



PREDICTION OF PETROPHYSICAL PROPERTIES USING MACHINE LEARNING AND HIERARCHICAL MULTI-TASK LINEAR MODELS

Marcelo Ramalho Albuquerque

Dissertação de Mestrado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Mestre em Engenharia de Sistemas e Computação.

Orientador: Carlos Eduardo Pedreira

Rio de Janeiro
Setembro de 2020

PREDICTION OF PETROPHYSICAL PROPERTIES USING MACHINE
LEARNING AND HIERARCHICAL MULTI-TASK LINEAR MODELS

Marcelo Ramalho Albuquerque

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO INSTITUTO
ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE
ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO
COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO
GRAU DE MESTRE EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E
COMPUTAÇÃO.

Orientador: Carlos Eduardo Pedreira

Aprovada por: Prof. Carlos Eduardo Pedreira
Prof. Geraldo Bonorino Xexéo
Dr. Rodrigo Surmas

RIO DE JANEIRO, RJ – BRASIL
SETEMBRO DE 2020

Albuquerque, Marcelo Ramalho

Prediction of Petrophysical Properties Using Machine Learning and Hierarchical Multi-Task Linear Models/Marcelo Ramalho Albuquerque. – Rio de Janeiro: UFRJ/COPPE, 2020.

XII, 57 p.: il.; 29,7cm.

Orientador: Carlos Eduardo Pedreira

Dissertação (mestrado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2020.

Referências Bibliográficas: p. 54 – 57.

1. petrophysics. 2. machine learning. 3. hierarchical linear regression. I. Pedreira, Carlos Eduardo. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

*I just wondered how things were
put together - Claude Shannon*

Agradecimentos

Aos meus pais, pelo cuidado e amor com que criaram todos os seus filhos.

Ao professor Pedreira, pelo compartilhamento de conhecimento, sabedoria e orientação.

Aos colegas de laboratório da Petrobras, pelo apoio e pelos conhecimentos compartilhados e construídos.

À Petrobras, pela oportunidade e incentivo ao desenvolvimento dos seus funcionários e da ciência.

À Bibi, pela paciência, carinho e amor!

Resumo da Dissertação apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

PREVISÃO DE PROPRIEDADES PETROFÍSICAS UTILIZANDO MODELOS
DE APRENDIZADO DE MÁQUINA E MODELOS LINEARES
HIERÁRQUICOS MULTI-TASK

Marcelo Ramalho Albuquerque

Setembro/2020

Orientador: Carlos Eduardo Pedreira

Programa: Engenharia de Sistemas e Computação

A caracterização petrofísica de rochas reservatório é uma etapa fundamental na avaliação de reservatórios de petróleo e é normalmente realizada através de ensaios laboratoriais que incorrem em custos e prazos significativos. Neste trabalho, foram avaliadas técnicas estatísticas e de aprendizado de máquina na estimativa de permeabilidade absoluta, curvas de pressão capilar óleo-água e curvas de permeabilidade relativa água-óleo, a partir de dados de porosimetria por intrusão de mercúrio, permeabilidade absoluta e porosidade. A partir da extração de diversas características das curvas de pressão capilar por intrusão de mercúrio, algoritmos de aprendizado de máquina e métodos estatísticos para a estimativa destas propriedades foram avaliados e comparados a métodos clássicos da literatura. Métodos lineares hierárquicos e multi-task foram avaliados para estimativa de curvas de pressão capilar óleo-água e permeabilidade relativa água-óleo. Foi demonstrado o efeito de regularização dos modelos lineares hierárquicos, que resultaram em modelos mais precisos, coerentes e com menor incerteza a posteriori.

Abstract of Dissertation presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc.)

PREDICTION OF PETROPHYSICAL PROPERTIES USING MACHINE LEARNING AND HIERARCHICAL MULTI-TASK LINEAR MODELS

Marcelo Ramalho Albuquerque

September/2020

Advisor: Carlos Eduardo Pedreira

Department: Systems Engineering and Computer Science

Petrophysical characterization of reservoir rocks is a fundamental step in the evaluation of oil reservoirs, and is usually executed through laboratory experiments that incur in large costs and schedules. In this work, statistical techniques and machine learning models were evaluated for the estimation of absolute permeability, oil-water capillary pressure and water-oil relative permeability curves, using mercury injection porosimetry, absolute permeability and porosity data. Through the exploration of several feature engineering and modeling strategies, machine learning and statistical models were assessed and compared to classical linear methods. Multi-task and hierarchical linear models were also evaluated for the estimation of special core analysis parameters from mercury injection porosimetry and routine core analysis data. On the evaluated dataset, hierarchical linear models were shown to have better precision, consistency and lower posterior uncertainty metrics when compared to simple linear regression models.

Contents

List of Figures	ix
List of Tables	xii
1 Introduction	1
2 Measuring Petrophysical Properties	3
2.1 Routine Core Analysis	3
2.2 Special Core Analysis	4
2.2.1 Mercury Injection Capillary Pressure	4
2.2.2 Centrifuge Capillary Pressure	9
2.2.3 Relative Permeability	12
3 Absolute Permeability Regression	18
3.1 Mercury Porosimetry Feature Engineering	18
3.1.1 Statistical Features	18
3.1.2 Linear Features	19
3.1.3 Pore Throat Size Class Distribution	19
3.1.4 Gaussian Mixture Fit Features	20
3.1.5 Dimensionality Reduction Features	22
3.2 Absolute Permeability Regression Models	24
3.2.1 Linear Regression Models	25
3.2.2 Black-Box Machine Learning Models	26
4 Hierarchical Multi-Task Linear Regression	31
4.1 Multi-Task Linear Regression	31
4.2 Hierarchical Linear Regression	33
Conclusion	52
Bibliography	54

List of Figures

1.1	Reservoir Engineering Simulation Models	1
2.1	Pore media grain and pore spaces	4
2.2	Contact Angle between fluid phase boundary and solid surface	5
2.3	AutoPore IV Series Mercury Porosimeter (left) and glass penetrometer (right) used in automated MICP acquisitions.	6
2.4	Mercury Injection Capillary Pressure Curves for several rock samples, colored by absolute permeability.	7
2.5	Accessible pore volume probability distribution curves $p(\log r)$, for the same MICP experiments depicted in Figure 2.4.	8
2.6	Relationship between median pore throat radius and absolute permeability of several different reservoirs, from most to least heterogeneous	9
2.7	Drainage Oil-Water Centrifuge capillary pressure experiment geometry.	10
2.8	Estimation of capillary pressure curve from centrifuge measurements.	11
2.9	Sample of the experimental capillary pressure curve dataset (points) and fitted capillary pressure model (lines).	12
2.10	Sample centrifuge capillary pressure curves, parameterized by equation (2.11), and corresponding pore volume probability distribution curves $p(\log r)$	13
2.11	Associated pore volume distributions $p(\log r)$ of correspondent Centrifuge Capillary Pressure (blue curve) and MICP (black curve) samples.	14
2.12	Correlations between parameters α , S_{wi} and median pore throat radius of correspondent rock fragment $p(\log r)$ distribution.	14
2.13	Unsteady-state relative permeability experimental setup.	15
2.14	Example of history matched experimental data.	16
2.15	Example of history matched relative permeability and fractionary flow curves compared to analytic JBN interpreted experimental results. . .	16
3.1	Pore-throat class distribution for two sample reservoirs (A and B). . .	20
3.2	Example of bimodal gaussian mixture fit of $p(\log r)$ distribution. . . .	20

3.3	Gaussian mixture (red lines) fitted to $p(\log r)$ distributions of MICP data samples (black lines).	21
3.4	First two principal components of each of the datasets MICP samples.	22
3.5	Negative correlation between absolute permeability and the first principal component.	23
3.6	Positive correlation between the interquartile-range iqr and the second principal component $PC2$.	23
3.7	MICP curves non-linear decomposition using UMAP.	24
3.8	Manifold maps (left) showing areas with similar associated pore volume distributions $p(\log r)$ (right).	25
3.9	Visual evaluation of the predicted and observed absolute permeability models using linear features.	26
3.10	Comparison of the R^2 metric for the fitted linear regression models.	27
3.11	Five-fold cross-validation procedure.	27
3.12	Boxplots for repeated five-fold cross-validated MAE, R^2 and RMSE metric results for each model.	29
3.13	Visual evaluation of the estimated absolute permeability models using Machine Learning and linear models.	30
4.1	Sample of the experimental capillary pressure curve dataset (points) and fitted capillary pressure model (lines).	32
4.2	Correlations between parameters α , Sw_i and median pore throat radius of correspondent rock fragment $p(\log r)$ distribution.	32
4.3	Visual comparison of experimental capillary pressure curves (blue lines), samples from the conditional distribution (grey lines) and average predicted curves (dashed black lines).	34
4.4	Experimental capillary pressure curve parameters (blue dots and line tendency) and predicted parameters estimated using the posterior mean (black dots).	35
4.5	Behaviour of predicted capillary pressure curves with absolute permeability.	35
4.6	Non-pooled simple linear regression.	37
4.7	Completely pooled simple linear regression models, grouped by reservoir.	38
4.8	Simple (top) and Hierarchical (bottom) linear regression models of Sw_i vs $\log(k_{abs})$, grouped by reservoir.	41
4.9	Intercept (top) and slope (bottom), simple and hierarchical linear regression Sw_i vs $\log(k_{abs})$ model parameters, grouped by reservoir.	42

4.10	Simple (top) and Hierarchical (bottom) linear regression models of Sor vs $\log(k_{abs})$, grouped by reservoir.	43
4.11	Intercept (top) and slope (bottom), simple and hierarchical linear regression Sor vs $\log(k_{abs})$ model parameters, grouped by reservoir. .	44
4.12	Simple (top) and Hierarchical (bottom) linear regression models of $k_{ro}@Swi$ vs $\log(k_{abs})$, grouped by reservoir.	46
4.13	Intercept (top) and slope (bottom), simple and hierarchical linear regression $k_{ro}@Swi$ vs $\log(k_{abs})$ model parameters, grouped by reservoir.	47
4.14	Simple (top) and Hierarchical (bottom) linear regression models of $k_{rw}@Sor$ vs $\log(k_{abs})$, grouped by reservoir.	49
4.15	Intercept (top) and slope (bottom), simple and hierarchical linear regression $k_{rw}@Sor$ vs $\log(k_{abs})$ model parameters, grouped by reservoir.	50
4.16	Example of multivariate posterior sample of relative permeability curves, fully incorporating the information from the available dataset.	51

List of Tables

3.1	Classical MICP linear correlation features	19
3.2	Median repeated five-fold cross-validated MAE, R^2 and RMSE met- rics for each model.	28
4.1	S_{wi} linear regression model metrics.	39
4.2	S_{or} linear regression model metrics.	40
4.3	$k_{ro}@S_{wi}$ linear regression model metrics.	45
4.4	$k_{rw}@S_{or}$ linear regression model metrics.	45
4.5	Multi-task linear regression model metrics.	48

Chapter 1

Introduction

Since the beginning of the 20th century, oil and the combustion engine have transformed many aspects of modern society in fundamental ways. As a cheap and reliable energy source, oil provided the means for revolutions in transportation, manufacturing, commerce and human behavior, in general. It remains a cornerstone of our society in 2020.

As knowledge of the warming effects of carbon in earth's atmosphere is further developed, conscious and economically sustainable exploitation of remaining oil reserves becomes paramount (Mackay 2009). The high volatility of oil prices induced by supply and demand shocks in recent years further complicates the assessment of oil and gas projects.

Reservoir engineering is a branch of petroleum engineering that studies fluid flow through porous hydrocarbon bearing rocks (Dake 2015). Assessment of oil and gas exploitation projects is commonly performed using numerical simulation models that incorporate geological, fluid and petrophysical information to predict oil and gas production curves. Reservoir uncertainties are particularly important at the beginning of a project, when few wells have been drilled, fluid and rock samples are scarce, and no production data is available. Petrophysical characterization entails the eval-

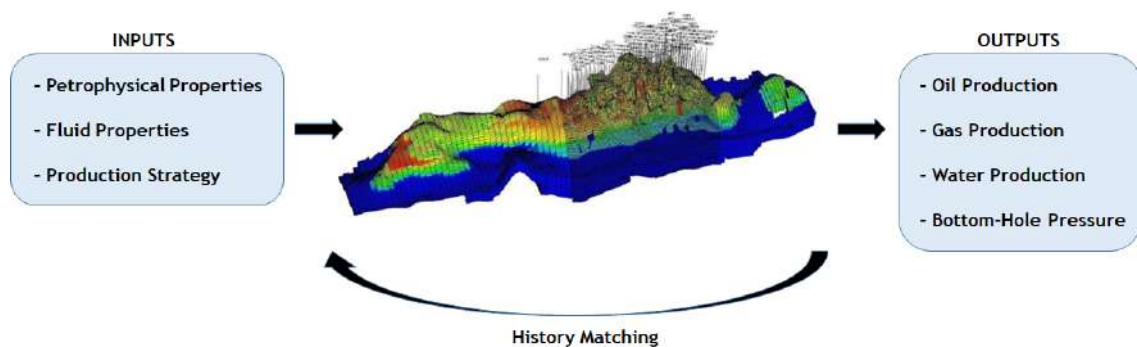


Figure 1.1: Reservoir Engineering Simulation Models

uation of physical and chemical properties of porous media, and its interaction with reservoir fluids. It is commonly performed through laboratory experiments, known in the petroleum industry as routine and special core analysis (Peters 2012). Routine petrophysical experiments characterize rock porosity and absolute permeability, are considered inexpensive and can be performed in little time, usually taking no more than a couple of days per sample. Special core analysis, characterize properties related to the balance of viscous and capillary forces acting at reservoir conditions and, in many situations, can be expensive and demand long experiments, which can take up to several months to be executed (Tiab and Donaldson 2004).

In this work, statistical methods were evaluated for the estimation of petrophysical properties using as input Mercury Injection Capillary Pressure (MICP) and routine core analysis data. This work is organized as follows: on chapter 2, petrophysical measurement methods and properties are briefly described; on chapter 3, black-box machine learning models for the prediction of absolute permeability from MICP data are evaluated and compared to classical linear models; on chapter 4, linear models for the prediction of special core analysis properties, using multi-task and hierarchical models, are studied; limitations and desirable features of the evaluated methods are further discussed on chapter 4.2.

Chapter 2

Measuring Petrophysical Properties

The measurement of petrophysical properties is performed through specialized laboratory experiments, which in the petroleum industry are classified as routine or special core analysis (SCAL), depending on the experiment complexity and on its investigated properties.

The analyzed rock samples may be extracted from well cores or from open borehole sidewall sampling tools, and can measure from one to several centimeters (Kennedy 2015). The rock samples are commonly extracted as cylindrical plugs, but can also have irregular shapes, case in which they are referred to as rock fragments.

2.1 Routine Core Analysis

Experimental measurements of a rock sample pore volume V_p , grain volume V_g , porosity ϕ and absolute permeability k_{abs} are commonly referred to as routine core analysis.

Porous media can be pictorially described as being constituted of grain and pore spaces (Figure 2.1). The percentage of the total space occupied by pores is known as rock porosity. V_p and ϕ measurements are important for the evaluation of hydrocarbon reserve volumes.

$$\phi = \frac{V_p}{V_p + V_g} \quad (2.1)$$

At reservoir conditions, porous space is filled by either water, oil or gas, each of which in a separate phase. The percentage of the porous volume occupied by water, oil or gas is known as water, oil or gas saturation, respectively.

$$S_f = \frac{V_f}{V_p}, \quad \text{for } f \text{ in } w, o, g \quad (2.2)$$

Darcy's law (2.3) describes the linear relationship between instantaneous flow rate q and a pressure gradient $\frac{\partial p}{\partial x}$ in laminar single-phase flow (Tiab and Donaldson

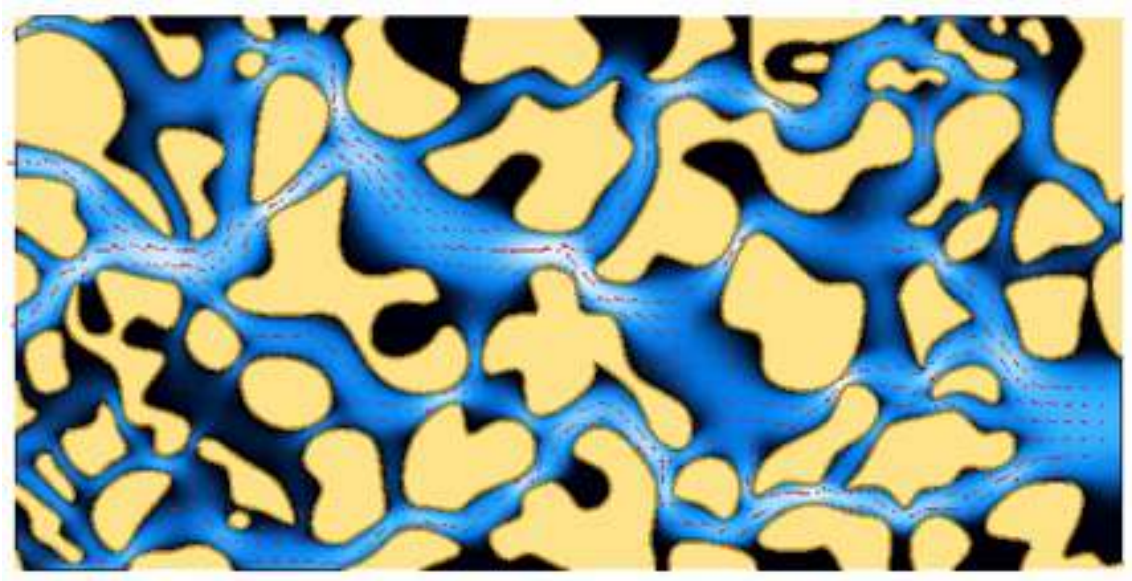


Figure 2.1: Pore media grain and pore spaces

2004). Given a stable steady-state single-phase flow, the porous media absolute permeability k_{abs} is directly proportional to the instantaneous flow rate q , fluid viscosity μ , and is inversely proportional to the pressure gradient $\frac{\partial p}{\partial x}$ and cross-sectional area A .

$$q = k_{abs} \frac{A}{\mu} \frac{\partial p}{\partial x} \quad (2.3)$$

2.2 Special Core Analysis

Measurement of multiphase properties of porous media, on which both viscous and capillary forces may be relevant, is usually more complex and incurs in longer and costlier experiments when compared to single-phase routine core analysis measurements.

2.2.1 Mercury Injection Capillary Pressure

In multi-phase porous systems, due to different interaction characteristics between fluid and solid phases, the attraction between the molecules of one of the fluid phases and the solid surfaces may be greater than that of the other fluid phase. This physical behavior is known as wettability, and it is responsible for capillary forces and related phenomena.

For a given multiphase porous system, with a surface tension σ between fluid phases, wettability may be quantified by the contact angle θ formed by the fluid phase boundary-solid interface, as depicted in Figure 2.2.

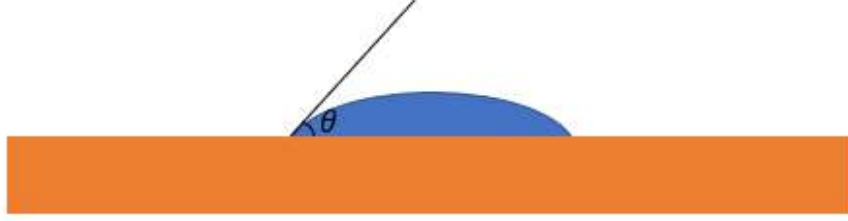


Figure 2.2: Contact Angle between fluid phase boundary and solid surface

The Young-Laplace equation describes the capillary pressure sustained across the interface between fluids at a pore throat constriction (Tiab and Donaldson 2004). A heterogeneous porous system is connected by pore throats of several different radii. For a system consisting of mercury and mercury vapor, with surface tension σ_{Hg-air} and contact angle θ_{Hg-air} with the rock surface, the relationship between capillary pressure and pore throat radius r may be described by equation (2.4). Capillary forces may strongly affect the distribution of fluid phases in the porous space.

$$P_c = \frac{2\sigma_{Hg-air} \cos \theta_{Hg-air}}{r} \quad (2.4)$$

A Mercury Injection Capillary Pressure (MICP) experiment consists of pre-specified and controlled mercury injection steps in a porous medium, previously subjected to a vacuum. It is usually performed on rock fragments with around one cubic centimeter, which allows it to be executed both in well core and sidewall sample fragments. Automated acquisition systems for MICP are commercially available (Micromeritics 2020)(Figure 2.3), making it a relatively fast and cheap experiment.

A fixed pressure stabilization time and a geometric progression sequence of increasing pressure steps is defined before each MICP experiment. At each pressure step, after the pre-defined pressure stabilization time, the volume of intruded mercury in the rock sample pore space is recorded. Given calibrated mercury-mercury-vapor surface tension σ_{hg-air} and contact angle θ_{hg-air} , each pressure step may be associated with a pore throat radius according to equation (2.4). The cumulative intruded mercury volume V_{hg} at each step, thus, corresponds to the porous volume accessible by pore throats with radius equal to or smaller than the relationship given by equation (2.4). Mercury saturation S_{hg} at each step may be calculated using equation (2.5).

$$S_{hg} = \frac{V_{hg}}{V_T} \quad (2.5)$$

The total fragment volume V_T is estimated during the experimental data interpretation procedure, using the difference of the total glass bulb volume and the intruded mercury volume necessary to “envelope” the rock fragment and start intruding the



Figure 2.3: AutoPore IV Series Mercury Porosimeter (left) and glass penetrometer (right) used in automated MICP acquisitions.

porous space. This happens at a pressure known as entrance capillary pressure P_e . The determination of the exact value of this entrance capillary pressure may introduce uncertainty on experimental results, particularly for very irregular shaped or vuggy rock samples.

The recorded capillary pressure P_c and mercury saturation S_{hg} values form a capillary pressure curve, whose format is directly related to the distribution of accessible pore volume in the porous medium of the rock sample. Figure 2.4 shows several capillary pressure curves acquired in MICP experiments.

A widely used transformation of capillary pressure curves is described by equation (2.6) (Lenormand 2003), which constructs a probability distribution of accessible pore volume as a function of the logarithm of the associated pore throat radius r . Figure 2.5 shows pore volume probability distribution curves $p(\log r)$, for the same MICP experiments depicted in Figure 2.4.

$$p(\log r) = \frac{dS_{hg}}{d \log P_c} \quad (2.6)$$

Several authors have proposed correlations between capillary pressure curve features and absolute permeability (Kolodzie 1980; Pittman 1992; Purcell 1949; Swanson 1981), among others. These correlations will be further explored on chapter 3.

In this work, a dataset of 2324 MICP curves acquired on rock fragments of sandstone and carbonate reservoirs from several Brazilian reservoirs was compiled. Routine and special core analysis results, performed on core sample plugs and sidewall samples extracted from the same well depths as the rock fragments were incorporated in the dataset.

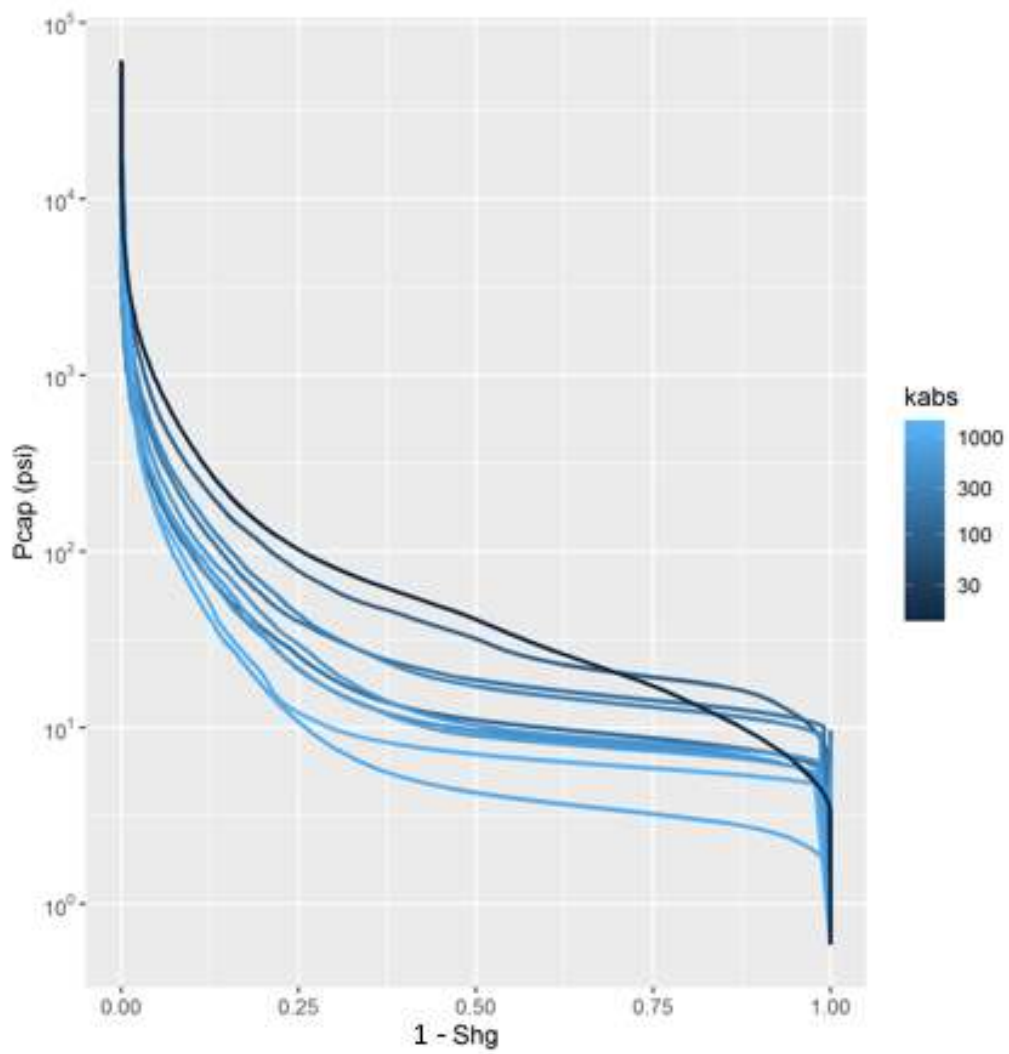


Figure 2.4: Mercury Injection Capillary Pressure Curves for several rock samples, colored by absolute permeability.

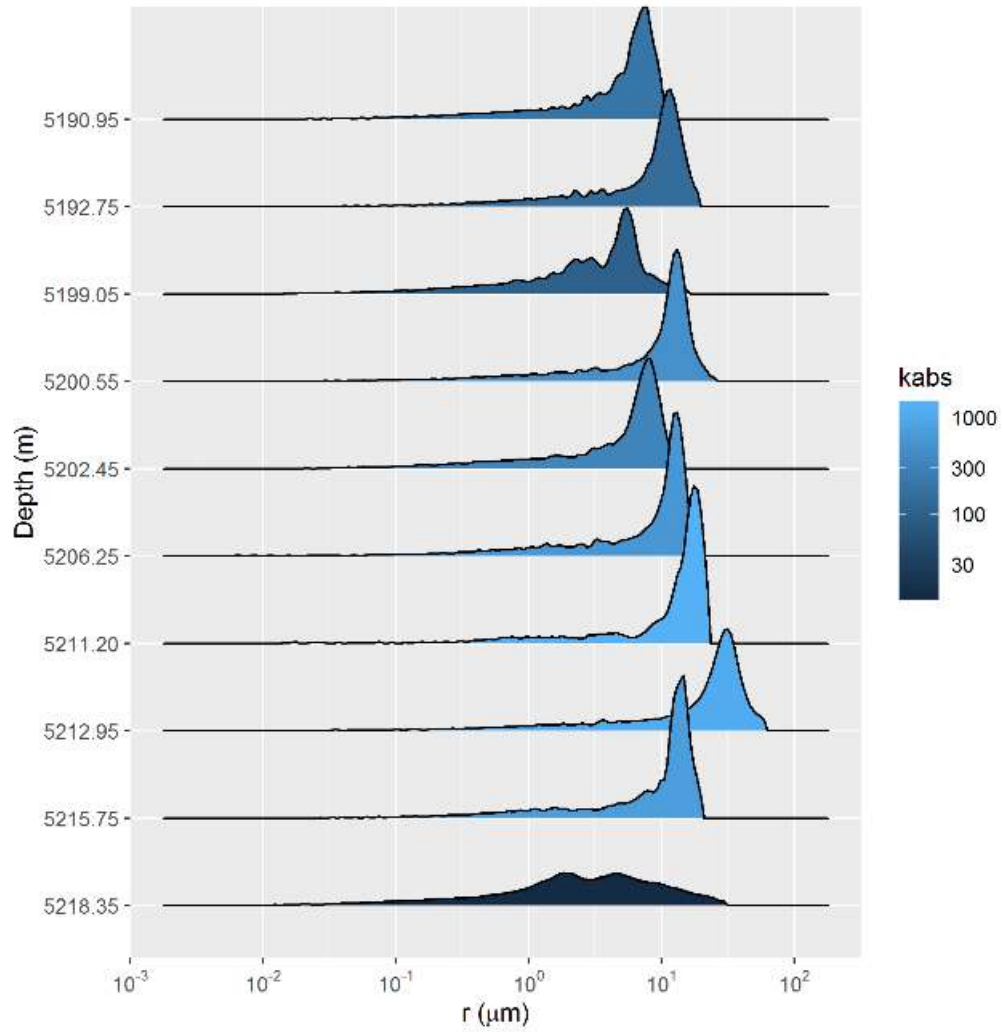


Figure 2.5: Accessible pore volume probability distribution curves $p(\log r)$, for the same MICP experiments depicted in Figure 2.4.

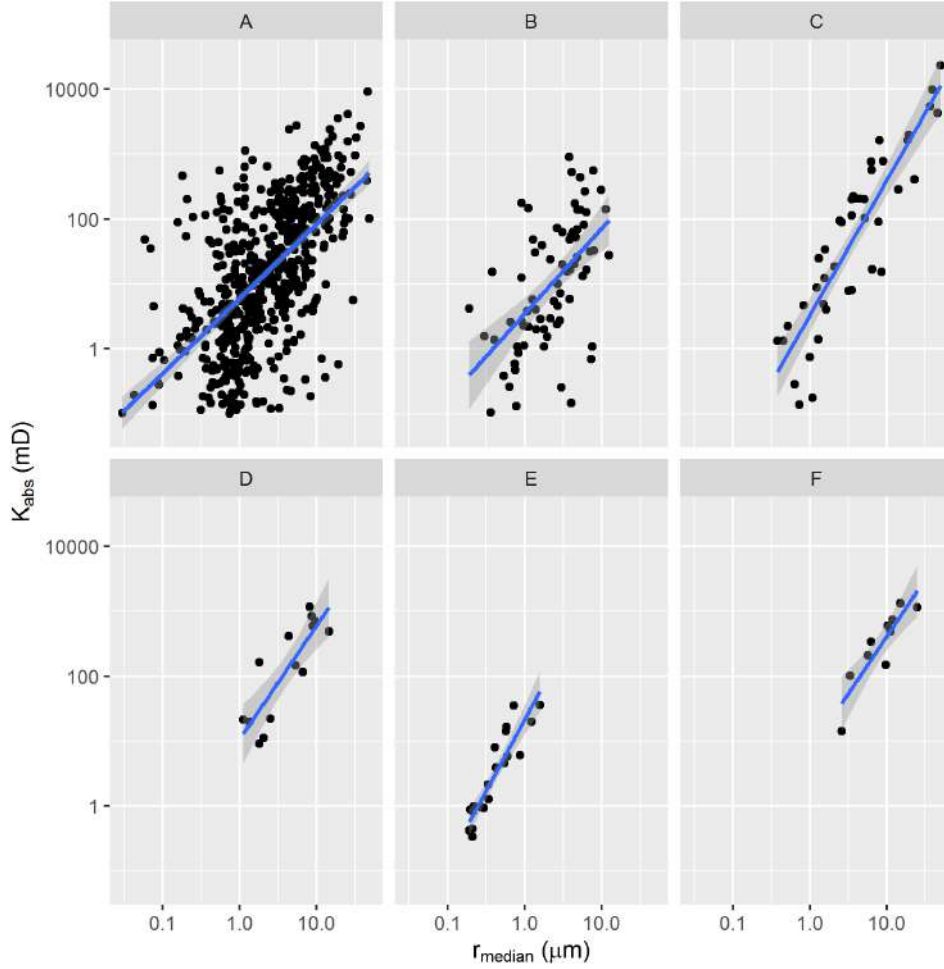


Figure 2.6: Relationship between median pore throat radius and absolute permeability of several different reservoirs, from most to least heterogeneous

Given that the size of the rock fragments utilized for MICP measurements were in many cases many times smaller than the size of the corresponding core sample plugs and sidewall samples, where routine and special core analysis were performed, it is very likely that, for many samples, the rock fragments may not constitute a representative volume of the porous space (Blunt 2017). Indeed, as illustrated in Figure 2.6, it is possible to assess that heterogeneous reservoirs display many more outlier samples on correlations between MICP and core analysis data, when compared to more homogeneous reservoirs.

2.2.2 Centrifuge Capillary Pressure

The centrifuge method is a standard technique for capillary pressure curve estimation of porous media. The method consists of imposing a capillary pressure profile on a rotating rock sample through centrifugal acceleration imposition in increasing rotation speed steps, and recording the average sample saturation at each step (Tiab

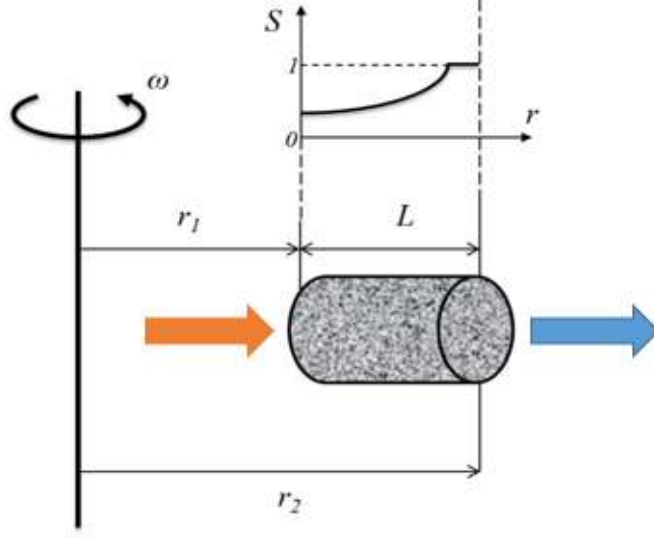


Figure 2.7: Drainage Oil-Water Centrifuge capillary pressure experiment geometry.

and Donaldson 2004). The fundamental equations that describe the method were developed by Hassler and Brunner (Hassler and Brunner 1945):

$$P_c = \frac{1}{2}\omega^2\Delta\rho(r^2 - r_2^2) \quad (2.7)$$

$$B = 1 - \left(\frac{r_1}{r_2}\right)^2 \quad (2.8)$$

$$S_w = \frac{V_w}{V_p} = \frac{V_w}{V_w + V_o} \quad (2.9)$$

$$\overline{S_w}(P_c) = \frac{(1 + \sqrt{1 - B})}{2} \int_0^1 \frac{S_w(xP_c)}{\sqrt{1 - Bx}} dx \quad (2.10)$$

where ω is the centrifuge rotation speed, $\Delta\rho$ is the difference in density between the displacing and displaced fluids, L is the rock sample length, r is the radius to the centrifuge rotation center, r_1 and r_2 are the radial distances to the sample extremities, V_p , V_w and V_o are the sample pore, water and oil volumes, $S_w(P_c)$ is the water saturation at a given point in the sample subjected to a P_c capillary pressure, and $\overline{S_w}$ is the sample average water saturation.

Several methods have been proposed for the estimation of capillary pressure curves from centrifuge experiments. These methods may be classified as direct or inverse according to how they solve the centrifuge saturation equation. Direct methods use several proposed differential and integral approximations of the saturation equation to directly estimate capillary-pressure curves from centrifuge experiment measurements at discrete capillary-pressure steps (Forbes 1994; Hassler and Brunner 1945; Skuse, Flroozabadl, and Ramey Jr. 1992). Inverse methods parameterize capillary-pressure curves and solve the saturation equation using non-linear regres-

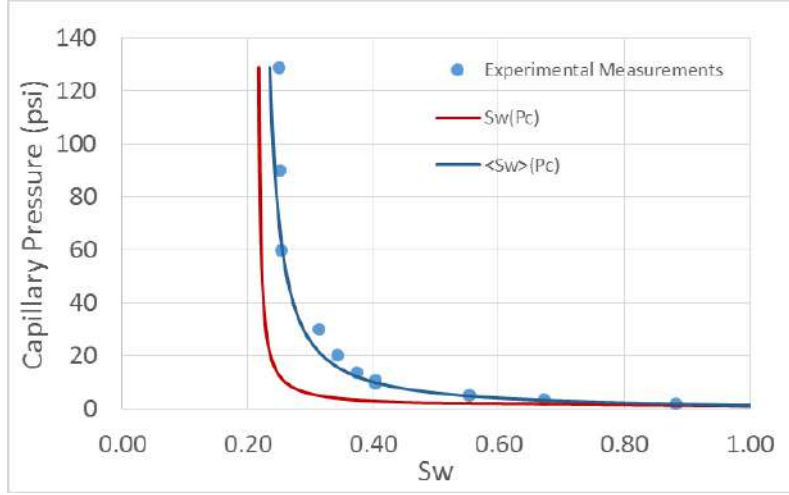


Figure 2.8: Estimation of capillary pressure curve from centrifuge measurements.

sion (Bentsen 1977) or linear regression with spline basis functions (Nordtvedt and Kolltvelt 1991). In this work, a dataset of 135 capillary pressure curves estimated from centrifuge experiments was compiled. Using the parameterization (2.11), proposed in (Albuquerque et al. 2018), each of the capillary pressure curves was fitted using a least-squares minimization procedure, with parameters S_{wi} , P_e , α and β .

$$S_w(P_c, S_{wi}, \alpha, \beta, P_w) = \frac{1 + \alpha S_{wi}(P_c - P_e)^\beta}{1 + \alpha(P_c - P_e)^\beta} \quad (2.11)$$

This parameterization is such that for high values of capillary pressure, water saturation S_w approaches S_{wi} , a parameter thus named irreducible water saturation. In a drainage capillary pressure model, the smallest pressure at which the saturation of the sample can be reduced from one hundred per cent is termed “entrance” capillary pressure (Peters 2012), modeled by parameter P_e in equation (2.11). The remainder parameters α and β are associated with the capillary pressure curve steepness and shape.

$$\lim_{P_c \rightarrow +\infty} S_w = S_{wi} \quad (2.12)$$

$$P_c(S_w = 1) = P_e \quad (2.13)$$

Considering a transformation analogous to the one utilized for mercury injection capillary pressure curves (2.6), pore volume probability distribution curves $p(\log r)$ may be estimated from the fitted centrifuge capillary pressure curve parameters, as described by equation (2.14).

$$p(\log r) = \frac{dS_w}{d \log P_c} = \frac{\alpha \beta S_{wi}(S_w - S_{wi})(P_c - P_e)^{\beta-1}}{1 + \alpha(P_c - P_e)^\beta} \quad (2.14)$$

As graphically displayed in Figure 2.10, the parameter α can be shown to be

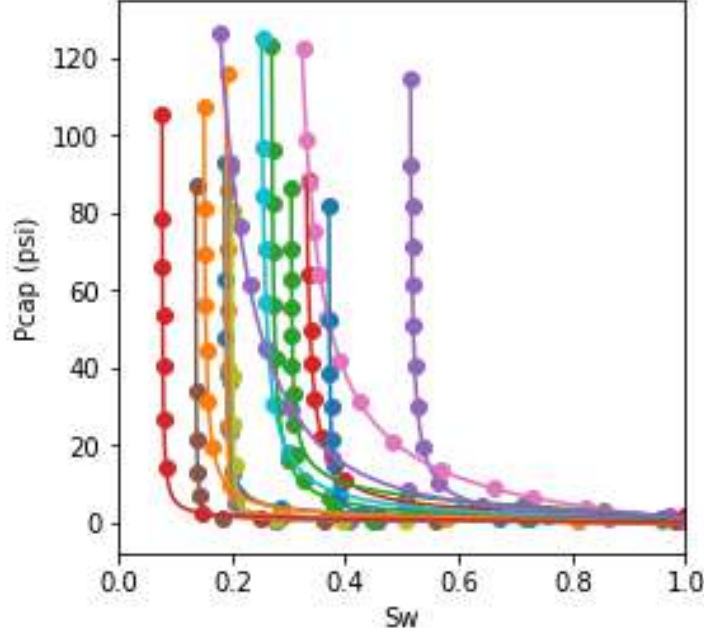


Figure 2.9: Sample of the experimental capillary pressure curve dataset (points) and fitted capillary pressure model (lines).

associated with the location of the distribution $p(\log r)$ and the parameter β can be shown to be associated with its scale or dispersion.

A dataset combining the fitted centrifuge capillary pressure curve parameters and several MICP features, detailed on chapter 3, of rock fragments extracted from corresponding core locations was assembled. As exemplified in Figure 2.11, associated pore volume distributions of corresponding centrifuge capillary pressure and MICP curves often display similar distributions, though to varying degrees as MICP curves may have been measured in rock fragments which may or may not be representative of its associated core sample pore volume. MICP experiments also cover a larger range of pore throat radius sizes and display significantly more detail, as a much tighter pressure and corresponding pore throat radius experimental grid is sampled. In Figure 2.12, it is possible to identify significant correlations between corresponding centrifuge capillary pressure curves α , S_{wi} and MICP median pore throat radius r_{median} parameters. Both α and S_{wi} parameters can be seen, thus, as correlated with the location of the distribution $p(\log r)$.

2.2.3 Relative Permeability

In porous media multiphase flow, relative permeability describes the linear proportion by which each fluid flow is penalized when compared to Darcy's law (2.3). For a water-oil two-phase flow, the following relations describe oil (k_{ro}) and water (k_{rw})

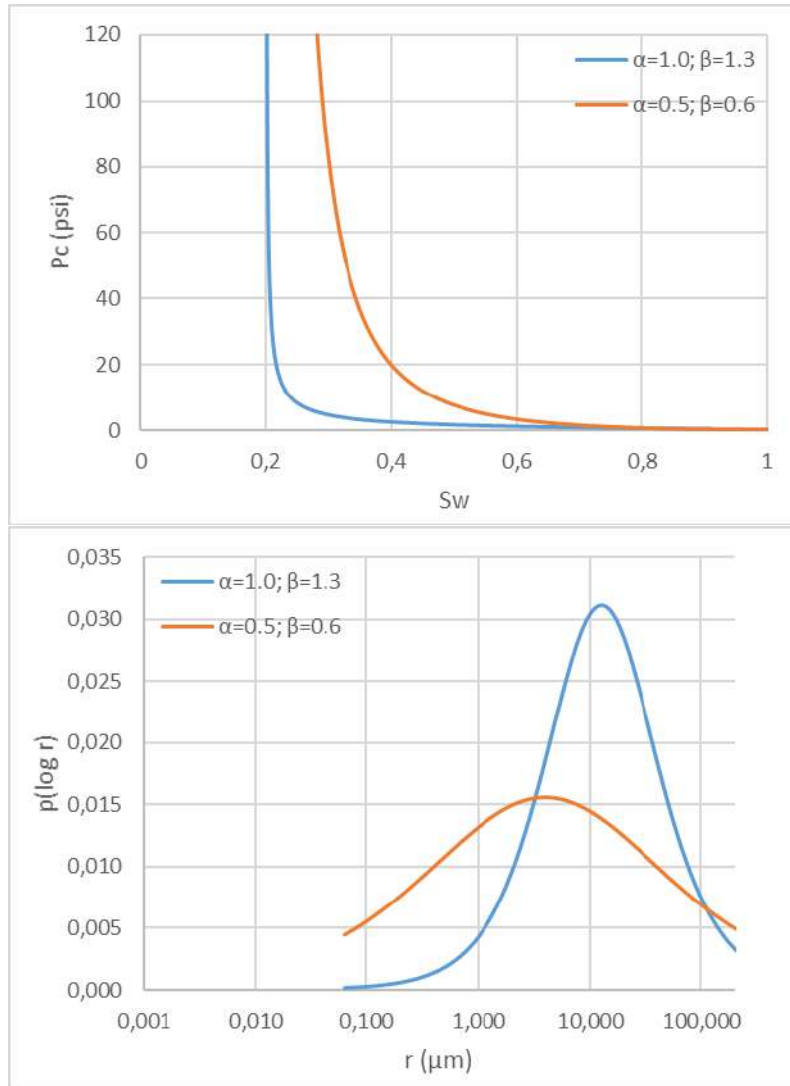


Figure 2.10: Sample centrifuge capillary pressure curves, parameterized by equation (2.11), and corresponding pore volume probability distribution curves $p(\log r)$.

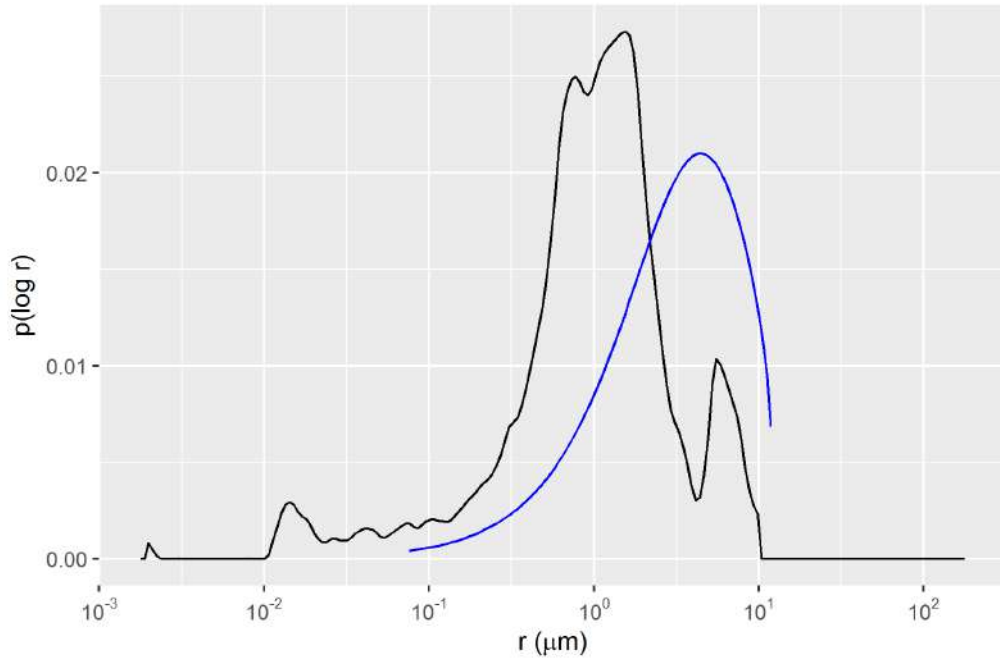


Figure 2.11: Associated pore volume distributions $p(\log r)$ of correspondent Centrifuge Capillary Pressure (blue curve) and MICP (black curve) samples.

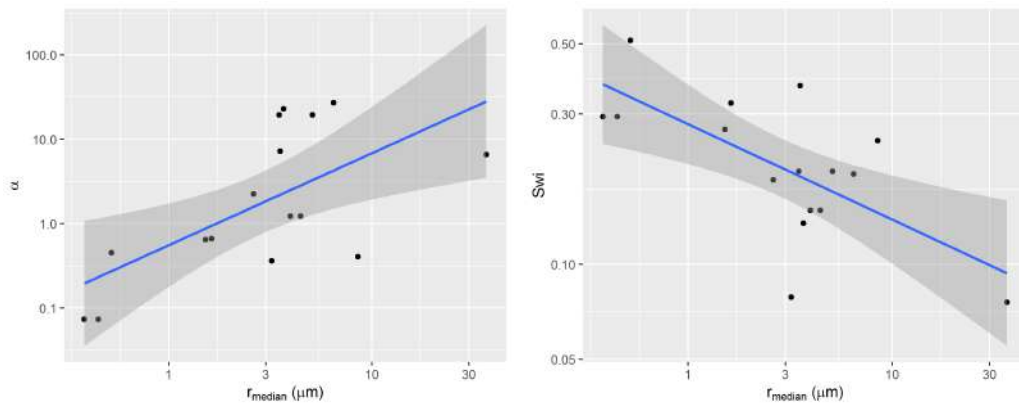


Figure 2.12: Correlations between parameters α , Sw_i and median pore throat radius of correspondent rock fragment $p(\log r)$ distribution.

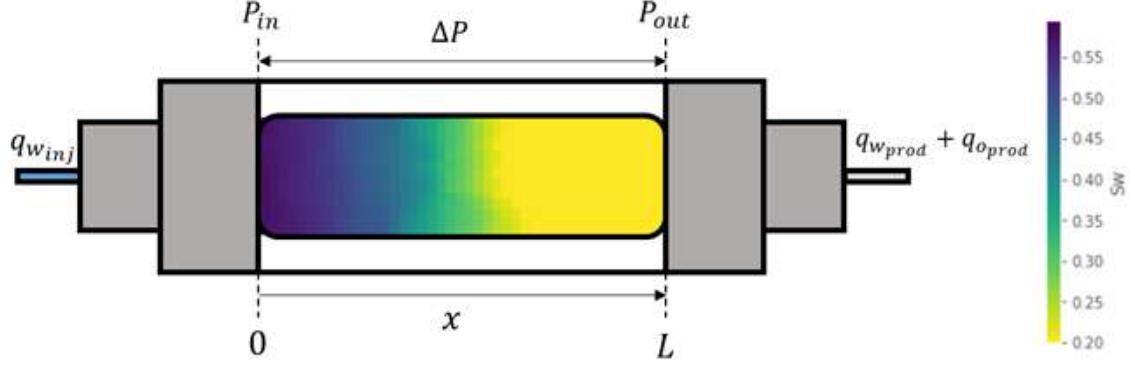


Figure 2.13: Unsteady-state relative permeability experimental setup.

relative permeability, and the their associated fractionary water flow (f_w).

$$q_o(S_w) = k_{ro}(S_w) \frac{k_{abs} A}{\mu_o L} \Delta p \quad (2.15)$$

$$q_w(S_w) = k_{rw}(S_w) \frac{k_{abs} A}{\mu_w L} \Delta p \quad (2.16)$$

$$f_w = \frac{\frac{k_{rw}}{\mu_w}}{\frac{k_{rw}}{\mu_w} + \frac{k_{ro}}{\mu_o}} \quad (2.17)$$

The most widely used experimental setup for water-oil relative permeability determination is the unsteady-state water-oil relative permeability experiment. In a unsteady-state water-oil relative permeability experiment, a sample initially saturated at irreducible water saturation S_{wi} condition is subjected to a constant rate water injection, simulating reservoir secondary recovery water injection.

During the experiment, water injection flow rate $q_{w_{inj}}$ and exit pressure P_{out} are kept constant, and cumulative oil production N_p and pressure differential Δp are recorded.

$$q_{w_{inj}} = cte \quad (2.18)$$

$$P_{out} = cte \quad (2.19)$$

Unsteady state water-oil relative permeability experiments are usually interpreted using either the JBN method (Johnson, Bossler, and Naumann 1959) or using history matched numerical simulated solutions (Lenormand and Lenormand 2016). Figure 2.14 displays an example of recorded and history matched cumulative oil production N_p and pressure differential Δp , obtained using maximum likelihood estimation of the relative permeability parameters that best fit the recorded experimental data. History matching is usually performed using parametric relative permeability models and finite difference partial differential equations simulation. The Corey model (Corey 1954), given by equations (2.20)(2.21)(2.22), is a widely used power-law

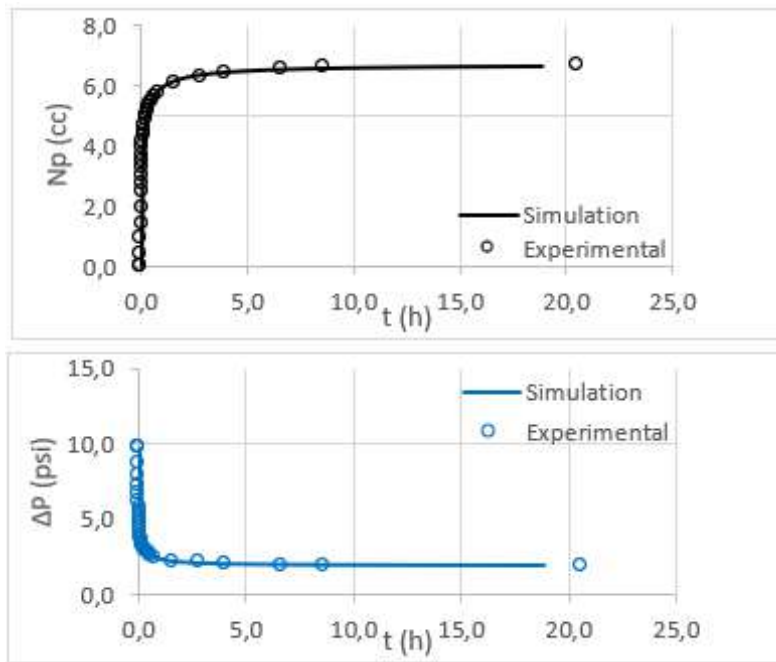


Figure 2.14: Example of history matched experimental data.

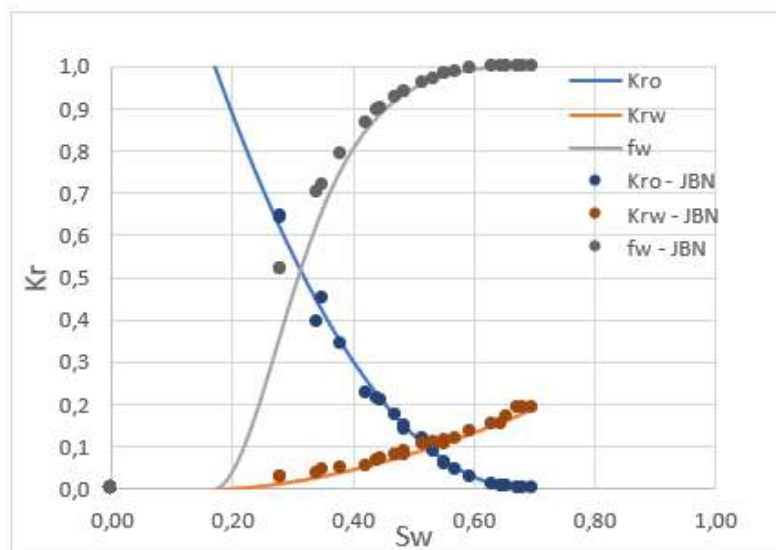


Figure 2.15: Example of history matched relative permeability and fractionary flow curves compared to analytic JBN interpreted experimental results.

model for relative permeability curves, whose parameters n_o and n_w are commonly referred to as Corey exponents.

$$S_{wD} = \frac{S_w - S_{wi}}{1 - S_{wi} - S_{or}} \quad (2.20)$$

$$k_{ro}(S_w) = k_{ro@S_{wi}}(1 - S_{wD})^{n_o} \quad (2.21)$$

$$k_{rw}(S_w) = k_{rw@S_{or}}(S_{wD})^{n_w} \quad (2.22)$$

The LET model (Lomeland, Ebeltoft, and Thomas 2005), given by equations (2.23)(2.24), was developed with six shape parameters L_o , E_o , T_o , L_w , E_w , T_w , and can accommodate much more flexible relative permeability curves. It is commonly used in reservoir simulation models, especially for history matching procedures that demand flexible representations of relative permeability models.

$$k_{ro}(S_w) = k_{ro@S_{wi}} \frac{(1 - S_{wD})^{L_o}}{(1 - S_{wD})^{L_o} + E_o(S_{wD})^{T_o}} \quad (2.23)$$

$$k_{rw}(S_w) = k_{rw@S_{or}} \frac{(S_{wD})^{L_w}}{(S_{wD})^{L_w} + E_w(1 - S_{wD})^{T_w}} \quad (2.24)$$

In this work, a dataset of 226 unsteady-state water-oil relative permeability curves was assembled. LET parameters for each of the 226 curves were fit using maximum likelihood estimation (DeGroot and Schervish 2012). Multivariate and multi-task linear regression models built using this dataset are described on chapter 4.

Chapter 3

Absolute Permeability Regression

Several authors have studied the use of MICP curves to estimate absolute permeability. Many of the proposed regression models use linear correlations between the logarithm of absolute permeability $\log k_{abs}$ and many different proposed capillary pressure curve features (Kolodzie 1980; Pittman 1992; Purcell 1949; Swanson 1981).

Comparison of linear regression and machine learning models for the estimation of absolute permeability using MICP data has been performed on datasets from middle-east reservoirs by several authors (Al Khalifah, Glover, and Lorinczi 2020; Nooruddin, Anifowose, and Abdulraheem 2013), showing promising results for non-linear machine learning models.

In this work, a comparison of linear regression and machine learning models for the estimation of absolute permeability using MICP data is performed on a dataset of 2324 MICP curves acquired on rock fragments of sandstone and carbonate reservoirs from several Brazilian reservoirs.

3.1 Mercury Porosimetry Feature Engineering

Several feature engineering and regression techniques were evaluated on the assembled MICP dataset, exploring both linear correlation features as well as non-linear transformations. The following sections describe each of the feature extraction procedures and the choice and training of the selected regression models.

3.1.1 Statistical Features

For each capillary pressure curve and associated pore volume probability distribution $p(\log r)$, mean pore throat radius r_{mean} and median pore throat radius r_{median} were calculated. Quantiles ranging from the 15th to the 90th pore throat radius values were estimated $\{r_{15}, r_{20}, \dots, r_{85}, r_{90}\}$. As a proxy for distribution heterogeneity, the interquartile-range iqr , given by the difference of pore-throat radius associated with

the 25th and 75th quartiles of the $p(\log r)$ distribution was estimated.

$$iqr = r_{25} - r_{75} \quad (3.1)$$

3.1.2 Linear Features

The following features, whose main authors are listed on Table 3.1 as feature names, were calculated for each one of the datasets MICP curves. Each of these features is associated with a linear equation, proposed by its authors as regression models for $\log k_{abs}$.

Table 3.1: Classical MICP linear correlation features

Feature Name	Feature	Linear Equation
Swanson (Swanson 1981)	$(\frac{S_{hg}\phi}{P_c})_{max}$	$\log k_{abs} = a + b.(\frac{S_{hg}\phi}{P_c})_{max}$
Purcell (Purcell 1949)	$\int_0^1 \frac{dS_{hg}}{P_c^2}$	$\log k_{abs} = a + b. \int_0^1 \frac{dS_{hg}}{P_c^2}$
Winland (Kolodzie 1980)	r_{35}	$\log k_{abs} = a + b.r_{35}$
Pittman (Pittman 1992)	r_{apex}	$\log k_{abs} = a + b.r_{apex}$
Dastidar (Dastidar 2007)	$r_{WGM} = [\prod_{i=1}^n r_i^{w_i}]^{1/\sum w_i}$	$\log k_{abs} = a + b.\phi + c.r_{WGM}$

3.1.3 Pore Throat Size Class Distribution

Pore throat size ternary class distributions were calculated for each $p(\log r)$ distribution, assigning classes micro, meso and macro to the percentage of the pore volume associated with pore throat sizes bellow $0.5 \mu m$, between $0.5 \mu m$ and $2.5 \mu m$, and above $2.5 \mu m$, respectively.

Figure 3.1 displays examples of ternary plots for micro, meso and macro pore volume percentage for $p(\log r)$ distributions of samples from reservoirs of two Brazilian fields A and B.

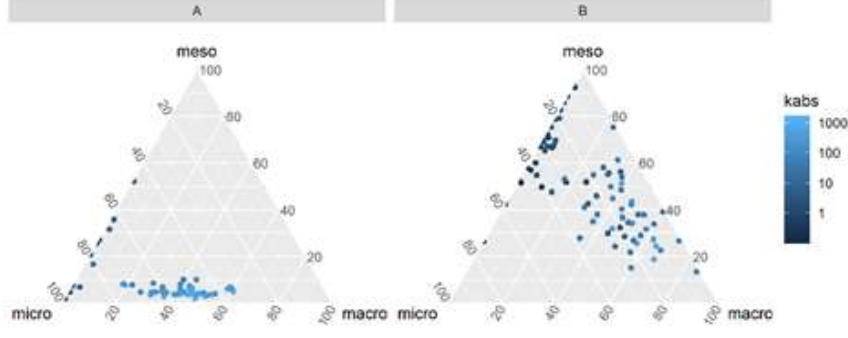


Figure 3.1: Pore-throat class distribution for two sample reservoirs (A and B).

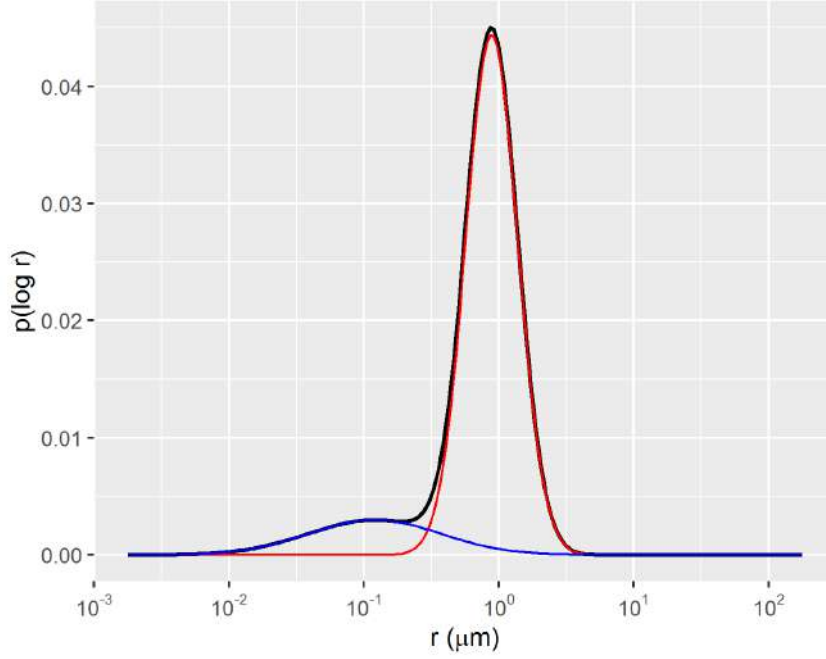


Figure 3.2: Example of bimodal gaussian mixture fit of $p(\log r)$ distribution.

3.1.4 Gaussian Mixture Fit Features

Following the methodology proposed by (Xu and Torres-Verdín 2013), bimodal gaussian mixture distributions, with parameters μ_1 , μ_2 , σ_1 , σ_2 , λ_1 and λ_2 , were fitted to each MICP associated pore volume distributions $p(\log r)$, using the following approximation.

$$p(\log r) \approx \lambda_1 \frac{1}{\sqrt{2\pi}\sigma_1} \exp \frac{-(\log r - \mu_1)}{2\sigma_1^2} + \lambda_2 \frac{1}{\sqrt{2\pi}\sigma_2} \exp \frac{-(\log r - \mu_2)}{2\sigma_2^2} \quad (3.2)$$

$$\lambda_1 + \lambda_2 = 1 \quad (3.3)$$

The parameters μ_1 , μ_2 , σ_1 , σ_2 , λ_1 and λ_2 , provide a compact representation of the MICP associated pore volume distributions $p(\log r)$, as exemplified in Figures 3.2 and 3.3.

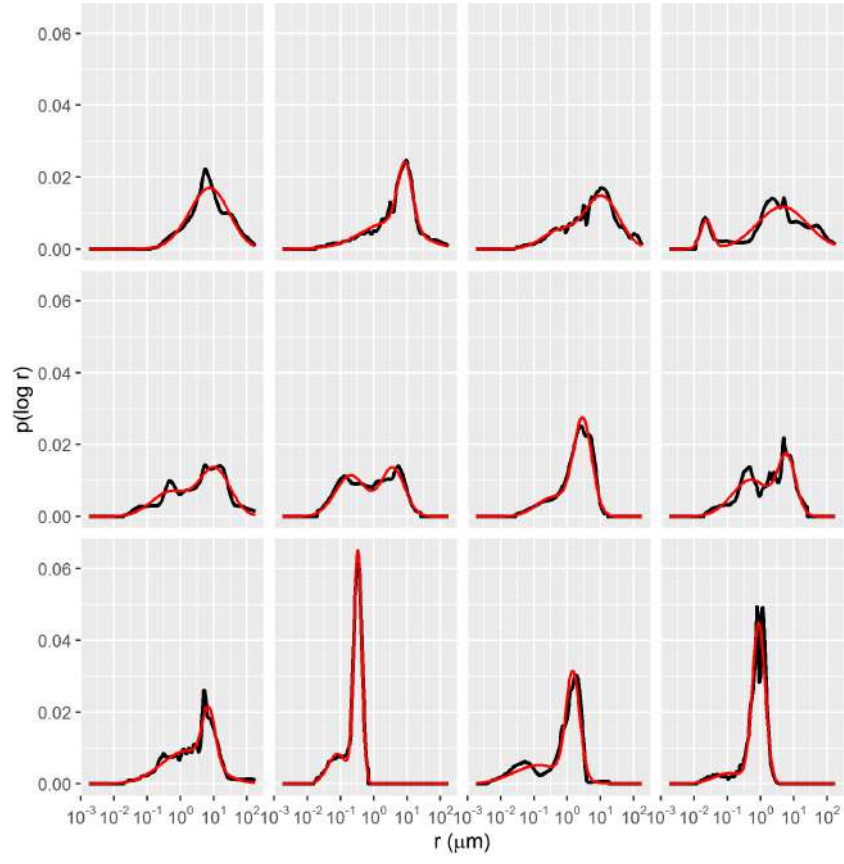


Figure 3.3: Gaussian mixture (red lines) fitted to $p(\log r)$ distributions of MICP data samples (black lines).

3.1.5 Dimensionality Reduction Features

Both linear and non-linear dimensionality reduction methods were used to extract features from the dataset of MICP curves. Dimensionality reduction techniques allow for the approximation of the dataset using a lower-dimensional representation, useful for data exploration and visualization and for data compression.

Using as input pore-throat radius quantiles $\{r_{15}, r_{20}, \dots, r_{85}, r_{90}\}$, the first two principal components PC1 and PC2 were estimated for each of the sample MICP curves using Principal Component Analysis (Hastie, Tibshirani, and Friedman 2009). The first two principal components represented a total of 96.1% of the variance in the dataset. A graphical depiction of the PC1 and PC2 features, colored by absolute permeability can be seen on Figure 3.4.

The first and second principal components were shown to be correlated with absolute permeability k_{abs} and interquartile range iqr , respectively, as displayed on Figures 3.5 and 3.6.

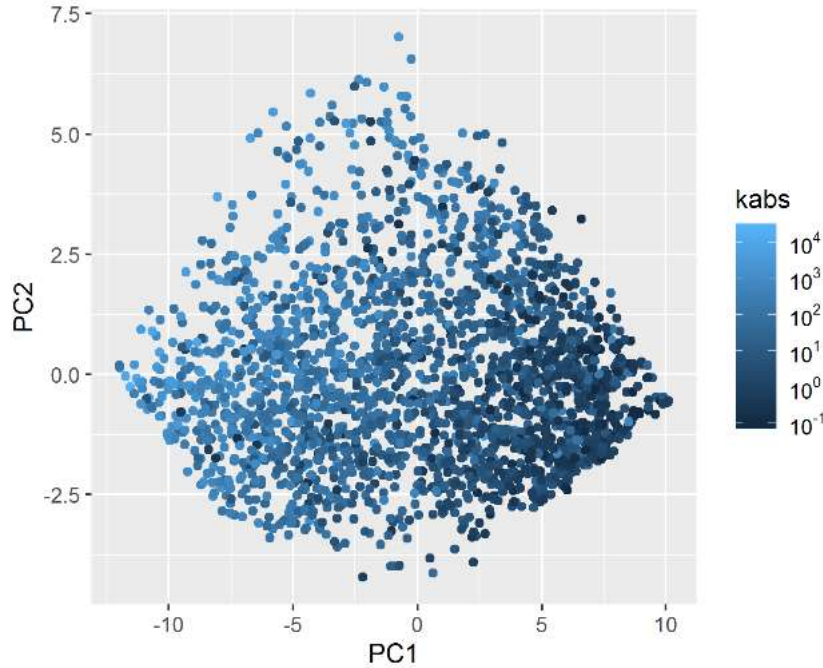


Figure 3.4: First two principal components of each of the datasets MICP samples.

Non-linear dimensionality reduction features were also estimated using Uniform Manifold Approximation and Projection, also known as UMAP (McInnes and Healy 2018), using as input pore-throat radius quantiles $\{r_{15}, r_{20}, \dots, r_{85}, r_{90}\}$.

The Uniform Manifold Approximation and Projection technique constructs a high dimensional graph representation of the data and optimizes a low-dimensional manifold to be as structurally similar to the high dimensional graph as possible (McInnes and Healy 2018). Due to its distance preserving properties, the low-

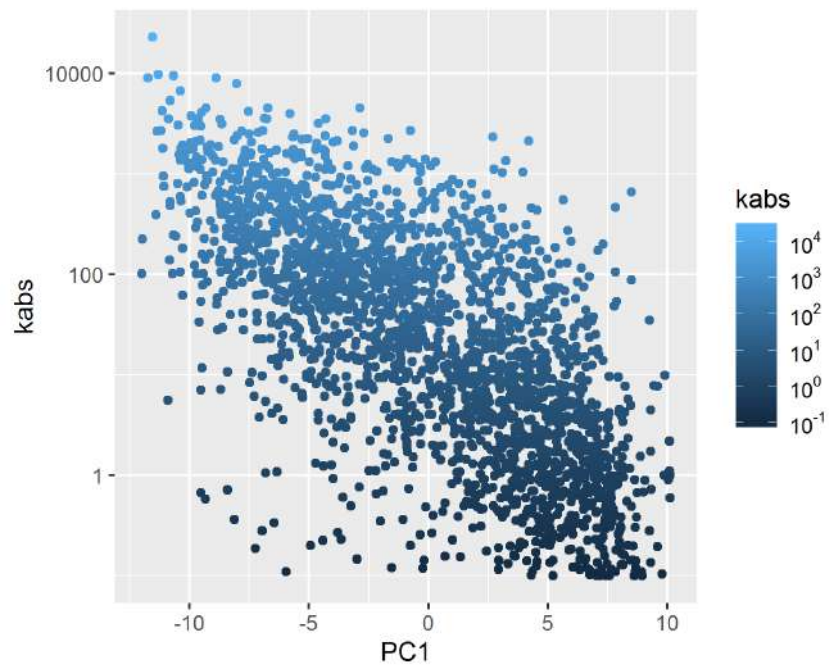


Figure 3.5: Negative correlation between absolute permeability and the first principal component.

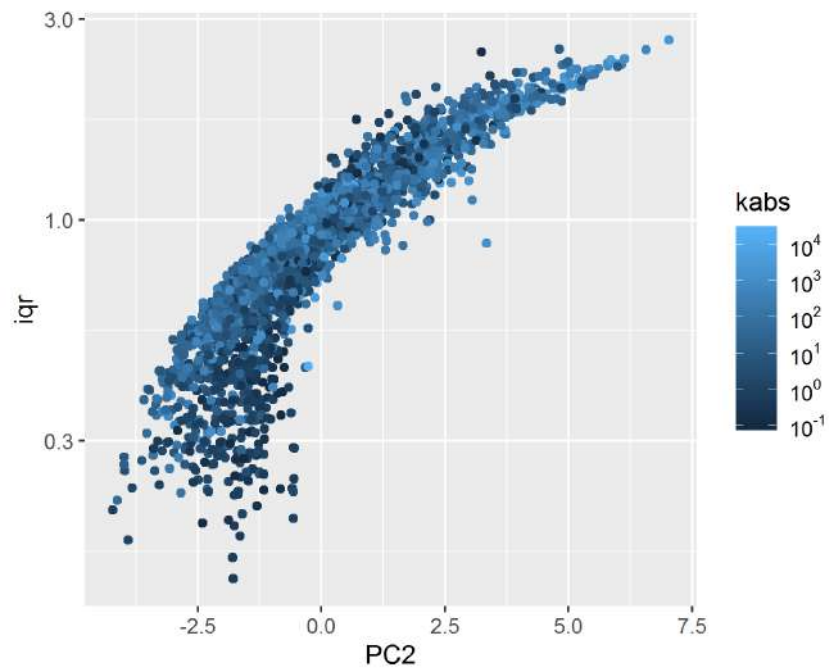


Figure 3.6: Positive correlation between the interquartile-range iqr and the second principal component PC2.

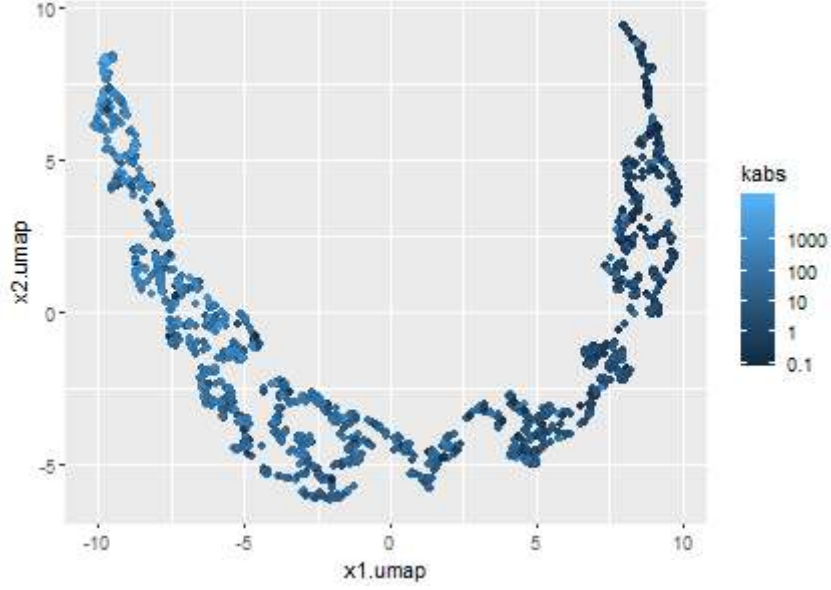


Figure 3.7: MICP curves non-linear decomposition using UMAP.

dimensional manifold, represented in Figure 3.7 by variables $x_1.umap$ and $x_2.umap$, is useful for visualization and identification of similar MICP curve samples.

In Figure 3.8, this property is exemplified by a sequence of three pairs of plots, showing that samples mapped to the same regions of the UMAP manifold, marked in the red in each of the plot pairs, correspond to distributions $p(\log r)$ with similar characteristics.

3.2 Absolute Permeability Regression Models

The regression metrics root mean squared error (3.4), median absolute error (3.5) and R^2 (3.6) were evaluated for the trained regression models. Due to the issues regarding the representativeness of the rock fragments with respect to the measured absolute permeability values, discussed on chapter 2, minimization of the median absolute deviation metric (3.5) was chosen to increase the robustness to outliers of the trained non-linear regression models, described in section 3.2.2. In both linear and non-linear regression methods, the R^2 metric (3.6) was used for further assessment of the trained regression models.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\log \hat{k}_{abs_i} - \log k_{abs_i})^2}{n}} \quad (3.4)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |\log \hat{k}_{abs_i} - \log k_{abs_i}|}{n} \quad (3.5)$$

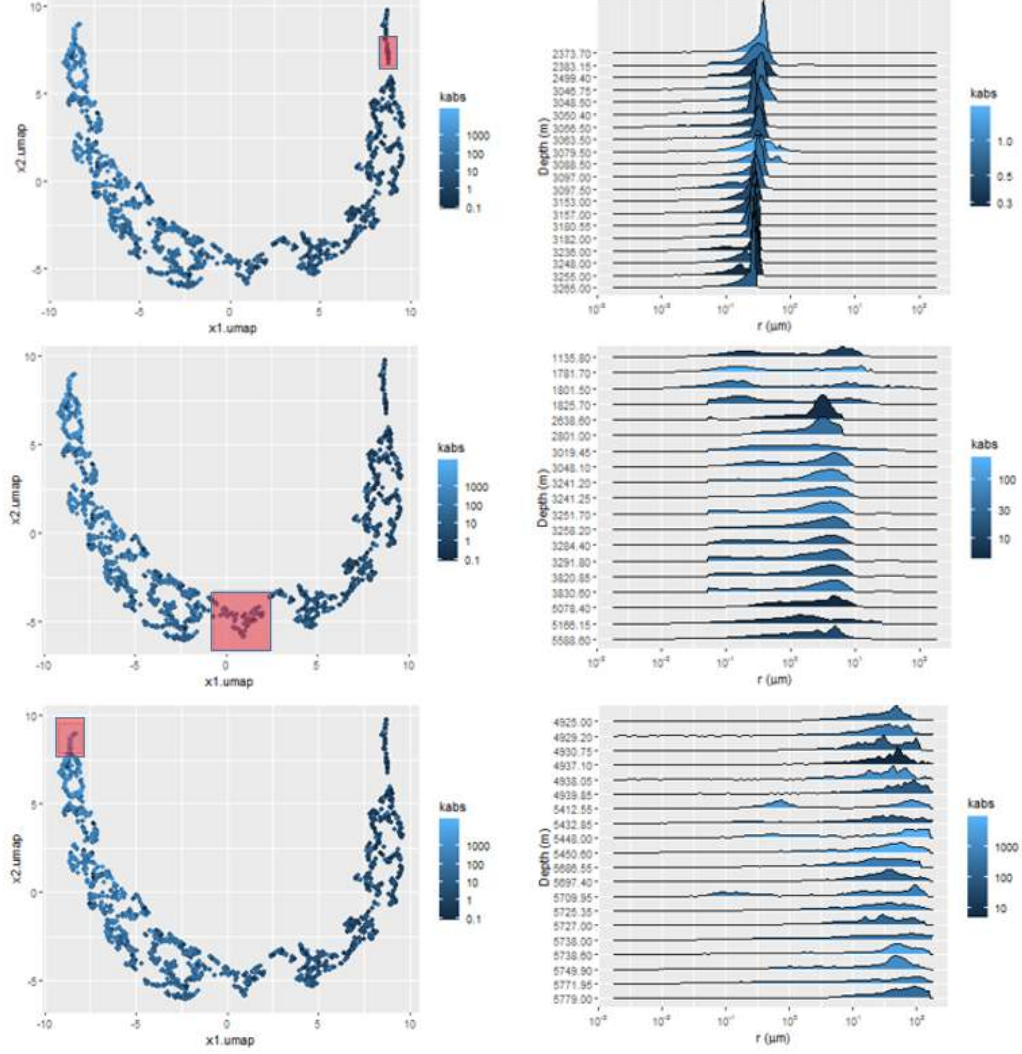


Figure 3.8: Manifold maps (left) showing areas with similar associated pore volume distributions $p(\log r)$ (right).

$$R^2 = 1 - \frac{\sum_{i=1}^n (\log \hat{k}_{abs_i} - \log k_{abs_i})^2}{\sum_{i=1}^n (\log k_{abs_i} - \overline{\log k_{abs_i}})^2} \quad (3.6)$$

3.2.1 Linear Regression Models

The linear features, described in section 3.1.2, were used to fit linear regression logarithmic absolute permeability $\log k_{abs}$ models.

Among the fitted linear regression models, the Winland and Swanson models obtained the lowest root mean squared error and highest coefficient of determination metrics.

Figure 3.9 shows a graphical representation of the fitted models, where a significant number of outlier logarithmic absolute permeability estimates $\log \hat{k}_{abs}$, with more than one order of magnitude errors, can be seen. The presence of this large

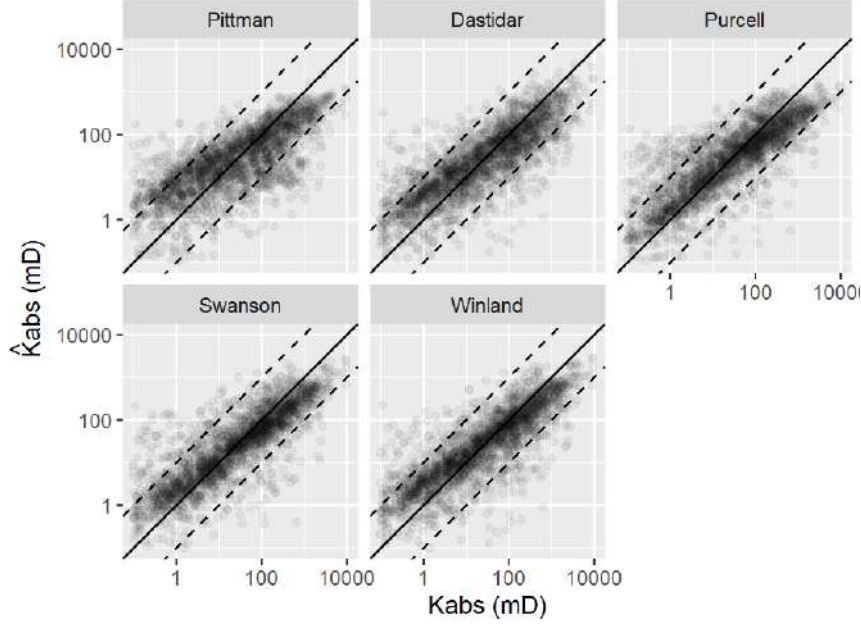


Figure 3.9: Visual evaluation of the predicted and observed absolute permeability models using linear features.

number of outliers can be explained by the limited rock fragment representativeness of the core sample pore volume for heterogeneous reservoirs, described in chapter 2.

3.2.2 Black-Box Machine Learning Models

Using the features described in section 3.1, six additional non-linear regression models were fit to the dataset using machine learning models. The 2324 MICP curves were split between training and test data, with a eighty to twenty percent ratio. Using the training set, each algorithm went through a hyper-parameter tuning procedure, using ten times repeated five-fold cross-validation error estimation, on a grid of selected hyper-parameters.

Two *knn* nearest neighbors models (P. Murphy 2012), were fit to the training set, one using features extracted using PCA (Hastie, Tibshirani, and Friedman 2009) and the other features extracted using UMAP non-linear dimensionality reduction (McInnes and Healy 2018).

A simple linear regression model was fit using the features extracted from the bi-modal gaussian mixture model proposed by (Xu and Torres-Verdín 2013), described in section 3.1.4.

A support vector regression (SVR) model, a random forest model and a gradient boosted trees model (Hastie, Tibshirani, and Friedman 2009) were fit to an expanded feature set, consisting of the features proposed by Swanson, Winland, Purcell, the sample porosity ϕ , the micropores and mesopores ternary class distributions, the

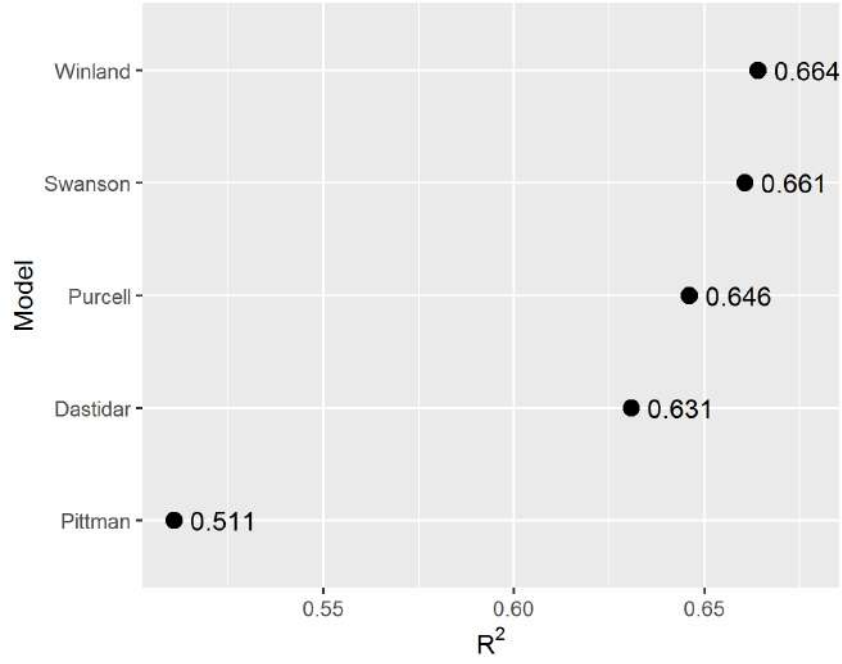


Figure 3.10: Comparison of the R^2 metric for the fitted linear regression models.

UMAP features and the bi-modal gaussian mixtures distribution features. A pre-processing standardizing step was applied to each feature. The gradient boosted trees model was fitted using the implementation provided by the XGBoost library (Chen and Guestrin 2016).

Table 3.2 and Figure 3.12 show median values of repeated five-fold cross-validated MAE, R^2 and RMSE metrics for each of the trained regression models. The regression models with the lowest median absolute error were the SVR, randomForest and gradient boosted trees XGB models. Through the use of a radial basis function kernel and an optimization procedure that minimizes hinge loss (Hastie, Tibshirani,

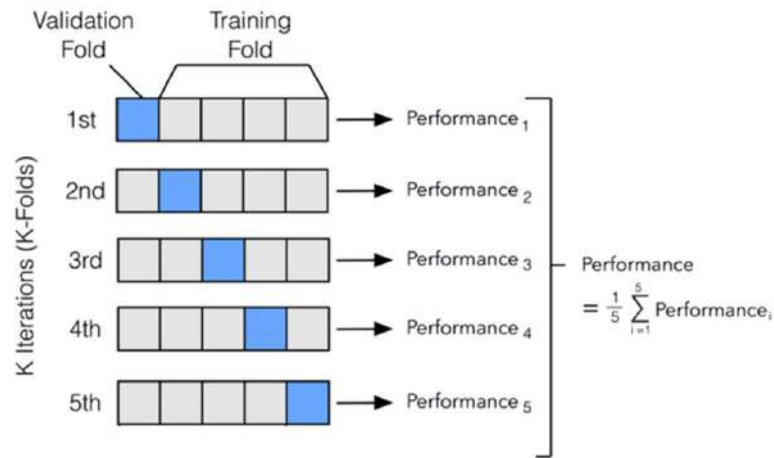


Figure 3.11: Five-fold cross-validation procedure.

and Friedman 2009), the SVR algorithm provides outlier robust predictions. The random forest and gradient boosted trees algorithms use randomized selection of features and data subsets, bagging and boosting techniques, respectively, to reduce prediction errors of ensembles of simpler decision tree models (Chen and Guestrin 2016).

Table 3.2: Median repeated five-fold cross-validated MAE, R^2 and RMSE metrics for each model.

Regression Model	MAE	R^2	RMSE
SVR	0.412	0.722	0.596
randomForest	0.437	0.712	0.604
XGB	0.439	0.711	0.605
Swanson	0.471	0.681	0.639
GaussianMixture	0.477	0.675	0.641
Winland	0.481	0.674	0.643
Purcell	0.487	0.661	0.654
Dastidar	0.520	0.642	0.673
knnUMAP	0.565	0.562	0.744
knnPCA	0.581	0.547	0.753
Pittman	0.613	0.523	0.775

Non-linear machine learning models presented consistently lower prediction errors, as can be seen on Figure 3.12. On Figure 3.13, a visual comparison between linear and non-linear machine learning models of the predicted versus measured data for the test dataset is shown. Although the non-linear machine-learning models did obtain better MAE, R^2 and RMSE metrics results, a significant number of outliers are still present, probably related to the issue of limited rock fragment representativeness of highly heterogeneous reservoirs described in chapter 2.

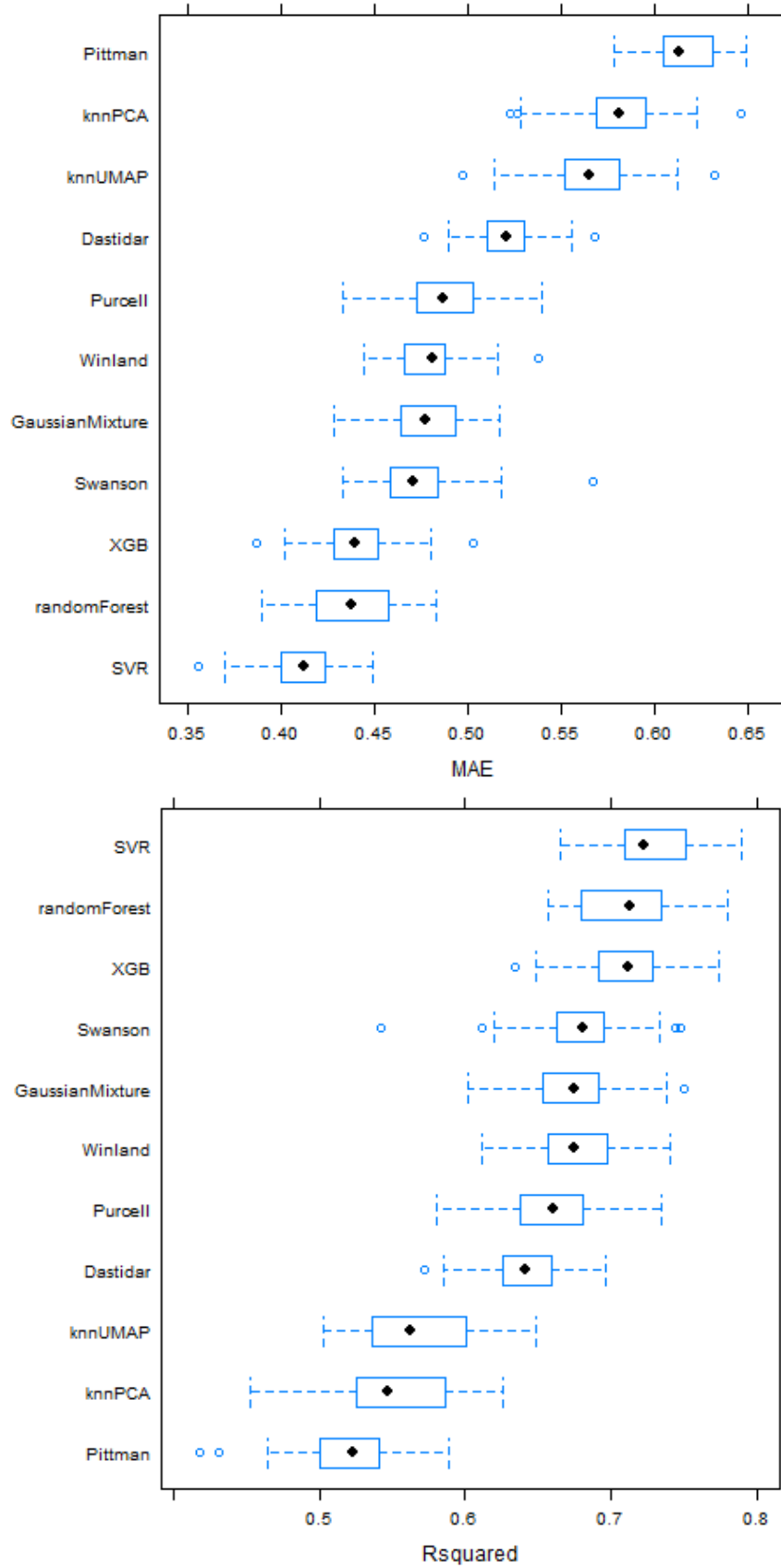


Figure 3.12: Boxplots for repeated five-fold cross-validated MAE, R^2 and RMSE metric results for each model.

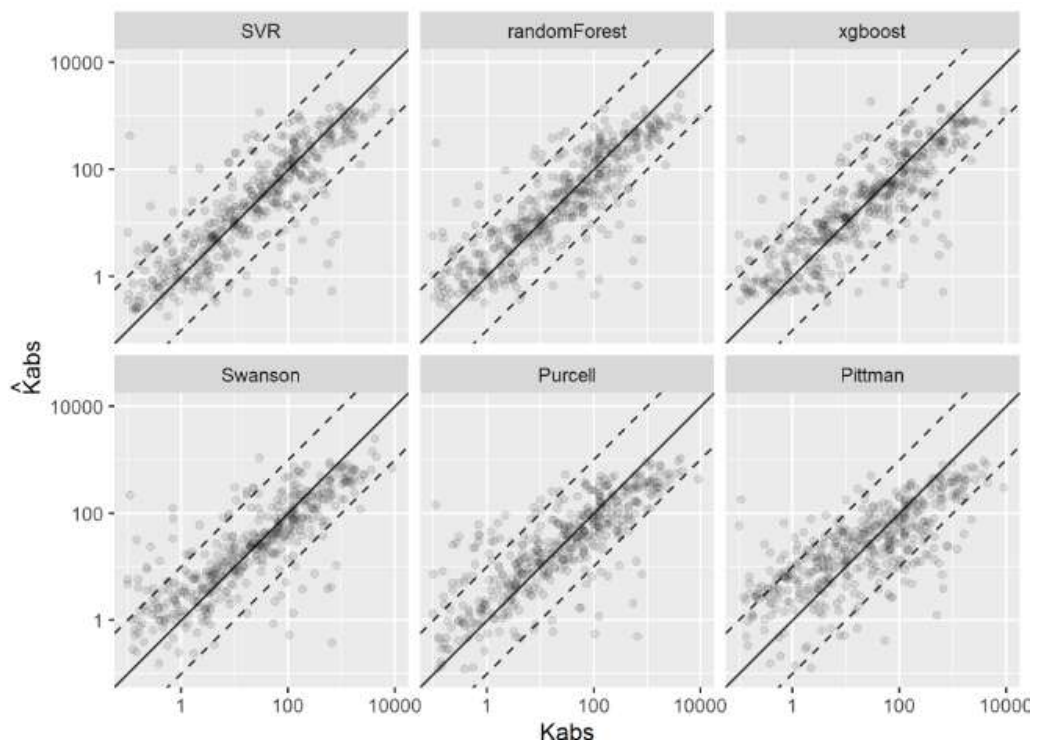


Figure 3.13: Visual evaluation of the estimated absolute permeability models using Machine Learning and linear models.

Chapter 4

Hierarchical Multi-Task Linear Regression

Linear regression methods are widely used in petrophysical characterization. Compared to non-linear machine learning models, they have as advantages simplicity, interpretability and their inherent linear behavior on extrapolated predictions. For the estimation of special core analysis properties, which are costly to acquire and thus usually scarce, simple models with a low number of parameters and that can deliver interpretable probabilistic inferences are highly desirable. In this work, two different linear regression techniques are evaluated on special core analysis datasets.

On a dataset containing 135 capillary pressure curves estimated from centrifuge experiments, a multi-task linear regression model is fit and evaluated, and some of its properties are analyzed.

On another dataset, containing 226 unsteady-state water-oil relative permeability curves, partially pooled hierarchical linear regression models are evaluated and compared to simple linear regression models.

4.1 Multi-Task Linear Regression

Using the parameterization (4.1), proposed by (Albuquerque et al. 2018), a dataset containing parameters S_{wi} , P_e , α and β fitted to each of the available 135 centrifuge capillary pressure curves, was assembled.

$$S_w(P_c, S_{wi}, P_e, \alpha, \beta) = \frac{1 + \alpha S_{wi}(P_c - P_e)^\beta}{1 + \alpha(P_c - P_e)^\beta} \quad (4.1)$$

To each of these samples, features extracted from MICP curves obtained on corresponding rock fragments were associated in a dataset containing $\{S_{wi}, P_e, \alpha, \beta, k_{abs}, r_{35}, iqr\}$ values for each MICP and centrifuge capillary pressure curve pair.

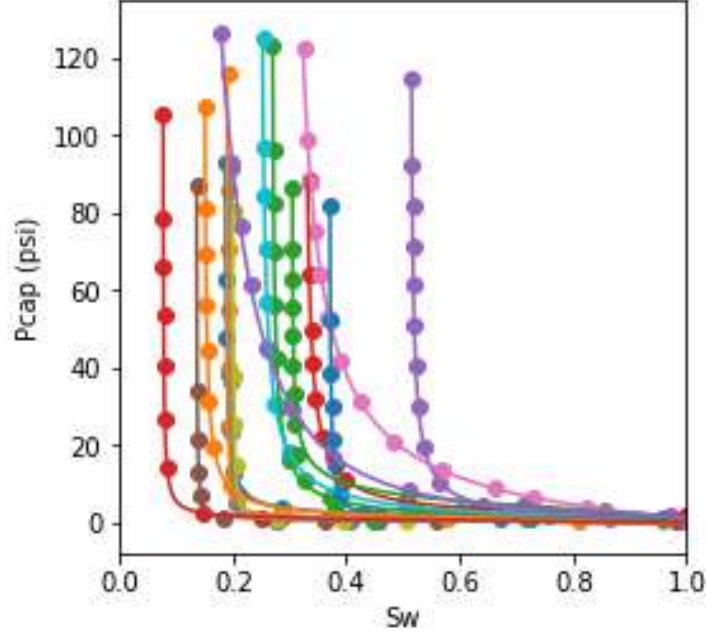


Figure 4.1: Sample of the experimental capillary pressure curve dataset (points) and fitted capillary pressure model (lines).

Significant correlations between corresponding centrifuge capillary pressure curves α , S_{wi} and MICP median pore throat radius r_{median} parameters can be visualized in Figure 4.2.

Given x a vector of input parameters $\{k_{abs}, r_{35}, iqr\}$ and y a vector of output parameters $\{S_{wi}, \alpha, \beta, P_e\}$, and considering a multivariate gaussian distribution described by equation (4.2), the conditional distribution of the output parameters given known input parameters $p(y|x)$ may be described by equations (4.3)(4.4)(4.5). Estimating mean and covariance matrix statistics of the multivariate gaussian distribution (4.2) on the assembled dataset, using maximum likelihood methods (De-

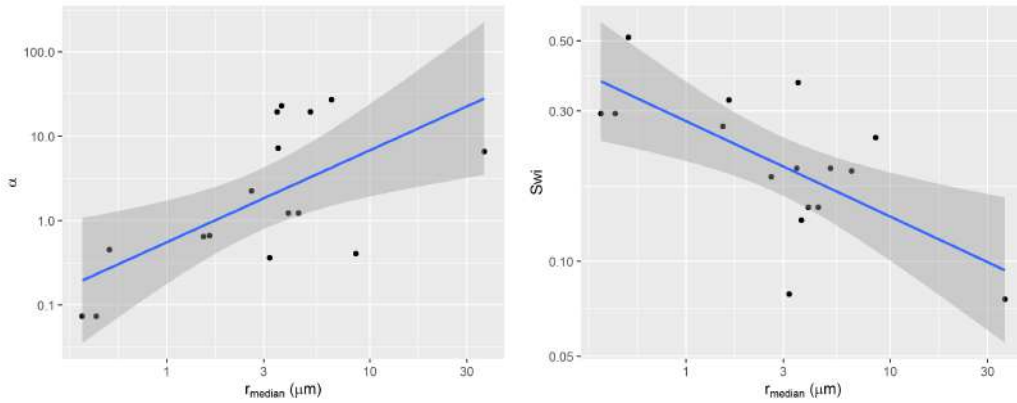


Figure 4.2: Correlations between parameters alpha, S_{wi} and median pore throat radius of correspondent rock fragment $p(\log r)$ distribution.

Groot and Schervish 2012), inference and uncertainty evaluation may be performed on desired new input x values.

$$\begin{pmatrix} x \\ y \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix} \right) \quad (4.2)$$

$$p(y|x) \sim \mathcal{N}(\mu_{y|x}, \Sigma_{y|x}) \quad (4.3)$$

$$\mu_{y|x} = \mu_y + (\Sigma_{xx}^{-1} \Sigma_{yx})^t (x - \mu_x) \quad (4.4)$$

$$\Sigma_{y|x} = \Sigma_{yy} - \Sigma_{xy} \Sigma_{xx}^{-1} \Sigma_{yx} \quad (4.5)$$

Given new $x = \{k_{abs}, r_{35}, iqr\}$ values, the expected capillary pressure curve parameters $E[y|x]$ may be obtained by the conditional mean $\mu_{y|x}$. Uncertainty evaluation of this prediction may be executed using samples of the conditional distribution $p(y|x)$. Figure 4.3 displays examples of experimental capillary pressure curves and predictions of these curves using as input associated $\{k_{abs}, r_{35}, iqr\}$ values. Samples from the conditional distribution $p(y|x)$ are shown as grey lines and illustrate prediction uncertainty.

On Figure 4.4, a comparison between experimental and predicted capillary pressure curve parameters is shown. Due to inherent noise associated with heterogeneous reservoir rocks and, as also observed in Figure 4.3, there is significant dispersion of experimental and predicted parameters values. Both in Figures 4.4 and 4.5, it is possible to visualize that capillary curve parameter predictions follow linear tendencies with absolute permeability. This property of linear model predictions is desirable, as it follows the expected physical behavior of reservoir rocks.

4.2 Hierarchical Linear Regression

In a reservoir model, the prediction of petrophysical properties on simulation cells distant from wells with available sampled cores is commonly performed using linear regression methods (Peters 2012). To account for sampling bias, special core analysis properties such as relative permeability and capillary pressure are commonly scaled according to reservoir wide available information, such as absolute permeability, porosity and geological facies models. For relative permeability, this procedure is commonly performed using simple linear regression (Gelman et al. 2014) of relative permeability parameters as a function of absolute permeability.

On this work, simple linear and hierarchical linear regression models were evaluated in a dataset containing 226 unsteady-state water-oil relative permeability curves from several Brazilian reservoirs. Parameters for each of the 226 curves were fitted using maximum likelihood estimation (Migon, Gamerman, and Louzada 2015)

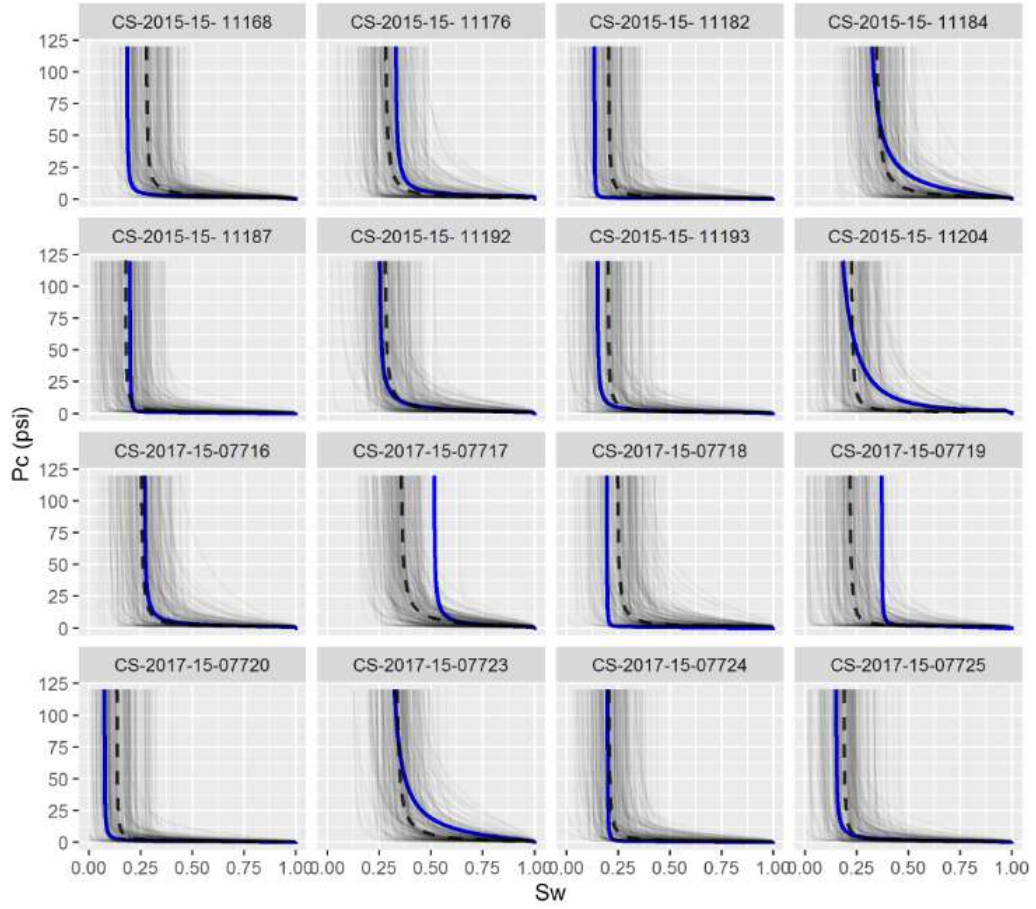


Figure 4.3: Visual comparison of experimental capillary pressure curves (blue lines), samples from the conditional distribution (grey lines) and average predicted curves (dashed black lines).

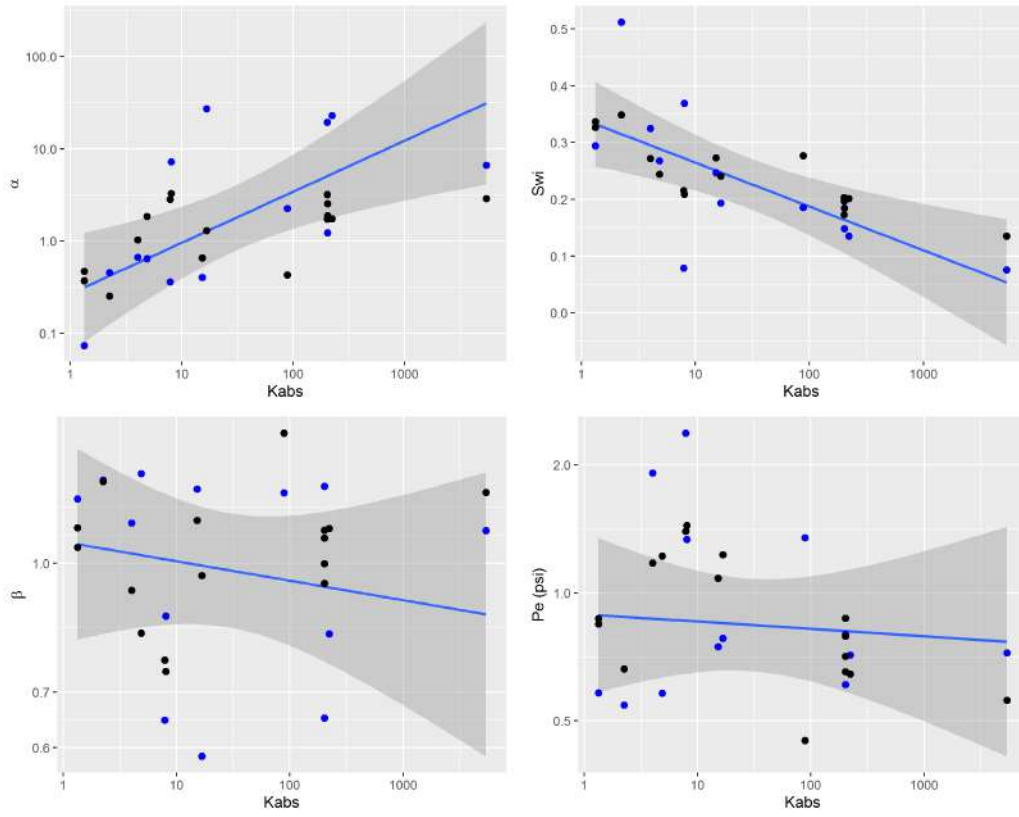


Figure 4.4: Experimental capillary pressure curve parameters (blue dots and line tendency) and predicted parameters estimated using the posterior mean (black dots).

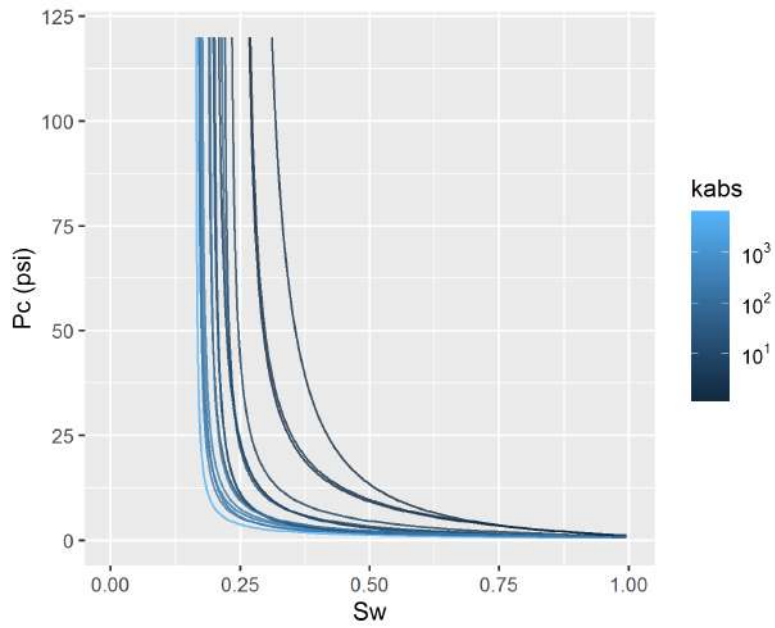


Figure 4.5: Behaviour of predicted capillary pressure curves with absolute permeability.

and the LET parameterization (4.6)(4.7)(4.8) proposed in (Lomeland, Ebeltoft, and Thomas 2005).

$$S_{wD} = \frac{S_w - S_{wi}}{1 - S_{wi} - S_{or}} \quad (4.6)$$

$$k_{ro}(S_w) = k_{ro@S_{wi}} \frac{(1 - S_{wD})^{L_o}}{(1 - S_{wD})^{L_o} + E_o(S_{wD})^{T_o}} \quad (4.7)$$

$$k_{rw}(S_w) = k_{rw@S_{or}} \frac{(S_{wD})^{L_w}}{(S_{wD})^{L_w} + E_w(1 - S_{wD})^{T_w}} \quad (4.8)$$

Categorical variables, such as field, reservoir or geological facies, may be used to group petrophysical models. A model that does not distinguish between groups, using constant intercept and slope parameters for all categories according to equation (4.9), may be referred to as a non-pooled model (Gelman et al. 2014). Figure 4.6 displays a simple linear non-pooled regression model of irreducible water saturation S_{wi} as the predicted y variable and the logarithm of absolute permeability $\log k_{abs}$ as the observed x variable. On equation (4.9), n represents the total number of data samples, indexed by the letter i , and α , β and σ_y^2 represent the intercept, slope and variance linear model parameters.

$$y_i \sim \mathcal{N}(\alpha + \beta x, \sigma_y^2), \quad \text{for } i = 1, \dots, n \quad (4.9)$$

Usually, though, separate linear regression models are fitted to each category of interest, in models that may be referred to as completely pooled. For each category j , independent α_j , β_j and $\sigma_{y_j}^2$ parameters are estimated, as described in equation (4.10).

$$y_i^j \sim \mathcal{N}(\alpha_j + \beta_j x_i^j, \sigma_{y_j}^2), \quad \text{for } i = 1, \dots, n; \quad \text{for } j = 1, \dots, J \quad (4.10)$$

Figure 4.7 displays a simple linear completely pooled regression model of irreducible water saturation S_{wi} and the logarithm of absolute permeability $\log k_{abs}$, grouped by reservoir. In completely pooled regression models, each linear regression model is independent of each other, with varying degrees of uncertainty on each model parameters α_j , β_j and $\sigma_{y_j}^2$. Categories with larger number of samples and smaller heterogeneities, usually display smaller uncertainties on model parameters.

Hierarchical or partially pooled linear regression models introduce information sharing and coupling between model parameters of different categories, modeling intercept and/or slope parameters as sampled from a latent parent distribution.

Varying intercept models, assume that the intercept of each category α_j is sampled from a common latent gaussian distribution (4.12). This information sharing, has a regularizing effect of shrinking the partially pooled parameters towards a com-

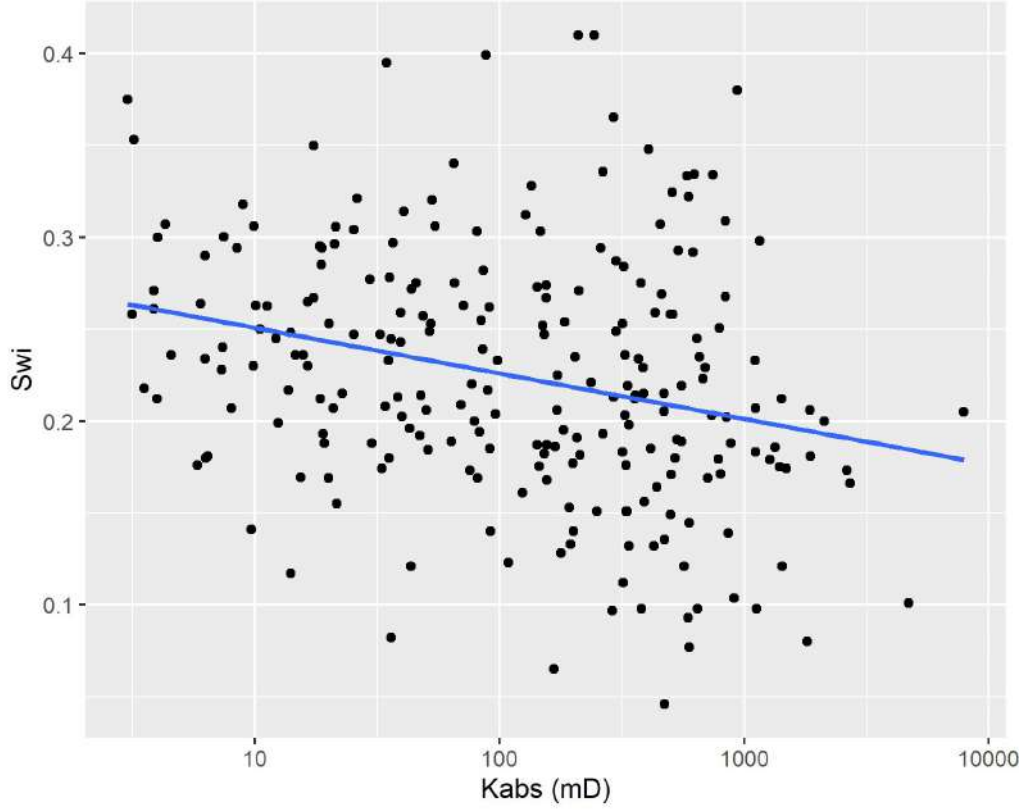


Figure 4.6: Non-pooled simple linear regression.

mon mean μ_α .

$$y_i^j \sim \mathcal{N}(\alpha_j + \beta_j x_i^j, \sigma_{y^j}^2), \quad \text{for } i = 1, \dots, n; \quad \text{for } j = 1, \dots, J \quad (4.11)$$

$$\alpha_j \sim \mathcal{N}(\mu_\alpha, \sigma_\alpha^2), \quad \text{for } j = 1, \dots, J \quad (4.12)$$

Varying slope models, assume that the slope of each category β_j is sampled from a common latent gaussian distribution (4.14), with the same regularizing effect of shrinking the partially pooled parameters towards a common mean μ_β .

$$y_i^j \sim \mathcal{N}(\alpha_j + \beta_j x_i^j, \sigma_{y^j}^2), \quad \text{for } i = 1, \dots, n; \quad \text{for } j = 1, \dots, J \quad (4.13)$$

$$\beta_j \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2), \quad \text{for } j = 1, \dots, J \quad (4.14)$$

Varying intercept and slope models, assume that both the intercept and slope of each category α_j and β_j are sampled from a common latent multivariate gaussian distribution (4.16).

$$y_i^j \sim \mathcal{N}(\alpha_j + \beta_j x_i^j, \sigma_{y^j}^2), \quad \text{for } i = 1, \dots, n; \quad \text{for } j = 1, \dots, J \quad (4.15)$$

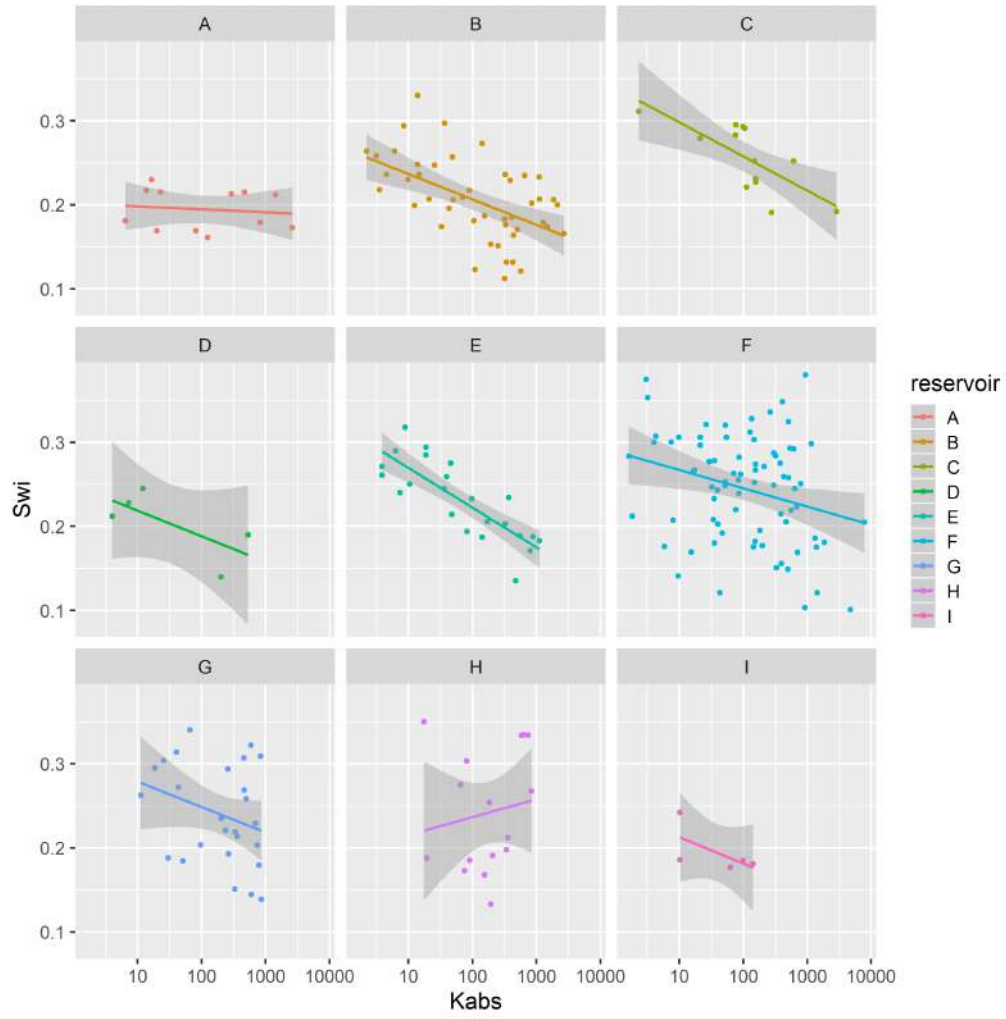


Figure 4.7: Completely pooled simple linear regression models, grouped by reservoir.

$$\begin{pmatrix} \alpha_j \\ \beta_j \end{pmatrix} \sim \mathcal{N} \left(\begin{pmatrix} \mu_\alpha \\ \mu_\beta \end{pmatrix}, \begin{pmatrix} \sigma_\alpha^2 & \rho\sigma_\alpha\sigma_\beta \\ \rho\sigma_\alpha\sigma_\beta & \sigma_\beta^2 \end{pmatrix} \right), \quad \text{for } j = 1, \dots, J \quad (4.16)$$

For the dataset of 226 unsteady-state water-oil relative permeability curve LET parameters, hierarchical varying slope models of relative permeability endpoint parameters S_{wi} , S_{or} , $k_{ro}@S_{wi}$ and $k_{rw}@S_{or}$, and logarithm of absolute permeability $\log k_{abs}$ were fit and compared to completely pooled simple linear regression models, grouped by reservoir.

Both hierarchical and simple linear regression model parameters were inferred using bayesian Hamiltonian Markov-Chain Monte-Carlo (Hoffman and Gelman 2014) and the software Stan (Carpenter et al. 2017). Default weakly informative model parameter priors were utilized, following the recommendations of (Gelman et al. 2008).

Comparison between hierarchical and simple linear regression models were performed using the root mean squared error RMSE, the Watanabe-Akaike Information Criteria (WAIC), the leave-one-out information criteria (LOOIC), and the bayesian R-squared R^2 and adjusted R-squared R_{adj}^2 metrics (Gelman et al. 2019)(Vehtari, Gelman, and Gabry 2017). The WAIC and LOOIC information criteria provide a trade-off between goodness-of-fit and model complexity, with lower WAIC and LOOIC values corresponding to lower cross-validation errors.

Hierarchical varying slope and simple linear regression models of irreducible water saturation S_{wi} and logarithm of absolute permeability $\log k_{abs}$ were evaluated on the assembled dataset.

Table 4.1 displays the obtained regression metrics for each fitted model. Hierarchical linear regression achieved slightly better WAIC, LOOIC and R_{adj}^2 metrics.

Table 4.1: S_{wi} linear regression model metrics.

Regression Model	RMSE	WAIC	LOOIC	R^2	R_{adj}^2
Simple Linear Regression	0.05	-681.48	-680.96	0.20	0.16
Hierarchical Linear Regression	0.05	-684.89	-684.66	0.24	0.17

Figure 4.8 displays simple and hierarchical linear regression S_{wi} vs $\log k_{abs}$ models, grouped by reservoir. Black dots represent observed samples, light blue lines represent samples from the posterior distribution of intercept and slope parameters, and dark blue lines represent mean intercept and slope parameters, for each reservoir.

Completely pooled, simple linear regression models display larger between-groups slope variations and uncertainty, as shown in Figure 4.9. Reservoirs with large num-

ber of samples, such as reservoir F, display only small changes between hierarchical and simple linear regression model posterior distributions. A stronger regularizing effect is displayed in reservoir H, which contains a small number of samples. Overall behavior consistency of S_{wi} with respect to $\log k_{abs}$ is increased in the hierarchical partially pooled linear model, as slopes are regressed towards a common mean. As information is shared between different reservoirs, posterior uncertainties are noticeably reduced in the hierarchical linear model.

Hierarchical varying slope and simple linear regression models of residual oil saturation S_{or} and logarithm of absolute permeability $\log k_{abs}$ were evaluated on the assembled dataset.

Table 4.2 displays the obtained regression metrics for each fitted model. Both models achieved similar WAIC, LOOIC and R_{adj}^2 metrics.

Table 4.2: S_{or} linear regression model metrics.

Regression Model	RMSE	WAIC	LOOIC	R^2	R_{adj}^2
Simple Linear Regression	0.09	-414.47	-413.64	0.18	0.11
Hierarchical Linear Regression	0.09	-414.40	-414.06	0.22	0.11

Figure 4.10 displays simple and hierarchical linear regression S_{or} vs $\log k_{abs}$ models, grouped by reservoir. Black dots represent observed samples, light blue lines represent samples from the posterior distribution of intercept and slope parameters, and dark blue lines represent mean intercept and slope parameters, for each reservoir.

Completely pooled, simple linear regression models display larger between-groups slope variations and uncertainty, as shown in Figure 4.11. Reservoirs with large number of samples, such as reservoir F, show only small changes between hierarchical and simple linear regression model posterior distributions. A stronger regularizing effect is displayed in reservoir I, which contains a small number of samples. Overall behavior consistency of S_{or} with respect to $\log k_{abs}$ is increased in the hierarchical partially pooled linear model, as slopes are regressed towards a common mean. As information is shared between different reservoirs, posterior uncertainties are noticeably reduced in the hierarchical linear model.

Hierarchical varying slope and simple linear regression models of oil relative permeability at irreducible water saturation condition $k_{ro}@S_{wi}$ and logarithm of absolute permeability $\log k_{abs}$ were evaluated on the assembled dataset.

Table 4.3 displays the obtained regression metrics for each fitted model. Hierarchical linear regression achieved slightly better WAIC, LOOIC and R_{adj}^2 metrics.

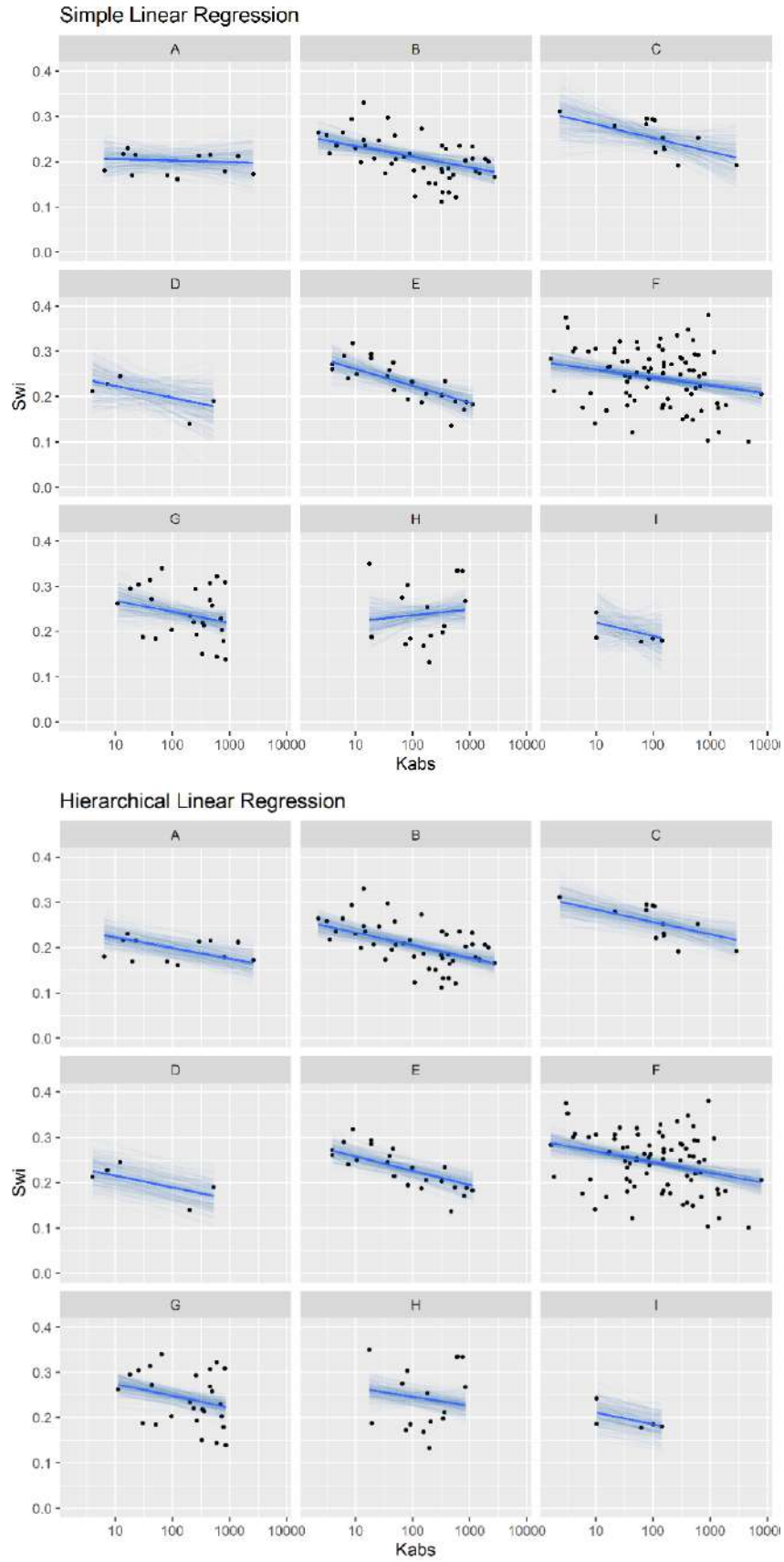


Figure 4.8: Simple (top) and Hierarchical (bottom) linear regression models of Swi vs $\log(\text{kabs})$, grouped by reservoir.

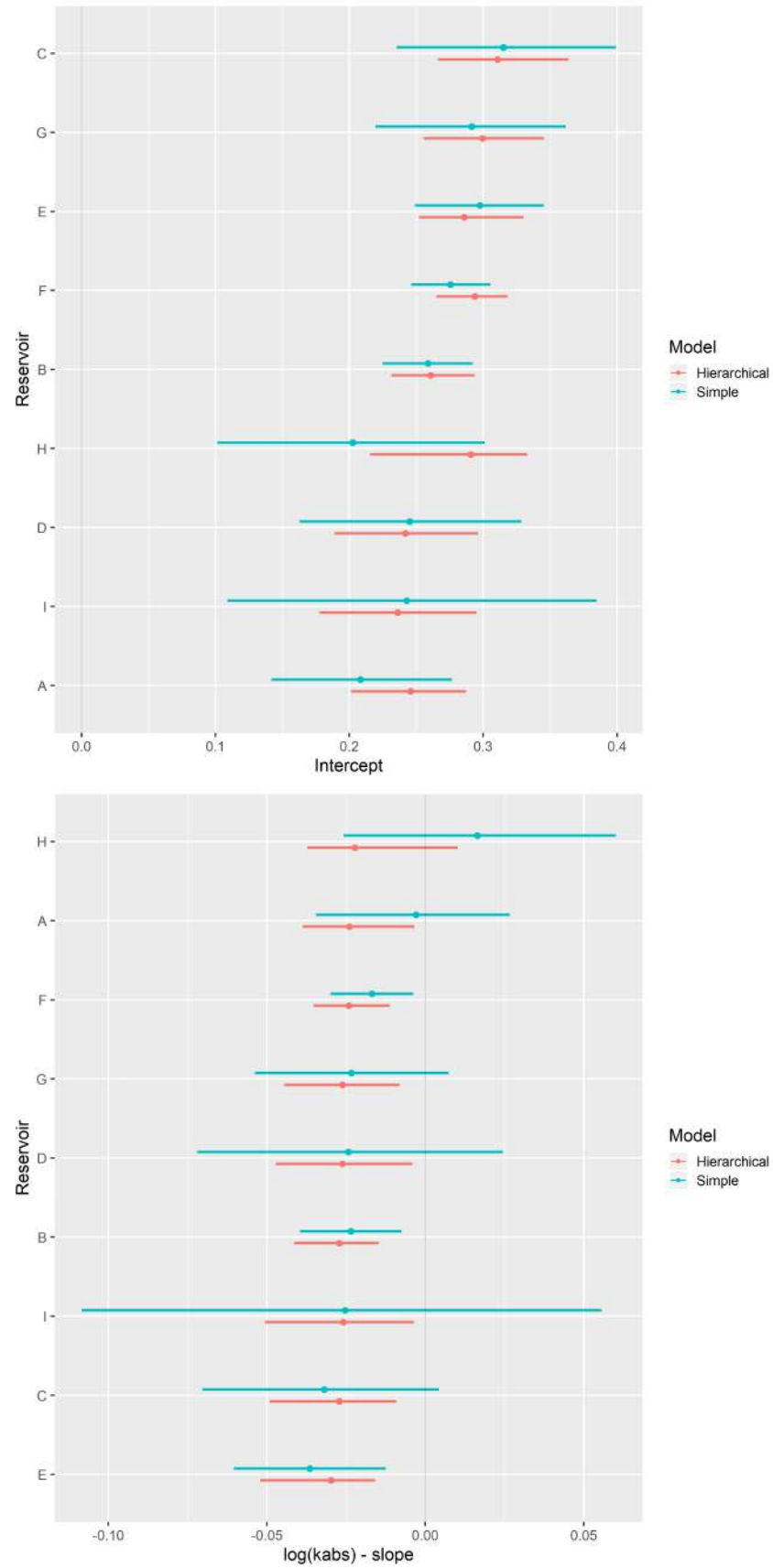


Figure 4.9: Intercept (top) and slope (bottom), simple and hierarchical linear regression Sw_i vs $\log(kabs)$ model parameters, grouped by reservoir.

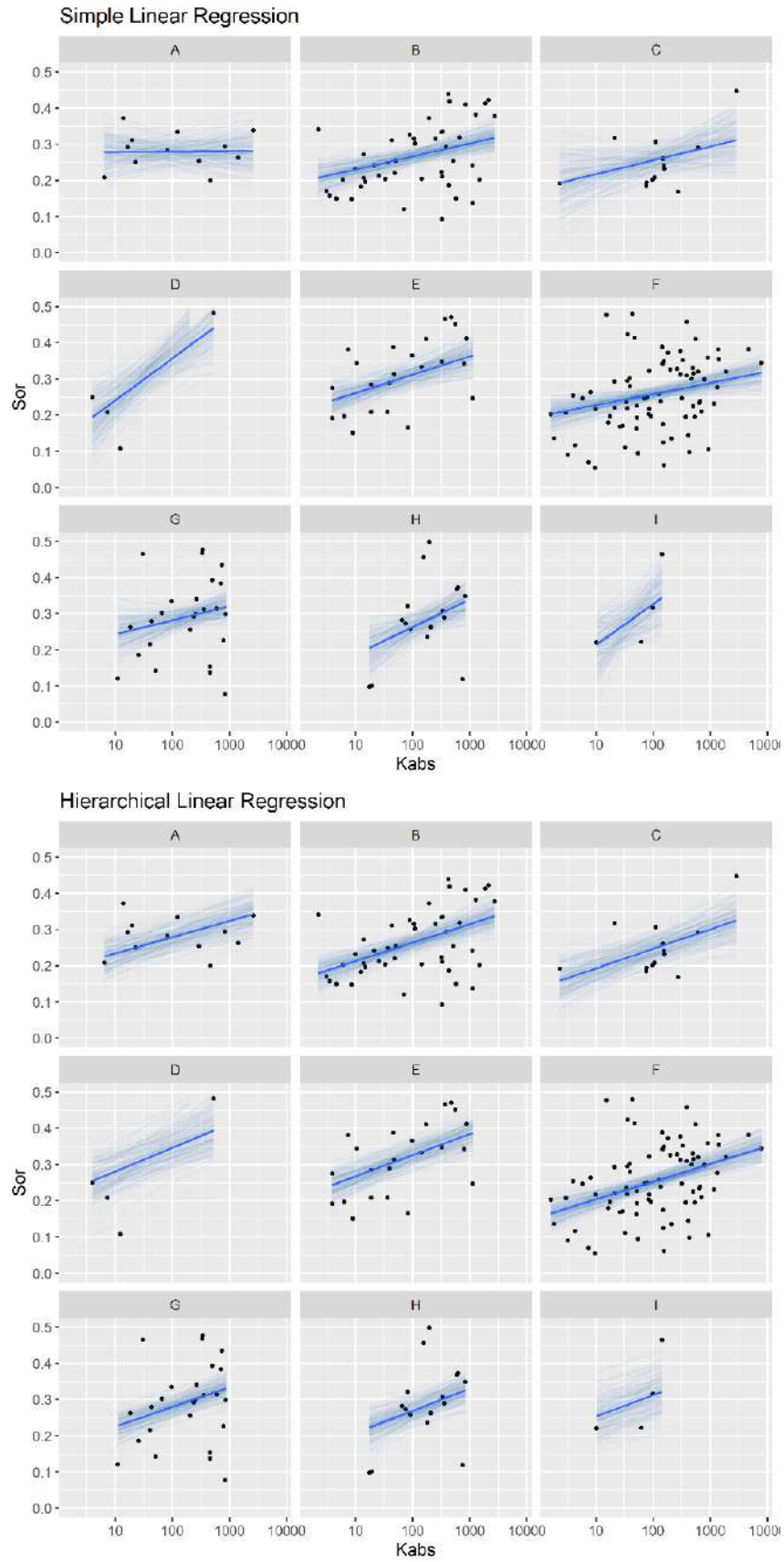


Figure 4.10: Simple (top) and Hierarchical (bottom) linear regression models of Sor vs $\log(k_{abs})$, grouped by reservoir.

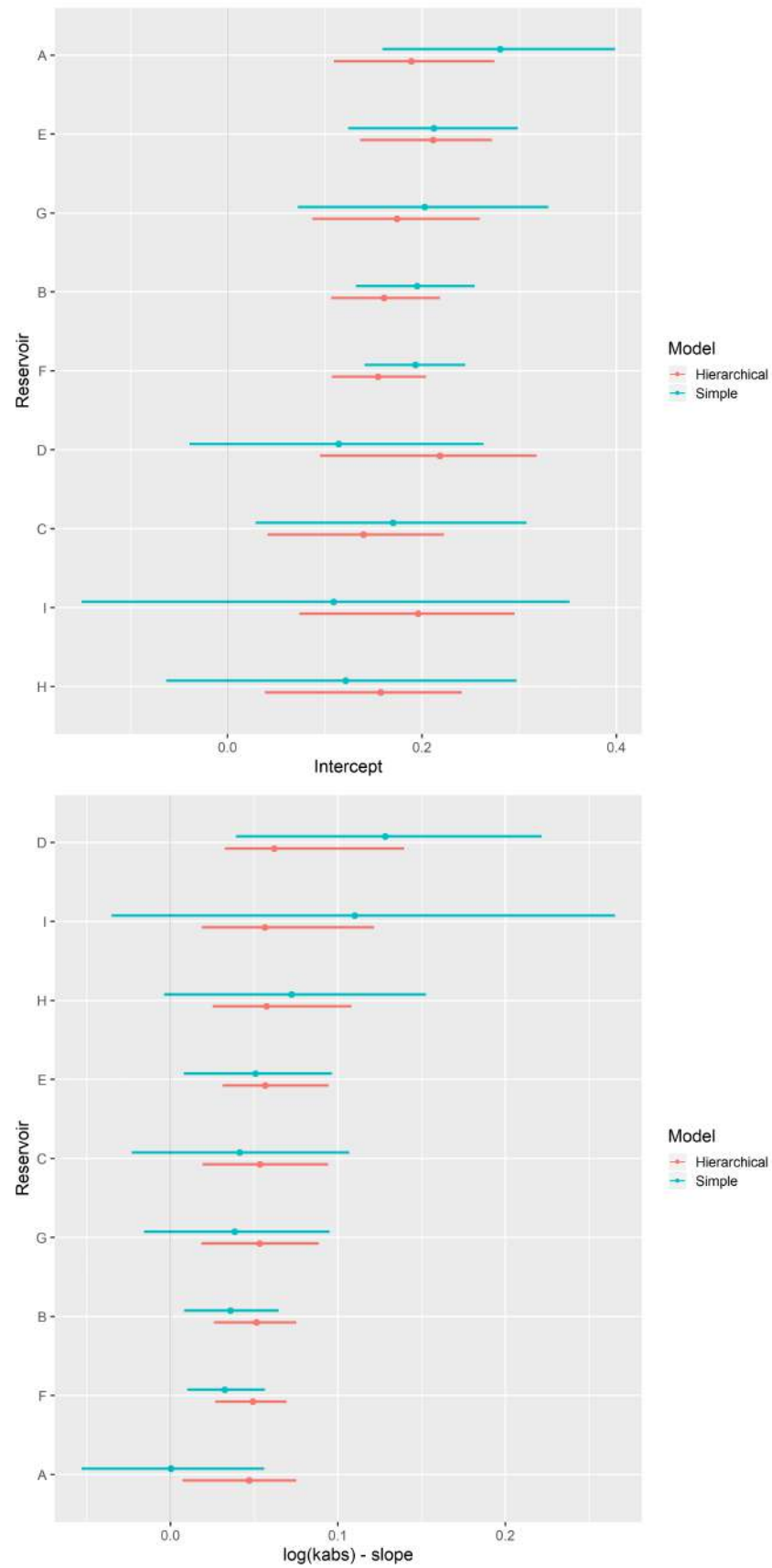


Figure 4.11: Intercept (top) and slope (bottom), simple and hierarchical linear regression Sor vs log(kabs) model parameters, grouped by reservoir.

Table 4.3: $k_{ro}@S_{wi}$ linear regression model metrics.

Regression Model	RMSE	WAIC	LOOIC	R^2	R_{adj}^2
Simple Linear Regression	0.20	-48.72	-47.50	0.34	0.28
Hierarchical Linear Regression	0.20	-62.02	-61.77	0.39	0.33

Figure 4.12 displays simple and hierarchical linear regression $k_{ro}@S_{wi}$ vs $\log k_{abs}$ models, grouped by reservoir. Black dots represent observed samples, light blue lines represent samples from the posterior distribution of intercept and slope parameters, and dark blue lines represent mean intercept and slope parameters, for each reservoir.

Completely pooled, simple linear regression models display larger between-groups slope variations and uncertainty, as shown in Figure 4.13. Reservoirs with large number of samples, such as reservoir F, show only small changes between hierarchical and simple linear regression model posterior distributions. A stronger regularizing effect is displayed in reservoir D, which contains a small number of samples. Overall behavior consistency of $k_{ro}@S_{wi}$ in respect to $\log k_{abs}$ is increased in the hierarchical partially pooled linear model, as slopes are regressed towards a common mean. As information is shared between different reservoirs, posterior uncertainties are noticeably reduced in the hierarchical linear model.

Hierarchical varying slope and simple linear regression models of water relative permeability at residual oil saturation condition $k_{rw}@S_{or}$ and logarithm of absolute permeability $\log k_{abs}$ were evaluated on the assembled dataset.

Table 4.4 displays the obtained regression metrics for each fitted model. Hierarchical linear regression achieved slightly better WAIC, LOOIC and R_{adj}^2 metrics.

Table 4.4: $k_{rw}@S_{or}$ linear regression model metrics.

Regression Model	RMSE	WAIC	LOOIC	R^2	R_{adj}^2
Simple Linear Regression	0.10	-366.62	-366.13	0.20	0.16
Hierarchical Linear Regression	0.10	-372.34	-372.18	0.24	0.17

Figure 4.14 displays simple and hierarchical linear regression $k_{rw}@S_{or}$ vs $\log k_{abs}$ models, grouped by reservoir. Black dots represent observed samples, light blue lines represent samples from the posterior distribution of intercept and slope parameters, and dark blue lines represent mean intercept and slope parameters, for each reservoir.

Completely pooled, simple linear regression models display larger between-groups

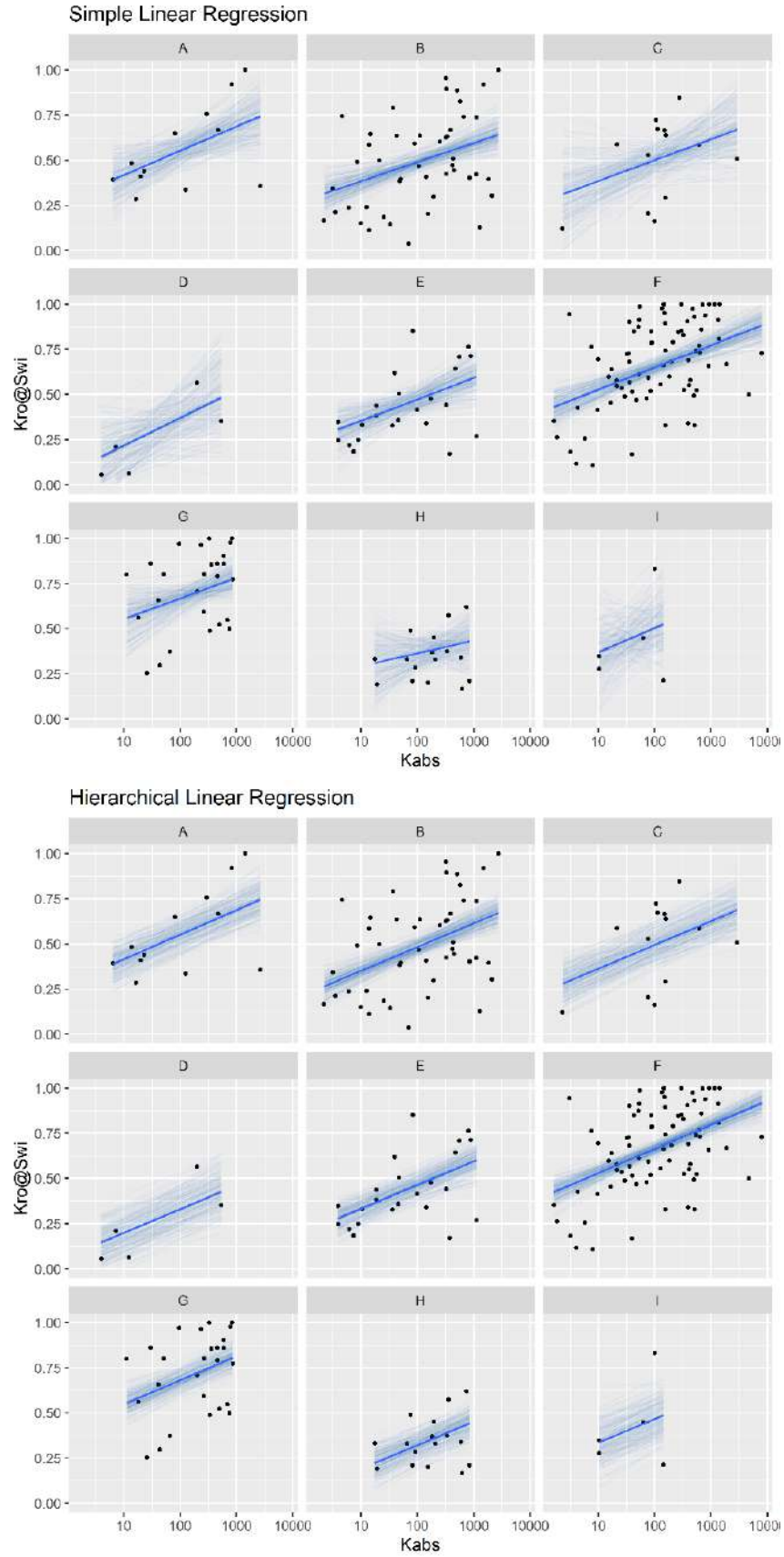


Figure 4.12: Simple (top) and Hierarchical (bottom) linear regression models of kro@Swi vs $\log(\text{kabs})$, grouped by reservoir.

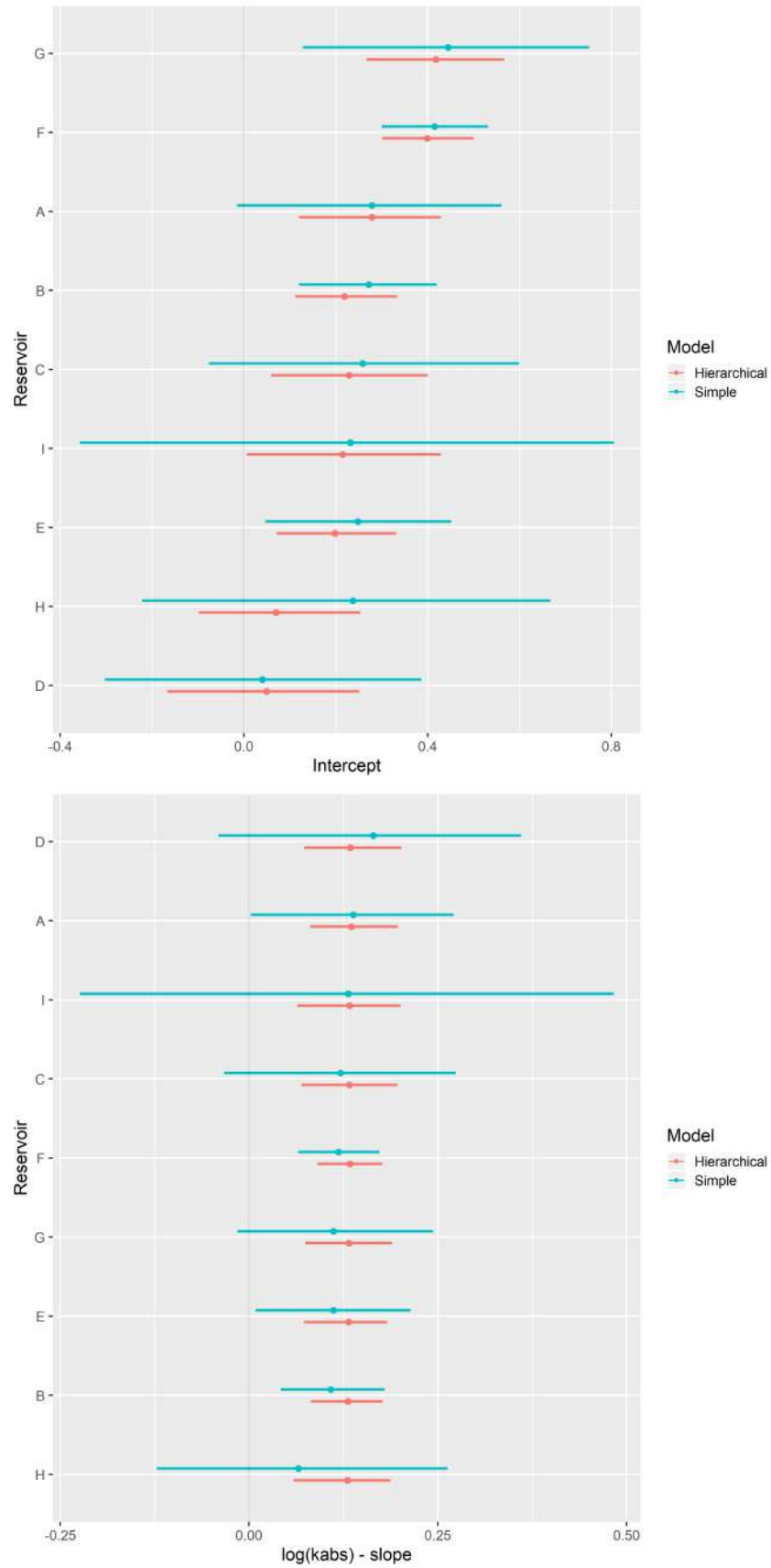


Figure 4.13: Intercept (top) and slope (bottom), simple and hierarchical linear regression kro@Swi vs $\log(\text{kabs})$ model parameters, grouped by reservoir.

slope variations and uncertainty, as shown in Figure 4.15. Reservoirs with large number of samples, such as reservoir F, show only small changes between hierarchical and simple linear regression model posterior distributions. A stronger regularizing effect is displayed in reservoir D, which contains a small number of samples. Overall behavior consistency of $k_{rw}@S_{or}$ in respect to $\log k_{abs}$ is increased in the hierarchical partially pooled linear model, as slopes are regressed towards a common mean. As information is shared between different reservoirs, posterior uncertainties are noticeably reduced in the hierarchical linear model.

Posterior distribution of latent μ_α and μ_β parameters of hierarchical linear regression models represent average behavior of model parameters across the different evaluated categories. Thus, they represent quantified petrophysical parameter model analogues, and may be used for preliminary characterization of reservoirs with similar characteristics as the ones used in the assembled model, but with no sampled data.

Multi-task simple and varying slopes hierarchical linear regression models, grouped by reservoir, were fitted to the assembled LET relative permeability parameter dataset. Comparison of WAIC and LOOIC metrics between them is shown in Table 4.5, displaying slightly better results for the multi-task hierarchical linear regression model.

Table 4.5: Multi-task linear regression model metrics.

Regression Model	WAIC	LOOIC
Simple Linear Regression	1119.6	1129.8
Hierarchical Linear Regression	1043.3	1048.0

The posterior distribution of the LET relative permeability parameters for a given reservoir and logarithmic absolute permeability may be used to sample relative permeability curves, fully incorporating the information from the available dataset, as exemplified in Figure 4.16.

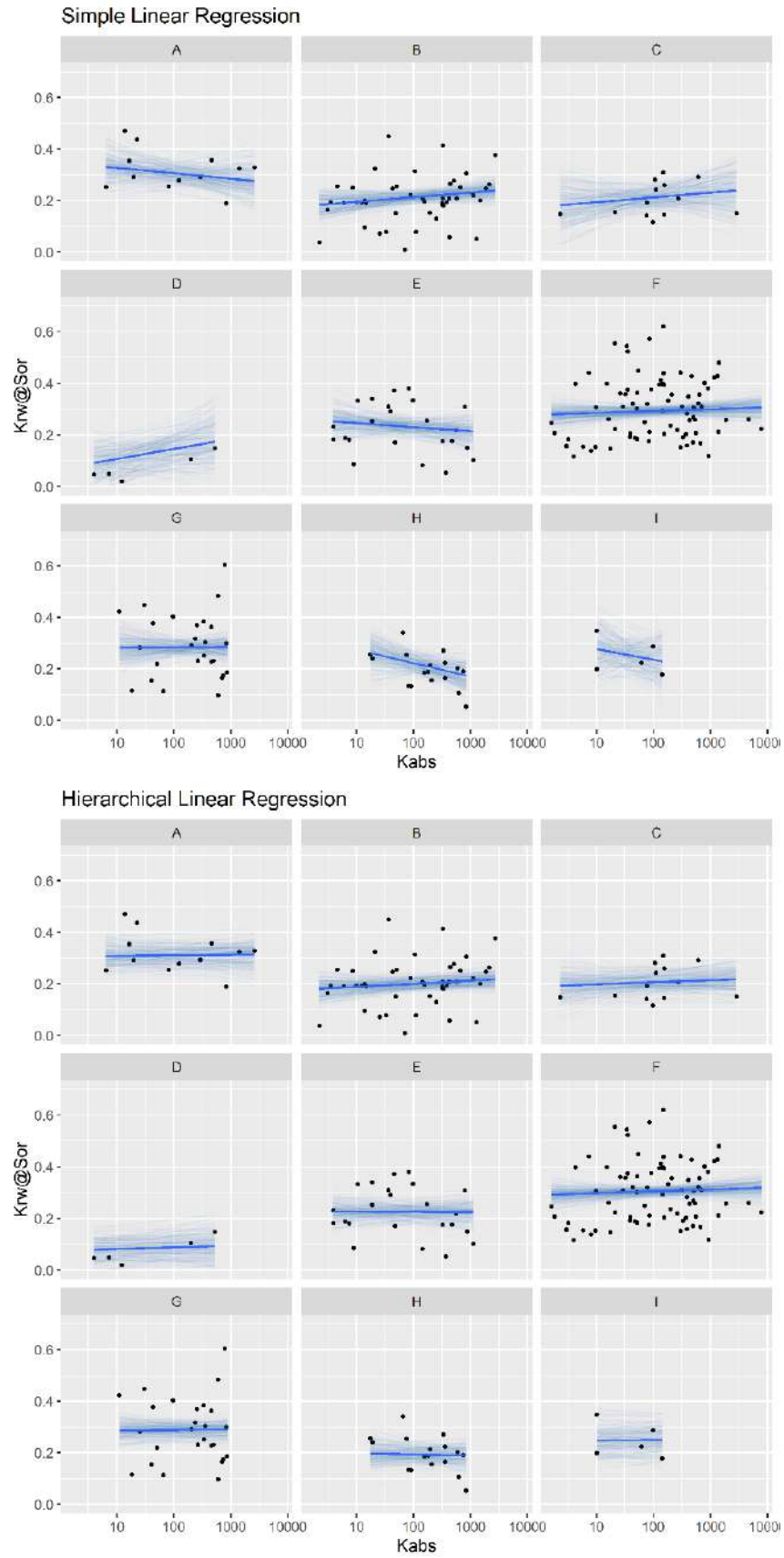


Figure 4.14: Simple (top) and Hierarchical (bottom) linear regression models of $krw@Sor$ vs $\log(kabs)$, grouped by reservoir.

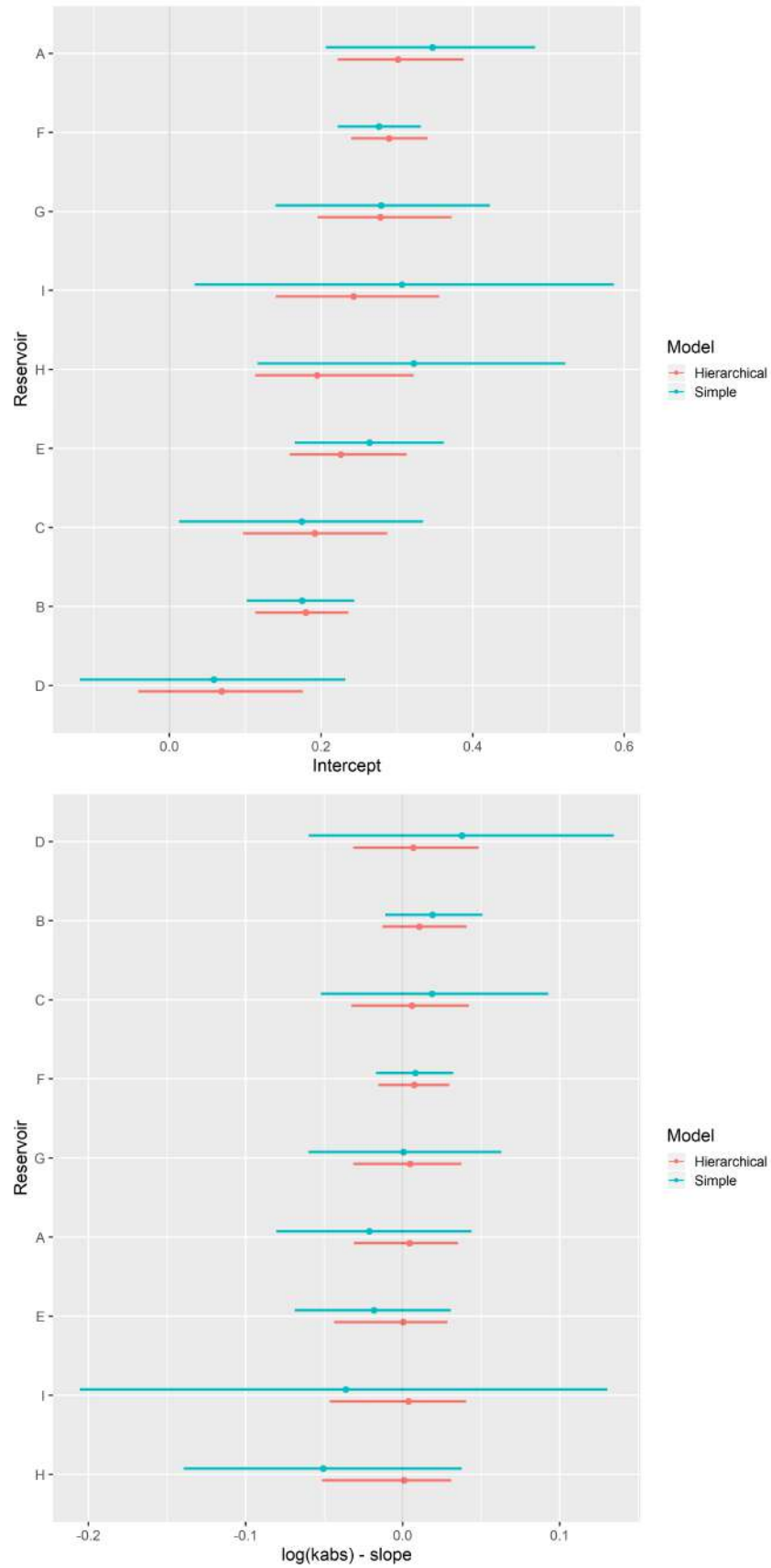


Figure 4.15: Intercept (top) and slope (bottom), simple and hierarchical linear regression $\text{krw}@\text{Sor}$ vs $\log(\text{kabs})$ model parameters, grouped by reservoir.

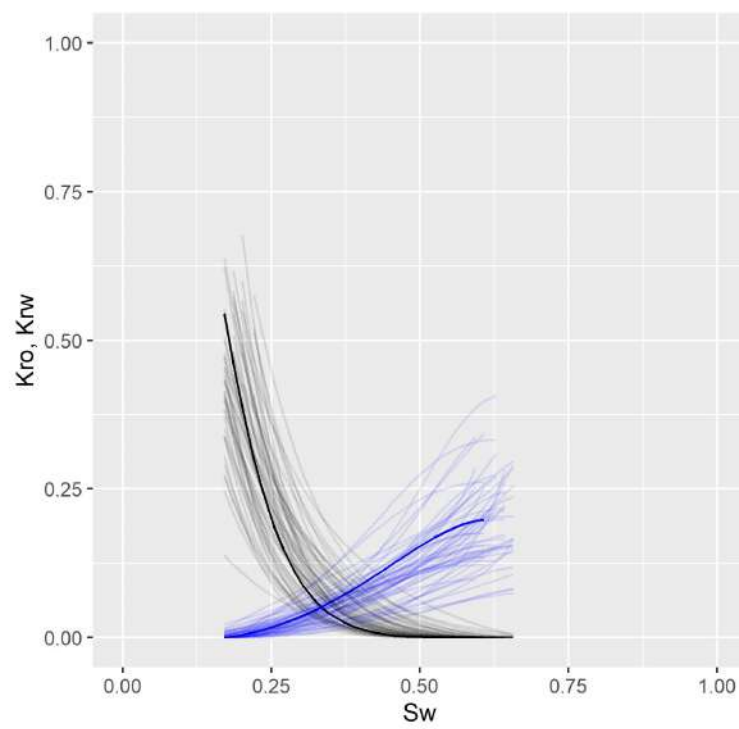


Figure 4.16: Example of multivariate posterior sample of relative permeability curves, fully incorporating the information from the available dataset.

Conclusion

In this work, machine learning and statistical regression models were evaluated for the prediction of routine and special core analysis petrophysical properties on datasets containing experimental results for rock samples from many Brazilian reservoirs.

Several feature engineering and machine learning techniques were evaluated for the estimation of absolute permeability from mercury injection capillary pressure curves. The absolute permeability regression models that achieved the lowest median absolute errors and largest correlation coefficients used the support-vector machine (SVR), random forest or gradient boosted decision trees algorithms, and several features extracted from the mercury porosimetry capillary pressure curve data. Among the linear models for prediction of absolute permeability, the models proposed by Swanson (Swanson 1981) and Winland (Kolodzie 1980) obtained the lowest root mean squared errors and highest correlation coefficients.

Using a parametric formulation, special core analysis capillary pressure and relative permeability curve regression problems were framed in a multi-task regression approach. For the estimation of capillary pressure curve parameters, an analytic formulation of the multi-task linear regression problem was considered using the multivariate gaussian conditional distribution. This model was evaluated on an experimental dataset, comparing average predictions and samples from the conditional distribution of parameters, to the observed capillary pressure curve parameters. Although significant dispersion of experimental and predicted parameters values were observed, it was possible to identify that predictions followed expected linear tendencies of capillary pressure curve parameters and absolute permeability.

Posterior distribution of partially pooled varying slopes hierarchical and completely pooled simple linear regression models, with respect to logarithmic absolute permeability, were inferred for relative permeability parameters. Hierarchical linear regression models displayed overall improved information criteria metrics, evaluated using bayesian Watanabe-Akaike and Leave-one-out information criteria. Due to the regularizing effect of the information sharing between different reservoir categories, posterior distribution of relative permeability parameters of hierarchical linear regression models displayed smaller uncertainties and greater consistency.

Posterior distribution of latent parameters of hierarchical linear regression models represent average behavior of model parameters across the different evaluated categories and may be used as quantified petrophysical parameter model analogues for petrophysical characterization of reservoirs with similar characteristics as the ones used in the assembled model, but with no sampled data.

Bibliography

- Albuquerque, Marcelo R, Felipe M Eler, Heitor V R Camargo, André Compan, Dario Cruz, and Carlos Pedreira. 2018. “Estimation of Capillary Pressure Curves from Centrifuge Measurements using Inverse Methods.” *Rio Oil & Gas Expo and Conference 2018*,
- Al Khalifah, H., P. W. J. Glover, and P. Lorinczi. 2020. “Permeability prediction and diagenesis in tight carbonates using machine learning techniques.” *Marine and Petroleum Geology* 112 (May 2019): 104096. <https://doi.org/10.1016/j.marpetgeo.2019.104096>.
- Bentsen, R G. 1977. “Using Parameter-Estimation Techniques To Convert Centrifuge Data Into a Capillary-Pressure Curve.” *Society of Petroleum Engineers Journal* 17 (1): 57–64. <https://doi.org/10.2118/5026-PA>.
- Blunt, Martin J. 2017. *Multiphase Flow in Porous Media: A Pore-Scale Perspective*. Cambridge University Press.
- Carpenter, Bob, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus A. Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. 2017. “Stan: A probabilistic programming language.” *Journal of Statistical Software* 76 (1). <https://doi.org/10.18637/jss.v076.i01>.
- Chen, Tianqi, and Carlos Guestrin. 2016. “XGBoost: A scalable tree boosting system.” *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 13-17-Aug: 785–94. <https://doi.org/10.1145/2939672.2939785>.
- Corey, A. T. 1954. “The Interrelation Between Gas and Oil Relative Permeabilities.” *Producers Monthly* 38-41.
- Dake, L. P. 2015. *Fundamentals fo Reservoir Engineering*. <https://doi.org/10.1016/B978-0-08-098206-9.00004-X>.
- DeGroot, Morris H., and Mark J. Schervish. 2012. *Probability and Statistics*. Edited by Addison-Wesley.

- Forbes, P. 1994. “Simple and Accurate Methods for Converting Centrifuge Data Into Drainage and Imbibition Capillary Pressure Curves.” *The Log Analyst*.
- Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2014. *Bayesian Data Analysis*. 3. <https://doi.org/10.1017/CB09781107415324.004>.
- Gelman, Andrew, Ben Goodrich, Jonah Gabry, and Aki Vehtari. 2019. “R-Squared for Bayesian Regression Models.” *The American Statistician* 73 (3): 307–9.
- Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu Sung Su. 2008. “A weakly informative default prior distribution for logistic and other regression models.” *Annals of Applied Statistics* 2 (4): 1360–83. <https://doi.org/10.1214/08-AOAS191>.
- Hassler, G. L., and E. Brunner. 1945. “Measurement of Capillary Pressures in Small Core Samples.” *Petroleum Transactions of AIME* 160 (1): 114–23. <https://doi.org/10.2118/945114-G>.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*.
- Hoffman, Matthew D., and Andrew Gelman. 2014. “The no-U-turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo.” *Journal of Machine Learning Research* 15: 1593–1623. <http://arxiv.org/abs/1111.4246>.
- Johnson, E. F., D. P. Bossler, and V. O. Naumann. 1959. “Calculation of Relative Permeability from Displacement Experiments.” *Trans. AIME*.
- Kennedy, Martin. 2015. *Practical Petrophysics*. Vol. 62.
- Kolodzie, Stanley. 1980. “Analysis of Pore Throat Size and Use of the Waxman-Smits Equation to Determine Ooip in Spindle Field, Colorado.” In *SPE Annual Technical Conference and Exhibition*, 10. Dallas, Texas: Society of Petroleum Engineers. <https://doi.org/10.2118/9382-MS>.
- Lenormand, R. 2003. “Interpretation of mercury injection curves to derive pore size distribution.” *International Symposium of the Society of Core Analysts* International: SCA2003–52. <https://doi.org/SCA2003-52>.
- Lenormand, Roland, and Guillaume Lenormand. 2016. “Recommended Procedure for Determination of Relative Permeabilities.” *International Symposium of the Society of Core Analysts*, 1–12. <https://www.scaweb.org/wp-content/uploads/SCA-2016-Technical-Papers-Tuesday.pdf>.
- Lomeland, Frode, Einar Ebeltoft, and Wibeke Hammervold Thomas. 2005. “A

- new versatile relative permeability correlation.” *International Symposium of the Society of Core Analysts, Toronto, Canada*, 1–12.
- Mackay, David J. C. 2009. *Sustainable Energy—without the Hot Air*.
- McInnes, L., and J. Healy. 2018. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.”
- Micromeritics. 2020. “AutoPore IV Series: Automated Mercury Porosimeters.” <https://www.micromeritics.com/product-showcase/autopore-iv.aspx>.
- Migon, Helio, Dani Gamerman, and Francisco Louzada. 2015. *Statistical Inference*. Second.
- Nooruddin, Hasan A., Fatai Anifowose, and Abdulazeez Abdulraheem. 2013. “Applying Artificial Intelligence Techniques to Develop Permeability Predictive Models using Mercury Injection Capillary-Pressure Data.” *SPE Saudi Arabia Section Technical Symposium and Exhibition*, 1–16. <https://doi.org/10.2118/168109-MS>.
- Nordtvedt, JE, and K Kolltvelt. 1991. “Capillary pressure curves from centrifuge data by use of spline functions.” *SPE Reservoir Engineering*, no. November: 497–501. <https://doi.org/10.2118/19019-PA>.
- Peters, E. J. 2012. *Advanced Petrophysics*. Austin: Live Oak Book Company.
- Pittman, E. D. 1992. “Relationship of porosity and permeability to various parameters derived from mercury injection-capillary pressure curves for sandstone.” <https://doi.org/10.1017/CB09781107415324.004>.
- P. Murphy, Kevin. 2012. *Machine Learning: A Probabilistic Perspective*. https://doi.org/10.1007/SpringerReference_35834.
- Purcell, W. R. 1949. “Capillary Pressures - Their Measurement Using Mercury and the Calculation of Permeability Therefrom.” *Journal of Petroleum Technology* 1 (02): 39–48. <https://doi.org/10.2118/949039-G>.
- Skuse, Brian, Abbas Flroozabadl, and Henry J Ramey Jr. 1992. “Computation and Interpretation of Capillary Pressure From a Centrifuge.” <https://doi.org/10.2118/18297-PA>.
- Swanson, B. F. 1981. “A Simple Correlation Between Permeabilities and Mercury Capillary Pressures.” *Journal of Petroleum Technology* 33 (12): 2498–2504. <https://doi.org/10.2118/8234-PA>.
- Tiab, Djebbar, and E. C. Donaldson. 2004. *Petrophysics: theory and practice of*

measuring reservoir rock and fluid transport properties. Boston: Gulf Professional Pub.

- Vehtari, Aki, Andrew Gelman, and Jonah Gabry. 2017. “Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.” *Statistics and Computing* 27 (5): 1413–32. <https://doi.org/10.1007/s11222-016-9696-4>.
- Xu, C., and C. Torres-Verdín. 2013. “Pore System Characterization and Petrophysical Classification Using a Bimodal Gaussian Density Function.” *Math Geosci*.