removing irrelevant logs.
- It can be used for reservoir evaluation and reservoir modelling to evaluate potential hydrocarbons.

8 Conclusion

This section summarises the main findings in the study.

The best calibrated model for the oil saturation is obtained from the original model using all petrophysical logs. The calibrated model presented the best oil estimation results with R^2 (Pearson) = 0.1, R^2 (Spearman)=0.17 and R^2 (Distance)=0.11 on well 15/12-4. The best calibrated model for water saturation is obtained from the stepwise regression method using 7 petrophysical logs: GR, SP, CAL, AC NPHI, RHOB and RD. The calibrated model presented the best water estimation results with R^2 (Pearson) = 0.29, R^2 (Spearman)=0.41 and R^2 (Distance)=0.40 on well 15/12-4. The best calibrated model of the porosity obtained from the second feature selection using 5 petrophysical logs: 5 logs: GR, CAL, NPHI, RHOB and RD. The calibrated model presented the best porosity estimation results with R^2 (Pearson) = 0.67, R^2 (Spearman)=0.45 and R^2 (Distance)=0.74 on well 15/12-4. The best models of the oil saturation and water saturation was obtained from the Pearson selection, while the best models for porosity prediction is obtained in the Spearman selection.

The prediction accuracy of the oil saturation, water saturation and porosity on the Varg field is very weak, weak and moderately strong, respectively. The prediction accuracy on the test set is moderately strong. If the models are trained using more wells and targeted petrophysical logs, it could potentially provide a fast and reliable method for predictions of reservoir parameters, which is very useful in both exploration and production in the petroleum industry.

The Distance correlation is usually better to evaluate the relationship between non-linear variables, such a petrophysical logs. The Pearson correlation and correlation results showed an insignificant correlation between the resistivity logs and the water and porosity saturation. In addition, the Spearman correlation showed an insignificant correlation between the resistivity logs and the water saturation. The oil saturation showed a significant correlation with the resistivity logs in all correlations.

The most relevant petrophysical logs can be obtained according to the PaD method, and the relative contribution of each input variable on the predictions. The resistivity logs (RD and RM) are found as the most important petrophysical log as an input variable in the estimations. The most irrelevant input variables are GR and AC for the oil saturation and porosity estimation and SP and AC for the water saturation estimation. The data quality of RM can slightly influence the results on porosity and oil saturation. It is stable as an input variable for water saturation.

The estimation and prediction results may not be reliable when the input variable increases up to 50%.

The findings from feature selection show that irrelevant petrophysical logs that reduce the accuracy of the models can be removed and improve the accuracy of the model (water saturation and porosity). This can help reduce the cost of well logging operations by targeting the petrophysical logs that improve the model's accuracy. Increasing the number of wells and re-evaluate the correlations between the parameters may help in further studies and well logging operations.

References

- Adamowski, J., Fung Chan, H., Prasher, S. O., Ozga-Zielinski, B., & Sliusarieva, A. (2012). Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water resources research*, 48(1), n/a. <https://doi.org/10.1029/2010WR009945>
- Ahmadi, M. A., & Chen, Z. (2019). Comparison of machine learning methods for estimating permeability and porosity of oil reservoirs via petro-physical logs. *Petroleum*, 5(3), 271–284. <https://doi.org/10.1016/j.petlm.2018.06.002>
- Alger, R. P., & Raymer, L. L. (1963). Formation Density Log Applications in Liquid-Filled Holes. *Journal of Petroleum Technology*, 15(03), 321–332. <https://doi.org/10.2118/435-PA>
- Anastassiou, G. A. (2011). *Intelligent Systems: Approximation by Artificial Neural Networks* (Bd. 19). Springer Berlin Heidelberg. <http://link.springer.com/10.1007/978-3-642-21431-8>
- Archie, G. E. (1942). The Electrical Resistivity Log as an Aid in Determining Some Reservoir Characteristics. *Transactions of the AIME*, 146(01), 54–62. <https://doi.org/10.2118/942054-G>
- Arif, J., Ray Chaudhuri, N., Ray, S., & Chaudhuri, B. (2009). Online Levenberg-Marquardt algorithm for neural network based estimation and control of power systems. 2009 *International Joint Conference on Neural Networks*, 199–206. <https://doi.org/10.1109/IJCNN.2009.5179071>
- Asquith, G., & Krygowski, D. (2004). *AAPG Methods in Exploration, No. 16, Chapter 3: Gamma Ray*. 31–35.
- Baker, R. O., Yarranton, H. W., & Jensen, J. L. (2015). Openhole Well Logs—Log Interpretation Basics. I *Practical Reservoir Engineering and Characterization* (s. 297–343). Elsevier. <https://linkinghub.elsevier.com/retrieve/pii/B9780128018118000092>
- Bebis, G., & Georgiopoulos, M. (1994). Feed-forward neural networks. *IEEE Potentials*, 13(4), 27–31. <https://doi.org/10.1109/45.329294>
- Beckers, R., Deneubourg, J. L., & Goss, S. (1992). Trails and U-turns in the selection of a path by the ant *Lasius niger*. *Journal of Theoretical Biology*, 159(4), 397–415. [https://doi.org/10.1016/S0022-5193\(05\)80686-1](https://doi.org/10.1016/S0022-5193(05)80686-1)
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Bojarski, M., Testa, D. D., Dworakowski, D., & Firner, B. (2016). *End to End Learning for Self-Driving Cars*. 10.
- Boulesteix, A.-L., Wright, M. N., Hoffmann, S., & König, I. R. (2020). Statistical learning approaches in the genetic epidemiology of complex diseases. *Human Genetics*, 139(1), 73–84. <https://doi.org/10.1007/s00439-019-01996-9>
- Chong, I.-G., & Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78(1–2), 103–112. <https://doi.org/10.1016/j.chemolab.2004.12.011>

- Cui, M., Zhao, Y., Xu, B., & Gao, X. (2017). A new approach for determining damping factors in Levenberg-Marquardt algorithm for solving an inverse heat conduction problem. *International Journal of Heat and Mass Transfer*, *107*, 747–754. <https://doi.org/10.1016/j.ijheatmasstransfer.2016.11.101>
- Cybenkot, G. (1989). *Approximation by superpositions of a sigmoidal function*. 12.
- Deegan, C. E., & Scull, B. J. (1977). *A standard lithostratigraphic nomenclature for the Central and Northern North Sea*.
- Demir-Kavuk, O., Kamada, M., Akutsu, T., & Knapp, E.-W. (2011). Prediction using step-wise L1, L2 regularization and feature selection for small data sets with large number of features. *BMC Bioinformatics*, *12*(1), 412. <https://doi.org/10.1186/1471-2105-12-412>
- Dimopoulos, I., Chronopoulos, J., Chronopoulou-Sereli, A., & Lek, S. (1999). Neural network models to study relationships between lead concentration in grasses and permanent urban descriptors in Athens city (Greece). *Ecological Modelling*, *120*(2–3), 157–165. [https://doi.org/10.1016/S0304-3800\(99\)00099-X](https://doi.org/10.1016/S0304-3800(99)00099-X)
- Dimopoulos, Y., Bourret, P., & Lek, S. (1995). Use of some sensitivity criteria for choosing networks with good generalization ability. *Neural Processing Letters*, *2*(6), 1–4. <https://doi.org/10.1007/BF02309007>
- Donaldson Erle C & Tiab Djebbar. (2012). *Basic Well-Log Interpretation*. 1–1.
- Duc, T. L., Leiva, R. G., Casari, P., & Östberg, P.-O. (2019). Machine Learning Methods for Reliable Resource Provisioning in Edge-Cloud Computing: A Survey. *ACM Computing Surveys*, *52*(5), 1–39. <https://doi.org/10.1145/3341145>
- Friedland, G., & Krell, M. (2018). A Capacity Scaling Law for Artificial Neural Networks. *ArXiv:1708.06019 [Cs]*. <http://arxiv.org/abs/1708.06019>
- Gavin, H. P. (2020). *The Levenberg-Marquardt algorithm for nonlinear least squares curve-fitting problems*. 19.
- Gevrey, M., Dimopoulos, I., & Lek, S. (2003). Review and comparison of methods to study the contribution of variables in artificial neural network models. *Ecological Modelling*, *160*(3), 249–264. [https://doi.org/10.1016/S0304-3800\(02\)00257-0](https://doi.org/10.1016/S0304-3800(02)00257-0)
- Glover, P. WJ. (2000). *Petrophysics*. University of Aberdeen, UK.
- Hagan, M., T., Demuth, H., B., & Beale, M., H. (2002). *Neural network design*.
- Hagan, M. T., & Menhaj, M. B. (1994). Training feedforward networks with the Marquardt algorithm. *IEEE Transactions on Neural Networks*, *5*(6), 989–993. <https://doi.org/10.1109/72.329697>
- Hamada, G., Sakka, A., & Nyein, C. (2020). Prediction of Porosity and Water Saturation Using Neural Networks in Shaly Sand Reservoirs, Western Deseret, Egypt. *Journal of Petroleum and Mining Engineering*, *0*(0), 1–12. <https://doi.org/10.21608/jpme.2020.36116.1040>
- Hecht-Nielsen. (1989). Theory of the backpropagation neural network. *International Joint Conference on Neural Networks*, 593–605 bd.1. <https://doi.org/10.1109/IJCNN.1989.118638>

- Helle, H. B., & Bhatt, A. (2002). Fluid saturation from well logs using committee neural networks. *Petroleum Geoscience*, 8(2), 109–118. <https://doi.org/10.1144/petgeo.8.2.109>
- Helle, H. B., Bhatt, A., & Ursin, B. (2001). Porosity and permeability prediction from wireline logs using artificial neural networks: A North Sea case study: Pore-perm prediction by neural nets. *Geophysical Prospecting*, 49(4), 431–444. <https://doi.org/10.1046/j.1365-2478.2001.00271.x>
- Hira, Z. M., & Gillies, D. F. (2015). A Review of Feature Selection and Feature Extraction Methods Applied on Microarray Data. *Advances in Bioinformatics*, 2015, 1–13. <https://doi.org/10.1155/2015/198363>
- Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5), 359–366. [https://doi.org/10.1016/0893-6080\(89\)90020-8](https://doi.org/10.1016/0893-6080(89)90020-8)
- Huang, Z., Shimeld, J., Williamson, M., & Katsube, J. (1996). Permeability prediction with artificial neural network modeling in the Venture gas field, offshore eastern Canada. *GEOPHYSICS*, 61(2), 422–436. <https://doi.org/10.1190/1.1443970>
- Huang, Z., & Williamson, M. A. (1997). Determination of porosity and permeability in reservoir intervals by artificial neural network modelling, offshore Eastern Canada. *Petroleum Geoscience*, 3(3), 245–258. <https://doi.org/10.1144/petgeo.3.3.245>
- Hunter, D., Yu, H., Pukish, I., Michael S., Kolbusz, J., & Wilamowski, B. M. (2012). Selection of Proper Neural Network Sizes and Architectures—A Comparative Study. *IEEE Transactions on Industrial Informatics*, 8(2), 228–240. <https://doi.org/10.1109/TII.2012.2187914>
- Isaksen, D., & Tonstad, K. (1989). *A revised Cretaceous and Tertiary lithostratigraphic nomenclature for the Norwegian North Sea*. Oljedirektoratet.
- Jabri, M., & Jerbi, H. (2009). Comparative Study Between Levenberg Marquardt And Genetic Algorithm For Parameter Optimization Of An Electrical System. *IFAC Proceedings Volumes*, 42(13), 77–82. <https://doi.org/10.3182/20090819-3-PL-3002.00015>
- Jazayeri, K., Jazayeri, M., & Uysal, S. (2016). Comparative Analysis of Levenberg-Marquardt and Bayesian Regularization Backpropagation Algorithms in Photovoltaic Power Estimation Using Artificial Neural Network. I P. Perner (Red.), *Advances in Data Mining. Applications and Theoretical Aspects* (Bd. 9728, s. 80–95). Springer International Publishing. https://doi.org/10.1007/978-3-319-41561-1_7
- Kakas, A. C., Cohn, D., Dasgupta, S., Barto, A. G., Carpenter, G. A., Grossberg, S., Webb, G. I., Dorigo, M., Birattari, M., Toivonen, H., Timmis, J., Branke, J., Toivonen, H., Strehl, A. L., Drummond, C., Coates, A., Abbeel, P., Ng, A. Y., Zheng, F., ... Tadepalli, P. (2011). Ant Colony Optimization. I C. Sammut & G. I. Webb (Red.), *Encyclopedia of Machine Learning* (s. 36–39). Springer US. https://doi.org/10.1007/978-0-387-30164-8_22
- Kavzoglu, T. (2009). Increasing the accuracy of neural network classification using refined training data. *Environmental Modelling & Software*, 24(7), 850–858. <https://doi.org/10.1016/j.envsoft.2008.11.012>

- Kayri, M. (2016). Predictive Abilities of Bayesian Regularization and Levenberg–Marquardt Algorithms in Artificial Neural Networks: A Comparative Empirical Study on Social Data. *Mathematical and Computational Applications*, 21(2), 20. <https://doi.org/10.3390/mca21020020>
- Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. *Proceedings of ICNN'95 - International Conference on Neural Networks*, 4, 1942–1948 bd.4. <https://doi.org/10.1109/ICNN.1995.488968>
- Kothari, S. C., & Oh, H. (1993). Neural Networks for Pattern Recognition. I *Advances in Computers* (Bd. 37, s. 119–166). Elsevier. [https://doi.org/10.1016/S0065-2458\(08\)60404-0](https://doi.org/10.1016/S0065-2458(08)60404-0)
- Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 13, 8–17. <https://doi.org/10.1016/j.csbj.2014.11.005>
- Li, J., Cheng, J., Shi, J., & Huang, F. (2012). Brief Introduction of Back Propagation (BP) Neural Network Algorithm and Its Improvement. I D. Jin & S. Lin (Red.), *Advances in Computer Science and Information Engineering* (Bd. 169, s. 553–558). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30223-7_87
- Mahmoudi, S., & Mahmoudi, A. (2014). WATER SATURATION AND POROSITY PREDICTION USING BACK-PROPAGATION ARTIFICIAL NEURAL NETWORK (BPANN) FROM WELL LOG DATA. *Journal of Engineering and Technology*, 8.
- Maniezzo, V., Gambardella, L. M., & Luigi, F. (2004). *Ant Colony Optimization*. https://doi.org/10.1007/978-3-540-39930-8_5
- Masters, T. (1993). Foundations. I *Practical Neural Network Recipes in C++* (s. 1–14). Elsevier. <https://doi.org/10.1016/B978-0-08-051433-8.50006-9>
- Mayyahi, A. A., Wang, W., & Birch, P. (2015). Levenberg-Marquardt optimised neural networks for trajectory tracking of autonomous ground vehicles. *International Journal of Mechatronics and Automation*, 5(2/3), 140. <https://doi.org/10.1504/IJMA.2015.075960>
- Mcculloch, W. S., & Pitts, W. (1943). *A LOGICAL CALCULUS OF THE IDEAS IMMANENT IN NERVOUS ACTIVITY*. 17.
- Minsky, M. L., & Papert, S. (1969). *Perceptrons; an Introduction to Computational Geometry*. MIT Press. <https://books.google.no/books?id=Ow1OAQAIAAJ>
- Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning, second edition*. MIT Press.
- Motamedie, M. H. (2020). *An Overview on Applications of Machine learning in petroleum Engineering*. 16.
- Murphy, K. P. (2012). *Machine learning: A probabilistic perspective*. MIT Press.
- Müller, A. C., & Guido, S. (2016). *Introduction to Machine Learning with Python: A Guide for Data Scientists*. O'Reilly Media, Incorporated. <https://books.google.no/books?id=qjUVogEACAAJ>

Newman, G. H., & Martin, J. C. (1977). Equipment And Experimental Methods For Obtaining Laboratory Compression Characteristics Of Reservoir Rocks Under Various Stress And Pressure Conditions. *All Days*, SPE-6855-MS. <https://doi.org/10.2118/6855-MS>

Nichols, J. A., Herbert Chan, H. W., & Baker, M. A. B. (2019). Machine learning: Applications of artificial intelligence to imaging and diagnosis. *Biophysical Reviews*, *11*(1), 111–118. <https://doi.org/10.1007/s12551-018-0449-9>

NPD, N. P. D. (2020). *Field—FactPages—NPD*. <https://factpages.npd.no/>

Patro, S. G. K., & sahu, K. K. (2015). Normalization: A Preprocessing Stage. *IARJSET*, 20–22. <https://doi.org/10.17148/IARJSET.2015.2305>

Pun, G. P. P., Batra, R., Ramprasad, R., & Mishin, Y. (2019). Physically informed artificial neural networks for atomistic modeling of materials. *Nature Communications*, *10*(1), 2339. <https://doi.org/10.1038/s41467-019-10343-5>

Pyle, H. C., & Sherborne, J. E. (1939). Core Analysis. *Transactions of the AIME*, *132*(01), 33–61. <https://doi.org/10.2118/939033-G>

Rosenblatt, F. (1957). *The Perceptron, a Perceiving and Recognizing Automaton Project Para*. Cornell Aeronautical Laboratory. https://books.google.no/books?id=P_XGPgAACAAJ

Rosenblatt, F. (1958). The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, *65*(6), 386–408. <https://doi.org/10.1037/h0042519>

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning Internal Representations by Error Propagation. I *Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Vol. 1: Foundations* (s. 318–362). MIT Press.

Rwechungura, R., Dadashpour, M., & Kleppe, J. (2011). Application of Particle Swarm Optimization for Parameter Estimation Integrating Production and Time Lapse Seismic Data. *All Days*, SPE-146199-MS. <https://doi.org/10.2118/146199-MS>

Saga Petroleum. (1999). *Final well report Well 15/12-A-8*.

Sammut, C., & Webb, G. I. (2011). *Encyclopedia of Machine Learning*. Springer Science & Business Media.

Santos, S. M., Valença, M. J. S., & Bastos-Filho, C. J. A. (2012). Comparing Particle Swarm Optimization Approaches for Training Multi-Layer Perceptron Neural Networks for Forecasting. I H. Yin, J. A. F. Costa, & G. Barreto (Red.), *Intelligent Data Engineering and Automated Learning—IDEAL 2012* (Bd. 7435, s. 344–351). Springer Berlin Heidelberg. http://link.springer.com/10.1007/978-3-642-32639-4_42

Saputro, O. D., Maulana, Z. L., & Latief, F. D. E. (2016). Porosity Log Prediction Using Artificial Neural Network. *Journal of Physics: Conference Series*, *739*, 012092. <https://doi.org/10.1088/1742-6596/739/1/012092>

Schlumberger Educational Services. (1989). *Schlumberger log interpretation charts*. Schlumberger Educational Services; /z-wcorg/.

- Schmidt, J., Marques, M. R. G., Botti, S., & Marques, M. A. L. (2019). Recent advances and applications of machine learning in solid-state materials science. *Npj Computational Materials*, 5(1), 83. <https://doi.org/10.1038/s41524-019-0221-0>
- Selley, R. C., & Sonnenberg, S. A. (2015). Methods of Exploration. I *Elements of Petroleum Geology* (s. 41–152). Elsevier. <https://doi.org/10.1016/B978-0-12-386031-6.00003-5>
- Sharma, S. K., & Wang, X. (2018). *Towards Massive Machine Type Communications in Ultra-Dense Cellular IoT Networks: Current Issues and Machine Learning-Assisted Solutions*. 38.
- Smith, G. (2018). Step away from stepwise. *Journal of Big Data*, 5(1), 32. <https://doi.org/10.1186/s40537-018-0143-6>
- Statoil. (1984). *Final well report Well 15/12-4*.
- Statoil. (1991). *Special core analysis, well 15/12-6*.
- Statoil. (1993). *Well completion report Well 15/12-9S PL 038*.
- Statoil. (2004). *Well completion report Well 15/12-6S PL 038*.
- Statoil-Esso. (1975). *Well completion report Well 15/12-1 PL 038*.
- Statoil-Esso. (1986). *Well completion report Well 15/12-5 PL 038*.
- Stiles, J. H., & Hutfilz, J. M. (1992). The Use of Routine and Special Core Analysis in Characterizing Brent Group Reservoirs, U.K. North Sea. *Journal of Petroleum Technology*, 44(06), 704–713. <https://doi.org/10.2118/18386-PA>
- Székely, G. J., Rizzo, M. L., & Bakirov, N. K. (2007). Measuring and testing dependence by correlation of distances. *The Annals of Statistics*, 35(6). <https://doi.org/10.1214/009053607000000505>
- Thompson, B. (1989). Why Won't Stepwise Methods Die? *Measurement and Evaluation in Counseling and Development*, 21(4), 146–148. <https://doi.org/10.1080/07481756.1989.12022899>
- Tixier, M. P., Alger, R. P., & Doh, C. A. (1959). *Sonic Logging*. 216, 9.
- Vandeginste, B. G. M., Massart, D. L., Buydens, L. M. C., De Jong, S., Lewi, P. J., & Smeyers-Verbeke, J. (1998). Artificial Neural Networks. I *Data Handling in Science and Technology* (Bd. 20, s. 649–699). Elsevier. <https://linkinghub.elsevier.com/retrieve/pii/S0922348798800543>
- Vollset, J., & Doré, A. G. (1984). *A Revised Triassic and Jurassic lithostratigraphic nomenclature for the Norwegian North Sea*. Oljedirektoratet.
- Wang, S.-C. (2003). Artificial Neural Network. I *Interdisciplinary Computing in Java Programming* (s. 81–100).
- Werbos, P., & John, P. (1974). *Beyond regression: New tools for prediction and analysis in the behavioral sciences* /.

- Widrow, B., & Hoff, M. E. (1960). *Adaptive switching circuits*. Defense Technical Information Center. <https://doi.org/10.21236/AD0241531>
- Wyllie, M. R. J., Gregory, A. R., & Gardner, G. H. F. (1958). AN EXPERIMENTAL INVESTIGATION OF FACTORS AFFECTING ELASTIC WAVE VELOCITIES IN POROUS MEDIA. *GEOPHYSICS*, 23(3), 459–493. <https://doi.org/10.1190/1.1438493>
- YEGNANARAYANA, B. (2009). *ARTIFICIAL NEURAL NETWORKS*. PHI Learning Pvt. Ltd.
- Yu, H., Samuels, D. C., Zhao, Y., & Guo, Y. (2019). Architectures and accuracy of artificial neural network for disease classification from omics data. *BMC Genomics*, 20(1), 167. <https://doi.org/10.1186/s12864-019-5546-z>
- Zayani, R., Bouallegue, R., & Roviras, D. (2008). Adaptive Predistortions Based on Neural Networks Associated with Levenberg-Marquardt Algorithm for Satellite Down Links. *EURASIP Journal on Wireless Communications and Networking*, 2008(1), 132729. <https://doi.org/10.1155/2008/132729>
- Özbek, F. S., & Fidan, H. (2009). Estimation of pesticides usage in the agricultural sector in Turkey using Artificial Neural Network (ANN). *Journal of Animal and Plant Sciences*, 4, 373–378.

APPENDIX A

Statistical indexes of all petrophysical logs (input variables) in each well.

Gamma-ray [API°]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	40.02	21.29	29.02	24.01	26.39	28.42	31.72	35.20
15/12-4	80.17	9.00	31.48	10.79	13.42	28.01	39.79	65.66
15/12-5	102.2	25.90	53.30	26.87	35.94	41.28	84.15	96.68
	0							
15/12-6S	66.39	21.40	40.06	26.18	30.64	39.07	49.56	55.87
15/12-9S	108.0	13.19	59.09	40.38	46.06	53.76	76.25	87.24
	3							
15/12-A8	152.1	35.02	80.72	49.21	56.44	68.71	113.0	126.6
	0						3	3
Self-potential [mV]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	100.4	54.05	63.87	57.40	58.92	61.82	68.81	73.15
	0							
15/12-4	126.1	97.24	107.0	98.49	102.2	104.9	107.7	120.1
	4		9		4	4	6	1
15/12-5	105.7	57.92	73.17	59.86	61.32	65.52	77.28	103.6
	1							1
15/12-6S	74.63	-7.04	35.73	2.38	12.30	40.50	58.14	67.53
15/12-9S	93.68	0.25	37.06	3.91	7.33	28.06	72.86	83.42
15/12-A8	121.7	100.0	110.7	104.0	105.7	108.7	116.8	119.3
	8	8	8	1	5	3	2	5
Caliper [in.]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	14.86	7.21	7.98	7.32	7.39	7.45	7.62	11.76
15/12-4	16.35	8.81	11.59	8.84	8.91	12.70	12.75	15.08
15/12-5	9.33	8.57	8.70	8.59	8.60	8.61	8.76	8.95
15/12-6S	18.27	8.50	10.96	8.60	9.00	10.70	11.71	14.50
15/12-9S	9.86	8.46	8.68	8.46	8.55	8.56	8.85	9.05
15/12-A8	10.43	8.55	8.71	8.57	8.60	8.63	8.83	8.89
Sonic [μsec/ft]								
Well	max	min	mean	P10	P25	P50	P75	P90

15/12-1	128.5	53.20	84.47	60.94	84.96	88.68	90.53	91.19
	6							
15/12-4	121.2	67.73	85.11	72.08	73.61	78.19	94.31	114.5
	9							9
15/12-5	116.3	77.16	88.78	81.20	81.88	87.24	93.03	99.45
	4							
15/12-6S	122.1	50.91	83.03	69.33	77.95	82.16	90.02	97.27
	6							
15/12-9S	129.3	54.14	88.57	80.43	82.74	88.28	94.25	98.78
	8							
15/12-	96.53	60.10	83.88	73.51	79.52	84.81	90.04	93.12
A8								
Neutron porosity [frac.]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	41.07	6.82	21.85	15.28	21.88	22.78	23.98	24.58
15/12-4	32.09	11.26	18.44	12.41	13.81	17.87	22.43	25.37
15/12-5	27.81	17.03	21.95	19.67	20.53	21.76	23.23	24.93
15/12-6S	64.34	3.74	19.79	12.93	16.47	19.24	22.14	25.75
15/12-9S	67.95	6.22	20.22	16.28	17.73	20.06	22.13	23.93
15/12-	31.58	5.55	19.40	14.38	16.58	19.53	22.06	24.76
A8								
Bulk density [g/cm³]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	2.70	1.92	2.38	2.29	2.33	2.35	2.40	2.56
15/12-4	2.54	2.03	2.37	2.15	2.26	2.43	2.48	2.50
15/12-5	2.51	2.12	2.29	2.18	2.21	2.25	2.43	2.48
15/12-6S	2.73	1.49	2.35	2.13	2.24	2.39	2.46	2.52
15/12-9S	2.66	1.31	2.25	2.11	2.14	2.23	2.38	2.42
15/12-	3.09	1.61	2.44	2.26	2.32	2.42	2.58	2.63
A8								
Log Deep resistivity [ohmm]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	2.77	-0.69	0.00	-0.53	-0.48	-0.13	0.16	0.74
15/12-4	1.42	-1.42	0.21	-1.17	-0.40	0.48	0.84	1.10
15/12-5	0.65	-1.61	-0.29	-1.43	-1.06	-0.11	0.51	0.57
15/12-6S	3.58	-0.36	1.02	0.22	0.52	0.78	1.34	2.30
15/12-9S	3.56	-1.66	0.43	-1.41	-0.94	0.36	1.52	2.36
15/12-	3.49	0.40	1.10	0.65	0.78	1.06	1.34	1.56
A8								

Log Medium resistivity [ohmm]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	2.29	-0.04	0.77	0.29	0.52	0.68	0.88	1.47
15/12-4	1.21	-1.63	0.14	-1.23	-0.45	0.47	0.76	0.93
15/12-5	0.63	-1.78	-0.34	-1.64	-1.13	-0.07	0.39	0.55
15/12-6S	3.90	0.15	1.28	0.64	0.71	1.12	1.57	2.22
15/12-9S	3.40	-1.54	0.49	-1.35	-0.72	0.48	1.52	2.23
15/12-A8	3.30	0.18	0.94	0.35	0.62	0.98	1.21	1.38

APPENDIX B

Statistical indexes of the oil saturation, water saturation and porosity in each well.

Oil saturation [frac.]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	43.30	4.30	14.38	8.76	10.48	11.70	15.63	25.56
15/12-4	89.70	0.00	5.42	0.00	0.00	0.85	3.60	13.43
15/12-5	17.20	0.00	4.22	0.00	0.00	0.90	9.00	13.11
15/12-6S	78.40	0.00	8.63	0.00	0.00	5.10	13.00	20.02
15/12-9S	46.90	0.00	10.91	0.60	2.70	9.10	17.10	24.85
15/12-A8	42.89	0.00	17.09	6.02	10.03	15.27	23.40	30.30
Water saturation [frac.]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	73.30	3.10	39.50	10.82	34.15	42.30	47.68	52.12
15/12-4	97.70	32.70	75.22	40.56	57.80	81.70	90.40	94.89
15/12-5	90.00	2.90	61.77	23.41	37.30	73.60	84.70	88.45
15/12-6S	91.70	13.30	62.96	35.36	44.50	66.40	83.85	88.24
15/12-9S	89.50	26.30	61.30	38.55	49.80	63.50	73.00	78.15
15/12-A8	97.23	26.52	65.54	44.55	55.48	67.05	78.15	82.28
Porosity [frac.]								
Well	max	min	mean	P10	P25	P50	P75	P90
15/12-1	28.00	1.70	18.28	7.90	15.20	20.85	23.00	23.83
15/12-4	31.60	8.90	17.08	10.01	11.25	13.70	23.30	29.19
15/12-5	27.81	17.03	21.95	19.67	20.53	21.76	23.23	24.93
15/12-6S	43.40	2.10	16.10	7.50	10.30	13.25	22.40	28.20
15/12-9S	36.10	6.70	21.77	12.38	15.00	22.40	28.80	31.10
15/12-A8	29.09	3.51	16.67	8.42	9.48	17.13	23.02	26.73

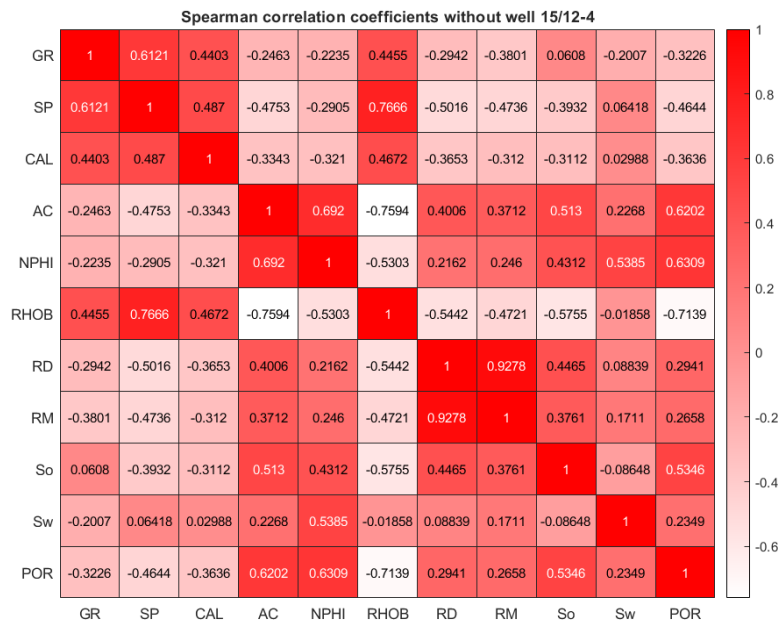
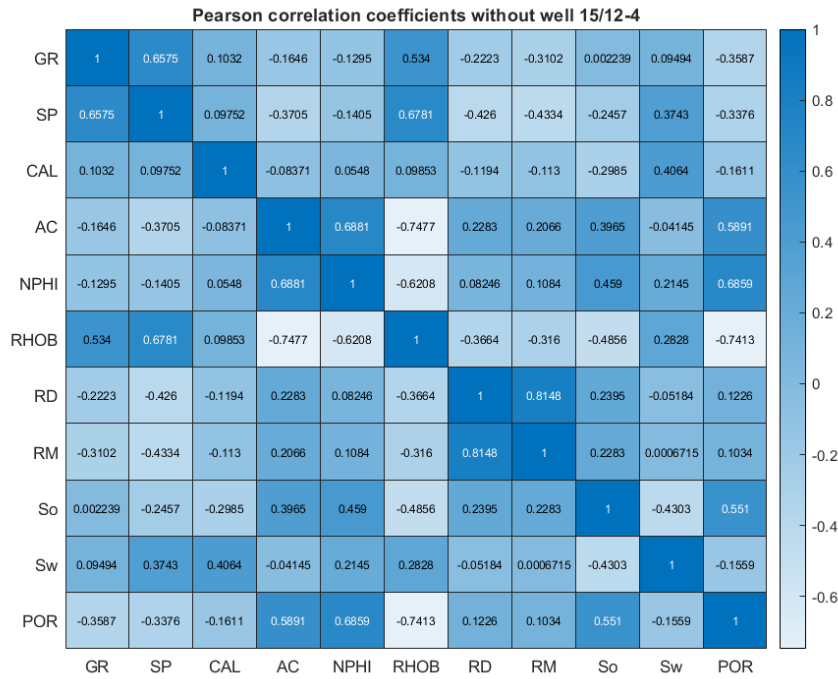
APPENDIX C

Statistical index of the data set including well 15/12-4

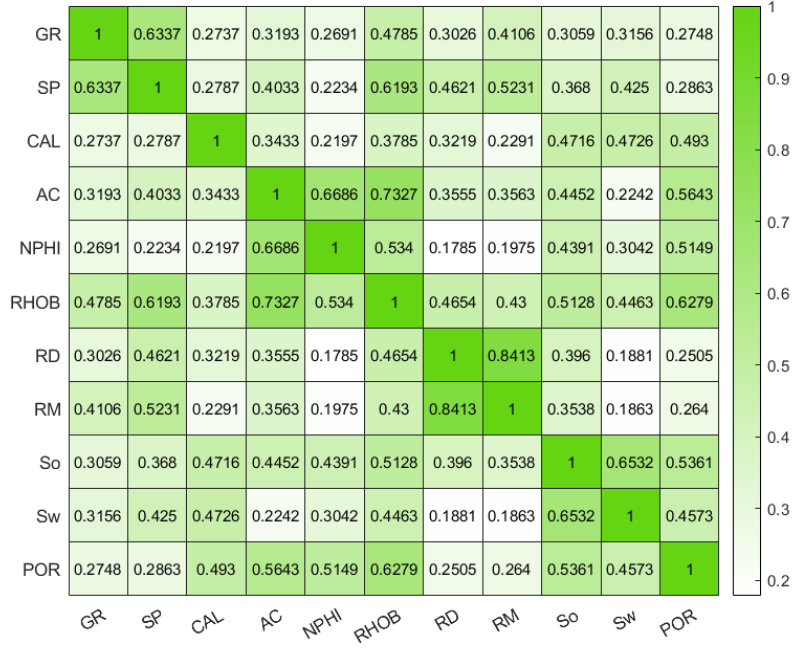
	max	min	mean	P10	P25	P50	P75	P90
GR [API°]	152.10	9.00	52.80	27.05	35.85	47.78	61.78	88.11
SP [mV]	126.14	-7.04	55.92	5.12	21.20	59.68	85.04	107.14
CAL [in.]	18.27	7.21	9.47	8.47	8.56	8.70	9.17	12.70
AC [μs/sec]	129.38	50.91	86.08	74.29	80.84	85.69	91.42	97.84
NPHI [frac.]	67.95	3.74	20.14	14.46	17.46	20.14	22.57	24.76
RHOB [g/cm ³]	3.09	1.31	2.32	2.13	2.20	2.32	2.44	2.50
Log (RD) [Ωm]	3.58	-1.66	0.56	-1.14	-0.11	0.58	1.20	2.04
Log (RM) [Ωm]	3.90	-1.78	0.66	-1.09	0.22	0.70	1.30	2.08
So [frac.]	97.70	2.90	34.64	17.00	18.68	22.20	47.20	79.10
Sw [frac.]	89.70	0.00	16.97	1.72	12.71	18.27	21.34	24.04
POR [frac]	67.95	1.70	18.59	9.60	13.77	19.47	22.51	25.50

APPENDIX D

The results from the Pearson, Spearman and distance correlation of the wells used in the training, validation, and test set.



Distance correlation without well 15/12-4



APPENDIX E

Here are the performance results from the calibrated models from the original models, first feature selection, second feature selection and the stepwise regression feature selection

A detailed overview of the performance result from the calibrated models using all petrophysical logs.

		Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Pearson selection	Training	25.57	5.06	20149.12	2.46	0.82	0.85	0.84	
	Validation	13.35	3.65	2015.11	1.96	0.89	0.91	0.90	
	Testing	29.90	5.47	3318.77	3.29	0.66	0.68	0.68	
	Well 15/12-4	285.86	16.91	12006.10	12.61	0.10	0.42	0.33	
			Water saturation						
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
		Training	22.14	4.71	17449.20	2.05	0.85	0.89	0.88
		Validation	12.89	3.59	1946.84	1.80	0.90	0.94	0.89
		Testing	19.31	3.59	2143.37	2.67	0.79	0.78	0.77
		Well 15/12-4	3625.45	60.21	152268.82	55.14	-0.26	-0.18	0.42
			Porosity						
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	0.00	0.00	13510.81	0.00	0.88	1.00	0.86	
	Validation	11.19	3.35	2843.09	2.10	0.87	0.89	0.86	
	Testing	25.31	5.03	4378.81	2.52	0.64	0.78	0.77	
	Well 15/12-4	478.44	21.87	55499.24	17.82	0.63	0.45	0.62	
Spearman selection			Oil saturation						
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
		Training	25.39	5.04	20008.22	2.45	0.82	0.87	0.85
		Validation	14.41	3.80	2176.44	2.05	0.88	0.93	0.88
		Testing	30.49	5.52	3383.93	3.03	0.64	0.73	0.70
		Well 15/12-4	250.34	15.82	10514.15	10.93	0.13	0.29	0.30
			Water saturation						
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
		Training	22.14	4.71	17449.20	2.05	0.85	0.89	0.88
		Validation	12.89	3.59	1946.84	1.80	0.90	0.94	0.89
		Testing	19.31	3.59	2143.37	2.67	0.79	0.78	0.77
		Well 15/12-4	3625.45	60.21	152268.82	55.14	-0.26	-0.18	0.42
		Porosity							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
	Training	0.00	0.00	13510.81	0.00	0.88	1.00	0.86	
	Validation	11.19	3.35	2843.09	2.10	0.87	0.89	0.86	
	Testing	25.31	5.03	4378.81	2.52	0.64	0.78	0.77	
	Well 15/12-4	478.44	21.87	55499.24	17.82	0.63	0.45	0.62	

Distance selection	Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	26.72	5.17	21056.67	2.53	0.81	0.86	0.84
	Validation	13.69	3.70	2066.50	2.03	0.89	0.93	0.90
	Testing	28.56	5.34	3170.54	2.92	0.67	0.75	0.75
	Well 15/12-4	356.65	18.89	14979.29	14.89	-0.07	-0.27	0.25
	Water saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	27.50	5.24	21671.38	2.76	0.81	0.85	0.82
	Validation	15.14	3.89	2286.71	2.17	0.88	0.92	0.90
	Testing	26.57	3.89	2949.07	3.17	0.69	0.76	0.73
	Well 15/12-4	4589.37	67.74	192753.60	63.94	-0.37	-0.47	0.49
	Porosity							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Training	0.00	0.00	13510.81	0.00	0.88	1.00	0.86	
Validation	11.19	3.35	2843.09	2.10	0.87	0.89	0.86	
Testing	25.31	5.03	4378.81	2.52	0.64	0.78	0.77	
Well 15/12-4	478.44	21.87	55499.24	17.82	0.63	0.45	0.62	

A detailed overview of the performance result from the calibrated models from the first feature selection

Pearson selection	Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	42.99	6.56	33875.57	3.91	0.67	0.63	0.62
	Validation	27.51	5.24	4153.81	3.28	0.76	0.81	0.74
	Testing	41.78	6.46	4637.60	3.91	0.43	0.57	0.53
	Well 15/12-4	226.79	15.06	9525.15	5.48	-0.11	-0.21	0.20
	Water saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	78.24	8.85	61649.42	5.86	-	-	-
	Validation	64.57	8.04	9750.16	5.42	-	-	-
	Testing	48.76	8.04	5412.24	4.74	-	-	-
	Well 15/12-4	3703.87	60.86	155562.45	57.67	-	-	-
	Porosity							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Training	0.00	0.00	28391.11	0.00	0.73	1.00	0.67	
Validation	19.69	4.44	5001.82	2.80	0.76	0.72	0.73	
Testing	26.61	5.16	4603.47	3.28	0.52	0.53	0.56	
Well 15/12-4	105.98	10.29	12293.55	9.54	-0.44	-0.40	0.48	

Spearman selection	Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	31.50	5.61	24822.07	3.09	0.77	0.81	0.77
	Validation	24.15	4.91	3647.15	2.64	0.79	0.86	0.80
	Testing	37.02	6.08	4108.67	3.55	0.53	0.66	0.63
	Well 15/12-4	233.72	15.29	9816.40	10.53	0.03	0.12	0.24
	Water saturation							

		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Distance selection	Training	42.99	6.56	33875.57	3.91	0.67	0.63	0.62	
	Validation	27.51	5.24	4153.81	3.28	0.76	0.81	0.74	
	Testing	41.78	6.46	4637.60	3.91	0.43	0.57	0.53	
	Well 15/12-4	226.79	15.06	9525.15	5.48	-0.11	-0.21	0.20	
	Porosity								
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training		78.22	8.84	61636.55	5.88	-	-	-
	Validation		64.58	8.04	9751.47	5.45	-	-	-
	Testing		48.78	8.04	5415.07	4.77	-	-	-
	Well 15/12-4		3718.04	60.98	156157.78	57.80	-	-	-
	Oil saturation								
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training		0.00	0.00	26114.72	0.00	0.75	1.00	0.70
	Validation		20.13	4.49	5112.04	2.73	0.76	0.75	0.75
	Testing		25.41	5.04	4395.84	3.06	0.54	0.58	0.60
	Well 15/12-4		59.66	7.72	6920.20	5.49	0.29	0.21	0.38
Water saturation									
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Training		109.97	10.49	86765.41	5.48	0.89	0.86	0.89	
Validation		62.68	7.92	9464.31	3.92	0.92	0.88	0.93	
Testing		226.67	7.92	25160.20	7.38	0.76	0.76	0.79	
Well 15/12-4		2009.16	44.82	84384.55	37.90	-0.30	-0.36	0.44	
Porosity									
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Training		0.00	0.00	26114.72	0.00	0.75	1.00	0.70	
Validation		20.13	4.49	5112.04	2.73	0.76	0.75	0.75	
Testing		25.41	5.04	4395.84	3.06	0.54	0.58	0.60	
Well 15/12-4		59.66	7.72	6920.20	5.49	0.29	0.21	0.38	

A detailed overview of the performance result from the calibrated models from the second feature selection

		Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	
Pearson selection	Training	26.27	5.13	20699.42	2.75	0.82	0.83	0.81	
	Validation	19.94	4.46	3010.24	2.67	0.83	0.85	0.81	
	Testing	28.41	5.33	3153.11	3.33	0.67	0.69	0.69	
	Well 15/12-4	982.87	31.35	41280.64	18.75	0.05	0.23	0.31	
	Water saturation								
			MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training		89.00	9.43	70217.69	5.21	0.91	0.84	0.91
	Validation		60.78	7.80	9177.85	4.34	0.93	0.82	0.93
	Testing		128.33	7.80	14244.20	6.36	0.87	0.78	0.87
	Well 15/12-4		2419.89	49.19	101635.34	42.16	-0.41	-0.49	0.54
Porosity									
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE	

Training	0.00	0.00	14098.95	0.00	0.88	1.00	0.85
Validation	12.69	3.56	3224.11	2.22	0.85	0.82	0.82
Testing	20.39	4.52	3527.98	2.54	0.69	0.76	0.77
Well 15/12-4	97.65	9.88	11327.50	7.14	0.59	0.57	0.68

Oil saturation							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	31.50	5.61	24822.07	3.09	0.77	0.81	0.77
Validation	24.15	4.91	3647.15	2.64	0.79	0.86	0.80
Testing	37.02	6.08	4108.67	3.55	0.53	0.66	0.63
Well 15/12-4	233.72	15.29	9816.40	10.53	0.03	0.12	0.24
Water saturation							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	94.45	9.72	74523.29	5.04	0.91	0.88	0.91
Validation	85.88	9.27	12967.16	3.96	0.89	0.92	0.92
Testing	208.88	9.27	23185.40	7.18	0.78	0.71	0.80
Well 15/12-4	999.66	31.62	41985.57	25.55	0.49	0.41	0.53
Porosity							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	0.00	0.00	14651.92	0.00	0.87	1.00	0.85
Validation	15.88	3.99	4034.12	2.52	0.81	0.83	0.80
Testing	22.24	4.72	3847.09	2.59	0.66	0.76	0.78
Well 15/12-4	45.55	6.75	5283.82	5.20	0.82	0.67	0.86

Oil saturation							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	30.77	5.55	24250.66	2.97	0.78	0.82	0.78
Validation	22.72	4.77	3430.17	2.64	0.81	0.86	0.82
Testing	36.92	6.08	4098.59	3.52	0.52	0.67	0.64
Well 15/12-4	1376.05	37.10	57794.25	31.67	0.23	0.47	0.29
Water saturation							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	109.97	10.49	86765.41	5.48	0.89	0.86	0.89
Validation	62.68	7.92	9464.31	3.92	0.92	0.88	0.93
Testing	226.67	7.92	25160.20	7.38	0.76	0.76	0.79
Well 15/12-4	2009.16	44.82	84384.55	37.90	-0.30	-0.36	0.44
Porosity							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	0.00	0.00	14098.95	0.00	0.88	1.00	0.85
Validation	12.69	3.56	3224.11	2.22	0.85	0.82	0.82
Testing	20.39	4.52	3527.98	2.54	0.69	0.76	0.77
Well 15/12-4	45.55	6.75	5283.82	5.20	0.82	0.67	0.86

A detailed overview of the performance result from the calibrated models using stepwise regression for feature selection.

Oil saturation							
	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	33.26	5.77	26207.71	3.21	0.76	0.77	0.75

Pearson selection	Validation	18.26	4.27	2757.74	2.56	0.85	0.86	0.81
	Testing	30.99	5.57	3439.49	3.24	0.61	0.68	0.65
	Well 15/12-4	332.75	18.24	13975.61	15.01	0.09	0.05	0.18
	Water saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	73.26	8.56	57798.74	4.34	0.93	0.87	0.93
	Validation	66.44	8.15	10032.55	4.13	0.92	0.88	0.92
	Testing	210.75	8.15	23393.18	7.03	0.79	0.74	0.81
	Well 15/12-4	5109.97	71.48	214618.84	61.92	-0.54	-0.64	0.63
	Porosity							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	0.00	0.00	23403.57	0.00	0.78	1.00	0.75
	Validation	15.35	3.92	3898.65	2.47	0.81	0.80	0.77
	Testing	23.33	4.83	4036.89	2.98	0.59	0.64	0.64
Well 15/12-4	95.56	9.78	11084.78	7.19	0.13	0.19	0.23	
Spearman selection	Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	37.01	6.08	29165.80	3.28	0.73	0.74	0.71
	Validation	20.79	4.56	3139.07	2.61	0.83	0.88	0.82
	Testing	42.56	6.52	4724.02	3.63	0.43	0.61	0.58
	Well 15/12-4	287.75	16.96	12085.37	11.37	-0.16	-0.12	0.22
	Water saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	56.03	7.49	44209.92	4.05	0.94	0.88	0.94
	Validation	94.68	9.73	14296.57	4.17	0.89	0.92	0.89
	Testing	155.59	9.73	17270.10	6.15	0.84	0.77	0.86
	Well 15/12-4	1122.04	33.50	47125.71	27.98	0.14	0.07	0.31
	Porosity							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	0.00	0.00	23403.57	0.00	0.78	1.00	0.75	
Validation	15.35	3.92	3898.65	2.47	0.81	0.80	0.77	
Testing	23.33	4.83	4036.89	2.98	0.59	0.64	0.64	
Well 15/12-4	95.56	9.78	11084.78	7.19	0.13	0.19	0.23	
Distance selection	Oil saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	37.01	6.08	29165.80	3.28	0.73	0.74	0.71
	Validation	20.79	4.56	3139.07	2.61	0.83	0.88	0.82
	Testing	42.56	6.52	4724.02	3.63	0.43	0.61	0.58
	Well 15/12-4	287.75	16.96	12085.37	11.37	-0.16	-0.12	0.22
	water saturation							
		MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
	Training	90.54	9.52	71435.86	5.46	0.91	0.79	0.91
	Validation	69.73	8.35	10528.49	4.00	0.91	0.79	0.93
	Testing	185.36	8.35	20575.11	6.84	0.81	0.71	0.83
	Well 15/12-4	821.80	28.67	34515.63	22.14	0.22	0.12	0.35
	Porosity							

	MSE	RMSE	SSE	MAE	PEARSON	SPEARMAN	DISTANCE
Training	0.00	0.00	17132.82	0.00	0.85	1.00	0.81
Validation	16.04	4.01	4074.77	2.62	0.81	0.80	0.78
Testing	18.75	4.33	3244.08	2.69	0.68	0.71	0.72
Well 15/12-4	862.97	29.38	100104.52	23.04	-0.11	-0.05	0.37

APPENDIX F

Summary of the results from variations of each variable and the statistical evaluations.

Oil saturation						
	Variation [%]	MSE	MAE	R pearson	R spearman	R distance
GR	10.00	222.00	12.09	0.00	0.02	0.06
	20.00	89.38	7.52	0.02	0.00	0.08
	30.00	88.54	7.44	0.00	0.01	0.05
	40.00	86.67	7.19	0.00	0.01	0.05
	50.00	85.05	6.97	0.00	0.01	0.07
SP	10.00	123.71	9.80	0.26	0.22	0.26
	20.00	151.74	11.04	0.21	0.13	0.22
	30.00	124.06	9.84	0.14	0.11	0.15
	40.00	111.30	8.70	0.03	0.00	0.08
	50.00	125.89	8.81	0.00	0.01	0.07
CAL	10.00	49.33	6.36	0.11	0.36	0.16
	20.00	50.68	6.35	0.12	0.36	0.16
	30.00	51.83	6.41	0.12	0.36	0.16
	40.00	53.54	6.49	0.10	0.34	0.14
	50.00	60.68	6.90	0.04	0.07	0.06
AC	10.00	466.99	19.52	0.04	0.02	0.07
	20.00	511.20	20.59	0.05	0.02	0.07
	30.00	546.53	21.74	0.07	0.08	0.09
	40.00	564.68	22.42	0.06	0.00	0.08
	50.00	524.27	22.00	0.26	0.18	0.28
NPHI	10.00	115.84	8.65	0.19	0.05	0.26
	20.00	110.53	8.42	0.17	0.04	0.26
	30.00	104.99	8.16	0.15	0.03	0.25
	40.00	99.47	7.88	0.12	0.03	0.23
	50.00	94.30	7.59	0.08	0.02	0.19
RHOB	10.00	96.88	7.96	0.07	0.01	0.12
	20.00	93.82	7.75	0.05	0.01	0.11

	30.00	91.07	7.56	0.04	0.00	0.11
	40.00	88.70	7.38	0.03	0.00	0.10
	50.00	86.75	7.19	0.02	0.00	0.10
RD	10.00	86.53	6.64	0.00	0.02	0.13
	20.00	86.12	6.63	0.00	0.02	0.13
	30.00	85.72	6.61	0.00	0.02	0.13
	40.00	85.33	6.59	0.00	0.02	0.14
	50.00	84.96	6.56	0.00	0.02	0.14
RM	10.00	81.81	6.49	0.00	0.01	0.15
	20.00	81.85	6.46	0.00	0.01	0.15
	30.00	81.93	6.45	0.00	0.01	0.15
	40.00	82.08	6.44	0.00	0.01	0.15
	50.00	82.29	6.45	0.00	0.01	0.15
Water saturation						
	Variation [%]	MSE	MAE	R pearson	R spearman	R distance
GR	10.00	290.80	15.74	0.11	0.03	0.18
	20.00	271.11	15.21	0.11	0.02	0.18
	30.00	255.22	14.76	0.10	0.02	0.17
	40.00	242.56	14.36	0.10	0.02	0.18
	50.00	232.39	14.01	0.10	0.01	0.18
SP	10.00	110.69	8.79	0.16	0.18	0.24
	20.00	131.72	10.11	0.19	0.16	0.25
	30.00	134.22	10.22	0.13	0.22	0.18
	40.00	162.24	10.81	0.02	0.03	0.03
	50.00	213.43	11.82	0.02	0.03	0.07
CAL	10.00	129.09	9.94	0.06	0.02	0.11
	20.00	159.10	11.06	0.01	0.01	0.09
	30.00	185.65	11.94	0.00	0.00	0.09
	40.00	203.64	12.46	0.01	0.00	0.10
	50.00	212.27	12.68	0.02	0.00	0.10
AC	10.00	218.49	12.76	0.21	0.22	0.27
	20.00	197.55	12.59	0.27	0.24	0.29

	30.00	178.70	12.09	0.21	0.10	0.27
	40.00	151.88	10.95	0.09	0.00	0.26
	50.00	105.45	9.16	0.02	0.00	0.11
NPHI	10.00	1156.73	26.44	0.23	0.15	0.32
	20.00	933.35	23.91	0.21	0.14	0.31
	30.00	721.88	21.34	0.21	0.15	0.29
	40.00	536.71	18.78	0.21	0.16	0.26
	50.00	391.18	16.40	0.20	0.19	0.24
RHOB	10.00	465.58	18.87	0.39	0.36	0.38
	20.00	399.31	17.02	0.36	0.34	0.35
	30.00	316.32	14.91	0.29	0.29	0.28
	40.00	233.21	12.35	0.21	0.19	0.21
	50.00	175.46	10.05	0.12	0.10	0.18
RD	10.00	222.63	12.70	0.08	0.01	0.16
	20.00	218.67	12.61	0.08	0.01	0.16
	30.00	214.93	12.52	0.08	0.01	0.16
	40.00	211.36	12.43	0.08	0.01	0.16
	50.00	207.96	12.34	0.07	0.01	0.15
RM	10.00	194.72	11.52	0.01	0.00	0.10
	20.00	190.10	11.28	0.02	0.00	0.11
	30.00	187.02	11.29	0.03	0.00	0.12
	40.00	185.18	11.35	0.03	0.00	0.12
	50.00	186.21	11.44	0.04	0.00	0.13